Perspective-Oriented Generation of Football Match Summaries: Old Tasks, New Challenges

NADJET BOUAYAD-AGHA, GERARD CASAMAYOR, and SIMON MILLE, DTIC, Universitat Pompeu Fabra
LEO WANNER, ICREA and DTIC, Universitat Pompeu Fabra

Team sports commentaries call for techniques that are able to select content and generate wordings to reflect the affinity of the targeted reader for one of the teams. The existing works tend to have in common that they either start from knowledge sources of limited size to whose structures then different ways of realization are explicitly assigned, or they work directly with linguistic corpora, without the use of a deep knowledge source. With the increasing availability of large-scale ontologies this is no longer satisfactory: techniques are needed that are applicable to general purpose ontologies, but which still take user preferences into account. We take the best of both worlds in that we use a two-layer ontology. The first layer is composed of raw domain data modelled in an application-independent base OWL ontology. The second layer contains a rich perspective generation-motivated domain communication knowledge ontology, inferred from the base ontology. The two-layer ontology allows us to take into account user perspective-oriented criteria at different stages of generation to generate perspective-oriented commentaries. We show how content selection, discourse structuring, information structure determination, and lexicalization are driven by these criteria and how stage after stage a truly user perspective-tailored summary is generated. The viability of our proposal has been evaluated for the generation of football match summaries of the First Spanish Football League. The reported outcome of the evaluation demonstrates that we are on the right track.

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Language generation*

General Terms: Human Factors Languages

Additional Key Words and Phrases: OWL, content selection, discourse structuring, information structuring, lexicalization, ontology, domain communication knowledge, user perspective, multilevel generation

ACM Reference Format:

Bouayad-Agha, N., Casamayor, G., Mille, S. and Wanner, L. 2012. Perspective-oriented generation of football match summaries: Old tasks, new challenges. *ACM Trans. Speech Lang. Process.* 9, 2, Article 3 (July 2012), 31 pages.

DOI = 10.1145/2287710.2287711 <http://doi.acm.org/10.1145/2287710.2287711>

1. INTRODUCTION

Generation of team sports commentaries is an old task in Natural Language Text Generation (NLG) in that it has been repeatedly a subject of investigation in the field; see, for instance, André et al. [1988]; Binsted [1998]; Tanaka-Ishii et al. [1998]; Voelz et al. [1999] for European football (or soccer), Robin [1994] for American football, Lareau et al. [2011] for Australian football, Barzilay and Lapata [2005]; Oh and Shrobe [2008] for baseball, or Kelly et al. [2009] for cricket. This is not by chance. On the one

This work was supported by the Spanish Ministry of Industry under grant CENIT-2007-1012.

Authors' address: Department of Information and Communication Technologies, Pompeu Fabra University, C/ Roc Boronat 138, 08018 Barcelona, Spain; email: nadjet.bouayad@upf.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2012 ACM 1550-4875/2012/07-ART3 \$15.00

DOI 10.1145/2287710.2287711 <http://doi.acm.org/10.1145/2287710.2287711>

hand, such commentaries constitute a popular genre and there is thus a demand for automation of their production. Also, the availability and richness of the data for experimentation play a role. However, there is an additional characteristic that makes them attractive to NLG research: they may express a more or less subtle bias towards one of the teams. Therefore, they call for techniques that are able to select content and generate wordings that reflect the affinity of the targeted reader to one of the teams and are challenging content selection, discourse organization and ordering, lexical choice, and even the architecture of the generator as a whole. This holds true even if their approaches to text generation are fundamentally different: some generate real-time verbal commentaries from game logs or real-time analysis of the game [André et al. 1988; Binsted 1998; Tanaka-Ishii et al. 1998; Voelz et al. 1999], others aim for generating a summary containing the most relevant events of the game based on a game knowledge base [Barzilay and Lapata 2005; Kelly et al. 2009; Lareau et al. 2011], while others are revision-based and work on an existing text [Robin 1994; Oh and Shrobe 2008]. However, it is not only team sports commentary generation that explores strategies for adapting content and wording to the user's previous knowledge, expertise, interest, and affinity. Among the pioneer generators in this respect are Hovy's Pauline [1988] and Paris' EES [1993]. Newer proposals on affective generation target texts that sound persuasive, appealing, or affective to each specific reader, varying the content, style and words; see, among others, Belz [2003] and de Rosis and Mellish [2007] for overviews.

The data used across the different state-of-the-art works not only differ greatly in nature and organization but also influence the whole approach. For instance, Barzilay and Lapata [2005]; Oh and Shrobe [2008]; Kelly et al. [2009]; Lareau et al. [2011] start from data that capture detailed accounts of match events in terms of logs or databases, often populated from online sources. The data are then aligned with linguistic corpora and used in machine learning (ML)-based or some other numerical method-driven linguistic realization. User model features are not considered. Similarly, rule-based real-time generators, as, for example, André et al. [1988]; Binsted [1998]; Tanaka-Ishii et al. [1998]; Voelz et al. [1999] use momentary data logs such as player location and orientation, ball location, game score, play modes, etc., which are grouped into higher level structures that serve as input to the linguistic generator, again without considering preferences of the target readers. In contrast, Hovy [1988]; Paris [1993]; Robin [1994]; Reiter et al. [2003] start from structured knowledge sources to which different ways of realization are explicitly assigned. These deep knowledge sources provide means for the addition of inferred knowledge not readily available in other data sources (e.g., knowledge on how user perspective is communicated), but are of limited size when compared to other data sources and unflexible due to the predetermined linguistic realization.

With the advances in semantic representation and creation of large scale ontologies over the last decade, such data-driven and knowledge-oriented techniques are no longer satisfactory: techniques are needed that are applicable to general purpose ontologies and large amounts of data, but which still take user preferences into account.

Our proposal attempts to take the best of both worlds in that it starts from raw domain data, but modelled in an application-independent OWL-ontology knowledge-base. It dynamically enriches the starting OWL-ontology by *domain communication knowledge* in the sense of Kittredge et al. [1991] in order to make it a richer and more complete source for generating coherent discourse from different perspectives, and then exploits the richness of the ontology at different levels of generation to generate perspective-oriented commentaries.

In what follows, we present this proposal and show its instantiation for generation of football match summaries of the First Spanish Football League in Spanish, which vary

Getafe-Levante, 2007-2008, Source: Terra *Un gol de Braulio Nóbrega en el último minuto mantiene la racha ascendente del Getafe, que ha sumado trece de los quince puntos disputados tras superar al Levante, en un encuentro envuelto en la amenaza de la zozobra por Iñaki Descarga, que quedó inconsciente tras sufrir dos golpes en la cabeza en el tramo final de la primera parte.*

lit. 'A goal by Braulio Nóbrega in the last minute maintains Getafe's upward streak, who has scored 13 of the 15 disputed points after beating Levante, in an encounter wrapped in anxiety for Iñaki Descarga, who was left unconscious after suffering two blows on the head in the final stretch of the first half.'

Racing de Santander-Osasuna, 2007-2008, Source: Mundo Deportivo *El conjunto cántabro logra una clasificación histórica para Europa tras derrotar por la mínima a un Osasuna que pese a perder, se mantiene en Primera.*

lit. 'The Cantabrian team manages a historical classification for Europe after defeating by the minimum Osasuna, who, despite loosing, stays in the First.'

Cádiz-Deportivo de La Coruña, 2005-2006, Source: ESPN *El argentino Lucas Lobos permitió al Cádiz salvar un punto El Dépor en un partido en el que el cuadro gallego acabó con nueve.*

lit. 'The Argentinian Lucas Lobos allowed Cádiz to save one point in front of The Depor in a match in which the Galician team ended with nine.'

Fig. 1. Samples of human-written football match summaries.

according to user preferences (the preferred team) from data obtained from the Web.¹ The generator has been developed in the framework of a larger project, which involved content extraction, ontology design, reasoning, and other information technologies-related tasks. However, it functions as a self-contained system, which does not use, any external modules apart from the ontology from which we start and the auxiliary module to scrape the data from predefined Web sources. For illustration, Figure 1 shows human-written summaries from different online sources with which we compete.²

In Section 2, we present the ontology from which we start: an application-independent ontology of the football domain. In Section 3, we discuss the extension of this base ontology by a second layer constituted by inferred knowledge and logico-semantic relations that allow, on the one hand, for the selection of one specific perspective on the content, and, on the other hand, the generation of a coherent discourse structure. In Section 4, we outline the individual tasks involved in perspective-oriented summary generation from such a dual layer ontology: content selection, discourse and information structure determination, and linguistic realization. Section 5 assesses the performance of the generator in action. Section 6 contains a brief summary of related work, and Section 7 presents some conclusions and an outline of future work we project in the area of perspective-oriented generation.

2. POINT OF DEPARTURE: THE OWL BASE ONTOLOGY

Given that the application we have chosen for illustration of our model is the generation of football summaries, the discussion in this section revolves around football

¹Note that despite the focus on team sports summaries, the proposed model is domain-independent: for any domain modelled in an OWL-ontology, domain communication knowledge can be inferred and added in terms of a separate extended ontology. The rich ontologies then allow taking into account user preferences from content selection all the way down to information structuring and linguistic realization.

²A summary in these online sources is typically a short paragraph appearing before the main article in a different font and/or layout.

ontologies. However, the argumentation with respect to techniques proposed to extend the application-neutral (or base) ontology is valid for any OWL ontology.

A number of ontologies that deal with sports, and more precisely, with European football (soccer) have already been developed, among them the SWAN Soccer Ontology by DERI,³ the sports fragment of the OpenCyc Ontology [Lenat 1995],⁴ the sports fragments in the DAML repository [Dukle 2003],⁵ and [Tsinaraki et al. 2005]. However, even the most detailed of them, such as SWAN, let alone generic ontologies, such as OpenCyc, still miss a large amount of specific football knowledge entities as referred to in newspaper reports on national football competitions—among others, “header”, “shot on goal” by a player, “block of the ball” by a field player or the goalkeeper, “approximation” of a team to a specific marked zone of the classification, and so on. Therefore, a base ontology has been developed from scratch by our project consortium.⁶

The base ontology describes the Spanish football league domain and is task-independent. It is composed of two different ontologies: an object ontology, which deals with the structural information of the competition (information on teams, competition phases, matches, players, etc.), and an event ontology which deals with information related to the events that happen in the match (information on penalties, goals, cards, etc.). To develop it, we followed the top-down strategy suggested by Uschold and King [1995]: the more abstract concepts are identified first and subclassified then into more specific concepts. This is done to control the level of detail wanted. A known drawback of this strategy is that it can lead to an artificial excess of high-level classes. We achieved a sufficient level of detail for our application domain (the First Spanish Football League) with a moderate number of classes. The object ontology contains 24 classes and 42 properties of the domain in question, the First Spanish Football League; the event ontology, 23 classes and 8 properties.⁷ For illustration of the organization of the base ontology, see Figure 2, which depicts some of the classes in the base ontology. The classes are shown as labeled boxes, while hierarchy relations are marked as arrows; for the sake of simplicity, not all subclasses in the base ontology are shown. Its population is discussed in Section 5.1.

3. EXTENDING THE BASE ONTOLOGY

While our base ontology is more comprehensive than the currently available ontologies, its nature as an application-independent resource implies the absence of a number of concept configurations explicitly or implicitly communicated in the summaries, and, in particular, concept configurations required for flexible, perspective-oriented generation. This additional information was gathered from a preliminary manual analysis of some human-written summaries of the type presented in Figure 1, considering mainly the most common new knowledge that can be inferred from the basic knowledge on the First Spanish Football League.⁸ These are, first of all:

³<http://sw.deri.org/knud/swan/ontologies/soccer>

⁴<http://sw.opencyc.org/2009/04/07/concept/en/Soccer>

⁵<http://www.daml.org/ontologies/374>

⁶The ontology has been developed by the Information Retrieval Group of the Autonomous University of Madrid. The distribution of the ontology and the corresponding knowledge bases is restricted to the i3media project consortium <https://i3media.barcelonamedia.org/>

⁷The top level classes of the object ontology are Competition, Match, Period, Person, Result, Season, Team, TeamCompositionRelation, and Title. The top level classes of the event ontology are ActionFault, Card, Corner, Fault, FaultKick, Goal, GoalKick, Interception, OffSide, Pass, Stop, Throw-in, Shot, and Substitution.

⁸Statistical information about matches within a season and across seasons (best scorer, consecutive wins, first victory in a given stadium, etc.), although mentioned in human-produced summaries, has been excluded for now.

Fig. 2. Organization of the football base ontology.

- (1) configurations inferrible from the explicit facts—for instance, the defeat of the local team and win of the guest team from the match result 2:3, or relegation of the three last teams at the bottom of the classification table to the Second League at the end of the season;
- (2) logico-semantic relations (in the sense of systemic-functional linguistics) between concepts or concept configurations—for instance, the Contrast relation between the placement in the Champion League zone and the placement in the relegation zone of the classification table of the two teams participating in a match, or the implication relation between a goal and a shot on goal.

The extension of the base ontology in these terms is akin to the task of Data Assessment and Interpretation in generators from time-series [Gatt et al. 2009; Wanner et al. 2010].

The most straightforward approach to the extension of the base ontology is to add to it the required concept configurations. This approach has been followed, for instance, by Bouttaz et al. [2011], who add communication-oriented configurations (*policies* in their case) to the *RDF-metastructure* of the *Virtual Research Environment*⁹ ontologies. The same approach can be said to be followed by Bontcheva and Wilks [2004] in the MIAKT ontology, and by Bateman [1997] in the Upper Model design and integration—with the difference that they introduce into the ontologies the necessary configurations when creating the ontologies. The disadvantage of this approach is that it produces an ontology that intermingles application-neutral with application-oriented concept configurations, which is prohibitive when the same ontology serves as a resource for several different applications. Therefore, we follow another approach, namely, the introduction of a second layer (the extended ontology) in which the required configurations are specified: the base ontology model’s domain knowledge (e.g., players’ actions, teams’ classifications, match result), and the extended ontology

⁹<http://ourspace.net>

models domain communication knowledge. In other words, the extended ontology “relates domain knowledge to ... aspects of verbal communication” [Kittredge et al. 1991]: goals become hat-tricks, a team’s position in the league classification becomes ascending or makes the team a leader, minutes become first halves and second halves, a goal implies a temporary draw during a match interval, and so on. This is the kind of information that we aim to provide in the generated summaries in addition to the numerical and nominal data included in the domain knowledge.

3.1. Inferring Concepts for the Extended Ontology

The domain communication knowledge described in the extended ontology is obtained, directly or indirectly, by inference from the base ontology. This is motivated by theoretical and practical reasons. The theoretical reason is that the knowledge in the extended ontology is an interpretation on how the knowledge in the base ontology should best be communicated. The practical reason is that the knowledge of the extended ontology, contrary to the base ontology (see Section 5.1), is not readily available in tabular form and its extraction from texts is still a challenging task [Nédellec and Nazarenko 2005]. When using inferences, the content of the extended ontology can be more readily obtained and controlled. However, we trade then recall for precision in a small subset of knowledge on which we focus, for which we have to devise explicit manually devised inference rules.

We infer new concepts for events and states of a match (goals and expulsions, and results and classifications respectively) using the Jena inference engine.¹⁰ The inferred concepts are divided into five categories: (1) result, (2) classification, (3) set, (4) match time, and (5) send-offs. To facilitate the inference, for each of these categories, one or several classes are defined.

Result-related concepts are inferred from the names of the local and guest teams and the numerical result of the match available in the base ontology. They are captured by the classes “NominalResult” and “CompetitionResult”. “NominalResult” is further subclassified as “Draw” and “VictoryDefeat”. An instance of a “VictoryDefeat” concept can, in its turn, be realized in the text as a reference to the victory of one team or to the defeat of the other team, depending on the perspective taken for the summary.

Classification-related concepts model information related to the position of each team in the competition, its accumulated points and the relative zone. For the zone, in addition to the four official zones Champions, UEFA, neutral, or relegation, we introduce two internal zones—the classes “Lead” and “BottomOfLeague”. It is also of interest to obtain after each gameweek, a team’s tendency (ascending, descending, stable) and distance with respect to its previous classification. The class “Tendency” represents the team’s change of zone in the competition, while the class “Distance” represents a team getting closer to (or further from) a higher (lower) zone. In addition to the real tendency, teams are assigned a virtual tendency, which represents the team’s change of zone taking a virtual result that may be different from the actual match result (for instance, if the team would have drawn instead of winning, what would be the tendency of its classification in the league table).

Set-related concepts model sets of events or processes for a given team in a match or for a given match. This is because we need to be able to talk about events or processes together in accordance with their chronological occurrence (first goal, team was winning then it drew, etc.). For this purpose, we introduce the classes “Set” and “ConstituentSet”. These classes also allow us to simply refer to the number of constituents within it (cf. the team had two red cards).

¹⁰<http://jena.sourceforge.net/inference/>

Fig. 3. Extended goal class.

(Legend: partido=match, nombre=name, equipo=team, enContra=against, aFavor=for, minuto=minute, sujeto=subject, periodo=period, PrimeraParte=first period, SaqueDeMeta=kick-off by the keeper, TiroAPuerta=shot on goal, gol=goal)

Match time-related concepts model the state of the match along its duration, creating intermediate results after each goal. To capture this knowledge, we define the class “IntermediateResult”. Thus, a team could be winning after a goal, even though the final result is a draw. It is also possible to refer to specific reference time points such as “beginning of the match”, and “conclusion of the first period”.

Send-off related concepts include the expulsion of a player after a red card and the number of players left after an expulsion. This is captured by the classes “Expulsion” and “PlayersInField”.

A number of further classes and properties have been added to facilitate the navigation in the ontology for the mapping to linguistic realization and for the inference of new knowledge. For instance, *aFavor* “for”, and *enContra* “against”, properties were added to the “Goal” class (see dashed arrows in Figure 3) in order to know the team that scored and the team that was scored against respectively, given that this information is only indirectly available in the base ontology via the player marking the goal. This is illustrated in Figure 3, where dashed arrows indicate the inferred properties while continuous arrows correspond to properties in the base ontology.

We are aware that our approach to the creation of the extended ontology is not without its limitations. First, the manual elicitation of domain communication knowledge from experts is a difficult task and many concepts and inference rules can be missed. Other approaches, such as Gatt et al. [2009] and Lareau et al. [2011], use statistical analysis of time series. Second, the modeling of some concepts that can be inferred, such as “PlayersInField”, is unsatisfactory from a knowledge representation point of view.

3.2. Adding Logico-Semantic Relations to the Extended Ontology

Logico-semantic relations between concepts and concept configurations are needed for the generation of a coherent discourse in that they form the basis for content selection and the determination of the rhetorical structure of the summary to be generated [Kosseim and Lapalme 2000; Wanner et al. 2010; Bouttaz et al. 2011]. For instance, if a team scored five goals and a player X from this team scored two, we must capture the fact that X’s two goals are a subset of the team’s five (*meronymy* relation); if a player has been sent off, we must capture that his team is left with ten players (*implication* relation); and so on. Table I lists the relations implemented in the current version of our extended ontology.

Table I. List of Logico-Semantic Relations and their Inferred Instantiations in the Extended Ontology

Relation	Arguments	Informal Conditions	Verbalization
Cause	Last goal + final result	There are two or more goals $\wedge$ the last goal is scored within the last 20min $\wedge$ changes the nominal result	<i>The match ended with team A drawing against team B thanks to the goal by A's player X.</i>
	Tendency + final result	Gamesweek $\neq 1 \wedge$ the tendency is either ascending or descending	<i>Thanks to this victory, the team ascends in the classification.</i>
	Distance + final result	Gamesweek $\neq 1 \wedge$ tendency is stable $\wedge$ distance is not static	<i>Thanks to this victory, the team is getting close to securing their position in the league.</i>
	Set of goals by a team + final result	Nominal result is not a draw	<i>Thanks to the goals of players X and Y, the team won the game.</i>
Violation-of-expectation	Instance of PlayersInField + final result	Number of PlayersInField < 11 for a team $\wedge$ final result is a win for that same team	<i>Despite playing with 10, the team won.</i>
	The intermediate result prior to the last goal + final result	There are two or more goals $\wedge$ the last goal is scored within the last 20min $\wedge$ the last goal changes the nominal result	<i>Although it was losing, the team ended up drawing.</i>
	Tendency + final result	If team is winner (or loser) $\wedge$ tendency for that team is descendent (or ascendent).	<i>Despite winning, the team has gone down to the UEFA zone.</i>
	Constituent with order = 1 of the goal set for the whole match + final result	The losing team is identical with the team who marked the first goal	<i>Despite marking the first goal, the team has lost.</i>
	Team's classification + final result	Gamesweek $> 4 \wedge$ (classificationTeam1.position – classificationTeam2.position $\geq 10 \vee$ (zoneTeam1 $\geq$ Champions $\wedge$ zoneTeam2 $\leq$ Relegation)) and Team2 won	<i>Despite being in the relegation zone, the team won.</i>
	Contrast	Classification distances + tendencies or positions of the teams match	<i>While team A reaches European positions / is Leader, team B descends into relegation zone / is at the bottom of the table.</i>
Equivalence	Red card + corresponding send-off	There is a red card	<i>He was shown a red card and sent away by the referee.</i>
Implication	Send-off + instance of PlayersInField	There is a send-off	<i>The team was left with ten players after the expulsion.</i>
Negative cause	Final nominal result + instance of virtual tendency	Gamesweek $\neq 1 \wedge$ virtual tendency $\neq$ stable $\wedge$ virtual tendency $\neq$ actual tendency	<i>Thanks to this victory, the team won't descend.</i>
Meronymy	Set of goals + subset of its constituents	Number of goals ≥ 2	<i>The team scored 5 goals, 2 by player X.</i>
Precedence	Pairs of temporally ordered constituents in a set of events for the whole match and for each team	There is a goal in the first and a goal in the second half	<i>After scoring a first goal in the first half, team A scored another one in the second half.</i>

The relations are domain-independent, but the types of concepts they can link reflect how content is communicated in football summaries. The football summary-specific restrictions are dictated by the set of rules used during inference. For instance, a “Cause” relation is added between an instance of a goal and an instance of the result of a match if this goal changed the result from a draw to a win for one the teams. This particular instance of the “Cause” relation reflects the fact that some summaries mention a specific goal because it flips the result of the match.

Fig. 4. Representation of a Cause relation between a goal and a nominal result in the ontology.

Contrary to Mann and Thompson’s [1987] rhetorical relations, logico-semantic relations are pre-realizational: they do not take into account the intended reader’s previous knowledge and perspective or the information structure, and can be mapped to different rhetorical relations depending on the factors just cited. Rhetorical relations (see Section 4.1) on the other hand, may have a specification of the effect that they intend to have on the addressee; in other words, they have an intentional load [Moore and Paris 1993]. Thus, some logico-semantic relations can be used from the perspective of both teams—as, for example, Violation-of-expectation (perspective, pro team A: *Team A scored the first goal but lost, after all* vs. perspective, pro team B: *Although team A scored the first goal, team B won/drew against it*), depending on the type of data that they relate. Likewise, the effect of a logico-semantic relation can be dampened by mapping it to a weaker rhetorical relation. For instance, Cause can be mapped to a rhetorical cause for the realization of the perspective pro team A: *Thanks to the goals of the players X and Y, team A won*. But it can be also mapped to a temporal circumstance (*With the goals of the players X and Y, team A won*), if a more neutral perspective is desired.

In the extended ontology, the logico-semantic relations are captured as subclasses of the class “LogicoSemanticRelation”. Figure 4 shows a fragment of the base ontology and its extension by the Cause relation, which holds between a nominal result and a goal.

4. GETTING THE PERSPECTIVE ACROSS

With the extensions described in the previous section, the ontology is rich enough to serve as input to a generator that gets across to the reader the desired perspective on the content.

The generator proper consists of three main pipelined modules.¹¹

- (1) *Text planning*. Content selection (content bounding, and content evaluation and extraction, which is preceded by a technically motivated conversion of OWL to an internal graph representation format), and discourse structuring, that is, mapping

¹¹In generators that start from time series [Sripada et al. 2003; Gatt et al. 2009; Wanner et al. 2010], data assessment and interpretation (roughly, our ontology extension task) is considered part of the generation process. In what follows, we do not consider it a genuine generation task.

of logico-semantic relations to discourse relations, discourse unit determination, and ordering.

- (2) *Information structuring*. Mapping the discourse plan onto a conceptual structure [Sowa 2000] required as input by our linguistic generator, theme/rheme arrangement according to user profile and sentence packaging.¹²
- (3) *Linguistic generation*. Casting of the structures provided by the information structuring module into linguistic constructions, that is, a summary in Spanish.

4.1. Text Planning

As mentioned in the preceding section, the content encoded in the dual-layer ontology is perspective-independent; but it supplies the text planning tasks with the means needed to realize perspective in accordance with a user model and in a way that mimics how the content is communicated in summaries. The two text planning tasks, content selection and discourse structuring, are thus perspective-driven. Content selection does the coarse-grained selection of the content to be communicated in the summaries. Discourse structuring refines this selection and determines the discourse structure.

4.1.1. Content Selection. The content selection strategy relies on the logico-semantic relations in the extended ontology and on the chosen user perspective. It starts from the entire content stored in the two-layer ontology and selects a subset of this content for perspective-oriented communication in the summary. This is done in the following two stages:

- (1) preselection of the fragment of the ontology that constitutes the maximal set of data available for generating any sort of summary for a given match;
- (2) determination of the content that is of relevance to the chosen user perspective on the match.

The motivation of the first stage is eminently practical: given the large size of the ontology, it is convenient to filter out irrelevant content and thus make subsequent tasks more manageable. The task is not trivial since we cannot discard, for instance, any content that may serve as reference within the information on the match, although it has no direct relation to the match—as, for example, the ranks of the teams involved in the match in the classification prior to the match. Therefore, our solution is to realize the first stage by a limited set of seven domain-specific rules, which select from the ontology the content related to the match for which the summary will be generated. Each rule selects the content related to a specific topic elaborated in the summary. For instance, the rule related to classification selects:

- (1) the instances of the class “Classification” for the gameweek in which the match took place and for the previous gameweek;
- (2) the instances of the class “TeamClassification”, one for each team playing the match and each of the two selected gameweeks;
- (3) the instances of the class “Tendency”, one for each classification instance of each team (e.g., ‘ascendent’, ‘descendent’, or ‘stable’);
- (4) the instances of the class “Distance”, one for each classification instance of each team (e.g., reducing, extending, maintaining distance);
- (5) the instances of the “ClassificationZone” class, related to the classification instances of each team and to the tendencies or distances (e.g., Champions, UEFA, League, etc.); and

¹²Since the discourse plan–conceptual structure mapping is a purely practical task, we do not discuss it further.

- (6) the instances of the class “Position” (position in the classification table) for each team. In addition, all logico-semantic relations whose arguments are among the selected concept configurations are also selected.

In the second stage, the relevance of the content preselected in the first stage is assessed. This is done by taking into account (1) the preference of the user for a team, (2) heuristics, and (3) the logico-semantic relations that link individuals in the preselected parts of the ontology. Nodes linked to the user’s team of interest are considered relevant. When the user profile is set to “neutral”, nodes linked to any of the teams are considered relevant. Heuristics are available only for instances of selected classes of the ontology for which classifiers were trained. By making relevant the arguments of logico-semantic relations, we can be sure that the generated discourse is coherent.

The heuristic weights of the nodes are obtained by supervised learning trained on a corpus of semiautomatically aligned data and online articles. For this, we focused on the ontology classes “TeamClassification”, “RedCard” and “Goal”, training a classifier for each. The corpus consists of articles of eight seasons of the Spanish League, from 2002/2003 to 2009/2010, with a total of 3040 matches downloaded from different Web sources. The data categories for each match include the teams, stadium, referee, players, major actions such as goals, substitutions, red and yellow cards, and some statistical information such as number of penalties. The purpose of this aligned corpus is to provide the basis for the extraction of a rich set of features.

For modeling “TeamClassification” of a team, we used a total of 760 feature types obtained by contrasting the team’s classification details with the details of the other teams in the competition of the gameweek under consideration and in the previous gameweek. For the description of “RedCard” and “Goal”, we used a set of over 100 feature types, which include information about the event (minute, event number in the match), the player involved (name, position, proportion of goals/cards in the match and in the season up to the match, proportion of games played in the season up to the match, etc.), the game, gameweek, season and team (including classification and statistical information), and comparison of the current event with previous and next events of the same category. For both “Goal” and “RedCard”, the same features are used since we consider both to be match events.

In order to classify the data, we used Boostexter [Schapire and Singer 2000], a boosting algorithm that uses decision stumps over several iterations and that has already been used in previous works on training content selection classifiers [Barzilay and Lapata 2005; Kelly et al. 2009].¹³ For each of the three categories (“TeamClassification”, “RedCard”, “Goal”), we experimented with 15 different classifiers by considering a section dimension (title, summary, and title + summary) and a source dimension (ESPN, Marca, Terra, any one of them (any), and at least two of them). We divided the corpus each time into 90% of the matches for training and 10% for testing.

Table II shows the performance of the classifiers for the determination of the relevance of the three categories (“TeamClassification”, “RedCard”, “Goal”) with respect to their inclusion into the summary section, comparing it to the baseline, which is the majority class. For red cards, the results correspond to considering title and summary from a source together, given that the results are not significant when considering the summary section only (accuracy is 78.1%, baseline accuracy is 65.4%, and $t = 4.4869$ with $p < 0.0001$). In all cases, the best performance is obtained by considering the content from any of the online sources. See Bouayad-Agha et al. [2011a; 2011b] for more details.

¹³After a number of experiments, the number of iterations was set to 300.

Table II. Performance of the Best Classifiers (vs. Majority Baseline) for Content Selection on a Test Set for the Summary Section (+Title in Case of Red Cards)

category	source	sample size	classifier	baseline	paired t-test
goal	any	1123	64%	51%	t = 6.3360 (p < 0.0001)
	terra	1121	65%	59%	t = 3.4769 (p = 0.0005)
card	any	62	85%	53%	t = 4.4869 (p < 0.0001)
classif	any	295	75%	61%	t = 4.4846 (p < 0.0001)

Tendency = {point_of_reference1>Zone,
point_of_reference2>Zone,
point_of_reference1>Position,
point_of_reference2>Position,
team>Team,
gameweek>Gameweek}

Goal = {team>Team,
in_favor>Team,
against>Team,
subject>Player,
period>PeriodMatch}

Fig. 5. A discourse unit determination template for Tendency and for Goal (for convenience of the reader, Spanish terms have been substituted by their English translations).

4.1.2. Discourse Structuring. The task of discourse structuring consists of three sub-tasks: (1) discourse unit determination, (2) discourse unit ordering, and (3) mapping of logico-semantic relations selected during the content selection stage to discourse relations.

Discourse unit determination completes the process of content selection. It is realized in terms of domain-specific configuration templates. A template is defined for each concept in the ontology that can be an argument of a logico-semantic relation; consider Figure 5 for illustration.¹⁴ The purpose of the templates is to group facts that are closely related to the concept in question. For instance, it has been observed in target texts that mentions of a goal are usually accompanied by references to the player who scored the goal, the team that benefited from the goal, and the minute when the goal was scored. Consequently, a template for goal instances includes facts concerning the goal itself, the scorer, the team, and the minute of the goal. Each template is also designed such that the discourse unit resulting from its application is sure to contain enough content to produce a meaningful sentence out of it during the subsequent linguistic generation.

Content preselected in the content selection stage but not included in at least one discourse unit is removed from the text plan.

Discourse unit ordering and the mapping of logico-semantic relations to discourse relations are mechanisms for further refinement of the perspective from which the summary is to be generated. In the course of discourse unit ordering, the created discourse units are sorted in the order in which they will appear in the text. A simple user-oriented ordering rule is applied: units containing content related to the user's preferred team are placed before units whose content is related to the other team or to both teams; see Figure 6. Note that the order defined by this task affects discourse units but not the content within the individual units. The content within each unit is left unspecified at this stage: it is the modules of information structure packaging and linguistic generation that later specify the content elements within each discourse unit in accordance with the user profile and linguistic constraints.¹⁵

¹⁴The templates in Figure 5 follow the syntax $\text{CoreConcept} = \{\text{objectProperty1} > \text{Concept1} > \dots > \text{objectPropertyN} > \text{ConceptN}\}$, which allows the inclusion of ConceptN in the unit about CoreConcept, following the path from CoreConcept to ConceptN in the ontology.

¹⁵This gradual ordering of the selected content reflects how the user model guides all tasks in the generation process to progressively confer a perspective to the generated text.

Fig. 6. A sample discourse plan constructed by the discourse-structuring module.

During the mapping of logico-semantic relations to discourse relations, the arguments of each logico-semantic relation are mapped onto the arguments of rhetorical relations in the sense of the Rhetorical Structure Theory [Mann and Thompson 1987]: the first argument of the logico-semantic relation is mapped to the nucleus, the second to the satellite role of the corresponding rhetorical relation. In our prototypical implementation, we assume a one-to-one correspondence between logico-semantic and rhetorical relations. For example, a Cause relation holding between a set of goals of a team and a victory of that team is mapped onto a *VolitionalCause*, where the victory is the nucleus (indicating that this content is more prominent according to the relation) and the set of goals is the satellite, that is, secondary to the nucleus. During linguistic generation, this discourse relation is signalled, for instance, by the marker *gracias a* (thanks to).

4.2. Information Structuring

The information structure, and in particular the dimension of *thematicity* in the sense of Mel'čuk [2001], is crucial for perspective-based generation.¹⁶ Thematicity is defined in terms of three parameters: *theme*, *rheme* and *specifier*. Theme marks what a statement under generation will be about. Rheme determines what will be said about theme. Specifier sets out the context in which rheme is stated about theme.

In sports commentaries written in an accusative language (of which Romance and Germanic languages are examples), the specifier will be by default realized in terms of a sentential circumstantial, theme as grammatical subject, and rheme as the verbal phrase of the sentence. For instance, in (1a) *Despite the expulsion of Samuel Eto'o during the first half is the specifier, Almería is theme and lost 1-2 against FC Barcelona*

¹⁶There is a substantial literature in linguistics on the information structure and, in particular, on theme and rheme and related terms: topic and focus [Sgall et al. 1986], topic and comment [Gundel 1988] and ground and focus [Vallduví 1990]. See Lambrecht [1994] or Mel'čuk [2001] for introductions.

Fig. 7. Illustration of the thematicity in a conceptual structure.

is rheme, In (1b), *Almería* is again theme, and *lost 1-2 against FC Barcelona despite Eto'o's sending-off during the first half* is rheme. No specifier is available in (1b).

- 1(a). *Despite the expulsion of Eto'o during the first half, Almería lost 1-2 against FC Barcelona.*
- 1(b). *Almería lost 1-2 against FC Barcelona despite Eto'o's sending-off during the first half.*

When the statement in question is involved in a discourse relation such as Cause, the causee hosts the theme and the rheme, and the causer is considered specifier. Similarly, in a relation such as Violation-of-expectation (see Figure 6), the fact that violates the expectation will host the theme and the rheme, and the fact that creates the expectation is considered specifier. In other words, in the presence of a discourse relation, our generator always produces sentences like (1a), and (1b) otherwise. To be able to generate this variety of sentences, we need to introduce the thematicity parameters into the conceptual structure as produced by the text planning module as illustrated in Figure 7.

In accordance with the semantics of the parameters of thematicity, we define the following five rules for their introduction into the conceptual structure.

- (1) IF a node N is an element of a discourse relation,
THEN MARK N as specifier.
- (2) IF the user has a preference for one of the teams
AND the user's favorite team t_f is part of a statement S ,
THEN MARK t_f as theme.
MARK $S \setminus t_f$ as rheme.

This rule captures the information structure in Figure 7, which leads to the sentence: *Despite the expulsion of Eto'o, Almería lost 1-2 against FC Barcelona.*

- (3) IF the user has no preference for any of the teams,
 THEN MARK the node *ResultadoPuntual* ‘numerical result’ as theme
 AND
 IF a node n in the statement S is not yet marked as theme or specifier,
 THEN MARK n as rheme.

This rule does not choose any of the teams as theme. Since a full active sentence requires a theme (otherwise, it is ungrammatical), the most neutral concept of the statement that links the mentioned teams (namely numerical result) is determined as theme. The neutrality scale that we use for this purpose is (in decreasing neutrality): result > draw > victory > defeat.

The application of this rule to the conceptual structure in Figure 7 will lead to an information structure resulting in the sentence *In spite of the expulsion of Eto’o, the game ended with the victory of FC Barcelona over Almería (1–2)*.

- (4) IF team t_1 or team t_2 (but not both) appear in the statement S ,
 THEN MARK this team t_α as theme.

That is, if only one team appears in the statement that is to be verbalized, this team is chosen as theme. This is because we assume that if this team has been singled out during content selection, it is because there is interesting information to be communicated about this team.

- (5) IF team t_1 and team t_2 appear in the statement S
 AND t_1 appears in the subtree s_1 and t_2 appears in the subtree s_2 (with $s_1 \neq s_2$),
 THEN
 IF the result is a victory/defeat
 THEN MARK as theme the victorious team
 ELSE MARK as theme the guest team.

This rule states that if both teams are mentioned in a statement but in different subtrees, there is no neutral concept that links them. Then, we have to choose more or less arbitrarily which of the two teams will be highlighted as theme in the sentence. In the current implementation, we highlight the team to which something positive occurs, that is, either a victory or a draw on another team’s field.

4.3. Surface Realization

Our surface realizer, which has been developed using the graph transducer workbench MATE [Bohnet et al. 2000], is based on the multistratal model of the Meaning-Text Theory (MTM) [Mel’čuk 1988]. In an MTM, the generation consists of a series of mappings between structures of adjacent strata (from the conceptual stratum to the linguistic surface stratum); see Figure 8.¹⁷ From this perspective, our realizer is quite similar to the SURGE generator [Elhadad and Robin 1996], even if, for instance, syntactic and lexical paraphrasing, ordering, and agreement resolution are not performed at the same points in the pipeline. Thus, in an MTM, the lexicalization starts during the projection of the language-independent conceptual structure onto the language-specific semantic structure in that concepts are mapped onto semantemes.

¹⁷For each pair of adjacent strata S_i and S_{i+1} , a transition grammar $\mathcal{G}_{i+1}^i$ is defined; for the formal background of $\mathcal{G}_{i+1}^i$ and the syntax of the rules in $\mathcal{G}_{i+1}^i$, see, e.g., Lareau and Wanner [2007] or Wanner et al. [2010].

Fig. 8. Surface-realization in an MTM-based generator.

victory = {default = ganar 'win _V '	cause = {default = causar 'cause _V '
agent = ganar 'win _V '	positive = causer-positive
patient = perder 'lose'	negative = causer-negative

Fig. 9. Sample entries in the conceptual dictionary.

ganar = {polarity = positive	causar-positive = {lex = causar ‘cause _V ’
lex = victoria ‘victory _V ’	lex = gracias ‘thanks-to’
lex = vencer ‘defeat’	lex = ...
lex = ganar ‘win _V ’	
lex = derrotar ‘defeat _V ’	
lex = ...	

Fig. 10. Sample entries in the semantic dictionary.

The main means for the realization of the user perspective on the content encoded in a conceptual structure is *lexicalization*: semantemes and lexemes predetermine the syntactic structure of an utterance; they also possess a neutral, negative, or positive polarity. Each concept can, as a rule, be expressed by several alternative semantemes. Therefore, for each concept, lists of possible semanticizations are compiled and each of the semantemes is tagged with its polarity (in the semantic dictionary); in the entries of the conceptual dictionary, some mappings to semantemes are restricted with respect to communicative goal and/or polarity. Figure 9 shows the entries for the concepts VICTORY and CAUSE in the conceptual dictionary. Figure 10 shows the entries for the meanings 'victoria' 'victory', and 'causar-positive' 'cause-positive' in the semantic dictionary.

During the conceptual-semantics mapping, rules check the configuration of the conceptual structure and introduce the corresponding semanteme into the semantic structure. The semanticization depends on the theme of the statement and the polarity of the semantemes available to express a concept. For instance, in the case of the concept VICTORY, if the winner is theme, the concept will be mapped to the semanteme 'ganar'

Fig. 11. Sample Con-Sem rule (thematicity criterion).

‘win’ (see the rule in Figure 11); if the winner is in the rheme, the concept will be mapped to the semanteme ‘perder’ ‘lose’.

If a concept can be rendered by a positive and a negative polarity semanteme (as, e.g., CAUSE), and the concept has a direct relation in the conceptual structure with another concept which is realized by a semanteme that has a polarity (as, e.g., VICTORY, possibly realized as ‘ganar’ ‘win’), the concept will be mapped onto ‘gracias a’ ‘thanks to (the victory)’. If it can be rendered only by a negative polarity semanteme, the concept will be mapped onto ‘debido a’ ‘because of (the defeat)’. If nothing constrains the semanticization of a concept, an associated semanteme which is neither positive nor negative is chosen; if none is available, we first look for a positive semanteme, the negative semanteme being the last-resort option.

In the rule in Figure 11, three distinct spans are shown: leftside (what the rule maps in the input structure), rightside (what the rule builds in the output structure), and conditions. Variables appear with a question mark. The rule states that if in the input structure a node ?Xl possesses a relation ?r which is ‘agent’ or ‘patient’ with another node ?Yl which is the user’s team, and if this relation is described in the entry for ?Xl in the conceptual dictionary, then we introduce a specific name (as stated in the dictionary) to the node corresponding to ?Xl in the output structure.

For more details on our realizer and how the different mappings that produce the final text are performed, see Wanner et al. [2010].

5. EXPERIMENTS

Our generator relies on a knowledge base prepopulated with data crawled from sports Web sites and extended through inference to incorporate domain communication knowledge. To successfully generate a summary of a match, the linguistic generator requires that data for that match be present in the knowledge base and that it has been extended beforehand. Therefore, to test the potential of the extended ontology model and the performance of our generator, we performed some experiments that involved the population of the ontologies, generation of football summaries, and evaluation of the quality of the obtained results.

5.1. Ontology Population

The population of the extended ontology uses a set of rules written in the Semantic Web Rule Language (SWRL) and interpreted using the inference engine provided by Jena.¹⁸

¹⁸<http://jena.sourceforge.net/>

```
[rule2: (?rn rdf:type base:NumericResult)
(?rn base:localScore ?localScore) (?rn base:visitorScore ?visitorScore)
(?rn base:localScore ?localScore) (?rn base:visitorScore ?visitorScore)
(?localScore base:result ?local) (?visitorScore base:result ?visitor)
differenceAbs(?local, ?visitor, ?r) ge(?r, 3)
-> (?rn inference:resultDifference "important")]
```

Fig. 12. Sample knowledge inference rule.

The ontologies described in Section 2 were first automatically populated with data scraped from Web pages on the Spanish League seasons (general information about competitions, players, stadiums, etc., and specific information about matches) and expanded then into the extended ontology applying the inferences as described in Section 3. The scraping was done by wrappers written specifically to extract all the information required by the classes defined in the base ontologies from the tabular data available in Football Web portals.¹⁹ These data were then automatically imported into the ontology. An informal evaluation of the ontology data thus obtained showed a high accuracy. Currently, the ontology is populated with three seasons of the Spanish League: 2007/2008, 2008/2009, and 2009/2010, with 4041 instances of the object ontology and 63623 instances of the event ontology.

The Jena inference engine works with a set of user-defined rules that consist of two parts: head (the set of clauses that must be accomplished to fire the rule) and body (the set of clauses that is added to the ontology when the rule is fired). We defined 93 rules, with an estimated average of 9.62 clauses per rule in the head part. Consider an example of a rule for classifying the difference between the scores of the two teams as important if it is greater than or equal to three in Figure 12. The inference rules are organized into five groups corresponding to the five categories of inferred knowledge described in Section 2.

For the 38 gameweeks of the regular football season, the inference engine generates from the data in the base ontologies, a total of 55,894 new instances.

5.2. Generation of Football Summaries

Based on the populated ontology, texts for a large number of matches were generated with the three possible settings of the user perspective. Consider three sample summaries for all three different user perspectives in Figure 13. The first summary reflects the perspective of a fan of the club Depor(tivo), the second of a fan of Mallorca, and the third reflects a neutral perspective. Further generated summaries and their human-written equivalents extracted from the sports press are presented in Appendix A.

5.3. Evaluation

To evaluate the performance of our generator, we performed an evaluation of content selection and of the linguistic generation.

5.3.1. Evaluation of the Content Selection. As Belz and Reiter [2006] point out, automatic evaluation of NLG systems by comparison with human-written texts is an attractive approach, as it is quicker and cheaper than human-based evaluation, and also repeatable. The possible disadvantage is that “generated texts can be very different from a corpus text but still effectively meet the system’s communicative goal.”

¹⁹Object and event information were extracted from the Sportec (<http://futbol.sportec.es>) and AS (<http://www.as.com/futbol>) portals respectively.

[Perspective: Deportivo] *Derrota del Depor. El Depor perdió ante el Mallorca por 1-0 y acabó el partido con 10 jugadores por culpa de la expulsión de Riki en el minuto 93. En la trigésima jornada, se enfrentará al Murcia.*

lit. ‘Defeat of Depor. Depor lost against Mallorca by 1-0 and finished the match with 10 players because of a send-off of Riki in the minute 70. In the 30th gameweek, Depor will face Murcia.’

[Perspective: Mallorca.] *Victoria del Mallorca. El Mallorca ganó contra el Depor por 1-0 gracias a un gol de Arango en el minuto 22. Gracias a esta victoria, el Mallorca se distanció de la zona de descenso. En la trigésima jornada, se enfrentará al Valencia.*

lit. ‘Victory of Mallorca. Mallorca won against Depor by 1-0 thanks to a goal by Arango in the minute 22. Thanks to this victory, Mallorca moved away from the relegation zone. In the 30th gameweek, Mallorca will face Murcia.’

[Perspective: Neutral.] *Victoria del Mallorca sobre el Depor. El partido acabó con la victoria del Mallorca sobre el Depor por 1-0. El Mallorca se distanció de la zona de descenso, mientras que el Depor se mantuvo en la zona de permanencia. En la trigésima jornada, el Depor se enfrentará al Murcia y el Mallorca al Valencia.*

lit. ‘Victory of Mallorca over Depor. The match ended with the victory of Mallorca over Depor by 1-0. Mallorca moved away from the relegation zone, while Depor stays in the middle zone of the table. In the 30th gameweek, Depor will meet Murcia and Mallorca will face Valencia.’

Fig. 13. Three sample football summaries of the same match as produced by our generator for different user perspectives.

The evaluation of the content was done by comparing the content of generated summaries with the content of human-written summaries (the gold standard). Our test corpus consists of 36 randomly selected matches from the set of matches of the 2007–2008 season, each with three associated summaries from three different Web sources (namely ESPN espndeportes.espn.com, Marca www.marca.com, and Terra deportes terra.com).²⁰ We compiled a list of all facts considered for inclusion in the content selection and discourse unit determination modules, for which explicit references could be found in target texts, including instances of the logico-semantic relations, which were modelled as classes in the ontology. For each of the 108 (36×3) summaries, we manually annotated whether a fact was verbalized or not. We also annotated for each text its affinity to one of the teams, by checking whether the majority of the content units was rendered from the perspective of one of the teams. In case of equality, the user profile was considered neutral. This allowed us to compare the generated text of a given match for a given profile with the gold standard text(s) for the same profile. As a baseline, we always selected both teams and the final result regardless of profile since the result and the associated teams are very often included in the summaries. This baseline is likely to have high precision and lower recall.

We performed three runs: (1) a full run with relevance weights determined by the trained models (see Section 4.1.1); (2) a run in which the relevance of the instances is determined from the aligned texts, taking the profile into account; and (3) a run like (2), but without taking into account the user profile when determining relevance.

Table III shows the results of the evaluation for each of the three sources; “estimated” stands for run (1), “real w., prof.” for run (2), and “real w., no prof.” for run (3). Precision and recall are obtained by measuring the facts included in the content plan by the estimated or baseline model against the facts mentioned in the gold standard. The recall is predictably lower in the baseline than in the other runs. The F-measure in

²⁰The 2007–2008 season was neither used for the development of inference rules (Section 5.1) nor for the development and testing of the system in general.

Table III. Content Selection Evaluation Results

sc	#	baseline			estimated			real w., prof.			real w., no prof.		
		prec.	rec.	F1	prec.	rec.	F1	prec.	rec.	F1	prec.	rec.	F1
e	157	83.3	57.3	67.9	43.2	77.1	55.4	42.5	79.6	55.4	35.1	85.4	49.7
m	74	49.0	63.5	55.3	21.8	79.7	34.2	20.2	79.7	32.2	17.7	90.5	29.6
t	223	98.1	47.5	64.0	54.2	64.1	58.7	56.1	65.9	60.6	44.8	75.8	56.3

(“sc”=“source” (of the individuals), “#”=“number of individuals”; “e”=“espn”, “m”=“marca”, “t”=“terra”)

the source Marca is considerably lower for the three runs than for the baseline. This is because the summaries in this source are very much like short titles (for Marca, we had an average of 2 facts mentioned per summary vs. 4 for ESPN and 6 for Terra). The performance of content selection with empirically estimated relevance is comparable to the performance of content selection with relevance taken from the target texts—which indicates that there are benefits in using supervised learning for estimating relevance. The runs without profile understandably have a higher recall since content selection is less discriminative without a user profile (or rather with a neutral user profile). Nonetheless, they show a somewhat lower F-measure than those with profile, especially for the two sources with the longest summaries.

In general, we can state that the precision in all three runs is considerably lower than the baseline. This is partially because high precision lies in the nature of the baseline. Nonetheless, we still need to improve the precision of our model by investing more effort into feature engineering. On the other hand, the recall in the three runs is significantly higher than the baseline. In other words, our content selection module prefers to say more at the risk of saying irrelevant things.

This evaluation also shows that perspective is, together with an empirically determined relevance score and the coherence of the selected content, a decisive criterion in the selection of content—although our modeling of the perspective is still very coarse in that perspective is a binary measure based on the node’s direct link to the team of interest, and relevance scores were only determined for three broad categories subsuming various subcategories of content, for example, the classification category subsumes tendency, distance, position, and zone (see Section 4.1).

5.3.2. Evaluation of the Linguistic Generation. To assess the quality of the linguistic generation, a qualitative evaluation has been performed. Ten evaluators previously not involved in the project participated in the evaluation. The evaluators were given 51 randomly selected texts produced by the generator. Each evaluator assessed between 10 and 30 texts, such that each text was evaluated by three evaluators.

The evaluators were asked to assess the texts along two dimensions, Intelligibility/Grammaticality and Fluidity.

—Intelligibility/Grammaticality [1. . . 5]

5/5: The meaning of all sentences is crystal clear; grammar and/or lexical choices are all appropriate.

4/5: The meaning of all sentences is clear, but there are minor problems in some grammatical and/or lexical choices.

3/5: The meaning of most sentences is clear, but some parts are unclear because of grammar and lexical choice problems.

2/5: Some sentences contain serious grammatical and lexical choice problems, and the evaluator can only guess the meaning after a careful study, if at all.

1/5: The sentences cannot be understood at all; no amount of effort will produce any meaning.

—Fluidity [1. . . 5]

5/5: The text is very easy to read; it seems perfectly natural.

Fig. 14. Results of text quality evaluation (Black=Grammaticality, White=Fluidity).

Fig. 15. Distances between highest and lowest ratings in text quality evaluation (Black=Grammaticality, White=Fluidity).

- 4/5: The text is easy to read, although there are some details that seem unnatural.
- 3/5: The text is not too difficult to read, but there are annoying repetitions or abusive agglutination of information in the same sentence.
- 2/5: The text is difficult to read, but it is still worth an effort.
- 1/5: The text is not readable.

In Figure 14, the added-up results are shown for each dimension, giving a maximum grade of 15 for each of the 51 texts.

The Intelligibility and Grammaticality of the texts is good, with an average 4.52/5. There are actually only 6 texts for which the average rating is below 4/5 (12/15). The numbers are quite similar with respect to Fluidity, with an average rating of 4.42/5 and only 7 texts rated below 4/5. The results indicate that our generator produces few ungrammatical sentences, and that the quality of the output is generally more than acceptable. We also observe a certain expected correlation between Grammaticality and Fluidity.

To capture the interevaluator agreement rate, we counted how many times the majority of the evaluators (two or all three) coincide in their judgements. There are 51 texts and two dimensions, which means 102 ratings, each made by three evaluators. With our 5-level scale, we found out that in 41 cases (40.2%), all three evaluators coincide, in 58 cases (56.9%), two coincide, and only in 3 cases (2.9%), their opinions differ. In most of the cases, the discrepancy is between whether a text is perfect or good, that is, whether to rate it 5 or 4. On a common three-value evaluation scale ('good' vs. 'acceptable' vs. 'bad'), these discrepancies disappear: in 82.4% of the evaluations, all three evaluators make the same judgement, and in the remaining 17.6%, two evaluators coincide.

Figure 15 shows for each text, the difference between the highest grade and the lowest grade with respect to the 5-level scale. In total, 62 out of 102 ratings show a discrepancy between the evaluators. But only for 10 ratings do the judgements vary 2 or 3 points on the scale; 6 times concerning fluidity, and 4 times concerning grammaticality.

In a separate evaluation, we let the evaluators judge the perspective orientation of the generated texts. For this purpose, we removed 3 of the 51 texts used in the preceding evaluation because their poor quality would have made an objective evaluation

Fig. 16. Results of the user perspective identification (Black=Neutral, Dark Grey=Non-neutral, Light Grey=Discrepancy between real and interpreted perspective).

Fig. 17. Results of the evaluation of missing information with respect to perspective (Dark=texts are complete, Light=information missing).

very difficult. Seven evaluators were asked to look at each of the remaining 48 texts and indicate whether it expresses affinity to one of the teams, and if yes to which one. Furthermore, they were asked to indicate whether any information likely to be considered important by a fan of the corresponding team was missing. Figure 16 shows how many evaluators identified the intended perspective of a generated text (Team A, Team B, Neutral).

This evaluation shows that in 100% of the cases, the neutral perspective was confirmed by all seven evaluators. In the case of texts that are supposed to express affinity to one of the teams, there are a few deviating ratings, but with 6% the global deviation rate remains very low. The judgements of the evaluators coincide in 79.2%–100% with the intended affinity. All evaluators assigned to 33 texts the affinity intended by the generator; for 11 texts, only one of the seven annotators disagreed, and for the 4 remaining texts two or three annotators did not share the views of the others. The high percentages and the strong interannotator agreement confirm that the strategies we followed at different stages of the process actually allow us to produce perspective-oriented summaries.

The evaluation concerning missing information from the perspective of one of the teams led to a great variety of answers (from 0 to 17 texts were judged to lack important information); see Figure 17. To keep this evaluation informative, we took into account only the non-neutral texts and the opinions of the evaluators who coincided with the generator on the judgement of the affinity (this is why not all bars in Figure 17 have the same height). Only for seven texts does everybody agree that the text is complete, but in only four cases do a larger number of evaluators think relevant information is missing. More generally, 74.4% of the assessments (one assessment being one evaluator of one text) conclude that no important information is missing—which is surprisingly homogeneous for such a subjective topic as soccer reports.

The evaluation shows that our stratified approach to user-tailored linguistic realization allows us to: (1) control perspective generation at all levels of linguistic representation, (2) dedicate a stratum for dynamic syntactic aggregation, making redundancies avoidable, and (3) easily parametrize our generator to produce a variety of perfectly synonymous texts such that the output summaries look more natural.

In general, to start from nonlinguistic representations (conceptual structures) allows a great flexibility in sentence realization: the superimposed information structure decides on the structure of the sentences without any other linguistic constraint. As a result, for the same content, our generator can produce quite distinct sentences, depending on the user perspective. The drawback is the costs. First, there are extensive semantic, lexical, and grammatical resources to be compiled in the correct format. Second, the generator requires a lot of computational memory to work at a satisfactory speed, since it computes different correct combinations of units before randomly choosing one (with around 1GB of memory, generating one report usually takes around 15 seconds). We are currently addressing these issues within MATE by (1) learning transition grammars from multilevel annotated corpora [Mille et al. 2012] and (2) reimplementing the generator with the aim of reducing the time by a factor of 100. With regards to the evaluation, a more robust evaluation could be performed following a blind approach in which human-written, automatically generated, and baseline texts are presented indiscriminately to the evaluators for rating [Belz and Reiter 2006].

6. RELATED WORK

Several main strands of related work can be identified. In particular, these are: (1) generation from KBs/ontologies, (2) perspective-based generation, (3) content selection, (4) data to text summarization, and (5) generation of time-varying scenes.

6.1. Generation from Ontologies/KBs

As pointed out in Section 1, most approaches to the generation from ontologies or knowledge bases use, as a rule, small repositories crafted specifically for the purpose of generation; see, for example, Hovy [1993]; Paris [1993]; Reiter et al. [2003]; and Bontcheva and Wilks [2004]. It is only recently, with the advent of standard OWL-RDF representations, that the use of larger, application-neutral ontologies has been explored; see, for example, Mellish and Pan [2010]; Power and Third [2010]; and Bouttaz et al. [2011]. While Power and Third [2010] explore the feasibility of realizing axioms taken from any ontology in natural language, both Mellish and Pan [2010] and Bouttaz et al. [2011] focus on content determination. Mellish and Pan [2010] explore how to select axioms from an OWL DL knowledge base using non-standard inference guided by constraints relevant to NLG—in particular by constraints derived from the conversational maxims formulated by Grice. Their approach is implemented in a proof-of-concept system for generating subsumers for a given class of an ontology using a reasoner. The content selection of Bouttaz et al. [2011] is governed by an extensive ontological model of the user, including its social context and a task-specific context which consists of policies expressed in terms of obligations, prohibitions, and permissions related to RDF resources and actors in the Web-based environment. The approach of Bouttaz et al. is similar to ours in that its policies can be seen as domain communication knowledge extending the base RDF knowledge base available for generation.

6.2. Perspective-Based Generation

Perspective-based generation has been studied from two main angles: (1) Varying the wording while expressing the same content in different conversational settings to users with different profiles; (2) Varying the content while conveying information on the same topic to users with different profiles. A classical example of the first type (which

is the one that is closer to our work) is Pauline [Hovy 1988], and EES [Paris 1993] of the second. Pauline is also one of the few proposals that involve text planning, syntactic structure determination, and lexical choice, to produce a perspective-oriented text. It is based on a parameter-driven selection of linguistic constructions of different abstraction (discourse span templates, syntactic templates, and lexical items). The parameter set involves the attitude of the speaker and hearer towards the content, the emotional state of the speaker/hearer, the interpersonal relation between speaker and hearer, and so on. Fleischman and Hovy [2002] follow a similar approach in that they assign to each lexical item in the lexicon numerically-valued features, choosing at the moment of generation lexical items whose features match most closely. Isard et al. [2006] also focus mainly on lexical choice parametrized by features of the speaker's personality such as extraversion, openness, neuroticism, and so on, although their strategy for choosing the lexical items is very different in that it is based on n -gram models. Probably the most detailed proposal for user-driven lexical choice is Stede [1993], who identifies a number of criteria (among them, formality, euphemism, slant, and floridity), each of them with a scale (formality from -3 to 3 , euphemism from 0 to 3 , etc.). Thus, *flick* would be placed on the scale of formality at -3 , and *motion picture* at 3 ; *toilet* is likely to receive the value 1 , *washroom* the value 2 and *restroom* the value 3 at the euphemism scale; and so on. Depending on the user, the desired (or most adequate) value for each scale can be chosen to determine the lexical style. A detailed study of the change of perspective by lexical means is also studied by Wanner [1997]. Oh and Shrobe [2008] address the interesting problem of perspective-based regeneration of baseball game summaries. They take a neutral baseball game summary and produce two articles out of it, one for each of the teams. They notice that in games where one of the teams dominates, there is not much overlap of content between the summaries. So they concentrate on studying games with high content overlap. Different perspectives are due to different groupings of events by different features. To turn a neutral article into a local team perspective article, they take the at-bats that should appear in the article, look at the feature values that are shared among them, and find the highest-ranked feature value for that team. Any remaining at-bats are arranged in chronological order. In contrast, our approach aims to address the realization of perspective at all levels of generation, from content selection to expression of rhetorical relations, ordering, information structuring and lexicalization.

6.3. Content selection

Works on content selection in general and on empirical statistical selection in particular [Barzilay and Lapata 2005; Kelly et al. 2009] select content in a way that in the best of cases forms a coherent whole in terms of the content itself; see “collective content selection” in Barzilay and Lapata [2005]; and Kelly et al. [2009]. However, the content plans produced by these approaches do not give any hint of how to derive a discourse plan from them, or for that matter, on how to imprint the discourse with a particular perspective.

6.4. Data to Text Summarization

There is a trend of work on summarizing large time-series data [Sripada et al. 2003; Gatt et al. 2009; Lareau et al. 2011]. For instance, Gatt et al. [2009] generate textual summaries from neonatal intensive care data. They start with a signal analysis in which patterns and trends in the data are identified. Then, basic reasoning is applied to infer relations between events (e.g., “A” causes “B”), before content selection, aggregation, microplanning, and linguistic realization are performed. An ontology is drawn upon, which captures some of the domain knowledge required by the reasoning

and data interpretation module and some of the linguistic knowledge. However, the ontology is not used to store data or perform reasoning.

Signal analysis would have been beneficial in our application as well—although we argue that it is not a generation task and should come before the generation proper. It is used, for instance, by Lareau et al. [2011] in the Australian football domain.

6.5. Generation of Time-Varying Scenes

As mentioned in the Introduction, there is an important body of work on generation of real-time commentaries [André et al. 1988; Binsted 1998; Tanaka-Ishii et al. 1998; Voelz et al. 1999]. These works address the challenge of selecting the most important events as they are occurring, while taking into account strict time constraints. Emotions (for example, speaker is sad because team A just scored) are usually expressed using intonation [Voelz et al. 1999] or facial expressions [Binsted 1998], rather than information, syntactic, and lexical means as we would do.

7. CONCLUSIONS AND FUTURE WORK

In the light of the increasing availability of large-scale application-independent ontologies, the question of perspective-oriented generation must be readdressed. First, strategies are needed that dynamically extend these ontologies by domain communication knowledge needed for flexible perspective-tailored generation, leaving them intact for other applications. For this purpose, we advocate a two-layered ontology, in which the first layer is composed of a general, application-neutral ontology and the second layer its NLG-oriented communication knowledge extension. Second, a generator is needed that is capable of rendering the content from the perspective of the user. We identified content selection, discourse structuring, information structuring, and lexicalization as instrumental for perspective-realization and presented the design and realization of these four tasks within an operational concept-to-text generator. The generator operates on a knowledge base populated and extended beforehand, and can thus generate commentaries for those matches for which data are available in the knowledge base. The evaluation of the performance of the individual facets of the generator shows that our proposal is viable. However, our work is merely a proof-of-concept, and further work is needed in the following areas if we are to make our approach portable to other domains.

Reasoning-driven content preselection. This is currently based on a manually crafted set of inference rules, which are applied offline to the whole ontology. In the future, the rules should be acquired empirically via time-series analysis and applied online, taking into account the user model (perspective in our case), for selecting the content to which inferences are to be applied.

Content selection. The empirical relevance weights of content units, which are the basis for our content selection, are currently obtained in our implementation only for a small number of categories from a corpus of texts aligned with data. However, not all domains have text corpora available and even when they do, the target texts for the generator might be quite different. Therefore, other strategies might be needed for obtaining a corpus of selected content—for instance, based on interactive selection of content units from a knowledge base in which the user perspective, among other preferences, is recorded.

Text planning. In our current implementation, the text planning tasks are template-based and thus of limited flexibility. Flexibility could be added by defining rules for relation mapping, unit determination and ordering in an ontology-friendly query formalism (e.g., SPARQL) that takes into account content and user perspective and thus allows for more options. Furthermore, a constraint-based approach

to ordering should be explored that takes into account not only perspective but also coherence and relevance weights, thus performing content selection and text planning simultaneously.

Relation between information structure and perspective realization. The involvement of further dimensions of the information structure, for instance, backgrounding/foregrounding, emphasis or focalization [Mel'čuk 2001], should be considered since it is not only thematicity that influences perspective realization.

Syntactic structuring. Lexicalization is the most prominent means for expressing a specific perspective, but syntactisization is certainly also of relevance and should be factored in.

User modeling. A more elaborate user model that records, for example, (in the football domain) the interest for specific teams, games, and players, the desired length or level of detail of summaries, the user's expertise, or the desired style, and so on would be of advantage and should be explored.

Automation of knowledge-base population. In our implementation, the knowledge base was populated and extended once before any queries were run. Further automation could be achieved if data was scraped from online sources and extended through inference on-demand for each query received by the system.

We are currently addressing some of these areas in the environmental domain, although the focus is not on user perspective but on a user model that takes into account user activity and sensitivity to specific pollutants or pollen [Bouayad-Agha et al. 2012]. In this new application domain, we also make a distinction between a base task-independent ontology and an extended task-dependent ontology. Inferences are performed on-demand upon receiving the user request, template-based content selection and discourse structuring are addressed using SPARQL queries, and lexical and grammatical resources are adapted to cover terminology and idiomatic expressions and constructions of the new domain. In addition to template-based content selection, an interface is being developed for obtaining new empirical evidence for content selection via domain expert interaction.

APPENDIXES

A. SAMPLE GENERATED SUMMARIES

In what follows, we cite, for illustration, a number of perspective-oriented summaries as produced by our generator and the corresponding human-written summaries as appeared in the press.

A.1. Villarreal-At. Madrid, 2007-2008

A.1.1. Generated summaries

*Perspective At. Madrid (31).*²¹ *Derrota del Atlético de Madrid. El Atlético de Madrid perdió ante el Villarreal por 3-0 y por culpa de esta derrota, se distanció de la zona UEFA. En la trigésimo primera jornada, se enfrentará al Almería.*

lit. "Defeat of Atlético of Madrid. Atlético of Madrid lost against Villareal by 3-0 and because of this defeat it moved away from the UEFA zone. In the 31st gameweek, it will face Almería."

Perspective Villarreal (32). *Victoria del Villarreal. El Villarreal ganó contra el Atlético de Madrid por 3-0 gracias a un gol de Cazorla en el minuto 38 y 2 de Nihat en el minuto 42 y el minuto 66. En la trigésimo primera jornada, se enfrentará al Sevilla.*

²¹The number in parentheses after the team perspective of each generated summary refers to the number of the summary given to the evaluators. Please refer to the evaluation section to see the corresponding ratings.

lit. "Victory of Villareal. Villareal won against Atlético of Madrid by 3-0 thanks to a goal by Cazorla in the minute 42 and 2 by Nihat in the minute 42 and the minute 66. In the 31st gameweek, it will face Sevilla."

Perspective Neutral (33). Victoria del Villarreal sobre el Atlético de Madrid. El partido acabó con la victoria del Villarreal sobre el Atlético de Madrid por 3-0. El Villarreal se mantuvo en la zona de champions mientras que el Atlético de Madrid se distanció de la zona UEFA. Subió a la segunda posición mientras que el Atlético de Madrid se mantuvo en la cuarta posición. En la trigésimo primera jornada, el Atlético de Madrid se enfrentará al Almería y el Villarreal al Sevilla.

lit. "Victory of Villareal over Atlético of Madrid. The match ended with the victory of Villareal over Atlético of Madrid by 3-0. Villareal stayed in the Champions zone, while Atlético of Madrid moved away from the UEFA zone. It ascended to the second position, while Atlético of Madrid stayed at the fourth position. In the 31st gameweek, Atlético of Madrid will face Almería and Villareal Sevilla."

A.1.2. Human summaries

EPSN. El Villarreal golea al Atlético Madrid y demuestra es candidato real al título de la Liga.

lit. 'Villareal hammered Atlético Madrid and shows it is a real candidate for the title of the league.'

Marca. El Villarreal mete la quinta. lit. "Villareal shifts into the fifth (gear)."

Terra. El 'submarino' ya es segundo en la Liga, tras deshacerse con eficacia de un gris Atlético. Nihat, por dos veces, y Cazorla certificaron la superioridad amarilla.

lit. "The 'submarine' is already the second in the League, after getting rid with efficiency of a grey Atlético. Nihat, twice, and Cazorla certified the yellow superiority."

A.2. Real Madrid-Recreativo Huelva, 2007-2008

A.2.1. Generated summary

Perspective R. Madrid (14). Victoria del Real Madrid. El Real Madrid ganó contra el Recreativo por 2-0 en el descuento gracias a un gol de V. Nistelrooy en el minuto 72 y un de Higuaín en el minuto 91. Gracias a esta victoria, no bajó a la zona de champions. En la octava jornada, se enfrentará al Espanyol.

lit. "Victory of Real Madrid. Real Madrid won against Recreativo by 2-0 during the injury time thanks to a goal by V. Nistelrooy in the minute 72 and one by Higuaín in the minute 91. Thanks to this victory, it did not descend to the Champions zone. In the 8th gameweek, it will face Espanyol."

A.2.2. Human summaries

EPSN. Real Madrid sumó fases de buen fútbol a su eficacia liguera para obtener su sexta victoria, con goles de Ruud Van Nistelrooy y Gonzalo Higuaín.

lit. "Real Madrid added phases of good football to its League efficiency to obtain its sixth victory, with goals by Ruud Van Nistelrooy and Gonzalo Higuaín."

Marca. Trabajada victoria para seguir líder.

lit. "Overworked victory to continue (as) leader."

Terra. El Real Madrid, líder de Primera división, sumó fases de buen fútbol a su eficacia liguera para obtener su sexta victoria del campeonato, con un gol de Van Nistelrooy, a pase de Raúl, y otro de Higuaín en los últimos segundos, ante un Recreativo valiente a pesar de sus bajas.

lit. "Real Madrid, leader of the First League, added phases of good football to its League efficiency to obtain its sixth victory of the championship, with a goal by

Van Nistelrooy, upon a pass by Raúl, and another by Higuaín in the last seconds, against Recreativo, brave despite its drop outs.”

A.3. Valladolid-Real Madrid, 2007-2008

A.3.1. Generated summary

Perspective Valladolid (26). Empate del Valladolid. El Valladolid empató contra el Real Madrid (1-1) gracias a un gol de Saviola en el minuto 86 a pesar de un gol de Pedro López en el minuto 69. En la quinta jornada, se enfrentará al Mallorca.

lit. “Draw of Valladolid. Valladolid drew against Real Madrid (1-1) thanks to a goal by Saviola in the minute 86 despite a goal by Pedro López in the minute 69. In the 5th gameweek, it will face Mallorca.”

A.3.2. Human summaries

EPSN. El Real Madrid sólo pudo empatar tras enredarse en la presión de un Valladolid voraz, ordenado, rápido y profundo, que gobernó el partido.

lit. “Real Madrid could only draw after getting entangled in the pressure of a voracious, ordered, fast and deep Valladolid, which ruled the match”.

Marca. Petróleo blanco en Valladolid.

lit. “White petroleum in Valladolid.”

Terra. El Real Madrid sin Sneijder es otro. El Valladolid aprovechó la ausencia del tulipán y superó al conjunto de Schuster en juego y ganas pero no en goles. Un zapatazo de Pedro López dejó helados a los jugadores blancos, pero Saviola, recién entrado al campo, logró el empate a uno definitivo. El Madrid sigue líder.

lit. “Without Sneijder, Real Madrid is different. Valladolid took advantage of the absence of the tulip and beat the team of Schuster in the game and zest, but not in goals. A blow with the shoe by Pedro López left the white players frozen, but Saviola, (who) recently entered the field, reached the final draw by one. Madrid continues leader.”

A.4. Zaragoza-Recreativo Huelva, 2007-2008

A.4.1. Generated summary

Perspective Recreativo (40). Derrota del Recreativo. El Recreativo perdió ante el Zaragoza por 3-0 y a causa de esta derrota, bajó a la zona de descenso. Acabó el partido con 10 jugadores a causa de la expulsión de Marco Rubén en el minuto 40. En la trigésimo cuarta jornada, se enfrentará al Levante.

lit. “Defeat of Recreativo. Recreativo lost against Zaragoza by 3-0 and because of this defeat descended to the descent zone. Ended the match with 10 players because of the send-off of Marco Rubén in the minute 40. In the 34th gameweek, it will face Levante.”

A.4.2. Human summaries

EPSN. El Zaragoza se sube al último tren de la salvación al derrotar al Recreativo como local.

lit. “Zaragoza boards the last train of salvation by defeating Recreativo as a local.”

Marca. El Recre llora en La Romareda.

lit. “Recre cries in La Romareda.”

Terra. El Zaragoza, tras su victoria sobre el Recreativo, se sube al último tren de la salvación que tenía y consigue salir de la zona de descenso para meter en ella al conjunto onubense, aunque empatado a puntos con los aragoneses.

lit. “Zaragoza, after its victory over Recreativo, boards the last train of salvation it had and manages to get out of the descent zone (in order) to put into it the Huelva group, although equalized with the Aragonese.”

ACKNOWLEDGMENTS

The authors would like to thank Pablo Castells, Fernando Díez, Sergio López Hernández, and David Vallet from the Autonomous University of Madrid for their generous help in all tasks related to the i3media ontology. Many thanks are also due to the three anonymous reviewers of the ACM TSLP for their very helpful and insightful comments on the previous versions of the article.

REFERENCES

- ANDRÉ, E., HERZOG, G., AND RIST, T. 1988. On the simultaneous interpretation of real world image sequences and their natural language description: The system SOCCER. In *Proceedings of the 8th European Conference on Artificial Intelligence (ECAI)*. 449–454.
- BARZILAY, R. AND LAPATA, M. 2005. Collective content selection for concept-to-text generation. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (EMNLP)*.
- BATEMAN, J. 1997. Enabling technology for multilingual natural language generation: The KPML development environment. *Natural Lang. Engin.* 3, 2, 15–55.
- BELZ, A. 2003. And now with feeling: Developments in emotional language generation. Tech. rep. ITRI-03-21, Information Technology Research Institute, University of Brighton.
- BELZ, A. AND REITER, E. 2006. Comparing automatic and human evaluation in NLG. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. 313–320.
- BINSTED, K. 1998. A talking head architecture for entertainment and experimentation. In *Proceedings of the AAAI Fall Symposium (Emotional and Intelligent: The Tangled Knot of Cognition)*.
- BOHNET, B., LANGJAHR, A., AND WANNER, L. 2000. A development environment for an MTT-based sentence generator. In *Proceedings of the 1st International Natural Language Generation Conference (INLG)*.
- BONTCHEVA, K. AND WILKS, Y. 2004. Automatic generation from ontologies: The MIAKT approach. In *Proceedings of the International Conference on Applications of Natural Language to Information Systems*. 324–335.
- BOUAYAD-AGHA, N., CASAMAYOR, G., MILLE, S., ROSPOCHER, M., SERAFINI, L., SAGGION, H., AND WANNER, L. 2012. From ontology to NL: Generation of multilingual user-oriented environmental reports. In *Proceedings of the 17th International Conference on Applications of Natural Language Processing to Information Systems (NLDB)*.
- BOUAYAD-AGHA, N., CASAMAYOR, G., AND WANNER, L. 2011a. Content selection from an ontology-based knowledge base for the generation of football summaries. In *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG)*.
- BOUAYAD-AGHA, N., CASAMAYOR, G., WANNER, L., DÍEZ, F., AND LÓPEZ HERNÁNDEZ, S. 2011b. FootbOWL: Using a generic ontology of football competition for planning match summaries. In *Proceedings of the 8th Extended Semantic Web Conference (ESWC)*. 230–244.
- BOUTTAZ, T., PIGNOTTI, E., MELLISH, C., AND EDWARDS, P. 2011. A policy-based approach to context dependent natural language generation. In *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG)*. 151–157.
- DE ROSIS, F. AND MELLISH, C. 2007. Tutorial on affective natural language generation. In *Proceedings of the 11th International Conference on User Modeling (UMAP)*.
- DUKLE, K. 2003. A prototype query-answering engine using semantic reasoning. M.S. thesis, University of South Carolina, Charleston.
- ELHADAD, M. AND ROBIN, J. 1996. An overview of SURGE: A reusable comprehensive syntactic realization component. In *Proceedings of the 8th International Natural Language Generation Workshop (INLG)*.
- FLEISCHMAN, M. AND HOVY, E. 2002. Emotional variations in speech-based natural language generation. In *Proceedings of the International Natural Language Generation Conference (INLG)*.
- GATT, A., PORTET, F., REITER, E., HUNTER, J., MAHAMOOD, S., MONCUR, W., AND SRIPADA, S. 2009. From data to text in the neonatal intensive care unit: Using NLG technology for decision support and information management. *AI Comm.* 22, 153–186.
- GUNDEL, J. 1988. Universals of Topic-Comment Structure. In *Studies in Syntactic Typology*, M. Hammond, E. Moravcsik, and J. Wirth, Eds., John Benjamins Publishing Company, Amsterdam, 209–239.

- HOVY, E. 1988. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum, Hillsdale, NJ.
- HOVY, E. 1993. Automated discourse generation using discourse structure relations. *Artif. Intell.* 63, 1–2, 341–386.
- ISARD, A., BROCKMANN, C., AND OBERLANDER, J. 2006. Individuality and alignment in generated dialogues. In *Proceedings of the International Natural Language Generation Conference (INLG)*.
- KELLY, C., COPESTAKE, A., AND KARAMANIS, N. 2009. Investigating content selection for language generation using machine learning. In *Proceedings of the European Workshop on Natural Language Generation (EWNLG)*.
- KITTREDGE, R., KORELSKY, T., AND RAMBOW, O. 1991. On the need for domain communication knowledge. *Comput. Intell.* 7, 4, 305 – 314.
- KOSSEIM, L. AND LAPALME, G. 2000. Choosing rhetorical structures to plan instructional texts. *Comput. Intell.* 16, 3, 408–445.
- LAMBRECHT, K. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge Studies in Linguistics 71., Cambridge University Press, Cambridge, UK.
- LAREAU, F., DRAS, M., AND DALE, R. 2011. Detecting interesting event sequences for sports reporting. In *Proceedings of the 13th European Workshop on Natural Language Generation (EWNLG)*. 200–205.
- LAREAU, F. AND WANNER, L. 2007. Towards a generic multilingual dependency grammar for text generation. In *Proceedings of the Grammar Engineering Across Frameworks Workshop (GEAF07)*, T. King and E. Bender, Eds., CSLI, Stanford, CA, 203–223.
- LENAT, D. B. 1995. CYC: A large-scale investment in knowledge infrastructure. *Comm. ACM* 38, 11, 33–38.
- MANN, W. AND THOMPSON, S. 1987. Rhetorical structure theory: A theory of text organization. In *The Structure of Discourse*, L. Polanyi, Ed., Ablex Publishing Corporation, Norwood, NJ.
- MELLISH, C. AND PAN, J. 2010. Natural language directed inference from ontologies. *Artif. Intell.* 172, 10, 1285–1315.
- MEL’ČUK, I. 1988. *Dependency Syntax: Theory and Practice*. SUNY Press, Albany, NY.
- MEL’ČUK, I. 2001. *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*. Benjamins Academic Publishers, Amsterdam.
- MILLE, S., WANNER, L., AND BURGA, A. 2012. Treebank annotation in the light of the Meaning-Text Theory. *Ling. Issues Lang. Technol.* 7, 16.
- MOORE, J. AND PARIS, C. 1993. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Comput. Ling.* 19, 4, 651–694.
- NÉDELLEC, C. AND NAZARENKO, A. 2005. Ontology and information extraction: A necessary symbiosis. In *Ontology Learning from Text: Methods, Evaluation and Applications*, P. Buitelaar, P. Cimiano, and B. Magnini, Eds., IOS Press Publication.
- OH, A. AND SHROBE, H. 2008. Generating baseball summaries from multiple perspectives by reordering content. In *Proceedings of the International Conference on Natural Language Text Generation (INLG)*.
- PARIS, C. 1993. *User Modelling in Text Generation*. Frances Pinter Publishers, London.
- POWER, R. AND THIRD, A. 2010. Expressing OWL axioms by English sentences: Dubious in theory, feasible in practice. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*.
- REITER, E., ROBERTSON, R., AND OSMAN, L. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artif. Intell.* 144, 41–58.
- ROBIN, J. 1994. Revision-Based Generation of Natural Language Summaries Providing Historical Background. Ph.D. thesis, Graduate School of Arts and Sciences, Columbia University, NY.
- SCHAPIRE, R. AND SINGER, Y. 2000. Boostexter: A boosting-based system for text categorization. *Mach. Learn.* 39, 2/3, 135–158.
- SGALL, P., HAJIČOVÁ, E., AND PANEVOVÁ, J. 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Reidel Publishing Company, Dordrecht, Germany.
- SOWA, J. 2000. *Knowledge Representation*. Brooks Cole, Pacific Grove, CA.
- SRIPADA, S., REITER, E., AND DAVY, I. 2003. SumTime-Mousam: Configurable marine weather forecast generator. *Expert Update* 6, 3, 4–10.
- STEDE, M. 1993. Criteria for lexical choice in text generation. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.
- TANAKA-ISHII, K., HASIDA, K., AND NODA, I. 1998. Reactive content selection in the generation of real-time soccer commentary. In *Proceedings of the International Conference on Computational Linguistics (COLING)*. 1282–1288.

Perspective-Oriented Generation of Football Match Summaries: Old Tasks, New Challenges

- TSINARAKI, C., POLYDOROS, X., KAZASIS, F., AND CHRISTODOULAKIS, S. 2005. Ontology-based semantic indexing for mpeg-7 and tv-anytime audiovisual content. *Multimedia Tools Appl.* 26, 3, 299–325.
- USCHOLD, M. AND KING, M. 1995. Towards a methodology for building ontologies. In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, held in conjunction with *IJCAI*. 6.1–6.10.
- VALLDUVI, E. 1990. The information component. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.
- VOELZ, D., ANDRE, E., HERZOG, G., AND RIST, T. 1999. Rocco: A robocup soccer commentator system. In *Proceedings of RoboCup-98: Robot Soccer World Cup II*. Springer-Verlag, 50–60.
- WANNER, L. 1997. Exploring lexical resources for text generation in asystemic functional language model. Ph.D. thesis, Saarland University, Saarbrücken, Germany.
- WANNER, L., BOHNET, B., BOUAYAD- AGHA, N., LAREAU, F., AND NICKLASS, D. 2010. MARQUIS: Generation of user-tailored multilingual air quality bulletins. *Appl. Artif. Intell.* 24, 10, 914–952.

Received December 2011; revised March 2012; accepted May 2012