Natural Language Generation in the context of the Semantic Web

Article *in* Semantic Web · January 2014

DOI: 10.3233/SW-130125

3 authors:

Nadjet Bouayad-Agha
University Pompeu Fabra

Gerard Casamayor
University Pompeu Fabra

Leo Wanner
Catalan Institution for Research and Advanced Studies

Undefined 00 (2012) 00
IOS Press

Natural Language Generation in the context of the Semantic Web

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Nadjet Bouayad-Agha [a,*], Gerard Casamayor [a] and Leo Wanner [a,b]
[a] *Department of Information and Communication Technologies, Pompeu Fabra University, C/ Roc Boronat, 138, 08018 Barcelona, Spain*
[b] *Catalan Institute for Research and Advanced Studies, Passeig Lluís Companys, 23, 08010 Barcelona, Spain*
E-mail: {nadjet.bouayad|gerard.casamayor|leo.wanner}@upf.edu

Abstract. Natural Language Generation (NLG) is concerned with transforming some formal content input into a natural language output, given some communicative goal. Although this input has taken many forms and representations over the years, it is the semantic/conceptual representations that have always been considered as the "natural" starting ground for NLG. Therefore, it is natural that the semantic web, with its machine-processable representation of information with explicitly defined semantics, has attracted the interest of NLG practitioners from early on. We attempt to provide an overview of the main paradigms of NLG from SW data, emphasizing how the Semantic Web provides opportunities for the NLG community to improve their state-of-the-art approaches whilst bringing about challenges that need to be addressed before we can speak of a real symbiosis between NLG and the Semantic Web.

Keywords: semantics, natural language text generation, semantic web formalisms, web resources

1. Introduction

Natural Language Generation (NLG) is concerned with transforming a given formal content input into a natural language output, given some communicative goal in a specific context [118]. This input has taken many forms and representations over the years, from linguistic surface-oriented structures over semantic or conceptual representations to raw numeric data. However, it is the semantic/conceptual representations that have always been considered to be the "natural" starting ground for NLG: linguistic surface-oriented structures predetermine, at least partially, the linguistic form of the output, which is clearly undesirable for flexible NLG, and raw numeric data require prior preprocessing that is not related to NLG, e.g., [111,126]. Therefore, it is not surprising that the Semantic Web (SW), with its machine-processable representation of information with explicitly defined semantics, has attracted the interest of NLG practitioners from early on. The objective of this article is to provide an overview of the main paradigms of NLG from SW data, emphasizing how the Semantic Web provides opportunities for the NLG community to improve their state-of-the-art approaches whilst bringing about challenges that need to be addressed before we can speak of a real symbiosis between NLG and the Semantic Web.

We begin with a brief overview of NLG that delimits its scope, introduces its key tasks, modules, architectures and methodologies and summarizes its long term issues (Section 2). This is then followed by an overview of the main approaches and paradigms to NLG from SW data, which we hope will orient the SW researcher/engineer interested in using NLG technol-

*Corresponding author

ogy (Section 3). Next we discuss what we consider to be the most prominent "burning" issues for a successful symbiosis between NLG and SW (Section 4) before concluding (Section 5).

2. A brief overview of NLG

Our objective in this section is twofold: firstly, to introduce the field of NLG to the SW researcher/engineer, and secondly to pave the way to our arguments in the rest of the paper. After giving a bird's eye view of NLG (Section 2.1), we discuss semantically oriented NLG tasks (Section 2.2) which are especially relevant when dealing with SW data and thus for understanding NLG approaches to the Semantic Web presented in Section 3. We then discuss long-term NLG issues (Section 2.3) which become critical to address in the context of NLG from SW data as we will argue in Section 4. To avoid any overlap with discussions later in the article, we deliberately avoid in this brief overview of NLG any mention of the approaches that are specific to SW data.

2.1. A bird's eye view of NLG

As mentioned above, the global task of NLG is to map a given formal input onto a natural language output to achieve a given communicative goal in a specific context. The context can be entirely implicit if the generator focuses on one specific type of report for one specific type of user (as, e.g., in the case of the generation of clinical narratives for medical personnel), or allow for an explicit parameterization of only one or several dimensions, such as the level of expertise of the user (e.g., laypeople vs expert in the domain), his/her specific needs or preferences (e.g., user is allergic to a type of pollen, or is in favour of a particular football team), discourse history (e.g., what the user already knows) or physical location and surroundings (e.g., people in the vicinity of the user at a conference or exhibits the user is looking at in a museum for an advice or narration generation wearable device).

In Figure 1, we provide a summary of the main characteristics of an NLG system's input, output and context. The input can be characterized with respect to its type (i.e., whether it is structured or not, and what language representation it uses, if any), size, domain and (NLG) task (in)dependence[1]; the output with respect to its size, coherence, fluency[2], language and modality; and the context with respect to the profile and information need of the target audience and the communicative goal of the generator.

The range of admissible (or desired) characteristics of the input, output and context determines, to a certain extent, the complexity of the generator. Thus, a generator that accepts as input small unstructured sets of data and generates out of them short monolingual messages will have a simpler architecture than a generator that takes as input large semantic graphs to generate multilingual texts that vary in content, language and style according to the user profile and request. The format of the input may also vary, depending on whether a generator is used as a stand-alone application or is part of a larger automatic information processing application such as a Dialogue, Question Answering, Summarization, or Machine Translation system. This highlights the decisive difference between NLG and, for instance, parsing: NLG cannot always start from the same input (while parsing always starts from the language surface).[3] The consequence of this difference is that, in NLG research, no consensus has been achieved so far on what a generation application is supposed to start from and what the standard input representation for generation should exactly look like—although it seems clear that "some sort of" semantic representation is the most appropriate starting ground. Over the years, different types of semantic representations have been experimented with—including first-order logics and related formalisms [6,30,109], Schank's scripts [69,102], Sowa's conceptual graphs [107], a variety of frame representations such as KL-ONE and LOOM [68,75,116], up to Semantic Web representations; see the following sections.

Fully-fledged NLG traditionally implies a number of tasks. The six most central of them are: 1) *content selection* that determines which parts of the content received as input are to be verbalized according

[1]A linguistically independent conceptual structure is both domain- and task- independent, a linguistically motivated conceptual representation is domain independent- but task- dependent, and a relational database is typically domain- dependent but task- independent.

[2]Coherence is concerned with the sentences and clauses of the text forming a meaningful, unified whole that "hangs together". Although coherence contributes to fluency, in this paper, we restrict fluency to the sentence level, where it is affected mainly by grammaticality issues.

[3]To illustrate this problem, a famous statement by Yorick Wilks that "the difference between Natural Language Understanding and Natural Language Generation is like the difference between counting from one to infinity and from infinity to one" is often quoted.

Input	
Type	input data representation (e.g., semantic graph, database, tabular, template); input representation language (first-order logic, OWL-DL, etc); single or multiple input.
Size	small (e.g., a small RDF graph), large or very large input (e.g., hundreds of thousands of measurements or database entries).
Domain independence	input representation domain- dependent or independent
Task independence	input representation independent or dependent of the task of text generation.
Context	
Targeted genre	term definition, report, commentary, narrative, etc.
Targeted audience	lay person, informed user, domain expert, etc.
Request	information solicitation, decision support request, etc.
Communicative goal	exhaustive information on a theme, advice, persuasion, etc.
User profile	user preferences, needs or interests in the topic, individual expertise, previous knowledge, discourse history, etc.
Output	
Size	single sentence, paragraph or a multi-paragraph text
Coherence	set of disconnected sentences or a coherent text
Fluency	fluent NL, stilted NL, telegraphic style
Language	monolingual (English, French, German, Spanish, . . .) or multilingual
Modality	textual only (written or spoken) or multimodal (e.g., text or speech with table or figures) and the degree of the multimodality

Fig. 1. Summary of dimensions of NLG system with respect to input, output and context

to the context; 2) *discourse planning* that organizes the content so that it is rendered as a coherent text; 3) *lexicalization* that maps conceptual (or language-independent semantic) configurations onto language-specific senses or words; 4) *aggregation* that merges partially overlapping content and linguistic structures to avoid repetition and to improve the fluency of the output[4]; 5) *generation of referring expressions* (GRE), i.e., generation of anaphora and generation of references to entities supposedly already present in the reader's world model; and 6) *linguistic realization* that deals with mapping the discourse or sentence specifications obtained from the preceding tasks onto a syntactically, morphologically and orthographically correct text.

These tasks are conveniently separated and usually occur in staggered fashion along three main pipelined modules in a working, so-called *pipeline* architecture [98,117]: 1) *document or text planning* (sometimes also referred to as *macro-planning*), 2) *sentence planning* (otherwise known as *micro-planning* in opposition to the previous macro-planning module), and 3) *surface realization*. The document planning module is in charge of deciding *what to say* and organizing the chosen content into a coherent whole. The sentence planner is in charge of mapping the text plan to the linguistic structure, grouping information into sentences and performing aggregation, and lexicalization along the way. Finally, the surface realizer is responsible for rendering each sentence plan into a sentence string. It is therefore obvious that it is the document planning module and, to some extent, the sentence planning module that must be able to cope with semantic representations.

The sequential nature of decision-making in a pipeline architecture means that there is no mechanism to change decisions taken in an earlier module when in a later module. However, it has long been known that the tasks that make up the modules of the pipeline are not independent of each other. For instance, micro-planning tasks like lexicalization, syntactic realization and referring expression generation may require micro-level content selection. For instance, Nicolov et al. [105] address the problem of sentence generation from a non-hierarchical semantic graph in which only a subset of the input might be linguistically realizable and hence included in the final text. Jordan and Walker [73] address the issue of deciding what con-

[4]The aggregation of the syntactic representations corresponding to the following 2 sentences "`The wind will be weak. The rain will be light`" will produce the more fluent text: "`The wind will be weak and the rain light`"

tent to include in a description for the generation of referring expressions[5].

Other NLG architectures have been proposed that provide alternatives to the linear decision-making space of the pipeline architecture. Generate-and-select (or revision-based) approaches, e.g., [11,85], follow various paths at decision-making points, postponing the selection of the best action until enough evidence has been gathered. Optimization approaches [44,88] model decision-making as a network of states connected by the outcomes of individual actions, and search the decision space for a set of actions that minimize a cost function. NLG has also been approached in [79] as direct content to text mapping in the context of database records aligned with texts, therefore doing away with all intermediate stages.

The methodologies applied in NLG to map the given input onto the natural language output range from the use of simple fill-in templates and canned text for straightforward verbalization of messages of limited complexity to the exploitation of strategies that implement informed projections between theoretically sound representations for each individual generation task. However, as van Deemter et al. [41] point out, the distinction between practical application-oriented template-based NLG and "real" research NLG is becoming increasingly blurred in that template-based generation becomes quite sophisticated and research-oriented NLG experiments often bridge tasks that are not in focus by simple realizations or omit them altogether. As a consequence, the theoretical justification, maintainability, and output quality and variability cannot be predicted based only on the fact of whether templates are used or not. Until recently, rule-based realizations prevailed for methodologies of all complexities. However, statistical/heuristic-based realizations of different NLG tasks have become increasingly popular (e.g., see [82] for some recent approaches), from sentence planning and realization [11,85,131], text planning [95], ordering of content units [49] to content selection [5].

The evaluation of the performance of the different NLG modules or tasks has been given increasing prominence in the NLG community. Typically, evaluation is done by asking human subjects to read and judge automatically generated texts and comparing those judgments with those of a gold standard corpus, a baseline corpus or a corpus obtained using some other NLG approach [13]. The use of corpus-based evaluation metrics obtained by automatically comparing the generated output (i.e., text or intermediate representation) against a gold standard is also becoming increasingly popular [5,12,14,15,48]. For example, to evaluate their approach for ordering semantic structures, Duboue and McKeown's [48] use a rule-based planner as a gold standard reference to compare their ordering with several random baseline orderings.

2.2. *Semantically oriented NLG-tasks*

Among the generation tasks presented in Section 2.1, essentially content selection, discourse structuring, and lexicalization depend on the type of semantic input structure used by the generator since they operate directly on the input structure or on a fragment of it.[6] The first two tend to output the same type of semantic structure as they take as input, while the third (lexicalization) tends to output a lexicalized structure whose constituents differ in type from the semantic structure. Let us discuss each of these three tasks in turn.

2.2.1. *Content selection*

Content selection addresses the problem of selecting a subset of content from a larger set, whether at the document (macro) or sentence (micro) level.]. In this section we mainly discuss approaches to macro-level content selection.[7]

Determining what content to convey depends largely on the context dimensions detailed in Figure 1. This is traditionally done in terms of rule-based templates, adopting the so-called "closed planning" paradigm [34]. Some approaches, however, adopt a network representation of the input data and exploit the network topology to perform an informed search of the most relevant nodes, adopting the so-called "open planning" paradigm [34], whether for macro-level content selection, as, e.g., in the approaches of O'Donnell et al. [108] and Demir et al. [42], or micro-level content selection, as e.g., in Krahmer et al's [81] approach for referring expression generation.

[5]For example, the same rug can be referred to as "the yellow rug", "the $150 rug" which implies the selection of different attributes to refer to that object.

[6]Conceptual aggregation also operates on the input structure. However, as it tends to be addressed in an ad-hoc way we did not think it deserved to be discussed in this section.

[7]A couple of approaches to micro-selection are discussed in section 2.1 above, namely Nicolov et al.'s [105] and Jordan and Walker's [73] approaches.

In the approaches that follow the open planning paradigm, content is often seen as forming a content graph where nodes correspond to content atoms (e.g., facts in a KB or database cells), while edges indicate selection constraints between pairs of nodes. In some cases, the selection constraints are derived from links between data found in the content. The links serve as indicators of related content that can be selected together. In other cases, constraints are elicited from special relations which indicate potential discourse relations between facts when placed in a discourse plan. Thus, potential rhetorical relations are established between sets of facts in O'Donnell et al's [108] ILEX system, while in the work by Demir et al. [42] *attractor* and *repeller* relations indicate discourse compatibility (or incompatibility) between facts.

The nodes and edges of the content graph on which open planning strategies operate can be assigned weights that modulate the strength of the constraints or quantify the degree of interest of the user for certain types of content, as encoded in a user model. Weights warrant the application of optimization- and graph-based algorithms to solve the content selection problem. They can be assigned either manually as in, e.g., Demir et al.'s approach [42] or be statistically inferred from a corpus of texts aligned with data as in, e.g., Barzilay and Lapata's approach [5].

During content selection, message determination can also be performed, which groups content units into messages that can later facilitate linguistic expression.[8] These messages may correspond to fine-grained content units such as facts or events (e.g. [108,111]) or to topic-related groupings of fine-grained content units (e.g. [118]), and be realized in the text as constituents, clauses or sentences. Message determination has been addressed mainly using templates in a pipeline architecture [118], thus ensuring that each message can be rendered in natural language. An example of a message given by Reiter and Dale [118] is the RainSpellMsg message in a weather reporting generator which is built only if rain occurs on more than a specified number of days in a row:

```
((message-id msg096)
 (message-type rainspellmsg)
 (period ((begin ((day 04)
                  (month 02)
                  (year 1995)))
          (end ((day 11)
                (month 02)
                (year 1995)))
          (duration ((unit day)
                     (number 8)))))
 (amount ((unit millimetres)
          (number 120))))
```

[8]This content selection is sometimes referred to as content determination [118], which we think has a broader scope than content selection.

Message determination can also occur after content selection and prior to discourse structuring and the Elementary Discourse Units (EDUs) thus built used as basic units for discourse structuring (e.g., [134]).

2.2.2. *Discourse structuring*

Although often handled with content selection in one task [118] using text schemas in the sense of McKeown [93], it is at least theoretically undisputed that discourse structuring is an NLG task on its own. Discourse structuring is concerned with the derivation of a coherent discourse structure of the content that is to be turned into natural language. It can be carried out either after content selection [70,89,120]) or interleaved with content selection [92,104,108], such that during the relevant content search through the graph only those nodes are taken into account that can be connected to the already selected nodes via a discourse relation—which ensures discourse coherence. The discourse relations between content nodes are introduced either directly into the content graph prior to NLG proper [120], via *plan operators* during text planning [70,92,104] or via a projection from semantic relations established between content nodes [24,80].

A very popular discourse structure theory in NLG is *Rhetorical Structure Theory* (RST) [87] because of its pre-realizational definition of rhetorical relations in terms of speaker's intentions and effects on the hearer on the one hand, and the distinction between the main (nucleus) and supporting (satellite) arguments of the asymmetric discourse relations on the other hand. This asymmetric property has been formalized and exploited by some researchers, in particular by Marcu [89], to build up coherent discourse structures.

In some approaches, the problem of discourse structuring is reduced to the problem of determining the best order between selected facts using empirical approaches [45,49].

2.2.3. *Lexicalization*

The strategies employed to realize the mapping between the semantic and lexical entities can be broadly divided between three main approaches: discrimination networks [59], sequential graph-rewriting approaches [50,106,107,127] and structure mapping approaches [78,86,110,133].

Goldman [59] uses discrimination networks that select the most specific lexical unit that subsumes the target object or event, e.g., the concept `Ingest` with value restrictions `actor:John` and `theme: Milk027` has as its most specific subsuming lexical unit `Ingest` with value restriction `theme:Liquid` and is thus lexically realized as `'drink'`.

Graph-rewriting and structure mapping approaches are especially relevant for multilingual generation and for taking into account collocational and other realizational constraints. Graph-rewriting approaches are essentially pattern matching approaches. The words and their senses are expressed by the same formalism and lexicalized items are selected provided that their underlying content covers all or parts of the conceptual input, whether through unification [50,107,127] or some matching metric [106].

Structure mapping approaches map a given input structure into a lexicalized output by using a semantic dictionary that maps a sense onto one or more lexical entries and a lexicon that provides the lexico-syntactic constraints for each individual entry (e.g., [86]).

In addition to these three main approaches, some proposals also treat the items of the semantic representation as lexical items, such that they do not change the type of the output; see e.g., [116].

In order to facilitate the mapping from a language-independent domain representation onto a task-dependent (i.e., NLG-dependent) linguistic representation whilst giving sufficient room for flexible verbalization, domain- and task-independent language-oriented ontologies have been introduced into generation. The most prominent of these is, without doubt, the *Penman Upper Model* (UM) [7]. Originally used in the context of the systemic-functional generators PENMAN [91] and KPML [8], the UM evolved over the years into a major multilingual, linguistically oriented ontology known as the *Generalized Upper Model* (GUM); see, for instance, [9,10].

2.3. *Long-term NLG issues*

Some of the main well-known issues that applied NLG systems need to address are 1) portability across domains, so that the same system can be applied to new datasets with minimal effort, 2) robustness, so that NLG systems can scale up to large datasets, and 3) evaluability, so that competitive evaluation of NLG modules and tasks can be achieved.

Key to portability is the ability to reuse NLG resources, that is modules, technologies and task-related knowledge. Whilst off-the-shelf surface realizers are the most reused of NLG resources, e.g., SimpleNLG [58], much remains to be done to promote reusability of NLG modules and technologies. Reuse of task-related knowledge is also scarce. There is not even a consensus as to what the input and output of each individual module/representation should be and how they should be mapped from one to the other, e.g., the different approaches to mapping domain information structures onto language-oriented structures that render this information described in Section 2.2.3. Efforts such as RAGS (“Reference Architecture for Generation Systems”) [98] to come up with a formal specification of consensus interface representations of the different NLG modules have so far failed to be taken up by the NLG community.

Data and technology reuse and sharing is also important for the evaluability of NLG systems/modules. Common datasets such as the TUNA corpus for referring to objects in a visual domain have been recently developed and used for NLG shared tasks and their evaluation [57]. Furthermore, any new NLG task or module need to be plugged in with existing ones so as to produce an evaluable output (i.e., text) [119].

Although some NLG applications require the use of existing domain knowledge bases and datasets developed outside the scope of the NLG application – e.g., data-to-text systems (e.g., [111]) or systems that use existing medical ontologies (e.g. [28]) – many NLG systems still develop, out of necessity, the required domain knowledge manually from scratch for the purpose of the application. The side-effect of this approach is that the datasets used tend to be small and the approaches taken idiosyncratic – mainly template-based. However, as we mentioned in Section 2, statistical approaches that tend to require the use of larger datasets are increasingly popular in NLG. Other approaches such as open text planning discussed in Section 2.2.1 are also suitable for use on large semantic networks. Another contribution towards robustness is the ability to deal with noisy and incomplete domain knowledge, for example by enriching it with so-called domain communication knowledge [78] (i.e., domain- and task-dependent knowledge), e.g., in the context of the generation of air quality bulletins [134], by interpreting raw measurements to find out minima and maxima, inferring a rating from a measurement or inferring a cause relation between pollutant rating(s) and air quality index.

3. Overview of approaches to NLG from SW data

NLG approaches to the Semantic Web have received several overlapping functional classifications. For instance, Bontcheva et al. [19,38] distinguish between SW-oriented NLG applications "[to help] users who are not knowledge engineers to understand and use ontologies", and to present "formal knowledge [. . .] expressed in natural language in order to produce reports, letters, etc. [that is to] present structured information in a user-friendly way". A similar distinction is made between Natural Language Interfaces (NLIs) for ontology engineering and publication of knowledge from a specific communicative goal using NLG technologies [124], or between applications for querying, authoring and verbalizing ontologies (i.e., this latter subsuming both the tasks of documenting ontologies and publishing knowledge in an end-user application) [56]. Of course, these uses of NLG predates the Semantic Web. For example, Sevcenko's [123] presents a template-based approach for the verbalization of the logical axioms of the SUMO language-independent upper-ontology. These verbalized axioms are presented in an application that allows the user to browse the SUMO ontology and its alignment with the WordNet lexicon.

The aim of this section is to provide the SW engineer/researcher with an overview of the main existing paradigms and approaches for text planning ("what to say", section 3.1) and sentence planning ("how to say it", section 3.2) from SW data. Some of the reviewed approaches and paradigms use standard NLG techniques with or without SW-specific technology whilst others are more attuned to the characteristics of the SW data.

In order to support our analysis, we will refer to some parts of Table 1 that shows a summary of the most representative NLG applications on which this overview is based according to the NLG dimensions and features introduced in Section 2.1. The table is divided between approaches that use NLG for ontology engineering (i.e., NLIs) in the upper half and approaches that use NLG for knowledge publishing in the bottom half.[9] All the features, apart from the input size and input domain independence in Figure 1 are considered. Input size and input domain independence are not considered because (nearly) none of the approaches has a restriction on the size of the input and the input is always domain-dependent. [10]

3.1. Text planning for NLG from SW data

Table 1 distinguishes between four main communicative goals: (i) to say(almost) all there is to say about some input object (i.e., class, query, constraint, whole graph), e.g., [2,21,65] ; (ii) to verbalize the content interactively selected by the user, e.g., [47,66, 112]; (iii) to verbalize the most typical facts found in (real or virtual) target texts, e.g., [37,135]; or (iv) to verbalize the most relevant facts according to the context, e.g., [34,54].

Whereas (iii) is typically achieved by closed planning using straightforward templates, (iv) can be achieved either by open planning using the input graph topology and pondering the traversed nodes based on some criteria such as a distance metric or a user model [34,54], or by closed planning with restrictions on the applications of each template [24,27]. Some approaches [23,24,27] combine the requirement to communicate the most typical facts in a target text (e.g., result and team names for football match summaries) with the requirement to communicate the most relevant facts.

In what follows we describe the main characteristics of these four main approaches, focusing on closed planning approaches for saying the most typical things, and on open planning approaches for saying the most relevant things. When relevant, we detail some particularly outstanding approaches.

3.1.1. Say (almost) everything

This communicative goal subsumes approaches in which there is a verbalization request for the entire input graph (e.g., [20,21,129]), the entire set of

[9] In this overview we only consider NLIs that have a true NLG component. Indeed, although some usability studies have shown the users preferences for NLIs rather than graphical or formal language interfaces (e.g., [76]), not all NLIs use NLG. For instance, many querying systems comprise parsing capabilities that allow users to input queries in natural language but the results are conveyed graphically or using tables; see, e.g., [77,132]. In some cases, some generation capabilities exist but are very limited, such as the generation of sentences or phrases from formal queries for presenting to the user using the same realization grammar as for parsing the sentences entered by the user into queries [17]. Some NLI tools for authoring are focused on the creation of new contents and therefore do not need to render any previously existing contents. These tools present contents using the same text introduced by the user when authoring them, without performing any NLG, e.g., [43,83,122].

[10] One exception with respect to input size is Sun and Mellish [129], whose input RDF graph is limited to 10 triples, because their system is a *microplanner* and the target text is a single sentence.

Approach	Input		Context					Output			
	Type	TI	T. Texts	T. Aud.	Verb. Req.	Com.Goal	UP	Flu.	Size	Coh.	Lang.
Hewlett [65]	OWL-DL ont.	Yes	No	LP, DE	Class description	Say all	No	+	P	Yes	Eng
Jarrar [72]	ORM	Yes	No	DE	Constraint	Say all	No	– –	S	N/A	Multi.
Sun [129]	RDF Graph	Yes	No	LP,DE	Graph	Say all	No	+	S	N/A	Eng
Kaljurand [74]	OWL-DL ont.	Yes	No	DE	Class description	Say all	No	–	P	Yes	Eng
Ang [2]	OWL-DL ont.	Yes	No	DE	Queries of a term	Say all	No	+	P	N/A	Eng
Mellish [96,97]	OWL-DL ont.	Yes	No	LP,DE	Class subsumers	Say most relevant facts	No	N/A	N/A	N/A	N/A
Davis [39]	OWL-DL ont.	Yes	No	DE	Classes, Properties Instance	Say all	No	–	P	No	Eng
Stevens [128]	(Subset of) OWL-DL ont.	Yes	No	DE	Class description	Say all	No	–	P	Yes	Eng
Power [112]	(restricted) OWL-DL graph	Mixed	No	DE	Class axiom Individual assertion	Say what user selects	No	–	S	N/A	Eng
Hielkema [66,67]	OWL-Lite/RDF graph	No	No	DE	Metadata description about intellectual artifact	Say what user selects	No	–	T	Yes	Eng
Dongilli [47]	OWL/RDF graph	Yes	No	DE	Query	Say what user selects	No	+	S	Yes	Eng
Wilcock [137]	RDF/XML graph, DAML-OIL ont.	Yes	No	LP	Answer to user question about individual or class	Say all about individual, Summarize class	No	+	P	Yes	Eng
Bontcheva [21]	DAML-OIL/RDF graph	Yes	Clinical case reports	DE	Graph	Say all	No	+	P	Yes	Eng
Bontcheva [20]	OWL/RDF graph	Yes	No	LP	Graph	Say all	Yes	+	T	Yes	Eng
Argüello [3]	OWL/RDF graph	No	Clinical narratives	DE	Description of current illness	Say typical facts	No	+	P	Yes	Eng
Galanis [54]	OWL-DL/RDF ont.	Mixed	No	LP,DE	Specific entity	Say most relevant facts	Yes	+	T	Yes	Eng, Gre
Bouttaz [27]	OWL/RDF graph	Yes	No	DE	Metadata description about digital artifact	Say typical, most relevant facts	Yes	N/A	N/A	N/A	N/A
Bouayad-Agha [23]	OWL-DL ont.	Mixed	Football summaries	LP	Match summary	Say typical, most relevant facts	Yes	+	P	Yes	Spa
Bouayad-Agha [24]	OWL-DL ont.	Mixed	Environmental bulletins	LP,DE	Environmental info given date and location	Say typical, most relevant facts	Yes	+	T	Yes	Eng, Fin, Swe
Weal [135]	Protegé frames RDF ont.	Yes	Artist bios (summary, chronology)	LP	Artist name	Say typical facts	Yes	+	T	Yes	Eng
Dai [34]	Semantic network	Yes	No	LP	Specific entity in network	Say most relevant facts	No	+	P	Yes	Chn, Eng
Dannélls [37]	OWL/RDF ontologies	Yes	Description of museum artifacts	LP	Museum artifact description	Say typical facts	Yes	+	P	Yes	Eng, Swe

Table 1: Summary of the Most Representative SW-oriented NLG Applications
Columns: Approach (by first referenced author), Input Type, Task Independence (TI), Target Texts (T.Texts), Target Audience (T. Aud.)={LP=Lay Person, DE=Domain Expert}, Verbalization Request (Verb. Req), Communicative Goal (Com. goal), User Profile (UP)[a], Fluency (Flu.)[b]={High(+),Medium(–),Low(– –)}, Output Size={S(entence), P(aragraph), T(ext)}, Coherence (Coh.)[c], Language (Lang.) = {Eng=English, Multi=Multilingual, Gre=Greek, Spa=Spanish, Fin=Finnish, Swe=Swedish, Chn=Chinese}

[a] User profile refers to NLG systems that generate different texts for different users given the same input using an explicit user profile/model.
[b] Our assessment of the level of fluency relies not only on our knowledge of the performed NLG tasks and their scope, but also on the sample texts, when provided in the reference articles.
[c] Coherence can only apply above the sentence level.

constraints or axioms (abox or tbox) in the domain model (e.g., [39,72,74]), all the queries of a term (e.g., [2]) or all the axioms related to a class description, e.g., [65,128]. In the first two cases, the content selection element is minimal (if at all there) and consists in eliminating redundancies [20,21]. In the last two cases, the content selection consists mainly in selecting the queries or axioms to verbalize for the given term or class description. Thus, Ang et al. [2] retrieves all the transitive query paths from the query term across its object properties. Stevens et al. [128] generate OWL class descriptions in natural language from a bio-ontology by selecting only axioms in which the class is directly involved (as either subject or object).

3.1.2. *Say what the user decides*

One of the main text planning approaches in which the user decides what to say is conceptual authoring, a term coin by Hallett et al. [61]. In conceptual authoring, a supporting NLI guides the user through the authoring process, e.g., [47,52,66,112]. The user selects or authors the concepts or individuals from the ontology to formulate a query or to design or populate the ontology by editing the underlying knowledge representation displayed to her via an interface in terms of NL statements generated automatically from the knowledge representation. The editing is done through substitutions in specific place-holder points in the rendered text, where the list of possible substitutions presented to the user is delimited by the system according to the underlying knowledge representation. Figure 2 shows an illustrative interface for conceptual authoring taken from Hallett et al.'s [61] article, in which the user is presented with a list of possible events to replace the more generic event acting as placeholder (leftside of the figure). After selecting "examined", an instance of the event underlying this verb is added to the semantic model and the feedback text is generated again to express the new event and its (not yet specified) arguments (rightside of the figure, which also shows the submenu that would appear to fill one of the new placeholders "some person").

As with other NL interfaces, there is no need to know complex query or ontological languages and so the expertise and training needed for authoring is minimal. However, what makes conceptual authoring different from other ontology editors, including the ones based on Controlled Natural Languages, is that there is no need for language interpretation. Complex knowledge configurations can be authored without the interpretation bottleneck. In addition, in the case of querying, one can be sure that the input to the system matches a valid query, thus avoiding the pitfall of "linguistic vs. conceptual failure" [53], where the user does not know whether a query failed because no data was returned or because it was not consistent with the KB schema.

Though initially applied to relational databases and pre-SW KBs [60,61,114], the use of conceptual authoring for SW ontologies was a natural step forward [47,52,66,67,112]. In some of these approaches, the authoring process is automatically supported using OWL DL standard reasoning (i.e., class satisfiability, subsumption, consistency and instance checks [47, 112]. For example, the feedback text *Mary owns an animal* is not presented if the user already specified that *Mary owns a pet*) [112].[11]

3.1.3. *Say the most typical (using closed planning)*

As already mentioned, closed planning subsumes template-based and rule-based approaches and any other approaches that do not exploit a semantic network representation of the input data. Some of these approaches use SW representations and technologies. For example, Bouttaz et al.'s [27] use SPARQL queries as content selection rules. Dannélls et al. [37] retrieve relevant triples from multiple datasets and ontologies using a single SPARQL end-point from which queries about museum artifacts can be formulated.

Bouayad-Agha et al. [24] model content selection and discourse structuring objects such as schemas, linear precedence relations and elementary discourse units, in a linguistic ontology. They use SPARQL queries to implement template-based content selection and text planning tasks wherein the NLG ontological objects and relations are instantiated.

In other approaches, coherence is attained using topic/triple ordering templates (see, e.g., [21,37]), simple ordering rules (as, e.g., class first, then properties) for prototypical texts [128,137], or partial order constraints as annotations on the ontology [46,54].

Weal et al. [135] use a template-based approach for generating biographies of artists to combine text fragments with sentences generated dynamically from the facts. Both text fragments and facts are harvested from the web using information extraction technologies and

[11] Power et al. in earlier papers (e.g., [114]) refer to their conceptual authoring approach as WYSIWYM (*What You See Is What You Meant*) given that what the user sees is the underlying knowledge, displayed in natural language.

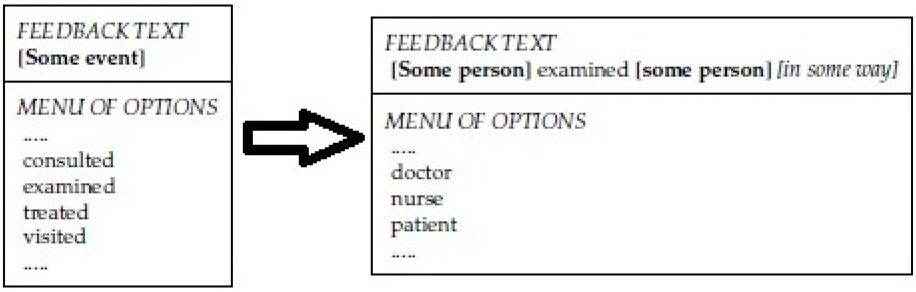

Fig. 2. Illustration of a Conceptual Authoring interface, where the user selects the event verbalized by “examined" (right) to replace the more generic placeholder (left) [61]

then stored in an ontology. In order to avoid repetition, the overlap of information between text fragments is monitored by a blackboard in which the already mentioned triples are added, such that no new text fragment that contains an already mentioned fact is included.

3.1.4. Say the most relevant (using open planning)

Given their focus on data-driven methods, i.e., section 2.2.1, open planning approaches are particularly promising in NLG from large and hetereogeneous SW datasets for which the creation of templates and rules becomes prohibitively expensive. Thus, Mellish and Pan [96,97] have addressed the problem of domain-independent content determination (i.e., selecting the content and organizing it into a coherent whole) from an OWL DL ontology for *tbox* verbalization. These authors argue that heuristic-based navigation, as used outside the scope of the SW, e.g., in O'Donnell et al.'s[108] approach (see Section 2.2.1), is not adapted to the challenge of *tbox* verbalization as it leads to overly complex sentences and possibly misleading information due to the false implicatures that can be drawn from some of the verbalizations of the ontological axioms [99]. They propose instead a new approach called “Natural Language Directed Inference" (NLDI) that selects axioms if their selection can be inferred from axioms already selected. Given an atomic concept in an OWL DL ontology, they perform a depth-first search for more general concepts that can be used to describe the concept. Their inference-based content selection method is based on criteria like preference of shorter formulas to long ones, realizability of selected axioms, and logical independence of the original axioms (such that they are included only once).

Dai et al. [34] present an approach which they call Semantic Network Language Generation (SNLG) which can be applied to the generation of texts from a generic semantic network. Starting from a node of interest in the input semantic network, they iteratively select additional nodes according to a distance function. The resulting set of nodes is then mapped to tree structures according to some patterns which can be linguistically realized as a sentence each. This mapping excludes nodes which cannot be mapped, a strategy that ensures the robustness of the system albeit at the expense of its coverage. The fluency and coherence of the resulting text are improved by applying pattern-based aggregation and Krahmer et al.'s [81] method for referring expression generation. An annotated corpus of texts is used to train the patterns used by the system components.

3.2. Sentence planning for NLG from SW data

Sentence planning approaches for NLG from SW data by and large use simple rules and templates. Thus, for packaging and aggregating information into sentences, semantic grammars [37], SPARQL rules [26], XSLT [136,137] or XML [39] templates, and aggregation patterns based on entity-sharing between triples or axioms [21,66,128] have been used.

For the mapping of content onto linguistic representations (i.e., lexicalization and choice of grammatical realization), the complexity of the approach is typically proportional to the fluency of the output, with the direct mapping of content labels onto linguistic ones on the one end of the continuum, e.g., [39, 72]. For example, Davis et al. [39] treat class names as proper names, as in “`SeniorResearcher` *attends* `Conference`". This assumption of linguistic expressibility of the ontology, whereby axioms are expressed by sentences specified by a grammar, one per axiom, and atomic terms involved in axioms are verbalized by entries in the lexicon, is called by Power [113] the *Consensus Model*. This consensus has been validated in two separate studies by Mellish and Sun [100] and Power [115] who analyzed in a collection of freely available ontologies the naming patterns of properties and classes on the one hand [100] and the complexity and hence linguistic expressibility of OWL axioms on the other hand [115]. According to these authors and other proponents of the consensus model, what these direct mapping approaches lose in output fluency they gain in domain independence and the simplicity of the engineering solution, thus reducing the cost of creating domain and ontology-specific lexicons; see e.g., [20]. Nonetheless non-strictly NLG tasks can further contribute to improve fluency, such as final spell checking [2], removal or monitoring of repetitions/redundancies [20,21,65,135] or presentation of sentence coordinated constituents in a bulleted list [65].

In what follows, we discuss the main different approaches to mapping SW content onto linguistic representation, elaborating first on the consensus model approach, including the use of Controlled Natural Languages in mapping grammars for axiom verbalization. We then discuss approaches that annotate content with linguistic knowledge and finally, approaches that use upper models as an intermediate representation between content and linguistic representation.

3.2.1. The consensus model

Under the consensus model, the fluency of the linguistic output can be somewhat improved by exploiting the linguistic patterns of properties and class names. For example according to Mellish and Sun [100], the following OWL constraint:

```
restriction(Onproperty(hasProducer),
            allValuesFrom(French))
```

can be verbalized as “has a producer who is French", instead of something like “has a property, hasProducer, which must have as its value, something that is in the class French", given that the syntax of the property and class names can be interpreted and their part of speech provided by WordNet. In addition to identifying patterns in ontologies, on-line lexical resources such as WordNet [130] or FrameNet [36] have also been used to associate these patterns with valency information (i.e., argument structure).

Some researchers suggest that automatic lexico-grammatical derivation from labels in ontologies can only be used as a fallback if no manual annotation is present in the ontology for a given property/concept [121,128,139] or that is should be accompanied by a revision from language experts to ensure quality [20]. Others prescribe that naming conventions that restrict the grammatical category and composition of terms should be enforced when authoring ontologies [65,112]. Meanwhile, a workaround for this lack or deficiency of naming conventions is to involve (expert) users of the NLG system in the acquisition of linguistic resources for each domain, as is envisaged in conceptual authoring [66,112] or some ontology and rule editors (e.g., [40]), which support user's creation of new entries in the lexicon for the newly added ontology concepts.

A popular strategy to reduce the amount of task-specific knowledge needed to generate language is the reduction of the generated language constructions to a controlled subset, so-called *Controlled Natural Language* or *CNL*, for which an unambiguous mapping from the formal languages used in the Semantic Web can be defined, e.g., [4,39,40,63,74]. This is especially desirable in ontology verbalization, where the generated text must be unambiguous, where there are no requirements for generating a coherent text and where “round trip” authoring is desirable. In “round trip" authoring, the ontology is verbalized into a CNL and can be edited in this way by an ontology engineer before being translated back into ontology axioms. Examples of CNLs used in bidirectional OWL$< - >$CNL verbalizers/parsers include Kaljurand et al.'s [40,74] *Attempto Controlled English* (ACE), a subset of which is used in the round-trip meaning preserving mapping with OWL, and Davis et al.'s [39] CLOnE, a very simple CNL with a bidirectional mapping to only a small subset of OWL.

Finally, it is important to note that the consensus model exploits patterns in ontologies developed essentially in English, and that the CNLs used are all subsets of English. Therefore, most approaches to ontology engineering verbalize in English, as Table 1 shows. There are of course some exceptions, such as Jarrar et al.'s [72] multilingual, albeit simple verbalization of logical theories using XML templates.

3.2.2. Annotating content with linguistic knowledge

Instead of keeping the lexicon separate from the domain data, some approaches opt for annotating the domain data with lexical information within the same SW representation. This is, for instance, the case of NaturalOWL [54], where classes and individuals in the OWL ontology are associated with noun phrases together with the gender of the head nouns and their singular and plural forms. Properties are assigned micro-plans for sentence planning. These micro-plans define an abstract clause structure with a verb as its head and information about the verb's inflection and valency. An example micro-plan for the property *#man-*

ufacturedBy is given in Figure 3. It is basically a template in the form of an RDF annotation which specifies that, in English, this property should be rendered by the slotted sentence "`<NP> was manufactured by <NP>`" (the information on the passiveness of the verb will become useful in case of aggregation to form longer sentences).

```
<owlnl:property rdf:about="...#manufacturedBy">
  <owlnl:order>1</owlnl:order>
  <owlnl:EnglishMicroplans ...>
    <owlnl:microplan ...>
      <owlnl:aggrAllowed>true</owlnl:aggrAllowed>
      <owlnl:slots ...>
        <owlnl:owner>
          <owlnl:case>nominative</owlnl:case>
          <owlnl:retype>re_auto</owlnl:retype>
        </owlnl:owner>
        <owlnl:verb>
          <owlnl:voice>passive</owlnl:voice>
          <owlnl:tense>past</owlnl:tense>
          <owlnl:val>was manufactured</owlnl:val>
        </owlnl:verb>
        <owlnl:text>
          <owlnl:val>by</owlnl:Val>
        </owlnl:text>
        <owlnl:filler>
          <owlnl:case>accusative</owlnl:case>
          <owlnl:retype>re_auto</owlnl:retype>
        </owlnl:filler>
      </owlnl:slots>
    </owlnl:microplan>
  </owlnl:EnglishMicroplans>
  <owlnl:GreekMicroplans ...>
  ...
</owlnl:Property>
```

Fig. 3. A micro-plan (or template) for the property *#manufacturedBy* for the NaturalOWL System [54]

For the purpose of annotating ontologies, the authors of NaturalOWL developed a tool [1,18,55] supported by reasoning that can be used to assign multilingual annotations to OWL ontologies.

This tight integration of linguistic and domain knowledge, although probably motivated on particular engineering grounds within single systems, raises some theoretical concerns and, on more practical grounds, limit the reusability of the linguistic data in different application settings which may require different mappings.

3.2.3. *Use of an upper model as an intermediate representation*

As discussed in Section 2.2.3, the use of task- and domain-independent linguistically-oriented representations such as the (Generalized) Upper Model [7,10] facilitates the mapping between content and linguistic representation, particularly if an off-the-shelf linguistic generator is to be used.

A practical approach taken by some researchers is not to use a full-blown upper model but instead to use a reduced set of upper concepts to support the mapping and portability of the NLG system. Thus, in MIAKT [21], one of the first implementations of NLG applications for report generation from existing SW medical ontologies, the surface generator HYLITE+ is applied after mapping all relations in the input ontologies to one of four generic and linguistically-motivated relations for which the surface generator has in-built support. This approach is extended in ONTOSUM [20], where the text and sentence planning modules can operate on any ontology that contains the same four upper relations. Bouayad-Agha et al. [23,24] use a linguistic generator based on a Meaning to Text Theory (MTT) model [101] that takes as input a conceptual representation in the sense of [125]. The authors suggest, for the future, to model the conceptual representation in the linguistic ontology already in use for content selection and text planning (see Section 3.1.3 above).

4. Towards a symbiosis between NLG and SW

The Semantic Web, with its large amount of heterogeneous task-independent data, demands portable and robust NLG approaches. Therefore, it is not surprising that, under the auspices of the Semantic Web, addressing the long-term issues for NLG presented in Section 2.3 becomes more critical. At the same time, the Semantic Web provides some support and opportunities to address these issues, in particular:

- The codification of NLG-related knowledge such as rules and templates has been facilitated by well-known APIs and standard query languages (e.g., SPARQL for RDF) that query data which is structured following a standard syntax (e.g., [24,27]). Furthermore both NLG-related (i.e., communication) knowledge and domain communication knowledge can be modeled in separate ontologies and integrated using the OWL import mechanism, which provides a limited form of modularization of knowledge.

- The availability of vast amounts of Linked Open Data (LOD), which use de-facto or standard vocabularies and ontologies (e.g., FOAF, Dublin Core, Geonames), should encourage knowledge reuse and the scaling up of NLG systems. As we have seen in Sections 2.2.1 and 3.1.4, open planning approaches are particularly suitable to handle these large graph-based datasets [34,42,54,96,97,108].

- The availability of vast amounts of hypertext documents, related or not to LO data, should encourage the development of empirical NLG approaches and promote better evaluability. So far, Dai et al.'s [34] generation system is the only SW-oriented NLG approach that performs an automatic evaluation of the generated texts by comparing them against their corresponding Wikipedia items using similarity metrics.

- Data assessment and interpretation of domain knowledge is greatly facilitated with the availability of off-the-shelf DL reasoners and rule engines, whether to perform standard DL reasoning operations such as consistency checks and instance classification over the input data (e.g., [18,52,53]), to apply domain-specific rules that extend domain knowledge with new domain communication knowledge (e.g., [25,26]) or to support content selection with subsumption reasoning (e.g., [96,97,99]).

However more needs to be done before we can speak of a real symbiosis between NLG and SW. These ongoing challenges are: (i) the codification and modeling of NLG-relevant knowledge using SW standards and formalisms, (ii) the use of linked open data and associated texts, (ii) the summarization of large volumes of data using NLG techniques, (iii) the combination of content distillation and generation, and (iv) more adaptation to context. We address each of these issues in turn below.

4.1. Codification and modeling of NLG-relevant knowledge in SW standards and formalisms

In generation from SW data, the interfacing between content and linguistic representations has mimicked traditional knowledge-based NLG, from using entity labels to derive lexical entries and mapping axioms to syntactic constructs (i.e., the so-called consensus model) to enriching the content with language-oriented annotations to the use of upper models (see Section 3.2). However, the codification and modeling of NLG-relevant knowledge in SW standards and formalisms can provide potential benefits in the interoperability between NLG modules and tasks and in the reuse of linguistic resources across applications.

Cross-domain linguistically motivated upper ontologies that predate the appearance of SW standards, have been published, at least partially, in the OWL language, e.g., the *Generalized Upper Model* [9][12] and the *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) [90][13]. Lexical databases which are commonly used by the NLP community have also been converted to RDF/OWL, i.e. WordNet [14] and Framenet [15]. In addition to these resources, recent efforts have been invested in engineering models that enforce a clear separation between (multilingual) linguistic information (i.e., lexicons and lexico-grammatical realizations) and its mapping to domain data, see for instance LexOnto [32], LexInfo [29],[16] and the *Linguistic Information Repository* (LIR) [103]. A concerted effort is needed to raise awareness of these new or old-as-new SW resources and to encourage their integration in SW-oriented NLG systems.

Closely related to the question of the codification and modeling of NLG-relevant information is the question of the input/output interface representations of individual modules in NLG whose standardization would help promote inter-operability between systems. However as we discussed in Section 2.3, efforts such as RAGS to provide an NLG reference architecture and more specifically the RAGS representation models have failed to be taken up by the NLG community. According to Mellish [94], one of the main authors of RAGS, this is due to the complexity of the framework, lack of tools to support its implementation (i.e., APIs in different programming languages, consistency checking, query engines, etc.), its idiosyncratic use of XML and its inability to define how to bring together RAGS representations with non-RAGS ones. Mellish suggests that these difficulties can be remedied by recasting RAGS data quite naturally in terms of RDF and formalizing the RAGS representations using SW-ontologies.

4.2. Use of linked open data and associated texts

Virtually all efforts in NLG from SW datasets are still restricted to isolated datasets, leaving much ground to cover before mature technologies are available for the production of text from Linked Data. When NLG taps the full potential of linked data, it will have to adapt to vast amounts of data belonging to multiple domains, described using multiple vocabular-

[12]http://www.ontospace.uni-bremen.de/ontology/gum.html
[13]http://www.loa.istc.cnr.it/DOLCE.html
[14]http://www.w3.org/TR/wordnet-rdf
[15]http://framenet.icsi.berkeley.edu
[16]http://lexinfo.net

ies encoded in knowledge representations of varying degrees of expressivity.[17] This will have an impact not only on the interpretation, assessment and selection of contents, but also on other NLG tasks which may operate or reason on the same input representation—for instance, ordering contents for their inclusion in the text or determining their lexical realization.

Recent NLG research is only beginning to address the use of multiple linked datasets. Dannélls et al. [37] approach multilingual generation from multiple independent datasets by adopting a unifying view based on the PROTON [18] upper ontology. This view allows them to treat domain-specific and generic datasets as a single body of knowledge with respects to querying and reasoning. For the verbalization of SPARQL queries, Ell et al. [51] exploit the mapping from DBPedia entities to types in the YAGO [19] ontology to lexicalize entities found in the queries. We believe that the use of upper ontologies as a means of integrating multiple datasets is likely to become common practice for NLG systems drawing from multiple sources.

A prominent feature of a number of open linked datasets is that the data they contain is related to Hypertext Web documents. For instance, entities in DBPedia and topics in Freebase are linked to the corresponding Wikipedia articles. The text contained in these documents constitutes potential training and evaluation material for empirical NLG methods. Texts could be used to learn what data is most relevant, how it is ordered, and which are the linguistic expressions that are used to communicate information. Some approaches outside the scope of the Semantic Web have explored the creation of corpora of text aligned with data and its use in some NLG tasks (as, e.g., [5] for the selection of contents from databases). Similar methods could also be applied in the context of Linked Data and Web documents. The recently announced content selection challenge from SW data paired with corresponding texts [22] is expected to advance the state-of-the-art in this field and bring the NLG and SW communities closer to each other.

4.3. Summarization of large volumes of data

We have already argued that the Semantic Web offers the chance and, at the same time, poses the challenge to scale up NLG-techniques to large volumes of data. Conversely, there is a growing need in the SW community for technologies that give humans easy access to the machine-oriented Web of data [64]. NLG provides the means for presenting semantic data in an organized, coherent and accessible way thanks to its capabilities of rendering content as multilingual or even multimodal information. NLG approaches have started to address the generation of information from large/multiple datasets (i.e., Sections 2.2.1 and 3.1.4). These approaches can be further developed on SW data to generate summaries that communicate the most important content in a dataset while filtering out the less relevant content.

Approaches to ontology summarization can also contribute to the generation of summaries from large datasets. While strategies for content selection in NLG take as criteria what content is communicated in target texts, approaches to ontology summarization employ topographical measures over RDF graphs in order to assess what content is most important; see, e.g. [31,141]. Ontology summarization could be used to leverage NLG and improve the suitability of text-based summaries as an accessible way of presenting data to humans.

4.4. Combining content distillation and generation

While vocabularies in the SW are typically manually crafted, often from the analysis or manual annotation of existing texts, datasets are often derived automatically from databases. Initiatives like the DBPedia also extract data from HTML markup of Wikipedia pages. Bontcheva et al. [19] and Wilks and Brewster [138] go further, arguing that, since the internet consists mainly of unstructured texts, the Semantic Web should and will tap into this vast pool of knowledge to build its datasets using techniques such as Information Extraction (IE) or Text Mining (TM).

The semantic content thus obtained, Bontcheva et al. [19] further argued, can be used to generate texts, thus giving rise to a so-called *language loop*. In fact, NLG plays an increasingly central role in the language loop because many original web texts require a para-

[17] More advanced forms of modularization will have to be used to integrate different ontologies and support the separation between different types of knowledge (i.e., domain and task-related knowledge), see for example the ϵ-connection framework [84]. Furthermore, in order for NLG systems to be able to reason over an entity across different datasets, identity of reference between entities of different datasets will have to be addressed by the SW community, i.e., Halpin et al.'s [62] discussion of the diverse use of the `owl:sameAs` property, which does not indicate only full identity.

[18] http://proton.semanticweb.org

[19] http://www.mpi-inf.mpg.de/yago-naga/yago/

phrase, summarization or translation in order to serve the needs of the targeted user. The Semantic Web acts as a sort of *interlingua* which helps in bridging the gap between the source and destination texts. NLG is thus to be seen here as regeneration guided by SW data, where NLG techniques are combined with, parsing, IE and TM techniques in order to produce new textual material. This material may contain both dynamically generated text and text fragments obtained from original Web documents. To the best of our knowledge, Weal et al. [135] (presented in Section 3.1.3) and Dai et al. [34] (presented in Section 3.1.4) are the only ones exploring regeneration.

4.5. Adapting to context

Tailoring the content and wording to the preferences and particularities of end users as well as to other contextual conditions is instrumental for personalized NLG. NLG has a long tradition of working on experimental techniques that adapt the output to the context by varying the information communicated in the text, its order and general organization, the language, the general writing style and lexical choices, and even the mode (e.g. text, images). As Table 1 shows, many of the SW-oriented NLG proposals are also, to a certain extent, context-sensitive. The contextual information they deal with includes the target language and type of user and specific request (as, e.g. [24]), user preferences and partial order restrictions for certain types of contents (as, e.g. [54]), or a history of the content communicated to a specific user (as, e.g. [35]). This information is used to influence common text planning tasks such as content selection or content ordering or to control the use of certain HTML-layout features (as in ONTOSUM [20]). However, the degree of their context orientation is, in general, considerably lower than in traditional (admittedly much more experimental) implementations.

However, as in the case of NLG-related knowledge (see Section 4.1 above), the Semantic Web constitutes an excellent means for the modeling of detailed user-oriented contexts. In this way, reusable models of the context can be seamlessly integrated with input data for querying and reasoning purposes, facilitating the application of adaptive NLG methods. This is illustrated in Bouttaz et al.'s approach [27], where complex contextual knowledge for content selection (i.e. policies based on provenance information) is encoded using SW technologies and is in part described using existing vocabularies such as FOAF. In Bouayad-Agha et al.'s approach [24], the modeling of the user is part of the ontology, and therefore allows for a unified ontology-based decision process. We are convinced that the codification of user models in SW formalisms can help to address adaptation of NLG systems to context.

5. Conclusions

In this survey paper we hope to have given the SW engineer/researcher a useful tour of the field of NLG from a semantic input representation in general and from SW data in particular. In our overview of existing paradigms for generating texts from Semantic Web data, we have presented several approaches that use existing trends in NLG, adapting them but also extending them to standard SW formalisms and technologies, therefore achieving a better inter-operability between the SW input and NLG-dependent representations.

We have argued that addressing the issues that have long weighed down on NLG research and applications, such as portability, robustness and evaluability, is critical for NLG systems aimed at Semantic Web and Linked Open Data. At the same time, the Semantic Web and Linked Open Data framework is an opportunity for NLG systems to address these issues, thanks to the availability of SW formalisms and technologies and vast amounts of task-independent inter-connected knowledge with or without associated texts.

One of the original and ultimate goals of the Semantic Web was to allow agents to carry out sophisticated tasks for users based on retrieving and processing meaningful content, and communicating the results in a user-friendly way [16]. Some applications of this kind have been developed in limited domains and for limited tasks; see, for instance Xu et al's system [140] that answers questions about pop trivia, Janzen and Maas' system [71] that answers questions about products for a virtual in-store shopping environment, or Weal et al.'s ArtEquAkt system [135] that combines SW knowledge and text fragments to generate adaptive artist biographies. The NLG components are specific to these applications and rely on relatively simple templates. We are still a long way from embodied characters that engage in conversation in a multimodal way to present informative content tailored to the context [33]. However this vision will only be attainable if the Semantic Web and Natural Language Generation community join their efforts to develop robust data-driven approaches and standardized representations of input (and also certain intermediate) structures, context, and processes.

References

[1] ANDROUTSOPOULOS, I., KALLONIS, S., AND KARKALETSIS, V. Exploiting OWL Ontologies in the Multilingual Generation of Object Descriptions. In *Proceedings of the 10th European Workshop on Natural Language Generation* (2005), pp. 150–155.

[2] ANG, W., KANAGASABAI, R., AND BAKER, C. Knowledge Translation: Computing the Query Potential of Bio-ontologies. In *Proceedings of the International Workshop on Semantic Web Applications and Tools for Life Sciences (SWAT4LS)* (2008).

[3] ARGÜELLO, M. F., DES, J., PRIETO, M. J. F., PEREZ, R., AND LEKKAS., S. An Ontology-Based Approach to Natural Language Generation from Coded Data in Electronic Health Records. In *Proceedings of the UKSim 5th European Symposium on Computer Modeling and Simulation* (2011), pp. 366–371.

[4] BAO, J., SMART, P., BRAINES, D., AND SHADBOLT, N. A Controlled Natural Language Interface for Semantic Media Wiki Using the Rabbit Language. In *Workshop on Controlled Natural Language (CNL '09)* (Marettimo Island, Italy, June 2009), pp. 206—225.

[5] BARZILAY, R., AND LAPATA, M. Collective Content Selection for Concept-to-Text Generation. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)* (2005), pp. 331–338.

[6] BASILE, V., AND BOS, J. Towards Generating Text from Discourse Representation Structures. In *Proceedings of the 13th European Workshop on Natural Language Generation* (Nancy, France, 2011), pp. 145–150.

[7] BATEMAN, J. Upper Modelling: Organizing Knowledge for Natural Language Processing. In *Proceedings of the 5th International Workshop on Natural Language Generation, 3-6 June 1990* (1990), pp. 54–60.

[8] BATEMAN, J. Enabling Technology for Multilingual Natural Language Generation: The KPML Development Environment. *Natural Language Engineering 3*, 1 (1997), 15–55.

[9] BATEMAN, J. A., HOIS, J., ROSS, R., AND TENBRINK, T. A Linguistic Ontology of Space for Natural Language Processing. *Artificial Intelligence 174*, 14 (2010), 1027 – 1071.

[10] BATEMAN, J. A., MAGNINI, B., AND FABRIS, G. The Generalized Upper Model Knowledge Base: Organization and Use. In *Towards Very Large Knowledge Bases*, N. Mars, Ed. IOS Press, Amsterdam, 1995, pp. 60–72.

[11] BELZ, A. Automatic Generation of Weather Forecast Texts using Comprehensive Probabilistic Generation-space Models. *Natural Language Engineering 14*, 4 (1997), 431–455.

[12] BELZ, A., BOHNET, B., MILLE, S., WANNER, L., AND WHITE, M. The Surface Realization Task: Recent Developments and Future Plans. In *Proceedings of the 7th International Conference on Natural Language Generation (INLG '12)* (Starved Rock, IL, 2012), pp. 136–140.

[13] BELZ, A., AND REITER, E. Comparing Automatic and Human Evaluation in NLG. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistic* (2006), EACL'06, pp. 313–320.

[14] BELZ, A., AND REITER, E. Comparing Automatic and Human Evaluation in NLG. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)* (2006), pp. 60–72.

[15] BELZ, A., WHITE, M., ESPINOSA, D., KOW, E., HOGAN, D., AND STENT, A. The First Surface Realization Shared Task: Overview and Evaluation Results. In *Proceedings of the 13th International Workshop on Natural Language Generation* (Nancy, France, 2011), pp. 217–226.

[16] BERNERS-LEE, T., HENDLER, J., AND LASSILA, O. The Semantic Web. *Scientific American* (2001), 29–37.

[17] BERNSTEIN, A., KAUFMANN, E., AND KAISER, C. Querying the Semantic Web with Ginseng: A Guided Input Natural Language Search Engine. In *15th Workshop on Information Technologies and Systems* (Las Vegas, NV, 2005), pp. 112–126.

[18] BILIDAS, D., THEOLOGOU, M., AND KARKALETSIS, V. Enriching OWL Ontologies with Linguistic and User-Related Annotations: The ELEON System. In *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence, ICTAI* (October 2007), vol. 2, pp. 464 –467.

[19] BONTCHEVA, D. K., AND CUNNINGHAM, H. The Semantic Web: A New Opportunity and Challenge for Human Language Technology. In *Workshop on Human Language Technology for the Semantic Web and Web Services. Held in conjunction with the Second International Semantic Web Conference (ISWC'03)* (Florida, USA, October 2003), H. Cunningham, Y. Ding, and A. Kiryakov, Eds., pp. 89–96.

[20] BONTCHEVA, K. Generating Tailored Textual Summaries from Ontologies. In *The Semantic Web: Research and Applications, Proceedings of the Second European Semantic Web Conference, ESWC* (Heraklion, Crete, Greece, 2005), pp. 531–545.

[21] BONTCHEVA, K., AND WILKS, Y. Automatic Report Generation from Ontologies: the MIAKT Approach. In *Natural Language Processing and Information Systems, Proceedings of the Nineth International Conference on Applications of Natural Language to Information Systems (NLDB)* (2004), vol. 3136 of *Lecture Notes in Computer Science*, Springer, pp. 324–335.

[22] BOUAYAD-AGHA, N., CASAMAYOR, G., MELLISH, C., AND WANNER, L. Content Selection from Semantic Web Data. In *Proceedings of the Seventh International Natural Language Generation Conference (INLG), Special Track on Future Generation Challenges Proposals* (2012), pp. 146–149.

[23] BOUAYAD-AGHA, N., CASAMAYOR, G., MILLE, S., AND WANNER, L. Perspective-Oriented Generation of Football Match Summaries: Old Tasks, New Challenges. *ACM Transactions on Speech and Language Processing (TSLP) 9*, 2 (2012), 3:1–3:31.

[24] BOUAYAD-AGHA, N., CASAMAYOR, G., ROSPOCHER, M., SAGGION, H., SERAFINI, L., AND WANNER, L. 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'2012). Lecture Notes in Computer Science series, vol. 7337* (Groningen, Netherlands, 2012), pp. 216–221.

[25] Bouayad-Agha, N., Casamayor, G., and Wanner, L. Content Determination from an Ontology-based Knowledge Base for the Generation of Football Summaries. In *Proceedings of the 13th European Natural Language Generation Workshop (ENLG)* (Nancy, France, 2011), pp. 27–81.

[26] Bouayad-Agha, N., Casamayor, G., Wanner, L., Díez, F., and Hernández, S. FootbOWL: using a Generic Ontology of Football Competition for Planning Match Summaries. In *Proceedings of the Extended Semantic Web Conference (ESWC)* (Heraklion, Crete, Greece, 2011), pp. 230–244.

[27] Bouttaz, T., Pignotti, E., Mellish, C., and Edwards, P. A Policy-Based Approach to Context Dependent Natural Language Generation. In *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG 2011)* (Nancy, France, 2011), pp. 151–157.

[28] Buchanan, B. G., Moore, J. D., Forsythe, D. E., Carenini, G., Ohlsson, S., and Banks, G. An Intelligent Interactive System for Delivering Individualised Information to Patients. *Artificial Intelligence in Medicine 7* (1995), 117–54.

[29] Buitelaar, P., Cimiano, P., Haase, P., and Sintek, M. Towards Linguistically Grounded Ontologies. In *Proceedings of the 6th European Semantic Web Conference (ESWC) on The Semantic Web: Research and Applications* (Heraklion, Crete, Greece, 2009), pp. 111–125.

[30] Bunt, H. Utterance Generation from Semantic Representations Augmented with Pragmatic Information. In *Natural Language Generation*, G. Kempen, Ed. Martinus Nijhoff Publishers, 1987, pp. 333–348.

[31] Cheng, G., Ge, W., and Qu, Y. Generating Summaries for Ontology Search. In *Proceedings of the 20th international conference companion on World wide web* (New York, NY, USA, 2011), WWW '11, ACM, pp. 27–28.

[32] Cimiano, P., Haase, P., Herold, M., Mantel, M., and Buitelaar, P. Lexonto: A Model for Ontology Lexicons for Ontology-Based NLP. In *Proceedings of the OntoLex (from Text to Knowledge: the Lexicon/Ontology Interface) Workshop at the International Semantic Web Conference (ISWC)* (2007), pp. 1–12.

[33] Cimiano, P., and Kopp, S. Accessing the Web of Data through Embodied Virtual Characters. *Semantic Web 1*, 1-2 (2010), 83–88.

[34] Dai, Y., Zhang, S., Chen, J., Chen, T., and Zhang, W. Semantic Network Language Generation Based on a Semantic Networks Serialization Grammar. *World Wide Web 13*, 3 (2010), 307–341.

[35] Dannélls, D. The Value of Weights in Automatically Generated Text Structures. *Computational Linguistics and Intelligent Text Processing* (2009), 233–244.

[36] Dannélls, D. Applying Semantic Frame Theory to Automate Natural Language Template Generation from Ontology Statements. In *Proceedings of the 6th International Natural Language Generation Conference* (Stroudsburg, PA, USA, 2010), INLG '10, Association for Computational Linguistics, pp. 179–183.

[37] Dannélls, D., Damova, M., Enache, R., and Chechev, M. Multilingual Online Generation from Semantic Web Ontologies. In *Proceedings of the 21st international conference companion on World Wide Web* (New York, NY, USA, 2012), WWW '12 Companion, ACM, pp. 239–242.

[38] Davies, J., Duke, A., Kings, N., Mladenic, D., Bontcheva, K., Grcar, M., Benjamins, R., Contreras, J., Civico, M., and Glover, T. Next Generation Knowledge Access. *Journal of Knowledge Management 9*, 5 (Oct. 2005), 64–84.

[39] Davis, B., Iqbal, A., Funk, A., Tablan, V., Bontcheva, K., Cunningham, H., and Handschuh, S. RoundTrip Ontology Authoring. In *Proceedings of the International Semantic Web Conference (ISWC)*, vol. 5318 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2008, pp. 50–65.

[40] De Coi, J., Fuchs, N., Kaljurand, K., and Kuhn, T. Controlled English for Reasoning on the Semantic Web. In *Semantic techniques for the web, the REWERSE Perspective*, F. Bry and J. Maluszynski, Eds. Springer, 2009, pp. 276–308.

[41] Deemter, K. V., Krahmer, E., and Theune, M. Real vs. Template-based NLG: a False Opposition? *Computational Linguistics 31*, 1 (2005), 15–24.

[42] Demir, S., Carberry, S., and McCoy, K. A Discourse-aware Graph-based Content Selection Framework. In *Proceedings of the 6th International Natural Language Generation Conference (INLG)* (2010), pp. 17–27.

[43] Denaux, R., Dimitrova, V., Cohn, A., Dolbear, C., and Hart, G. Rabbit to OWL: Ontology Authoring with a CNL-Based Tool. In *Workshop on Controlled Natural Language (CNL)* (Marettimo Island, Italy, 2009), pp. 246–264.

[44] Dethlefs, N., and Cuayáhuitl, H. Hierarchical Reinforcement Learning and Hidden Markov Models for Task-oriented Natural Language Generation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2* (Stroudsburg, PA, USA, 2011), HLT '11, pp. 654–659.

[45] Dimitromanolaki, A., and Androutsopoulos, I. Learning to Order Facts for Discourse Planning in Natural Language Generation. In *In Proceedings of the 9th European Workshop on Natural Language Generation (ENLG-2003) at EACL 2003* (2003), pp. 23–30.

[46] Dongilli, P. Natural Language Rendering of a Conjunctive Query. *KRDB Research Centre Technical Report No. (KRDB08-3). Bozen, IT: Free University of Bozen-Bolzano* (2008).

[47] Dongilli, P., and Franconi, E. An Intelligent Query Interface with Natural Language Support. In *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference* (Florida, USA, 2006), FLAIRS'06, pp. 658–663.

[48] Duboue, P. A., and McKeown, K. Content Planner Construction via Evolutionary Algorithms and a Corpus-based Fitness Function. In *In Proceedings of the 2 nd International Natural Language Generation Conference (INLG'02)* (2002), pp. 89–96.

[49] Duboue, P. A., and McKeown, K. R. Empirically Estimating Order Constraints for Content Planning in Generation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics* (Stroudsburg, PA, USA, 2001), ACL '01, Association for Computational Linguistics, pp. 172–179.

[50] Elhadad, M., McKeown, K., and Robin, J. Floating Constraints in Lexical Choice. *Computational Linguistics 23*, 2 (1997), 195–239.

[51] ELL, B., VRANDECIC, D., AND SIMPERL, E. SPARTIQULATION: Verbalizing SPARQL Queries. *Interacting with Linked Data (ILD 2012)* (2012), 50.

[52] FRANCONI, E., GUAGLIARDO, P., AND TREVISAN, M. An Intelligent Query Interface Based on Ontology Navigation. In *Proceedings of the Second International Workshop on Visual Interfaces to the Social and Semantic Web (VISSW)* (Hong Kong, 2010), vol. 10.

[53] FRANCONI, E., GUAGLIARDO, P., AND TREVISAN, M. Quelo: a NL-based Intelligent Query Interface. In *Proceedings of the Second Workshop on Controlled Natural Languages* (Marettimo Island, Sicily, Italy, 2010).

[54] GALANIS, D., AND ANDROUTSOPOULOS, I. Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies: the NaturalOWL System. In *Proceedings of the 11th European Workshop on Natural Language Generation* (2007), ENLG '07, pp. 143–146.

[55] GALANIS, D., KARAKATSIOTIS, G., LAMPOURAS, G., AND ANDROUTSOPOULOS, I. An Open-source Natural Language Generator for OWL Ontologies and its Use in Protégé and Second Life. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session* (Stroudsburg, PA, USA, 2009), EACL '09, Association for Computational Linguistics, pp. 17–20.

[56] GARDENT, C., BANIK, E., AND PEREZ-BELTRACHINI, L. Natural Language Generation and Natural Language Interfaces to Knowledge Bases, 2011. Tutorial at the Sixth International Conference on Knowledge Capture (K-CAP).

[57] GATT, A., AND BELZ, A. Introducing Shared Tasks to NLG: The TUNA Shared Task Evaluation Challenges. In *Empirical Methods in Natural Language Generation* (2010), pp. 264–293.

[58] GATT, A., AND REITER, E. SimpleNLG: a Realisation Engine for Practical Applications. In *Proceedings of the 12th European Workshop on Natural Language Generation* (2009), ENLG '09, pp. 90–93.

[59] GOLDMAN, N. Conceptual Generation. In *Conceptual Information Processing*, R. Schank, Ed. North-Holland Publishing Co, Amsterdam, 1975, pp. 5–21.

[60] HALLETT, C., SCOTT, D., AND POWER, R. Intuitive Querying of e-Health Data Repositories. In *Proceedings of the UK E-Science All-Hands Meeting* (Notthingham,UK, 2005).

[61] HALLETT, C., SCOTT, D., AND POWER, R. Composing Queries Through Conceptual Authoring. *Computational Linguistics 33*, 1 (2007), 105–133.

[62] HALPIN, H., AND HAYES, P. J. When OWL:sameAs isn't the Same: An Analysis of Identity Links on the Semantic Web. In *Linked Data on the Web WWW2010 Workshop (LDOW2010)* (2010).

[63] HART, G., JOHNSON, M., AND DOLBEAR, C. Rabbit: Developing a Control Natural Language for Authoring Ontologies. In *Proceedings of the 5th European Semantic Web Conference (ESWC)* (2008), pp. 348–360.

[64] HEATH, T., DOMINGUE, J., AND SHABAJEE, P. User Interaction and Uptake Challenges to Successfully Deploying Semantic Web Technologies. In *Proceedings of the Third International Semantic Web User Interaction Workshop (SWUI@ISWC2006)* (2006).

[65] HEWLETT, D., KALYANPUR, A., KOLOVSKI, V., AND HALASCHEK-WIENER, C. Effective NL Paraphrasing of Ontologies on the Semantic Web. In *Workshop on End-User Semantic Web Interaction, 4th Internatioanl Semantic Web conference (ESWC)* (Galway, Ireland, 2005).

[66] HIELKEMA, F., MELLISH, C., AND EDWARDS, P. Using WYSIWYM to Create an Open-ended Interface for the Semantic Grid. In *Proceedings of the Eleventh European Workshop on Natural Language Generation* (Stroudsburg, PA, USA, 2007), ENLG '07, Association for Computational Linguistics, pp. 69–72.

[67] HIELKEMA, F., MELLISH, C., AND EDWARDS, P. Evaluating an Ontology-driven WYSIWYM Interface. In *Proceedings of the Fifth International Natural Language Generation Conference* (2008), INLG '08, pp. 138–146.

[68] HORACEK, H. The Architecture of a Generation Component in a Complete Natural Language Dialogue System. In *Current Research in Natural Language Generation*, R. Dale, C. Mellish, and M. Zock, Eds. Academic Press, 1990, pp. 193–227.

[69] HOVY, E. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum, Hillsdale, New Jersey, 1988.

[70] HOVY, E. Approaches to the Planning of Coherent Text. In *Natural Language Generation in Artificial Intelligence and Computational Lingiustics*, C. Paris, W. Swartout, and W. Mann, Eds. Kluwer Academic Publishers, Dordrecht, 1991, pp. 83–102.

[71] JANZEN, S., AND MAASS, W. Ontology-Based Natural Language Processing for In-store Shopping Situations. In *Semantic Computing, 2009. ICSC '09. IEEE International Conference on* (2009), pp. 361–366.

[72] JARRAR, M., MARIA, C., AND DONGILLI, K. Multilingual Verbalization of ORM Conceptual Models and Axiomatized Ontologies. Tech. rep., Vrije Universiteit, Brussel, February 2006.

[73] JORDAN, P. W., AND WALKER, M. A. Learning Content Selection Rules for Generating Object Descriptions in Dialogue. *Journal of Artificial Intelligence Research 24* (2005), 157–194.

[74] KALJURAND, K., AND FUCHS, N. E. Verbalizing OWL in Attempto Controlled English. In *Proceedings of Third International Workshop on OWL: Experiences and Directions* (Innsbruck, Austria, 2007).

[75] KASPER, R. A Flexible Interface for Linking Applications to PENMAN's Sentence Generator. In *Proceedings of the DARPA Workshop on Speech and Natural Language* (1989), pp. 153–158.

[76] KAUFMANN, E., AND BERNSTEIN, A. Evaluating the Usability of Natural Language Query Languages and Interfaces to Semantic Web Knowledge Bases. *Web Semantics 8*, 4 (Nov. 2010), 377–393.

[77] KAUFMANN, E., BERNSTEIN, A., AND ZUMSTEIN, R. Querix: a Natural Language Interface to Query Ontologies based on Clarification Dialogs. In *Proceedings of the 5th International Semantic Web Conference ISWC* (2006), pp. 980–981.

[78] KITTREDGE, R., KORELSKY, T., AND RAMBOW, O. On the Need for Domain Communication Knowledge. *Computational Intelligence 7*, 4 (1991), 305 – 314.

[79] KONSTAS, I., AND LAPATA, M. Unsupervised Concept-to-text Generation with Hypergraphs. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (Montréal, Canada, June 2012), Associa-

tion for Computational Linguistics, pp. 752–761.

[80] KOSSEIM, L., AND LAPALME, G. Choosing Rhetorical Structures to Plan Instructional Texts. *Computational Intelligence 16*, 3 (2000), 408–445.

[81] KRAHMER, E., ERK, S., AND VERLEG, A. Graph-Based Generation of Referring Expressions. *Computational Linguistics 29* (2003), 53–72.

[82] KRAHMER, E., AND THEUNE, M. *Empirical Methods in Natural Language Generation*. Lecture Notes in Computer Science 5790. Springer Verlag, 2010.

[83] KUHN, T. Acewiki: A Natural and Expressive Semantic Wiki. In *Proceedings of Semantic Web User Interaction at CHI 2008: Exploring HCI Challenges* (2008).

[84] KUTZ, O., LUTZ, C., WOLTER, F., AND ZAKHARYASCHEV, M. ϵ-Connections of Abstract Description Systems. *Artificial Intelligence 156*, 1 (2004), 1–73.

[85] LANGKILDE, I., AND KNIGHT, K. Generation that Exploits Corpus-based Statistical Knowledge. In *Proceedings of the 17th international conference on Computational linguistics (COLING)* (Montreal, Québèc, 1998), pp. 704–710.

[86] LAREAU, F., AND WANNER, L. Towards a Generic Multilingual Dependency Grammar for Text Generation. In *Grammatical Engineering Across Frameworks (GEAF '07)*, T. H. King and E. Bender, Eds. CSLI Publications in Computational Linguistics, Stanford, 2007, pp. 203–223.

[87] MANN, W., AND THOMPSON, S. Rhetorical Structure Theory: A theory of text organization. In *The Structure of Discourse*, L. Polanyi, Ed. Ablex Publishing Corporation, Norwood, New Jersey, 1987.

[88] MARCINIAK, T., AND STRUBE, M. Discrete Optimization as an Alternative to Sequential Processing in NLG. In *Proceedings of the 10th European Workshop on Natural Language Generation (ENLG)* (Aberdeen, UK, 2005), pp. 101–108.

[89] MARCU, D. From Local to Global Coherence: A Bottom-up Approach to Text Planning. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference (AAAI/IAAI'97)* (1997), pp. 629–635.

[90] MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., AND OLTRAMARI, A. WonderWeb Deliverable D18, Ontology Library (final), 2009.

[91] MATTHIESSEN, C., AND BATEMAN, J. *Text Generation and Systemic-Functional Linguistics: Experiences from English and Japanese*. Frances Pinter Publishers and St. Martin's Press, London and New York, 1991.

[92] MAYBURY, M. Communicative Acts for Explanation Generation. *International Journal of Man-Machine Studies 37*, 2 (1992), 135–172.

[93] MCKEOWN, K. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press, Cambridge, England, 1985.

[94] MELLISH, C. Using Semantic Web Technology to Support NLG Case Study: OWL Finds RAGS. In *Proceedings of the 6th International Natural Language Generation Conference* (Stroudsburg, PA, USA, 2010), INLG '10, Association for Computational Linguistics, pp. 85–93.

[95] MELLISH, C., KNOTT, A., OBERLANDER, J., AND O'DONNELL, M. Experiments Using Stochastic Search for Text Planning. In *Proceedings of the 9th International Workshop on Natural Language Generation* (Niagara-on-the-Lake, Ontario, Canada, 1998), pp. 98–107.

[96] MELLISH, C., AND PAN, J. Finding Subsumers for Natural Language Presentation. In *Proceedings of the International Workshop on Description Logics* (The Lake District, England, 2006), pp. 127–134.

[97] MELLISH, C., AND PAN, J. Natural Language Directed Inference from Ontologies. *Artificial Intelligence 172*, 10 (June 2008), 1285–1315.

[98] MELLISH, C., SCOTT, D., CAHILL, L., EVANS, R., PAIVA, D., AND REAPE, M. A Reference Architecture for Natural Language Generation Systems. *Natural Language Engineering 12*, 1 (2006), 1–34.

[99] MELLISH, C., AND SUN, X. Natural Language Directed Inference in the Presentation of Ontologies. In *Proceedings of the 10th European Workshop on Natural Language Generation* (2005), pp. 118–124.

[100] MELLISH, C., AND SUN, X. The Semantic Web as a Linguistic Resource: Opportunities for Natural Language Generation. *Knowledge-Based Systems 19*, 5 (Sept. 2006), 298–303.

[101] MEL'ČUK, I. *Dependency Syntax: Theory and Practice*. SUNY Press, Albany, 1988.

[102] MIEZITIS, M. Generating Lexical Options by Matching in a Knowledge Base. Tech. Rep. CSRI-217, Dept. of Computer Science, University of Toronto, Toronto, 1988.

[103] MONTIEL-PONSODA, E., DE CEA, G. A., GÓMEZ-PÉREZ, A., AND PETERS, W. Enriching Ontologies with Multilingual Information. *Natural Language Engineering 17*, 3 (2011), 283–309.

[104] MOORE, J., AND PARIS, C. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics 19*, 4 (1993), 651–694.

[105] NICOLOV, N., MELLISH, C., AND RITCHIE, G. Sentence Generation from Conceptual Graphs. In *Conceptual Structures: Applications, Implementation and Theory, Third International Conference on Conceptual Structures, ICCS '95, Santa Cruz, California, USA, August 14-18, 1995, Proceedings* (1995), pp. 74–88.

[106] NIRENBURG, S., MCCARDELL, R., NYBERG, E., HUFFMAN, S., KENSCHAFT, E., AND NIRENBURG, I. Lexical Realization in Natural Language Generation. In *In Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages* (1988), pp. 18–26.

[107] NOGIER, J.-F., AND ZOCK, M. Lexical Choice as Pattern Matching. *Knowledge Based Systems 5*, 3 (1992), 200–212.

[108] O'DONNELL, M., MELLISH, C., OBERLANDER, J., AND KNOTT, A. ILEX: an Architecture for a Dynamic Hypertext Generation System. *Natural Language Engineering 7* (2001), 225–250.

[109] PHILLIPS, J. Generation of Text from Logic Formulae. *Machine Translation 8*, 4 (1993), 209–235.

[110] POLGUÈRE, A. Pour un Modèle Stratifié de la Lexicalisation en Génération de texte. *Traitement Automatique des Langues (T.A.L) 39*, 2 (1998), 57–76.

[111] PORTET, F., REITER, E., GATT, A., HUNTER, J., SRIPADA, S., FREER, Y., AND SYKES, C. Automatic Generation of Textual Summaries from Neonatal Intensive Care Data. *Artificial Intelligence 173*, 7-8 (2009), 789–916.

[112] POWER, R. Towards a Generation-based Semantic Web Authoring Tool. In *Proceedings of the 12th European Workshop on Natural Language Generation* (Stroudsburg, PA, USA,

2009), ENLG '09, Association for Computational Linguistics, pp. 9–15.

[113] POWER, R. Complexity Assumptions in Ontology Verbalisation. In *Proceedings of 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)* (Stroudsburg, PA, USA, 2010), ACLShort '10, pp. 132–136.

[114] POWER, R., SCOTT, D., AND EVANS, R. What You See Is What You Meant: Direct Knowledge Editing with Natural Language Feedback. In *Proceedings of the 13th Biennial European Conference on Artificial Intelligence (ECAI 98)* (Brighton, UK, 1998), pp. 675–681.

[115] POWER, R., AND THIRD, A. Expressing OWL Axioms by English Sentences: Dubious in Theory, Feasible in Practice. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (Stroudsburg, PA, USA, 2010), COLING '10, Association for Computational Linguistics, pp. 1006–1013.

[116] REITER, E. *Generating Appropriate Natural Language Object Descriptions*. PhD thesis, Aiken Computation Lab, Harvard University, 1990.

[117] REITER, E. Has a Consensus NL Generation Architecture Appeared, and Is It Psycholinguistically Plausible? In *Proceedings of the Seventh International Workshop on Natural Language Generation (INLGW)* (1994), pp. 163–170.

[118] REITER, E., AND DALE, R. *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, 2000.

[119] REITTER, D., AND CALLAWAY, C. Methods, requirements and licenses for shared NLG resources. In *Open Mic Session, Special Session on Sharing Data and Comparative Evaluation, Fourth International Natural Language Generation Conference* (2006).

[120] RÖSNER, D., AND STEDE, M. Customizing RST for the Automatic Production of Technical Manuals. In *Aspects of Automated Natural Language Generation*, R. Dale, E. Hovy, D. Rösner, and O. Stock, Eds., Lecture Notes in Artificial Intelligence. Springer Verlag, Berlin, 1992, pp. 199–214.

[121] SCHÜTTE, N. Generating Natural Language Descriptions of Ontology Concepts. In *Proceedings of the 12th European Workshop on Natural Language Generation (EWNLG)* (Athens, Greece, 2009), pp. 106–109.

[122] SCHWITTER, R., AND TILBROOK, M. Controlled Natural Language Meets the Semantic Web. In *Proceedings of the Australasian Language Technology Workshop* (2004), vol. 2004, pp. 55–62.

[123] SEVCENKI, M. Online Presentation of an Upper Ontology. In *Proceedings of Znalosti* (2003).

[124] SMART, P. Controlled Natural Languages and the Semantic Web. Technical Report ITA/P12/SemWebCNL, University of Southampton, July 2008.

[125] SOWA, J. *Knowledge Representation*. Brooks Cole, Pacific Grove, CA, 2000.

[126] SRIPADA, S., REITER, E., AND DAVY, I. SumTime-Mousam: Configurable Marine Weather Forecast Generator. *Expert Update 6*, 3 (2003), 4–10.

[127] STEDE, M. *Lexical Semantics and Knowledge Representation in Multilingual Text Generation*. Kluwer Academic Publishers, Dordrecht, 1999.

[128] STEVENS, R., MALONE, J., WILLIAMS, S., POWER, R., AND THIRD, A. Automating Generation of Textual Class Definitions from OWL to English. *Journal of Biomedical Semantics 2*, 2:S5 (2011).

[129] SUN, X., AND MELLISH, C. Domain Independent Sentence Generation from RDF Representations for the Semantic Web. In *Combined Workshop on Language-Enabled Educational Technology and Development and Evaluation of Robust Spoken Dialogue Systems, European Conference on AI (ECAI)* (Riva del Garda, Italy, 2006).

[130] SUN, X., AND MELLISH, C. An Experiment on "Free Generation" from Single RDF Triples. In *Proceedings of the 11th European Workshop on Natural Language Generation* (2007), Association for Computational Linguistics, pp. 105–108.

[131] WALKER, M., STENT, A., MAIRESSE, F., AND PRASAD, R. Individual and Domain Adaptation in Sentence Planning for Dialogue. *Journal of Artificial Intelligence Research* (2007), 413–456.

[132] WANG, C., XIONG, M., ZHOU, Q., AND YU, Y. PANTO: A Portable Natural Language Interface to Ontologies. In *The Semantic Web: Research and Applications*, E. Franconi, M. Kiefer, and W. May, Eds., vol. 4519 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2007, pp. 473–487.

[133] WANNER, L. *Exploring Lexical Resources for Text Generation in a Systemic Functional Language Model*. PhD thesis, Saarland University, Saarbrücken, 1997.

[134] WANNER, L., BOHNET, B., BOUAYAD-AGHA, N., LAREAU, F., AND NICKLASS, D. MARQUIS: Generation of User-Tailored Multilingual Air Quality Bulletins. *Applied Artificial Intelligence 24*, 10 (2010), 914–952.

[135] WEAL, M., ALANI, H., KIM, S., LEWIS, P., MILLARD, D., SINCLAIR, P., DE ROURE, D., AND SHADBOLT, N. Ontologies as Facilitators for Repurposing Web Documents. *International Journal of Human-Computer Studies 65*, 6 (June 2007), 537–562.

[136] WILCOCK, G. Pipelines, Templates and Transformations: XML for Natural Language Generation. In *Proceedings of the First NLP and XML Workshop* (Tokyo, 2001), pp. 1–8.

[137] WILCOCK, G., AND JOKINEN, K. Generating Responses and Explanations from RDF/XML and DAML+OIL. *Knowledge and Reasoning in Practical Dialogue Systems, IJCAI 2003* (2003), 58–63.

[138] WILKS, Y., AND BREWSTER, C. Natural Language Processing as a Foundation of the Semantic Web. *Foundations and Trends in Web Science 1*, 3–4 (2009), 199–327.

[139] WILLIAMS, S., AND POWER, R. Grouping Axioms for More Coherent Ontology Descriptions. In *Proceedings of the 6th International Natural Language Generation Conference* (Stroudsburg, PA, USA, 2010), INLG '10, Association for Computational Linguistics, pp. 197–201.

[140] XU, F., ADOLPHS, P., USZKOREIT, H., CHENG, X., AND LI, H. Gossip Galore: A Conversational Web Agent for Collecting and Sharing Pop Trivia. In *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART)* (Porto, Portugal, 2009), pp. 115–122.

[141] ZHANG, X., CHENG, G., AND QU, Y. Ontology Summarization Based on RDF Sentence Graph. In *Proceedings of the 16th international Conference on World Wide Web* (New York, NY, USA, 2007), WWW '07, ACM, pp. 707–716.

Reviewer 1: John Bateman

This paper presents a generally useful introduction both to NLG and to the particular application of NLG in the context of the Semantic Web. Since this is a growing area and will no doubt become increasingly important, it is worthwhile and timely presenting the range of work and issues involved in NLG to the semantic web community. This is the function of the paper as I see it and on this basis I would certainly recommend that it be published. The paper does however need to be worked over thoroughly in several ways, which I detail below. The paper's overview of the relevant field is really very extensive and the authors are to be congratulated on putting this all together in one accessible place.

Important note: The restructuring of the paper is explained in response to comments 1, 2 and 3 of Reviewer 2.

1	As a general style note, though, I find the form of referencing where only numbers appear to belong in the stone age and to be extremely reader-unfriendly: thus all references such as "[23] say that" should include the names of the authors. If this is a journal style policy, then the editors (and authors) should appeal to have it changed. There is no excuse for this form of reference these days.	The journal recommends using *plain* or *abbrv* bibtex style. We contacted the Journal editor and this is the answer we got: "Layout questions are up to the publisher, and IOS Press does not have a standing policy re. bibliographic references for submissions. I believe our published papers all use plain or abbrv or something that looks very similar. For the submission, it doesn't matter. For the final paper, the publisher (who does the final typesetting) may switch to their style." Therefore we have kept the plain or abbrv bibtex style, but used names of authors if in main argument position (subject or object): "Blog et al [23] say", or "as in Blog's [23] approach", but "qualitative evaluation is still popular [23]".
2	In addition, some of the references in the bibliography are incomplete. For example, Jarrar et al [69] and White [122] have no publication details beyond year; [98], [99], [109], [116] and possibly others are missing page numbers. The references should therefore be checked for completeness throughout.	All the references have been checked and completed (Note: unfortunately in a very reduced number of cases the page numbers could not be found no matter how extensive our search).
3	More minor points concerning the content and discussion, which the authors may want to respond to or adopt in their next version: p1, §2 "rendered as a coherent narration": of course there are many other text types, or genres, that the information could be rendered as (report, description, etc.); the restriction here to 'narration' is inappropriate.	We have replaced the term "narration" with the more neutral "text".
4	The issue of components for NLG has always been tricky; this is probably the main reason why RAGS has not found acceptance. For example, I would certainly have put lexicalization as a subtask of	We now present lexicalization and GRE in a more neutral way as separate tasks (as they are usually presented). In response to the reviewer's comment that the main reason why RAGS has not found acceptance has to do with the issue of components: we think that the

	generating referring expressions rather than the other way round!	proponents of RAGS were fully aware of the heterogeneity of NLG components and their varying use throughout the NLG architecture (e.g., see Cahill and Reape Technical Report, 1999, "Component Tasks in Applied NLG" for a study of these components in different systems), and that is why the focus of RAGS, rather than being on components, was on representations. We quote Mellish's (2010) justification as to why RAGS has not been widely accepted, which we believe is a fairly reasonable view (complexity of the framework together with lack of tools and lack of standardized representations).
5	§2.1: single genre generators: although it might be worth remembering that this was the original goal of *general purpose* generators, such as Penman (KPML), Mumble, SURGE and the others. And actually is still implicit even in components like SimpleNLG: the genre decision is a separate one.	We have removed the mention that "individual generators do not cover several genres", as it can lead to confusion, depending on what is meant by "generator" (fully fledged or linguistic realization components).
6	I am not sure that writing "for somewhat outdated detailed reviews" is beneficial: if something is outdated, then it certainly should not be mentioned here (unless this were a historical report, and it isn't).	We've removed these (outdated) references altogether.
7	Some of the references to 'staggered' architectures are pretty old; are there no newer references that have tried to breakdown the pipeline view (e.g.,: Marciniak/Strube; Dethlefs; Belz 2007?)	We have replaced the references to "staggered" architectures with more recent references (Belz2008, Marciniak/Stube2005, Dethlefs2011,Konstas2012).
8	p4, col.2 It may be relevant here to show more awareness of the limitations of Semantic Web formalizations as well: "OWL/RDF offers explicit support for the modularity..." This is limited to importing entire subontologies and so is only one kind of (quite restricted) modularity. There is an entire literature on more sophisticated modularities, as seen in the WoMo international workshop series (Workshop on modular ontologies: cf. http://www.informatik.uni-bremen.de/~okutz/womo5/) and some of these ideas may well be required for more advanced NLG capabilities and representations of resources in the future.	We limit the notion of modularity using owl:import (at the beginning of section 4) by saying that it provides a limited form of modularization of knowledge. We also include a brief mention of more advanced forms of modularizations using e-connections in a footnote in section 4.2 ("Use of linked data and associated text") with a reference to the paper of Kutz et al, AI, 2004, in response also to comment 14.
9	p5, col.1 here there is talk twice of "complete" something or others ("complete model about question answering", "complete ontological model"): it is unclear what this could mean. Either completeness (in the mathematical sense) has been proved, which does not appear to apply to either of these cases, or	Yes, some other word rather than "complete " was actually intended: 1) we now say "an explicit ontological model" instead of "a complete ontological model", and 2) an "extensive model of question answering" instead of "a complete model about question answering".

	the claim of being 'complete' is absurd. There are an almost infinite range of other facets of users that could be modelled, so to say that a complete user model exists is just silly. Perhaps this is a language problem and some other word rather than 'complete' was actually intended.	
10	It might also be well to reflect on the problem of similar formalisms encouraging merging things together that do not belong together: putting the lexicon in the knowledge base is, to my taste, a really bad decision (in general and on theoretical grounds, although of course it could be motivated on particular engineering grounds within single systems); thus the NaturalOWL approach messes up the entire modularity that could beneficially be enforced even within RDF/OWL.	We have made this theoretical concern (which we share with the reviewer) explicit. First we emphasize the tight integration of lexical info and domain data prevalent in NaturalOWL: "The uniform codification of domain data and lexical information has received particular attention. Thus, instead of keeping the lexicon separate from the domain data, some approaches opt for annotating the domain data with lexical information within the same SW representation. This is, for instance, the case of NaturalOWL". Then we point out the limitation of such approaches and use it as a bridge to introduce models that address better this separation: "This tight integration of linguistic and domain knowledge, although probably motivated on particular engineering grounds within single systems, raises some theoretical concerns and, on more practical grounds, limit the reusability of the linguistic data in different application settings which may require different mappings. For these reasons some lexical models enforce a clear separation between linguistic information and its mapping to domain data, thus allowing multiple mappings to coexist. For instance Buitelaar et al…".
11	p10, col.1 verbalization of ACE-sentences: it is not quite clear what happens to the first-order component; does this go into SWRL and then is *not* verbalised, or are only the components that have a translation in OWL (and so presumably falling within DL/SHOIN etc.) verbalised?	For the ACE<->OWL round-trip verbalization/parsing, the authors only consider a subset of ACE for which this round trip is possible. Therefore we have added the following precision (in subsubsection on "Controlled Natural Languages"): "Roughly speaking, Kaljurand et al. [..] require that the round-trip mapping between OWL and a subset of ACE be meaning preserving.
12	p11, col.1 "false implicatures that can be drawn from some of the ontological axioms": surely here you mean false implicatures can be drawn from the *verbalisations* of the ontological axioms! No implicatures follows from the axioms, just entailments within the logic they are couched in.	This has been fixed (now subsubsection on "Say everything" in overview of text planning).
13	And in this entire discussion of verbalisation, shouldn't the work by Adam Pease on verbalisations of SUMO/MILO using the connection to WordNet and generation templates be mentioned: or is this ruled as not being Semantic Web because of the first-order expressiveness of SUMO's specification language?	We now mention SUMO at the beginning of section 3, by pointing out that the use of NLG for ontology engineering predates the semantic web and using this particular verbalization application as an example..

14	p20, col.1 the mention of heterogeneous data sources is interesting but not taken very far. This relates to the discussion of modularity above; relating heterogeneous DL descriptions has been addressed in frameworks such as E-connections (Kutz etal 2004, AI Journal): there are few NLG frameworks that have begun to consider these more sophisticated possibilities however. A sketch of something moving in this direction is the Hois/Kutz paper in the 2008 Formal Ontologies and Information systems conference book (FOIS 2008). But the linking per se and getting inferences to work across such data sources is *not* an NLG problem necessarily.	See response to comment 8. We have included a mention of modularization but also of the problem of identity of references across datasets as a footnote in section 4.2, as this is not an NLG problem per-se but something that needs to be addressed by the SW community at large and that NLG would benefit from (hence the inclusion in the footnote and not in the main text).
15	Finally, the English is in many places very problematic and unfortunately not of a publishable quality. I give a list of basic corrections below, which may well not be exhaustive. After correction, the paper should be checked through again by a competent English reader.	The paper has been checked through by a competent English reader (amongst others, his main complaint was our tendency for over-long sentences) and his comments have been taken into account.
16	p1, §2.1: laymen --> lay people	Done
17	p2, col.1: prognosticated only based --> predicted based only	Done
18	"This is about to change." Delate (it has changed already) Better: "But statistical realizations of different NLG tasks are now also popular."	We have changed to: "However, statistical/heuristic-based realizations of different NLG tasks have become increasingly popular, from sentence planning and realization [..], text planning [..], ordering of content units [..] to content Selection [..]." See also our response to similar comment number 2 to reviewer 3.
19	"In the best of cases," Delete.	Done
20	"best for these dependences" --> "these dependencies"	The text about NLG architectures has been rephrased so this excerpt has now disappeared.
21	p2, col.2: "sufficient as a" --> "sufficient in many cases for a"	Same as comment 20.
22	"has been increasingly given more prominence" --> "has received increasing prominence"	The sentence has been changed as follows: "The evaluation of the performance of each module and of the individual tasks within each module has been given increasing prominence in the NLG community"
23	p3, col.1: "relevant on the research" --> "relevant for the research"	Done
24	"of the SW-technologies" --> "of SW-technologies"	Done
25	p3, §2.2 "The design and realization of nearly each of... of it." --> "Among the design and realization of the generation tasks listed at the beginning of Section 2,	Ok, this has been changed as follows: "Among the generation tasks presented in Section 2, especially...of it"

	especially content selection, discourse structuring, and lexicalization depend on the type of semantic input structure used by the generator since they ... of it."	
26	§2.2.1 "the so-called closed-planning" --> "so-called closed planning" (no determiner)	Done
27	§2.2.1 and wouldn't a reference to Nicolov be relevant here?	We now introduce Nicolov's work first as an example of content selection task interacting with lexico-syntactic realization in section 2.1. We then mention it again at the beginning of section 2.2.1 on the semantically oriented task of Content Selection as an example of micro-level content selection.
28	p3, col.2. "is the Rhetorical Structure Theory" no determiner.	Done
29	p4, col.1 "The most prominent of them" --> "The most prominent of these"	Done
30	"bridged the gap between LOOM representations and linguistic structures" : seems categorially confused. The gap bridged was, as far as I know, between domain representations and linguistic representations. LOOM refers to the formalisation language used to represent the information, not the status of the information itself.	OK, fixed.
31	"used in the context of systemic-functional generators" --> "used in the context of the systemic-functional generators"	OK, fixed.
32	"KMPL" --> "KPML"	OK, fixed.
33	Alongside [10], I suppose the rather more uptodate article in the Artificial Intelligence journal on the Generalized Upper Model would be relevant here Bateman, J. A.; Hois, J.; Ross, R. J. & Tenbrink, T. A linguistic ontology of space for natural language processing Artificial Intelligence, 2010, 174, 1027-1071	OK, added.
34	col.2 "As a consequence, the need for a separation between different types of knowledge..." : indeed: and this was, I take it, the principal motivation for the entire design of the Penman Upper Model, so a reference to this motivation here would be more appropriate.	In the new paper structure, there is now a subsubsection dedicated to the use of upper models in NLG for the semantic web. We also mention again upper models in passing when talking about the burning issue of "Codification and modelling of NLG-relevant knowledge in SW standards and formalisms", as an example of a linguistic resource that has a version in the OWL language.
35	p5, col.2 "In [86]", "[93,105] argue" yeeuch.	Following this comment and a similar one by another reviewer we have fixed all the references in argumentative position. See response to comment 1.
36	"Being appealing from" : not an English construction; try: "As it is appealing from the perspective of..."	N/A in the new version
37	"to ensure a certain quality" --> "to ensure quality"	Ok fixed.

38	"On the other hand" should only be used in pairs; use: "Alternatively", etc.	We checked the "on the other hand" in the article and changed the inappropriate ones
39	p6, col.2 "associated to their plan" --> "associated with their plan"	N/A anymore
40	"KL-ONE or LOOM)" add comma after closing paren.	N/A anymore
41	"not powerful enough to express all required meaning constellations" : again, senseless hyperbole, nothing is that powerful! (apart from people and even that is unclear). Tone down appropriately for a scientific article.	N/A anymore
42	"Semantic web offers" --> "The Semantic Web offers" (article obligatory)	N/A anymore
43	"can deal" --> "should be made able to deal" (not good, but better: since this is a challenge, there needs to be some causation expressed to bring something about: this is not present in "can")	N/A anymore
44	"As a matter of fact, it was extended by [55]." --> ", as shown in the extension provided by Name [55]."	N/A anymore
45	"network and stopping" --> "network, stopping"	N/A anymore
46	"et al. cope with large scale SW repositories in content selection" --> "et al. cope with content selection in large-scale SW repositories"	N/A anymore
47	p7, col.2 "developed so far and exploited" --> "developed so far have exploited"	N/A anymore
48	p8, col.2 "category captures Natural" --> "category covers Natural"	N/A anymore
49	p9, col.1 "the more needed and sophisticated is the NLG-technology" --> "the more sophisticated is the NLG-technology (...) required"	Rephrased as follows: "the more stilted and disconnected the output text, the more primitive (if used at all) is the NLG technology, whilst the more complex, lengthy and coherent the output text, the more sophisticated is the NLG technology"
50	p9, col.2 "certain idiosyncratic" --> "idiosyncratic"	N/A anymore
51	"She was then asked..." --> "The user was then asked to pick the one preferred based on the correctness..."	N/A anymore
52	p10, col.1 "a the Attempto" --> "the Attempto"	N/A anymore
53	p10, col.2 "and after having seen some examples" --> "and seeing some examples"	N/A anymore
54	"(Part of Speech)" --> "(part of speech: PoS)"	N/A anymore

55	p11, col.1 "where each minimal" --> "since each minimal"	N/A anymore
56	"joining all the neighbourhood" --> "joins all the neighbourhood"	N/A anymore
57	"structures is applied to form" --> "structures to form"	N/A anymore
58	p11, col.1/2 "as introduction of more specific concepts before more general ones and evaluation of" --> "as introducing more specific concepts before more general ones and evaluating"	Ok fixed.
59	p11, col.2 "two small ontologies, and although" --> "two small ontologies and, although"	Ok fixed.
60	p12, col.2 "The text planner that ensures" --> "The text planner then ensures"	N/A anymore
61	p13, col.1 "folskonomies" --> "folksonomies"	N/A anymore
62	paragraph beginning as follows is pretty awkward: "The authors justify the use of KMPL by ... Suggestion: "The authors justify the use of KPML by virtue of KPML's acceptance of ontology-oriented logical specifications formulated in terms of the Sentence Plan Language [72] as input and the system's provision of large-scale NLG resources (including grammars and task and domain-independent linguistically-motivated ontologies such as the Penman Upper Model [8]) for a number of languages. The system also offers a stable grammar development environment."	N/A anymore
63	p13, col.1 "test ontologies, the generated" --> "test ontologies the generated"	N/A anymore
64	p14, col.1 move the possessive marker after the names Galanis and Androustsopolous rather than after the [55]	N/A anymore
65	I'd suggeset making "Classes and individuals in the OWL ontology" a new paragraph.	N/A anymore
66	"associated to individuals" --> "associated with individuals"	N/A anymore
67	p14, col.2 "is provided such that" --> "is provided so that"	N/A anymore
68	"2) model" --> "2) a model"	N/A anymore
69	"3) model" --> "3) a model"	N/A anymore
70	"or he might" be consistent! Avoid the generic 'he', use 'she' if you must, or do you want to argue that principal	Ok fixed

	investigators (in opposition to the female 'user' adopted just after) are male! I'd report 'the investigator' personally and not use a pronoun at that point. Same for 'user'	
71	p15, col.1 "discourse (as" --> "discourse (such as"	N/A anymore
72	p15, col.2 "(FOHM) model" remove 'model' as this is already in FOHM.	N/A anymore
73	"They consist of ordered" --> "These templates consist of ordered"	N/A anymore
74	"realized as sentence" --> "realized as a sentence"	Fixed, this paragraph shortened and paraphrased in simpler terms following another reviewer's comment.
75	"which implements [77]'s" : add name	N/A anymore
76	p16, col.2 "As benefit" --> "As a benefit"	N/A anymore
77	"The obtained scores were significantly high" : in what sense? If statistical, say so; otherwise this is just noise and should be ommitted.	N/A anymore
78	p17, col.1 "we discussed in the previous sections" --> "we have discussed in the previous sections"	N/A anymore
79	p18: not all of the systems mentioned in the text are in the table, e.g., Schütte 2009 (as far as I can see): is this deliberate? If this is a selection, the grounds for inclusion should be clearer: what is 'condensed' on what criteria?	We've removed the word "condensed". We've changed the title in the table and refer to it as "a summary of the most representative NLG applications on which this overview is based"
80	p19, col.2 "discuss a few these isses" --> "discuss a few of these issues"	N/A anymore
81	the number of footnote 26 in the text occurs between its own commas, move after the comma and delete the spurious additional comma.	Fixed.
82	p20, col.1 "RAGS C. Mellish" the referencing is messed up here and a comma should come after RAGS if the name really appears in the body of the text.	Referencing fixed since this text was altered (now appears in long-term NLG issues and in section 4.1 on codification.
83	"recasting ... RAGS data" move the "RAGS data" immediately after "recasting" and move "quite naturally" after "remedied"!	Done.
84	"increase the performance of the generators": why does this follow? seems an independent dimension that could go either way depending on other issues.	N/A anymore
85	p20, col.2 "large data set(" --> "large data set ("""works on ontology" --> "work on ontology" ['work' in the plural is often clumsy, unless you are talking about	Fixed.

	paintings or something]	
86	p21, col.2 "The NLG-component used ... is" --> "The NLG-components used ... are" "data-driven approaches, and" no comma.	Fixed

Reviewer 2: Philipp Cimiano

This paper provides an overview of natural language generation approaches in the context of the Semantic Web. A review like this is clearly very useful and I would like to see it accepted for the journal. However, there are two issues that need to be addressed before this review paper can be published at the journal.

1	1) Lack of examples: One problem of the paper is that it lacks examples. The authors could include here and there some examples of input and output of particular NLG systems. This would make the article more accessible to a broader audience.	Following suggestions of reviewers 2 and 3, we have added examples of inputs/outputs and some other illustrations in place where we thought it was most necessary: - In section 2.1 when discussing pipeline architecture of NLG systems, example of 2 systems that address the inter-relation between NLG tasks (Nicolov et al. for an example of a sentence generator that takes into account content selection, and Palmer and Walker , who address GRE and content selection). - In section 2.1 when presenting NLG tasks, we have added an example of syntactic aggregation for people unfamiliar with the task (in a footnote) - In section 2.2.1 when presenting message determination we give an example of such message. - In section 2.3 (long-term issues), when discussing incomplete knowledge and the need for domain communication knowledge, we give an example of domain communication knowledge that can be inferred from domain knowledge. - Section 3.1.2: We provide an illustration (figure+explanations) of conceptual authoring/WYSIWYM. - When discussing the consensus model in section 3.2.1, we give an example of verbalization of axiom. - In section 3.2.2, as an example of annotation of content with linguistic knowledge, we show the RDF annotations for the NaturalOWL microplanner.
2	2) Challenges involved in NLG from Semantic Web data: the article does not emphasize sufficiently the technical challenges involved in scaling NLG to the Semantic Web. A section giving an overview of the main challenges could be added.	We now have a section at the end of the overview on NLG presenting long-term issues for NLG system (i.e., scalability, portability, knowledge and resources sharing and reus). After presenting the main paradigms/approaches to NLG, we argue that addressing these issues/challenges is critical for NLG to scale up to the semantic web, and that the SW provides opportunities/support to address them.
3	3) Too many details / no systematic overview: The article misses a systematic overview of different paradigms and approaches and an overview of the pro and cons of different NLG approaches. Instead, the article discussed many details of a number of system. The authors have	The paper has been restructured as follows: - We removed the systematic overview of systems, replacing it by the section giving an overview of paradigms of existing NLG approaches to the SW, occasionally focusing on a particular system if its approach was considered particularly interesting/relevant.

	clearly a lot of knowledge in the field, but discussing particular systems in detail is not particularly useful for somebody that wants to get an overview of the main approaches in the field or as a starting point for research in the field. I would propose that the authors try to abstract from details of systems and focus more on providing an overview of approaches, then mentioning systems that implement these approaches.	- As mentioned in the response to comment 2 above, we have also added an extra subsection to section 2 "A brief overview of NLG", which discusses long-term NLG issues that are especially relevant for NLG systems to scale up to SW. The motive for this section is that when discussing burning issues we kept making references to these long term issues so it seemed more natural to present them first where they belong in the overview of NLG. - We removed the summary section (that discussed the table summarizing the NLG systems). Instead, we present that table at the beginning of the section on the overview of NLG approaches to SW, and make references to the content of the table when relevant during the discussion the paradigms. - We have removed the section on "NLG for SW: What makes the difference" as there was a lot of overlap with 1) on the one hand the description of NLG paradigms for the SW, and 2) on the other hand the burning issues section. - Instead we provide a summary of the contributions SW has made so-far to NLG at the beginning of the section now called "Towards a symbiosis between NLG and SW". This section replaces the "Burning Issues" section. - Burning issues are now presented as subsections of that new section "Towards a symbiosis between NLG and SW". - We have merged the burning issue "Standardization of NLG-interfaces using SW" with "Codification of NLG-relevant knowledge in SW" as the former is about NLG interface representations, which are also NLG-relevant knowledge.
4	So my suggestion is to request a major revision of the paper by the authors. The article should be reduced in size to half of its current size. This is possible by removing the detailed discussion of particular systems and trying to provide a systematic overview of approaches / paradigms rather than discussing concrete systems. The article as it stands is clearly too long and fails to give a systematic overview of different approaches to NLG in the context of the SW.	See above (response to comment 3). These suggestions have been taken in. Excepting the bibliography, the article has been reduced to 15 pages (originally 22 pages). We couldn't achieve more reduction because some new material had to be added in order to address the reviewers' suggestions, in particular: - The conclusion now includes a summary of the paper (to address comment 3 of reviewer 3). - We have included examples as requested by two reviewers including two figures (see response to comment 1 of reviewer 2). - We have added a subsection on long-term issues (to address comment 2 of reviewer 2).
5	Section 3.2 „Use of SW technologies for	Section 3 has now disappeared (see

	reasoning in NLG": it is really not clear at all what this section tries to convey. It talks about many systems, about OWL 2.0, OWL-DL, SPARQL-RDF, SPIN-SPARQL the usage of SPARQL queries on OWL to get a content plan, etc. There are several issues here: first, it does not become clear here what kind of reasoning the authors talk about here and how it would help the NLG system. Is some sort of subsumption reasoning or other kind of reasoning in the focus her? Second, SPARQL is a language for querying RDF and not OWL. Finally, it seems that what the authors mainly discuss is the use of SPARQL in the content planning step. But then it is not clear why the title of the section talks about reasoning if the authors really only refer to using SPARQL to query a dataset and extract relevant information from which to generate NL. This needs to be clarified and the whole section needs to be reworked and the main points made more clear.	explanation of the restructuring of article in response to comment 3 above). There is now just a brief mention on the use of query languages and of the DL reasoners in two separate points at the beginning of section 4. These points summarize the current contributions of SW to NLG approaches.
6	Section 3.3 (Scaling up to large repositories of data): it is also not clear what this section aims at. The main point seems to be that „content selection strategies can deal with SW repositories". In this sense there seems to be an overlap with Section 3.2 to an extent that it is not clear anymore what Section 3.2 vs. Section 3.3 try to say. There is no real discussion on scalability and what the challenges/problems involved are, so also here the title of the section is quite misleading. The section is a good example for the point I made above: the section discusses specific work of O'Donnell, Bouayad-Agha, Dai et al, but clearly fails to provide a systematic overview of the main approaches to content selection from the Semantic Web, abstracting from concrete systems as well as formalisms. For example, using SPARQL is only a language to implement queries to an RDF dataset and not particulary interesting. The interesting thing is what content is retrieved. Whether this is done by SPARQL, SQL or in some other way is not really crucial IMHO.	Section 3 is no more (response to comment 3). There is now a brief reference to the need for NLG systems to scale up at the beginning of section 4 (contributions of SW to NLG) and the challenges involved in that scaling up are discussed in the burning issues 4.2 (use of linked open data and associated texts) and 4.3 (summarization of large volumes of data).
7	I would also advise to change the title of the article to: „Natural Language Generation in the	Done

	context of the Semantic Web“	
8	Some stylistic issues: The authors use references as nouns, i.e. as part of a sentence. This is bad style and should be avoided. Some examples with proposed rewritings follow: Section 2.2.1 : „in [24,27]“ => „in the systems proposed by Bouayad-Agha et al. [24] as well as by Bouttaz et al. [27] Section 2.2.2 „of [87]“ => „of McKeown [87] Section 3.1 „done in [36,55]“ => „in the approaches of Dannels [36] and Galanis and Androutsopolous [55]“ „[93,105] argue that“ => Melish and Sun [93] as well as Power and Third [105] have argued that...“ etc. etc.	Done (i.e., author(s)’ name(s) in addition to citation number in argumentative position. Otherwise just citation number).
9	Section 2.2.3 : what is „*semantemelexeme* association“ ???	We have rephrased the whole paragraph to make it clearer.
10	Section 2.1: representations have been experimented with – including model-theoretic semantics -> model-theoretic semantics is not a representaiton, but a way to define the semantics of a logical language (e.g. first-order logic, modal logic, descriptions logics, etc.). The authors probably mean first-order logics and related formalisms here.	Yes indeed. We have fixed that.
11	Section 3.5 „However, most NLG-applications developed so far and exploited...“ -> „However, most NLG-applications developed so far *have* exploited...“ ???	Not relevant anymore given restructuring of paper.
12	Section 5.5 „The vocabulary and data in the SW are often obtained automatically or semi-automatically from textual information in the Hyperdocument Web using Knowledge Acquisition, IE or Text Mining techniques“. => I do not agree with this. Most ontologies in the SW are modelled by hand and most data are generated automatically from existing databases. Could the authors provide evidence for this statement?	We agree, this was a slip of the tongue if we might say (it’s more of a future wish then current practice). We have fixed this as follows: “The vocabularies and asserted data in the SW are typically manually crafted, often from the analysis or manual annotation of existing texts, or obtained from databases. Initiatives like the DBPedia also extract data from textual markup on the web. Bontcheva et [..] and Wilks and Brewster [..] go even further, arguing that, since the WWW consists mainly of unstructured texts, the Semantic Web should and will tap into this vast pool of knowledge to build its datasets using techniques such as Information Extraction (IE) or Text Mining (TM).”

Reviewer 3: Ion Androutsopoulos

This article is a survey of Natural Language Generation (NLG) research related to Semantic Web (SW) data and technologies.

According to the journal's guidelines: "Survey articles should have the potential to become well-known introductory and overview texts. These submissions will be reviewed along the

following dimensions: (1) Suitability as introductory text, targeted at researchers, PhD students, or practitioners, to get started on the covered topic. (2) How comprehensive and how balanced is the presentation and coverage. (3) Readability and clarity of thepresentation. (4) Importance of the covered material to the broader Semantic Web community."

Starting from (4), the subject of the survey is particularly important for the SW community. As discussed in the article, NLG can be used during ontology engineering, to allow domain experts who are unfamiliar with SW formalisms to interact with natural language renderings of the ontologies; and it can also be used to present SW data (e.g., product descriptions) to end-users, possibly in different natural languages and tailoring the presentations to the users profiles. In both cases, NLG can help make SW data easier to understand for people who are not SW experts, and this may have a significant impact on the adoption of SW technologies.

Regarding (2), the article first provides an overview of established (largely pre-SW) concept-to-text NLG approaches, including a brief description of the most common processing stages of a typical pipeline architecture. The article then focuses on NLG especially for SW data and technologies, as well as "burning issues" that need to be addressed to maximize the positive interplay between NLG and SW.

(..) The use of English is very good; some typos and minor comments are listed below. The bibliography entries and figures are sufficient.

Regarding (1), I would definitely recommend the survey as an introductory text to researchers, PhD students, practitioners etc., especially if the suggested improvements are taken into account. This article may well become a standard introductory text for researchers interested in NLG and SW technologies.

Important note: The restructuring of the paper is explained in response to comments 1, 2 and 3 of Reviewer 2.

1	The overview of established NLG approaches is rather brief, and readers without an NLG background may get a limited view of the range of language phenomena and problems that NLG research has addressed over the last few decades; but sufficient pointers to relevant literature are provided, and interested readers could follow them to get a broader view of past NLG research.	We have extended the explanation about NLG architectures (and added some more recent pointers), we have also added (as requested in comment 2) specific references to statistical approaches. We have also added an extra subsection discussing long-term NLG issues which is also relevant later on when discussing the challenges of NLG to scale up to the Semantic Web.
2	There is also very little coverage of more recent, mostly statistical approaches to NLG, though this may be justified by the fact that most NLG work for SW data does not yet employ statistical methods.	We have provided more concrete pointers to some major works on statistical NLG for different NLG tasks (+ also works to integrated architectures), see section 2.1. Even though most NLG work for SW data does not yet employ statistical methods, it is worth knowing that there is some progress in this area.
3	The coverage of NLG work especially for SW data and technologies is very thorough, and this part of the article is also reasonably well organized and balanced. The “burning issues” section is potentially very useful for researchers wishing to get involved in NLG for the SW. By contrast, the concluding section is weak; a concise summary of the most important points discussed is needed.	We have extended the concluding section to include a concise summary of the most important points discussed in the article.
4	Regarding (3), the article is generally well written. My only complaint is that the descriptions of some of the methods are difficult to follow in the absence of examples illustrating how the methods work. I realize that including examples would increase the article’s length, but I would still recommend including examples at least at points that are otherwise very difficult to follow; see detailed comments below.	We have added extra examples (taking into account place where it was said to lack clarity), see response to similar comment 1 by Reviewer John Cimiano for a concrete list of what example were added.
5	There are also several acronyms, names of components or resources etc. that do not seem to be absolutely necessary, and I would recommend removing them; again, see detailed comments below.	We’ve taken this into account given the comments below but also checking the entire paper for acronyms and keeping only the ones we thought were useful/informative/necessary.
6	Section 1, lines 10-12, “linguistic surface-oriented structures already predetermine the linguistic form of the output beforehand”: Saying “ already predetermine, at least partially, the linguistic form ” might be more accurate.	OK done.

7	Section 1, line 13: It would be worth providing a few pointers to prominent work on NLG from numeric data.	Ok done, added 2 references (Portet et al 2009 , Sripada et al 2003).
8	Page 1, column 2, line 1, “machine-processable semantic data”: What is “semantic data”? Maybe say “machine-processable data with formally defined semantics.	We now say: “the semantic web with its machine-processable representation of information with explicitly defined semantics …”
9	Same paragraph, last line: What is “semantic NLG”?	This has been changed to:“[in order to improve the usability of] NLG when applied to SW data.”
10	Section 2, line 5, “discourse and information structuring”: “Text planning” or “discourse planning” would be more standard terminology.	We now say “discourse planning”.
11	Same paragraph, aggregation: It might be better to mention aggregation after lexicalization, since aggregation may also examine linguistic structures produced by lexicalization. Maybe also provide a brief example of aggregation, for the benefit of readers unfamiliar with NLG.	Ok we have put aggregation after lexicalization (though it can happen before in the case of a conceptual or semantic aggregation). We have also added an example for aggregation (as a footnote): e.g., the aggregation of the syntactic representations corresponding to the following 2 sentences `The wind will be weak. The rain will be light" will produce the more fluent text: ```The wind will be weak and the rain light”
12	Same paragraph, “onto language-specific semantemes and lexemes”: I would recommend using simpler terms, for example say “onto language-specific sentence representations”.	We replaced by “onto language-specific senses or words”.
13	Same paragraph, “morpho-syntactic realization and linearization”: “Surface realization” would be more standard terminology.	We’ve put “linguistic realization” to make sure it is clear that it is not just about linearization but also about mapping to morpho-syntactic form.
14	Same paragraph, (5), “projection of the discourse or sentence plan obtained from the preceding tasks onto the surface”: I recommend rephrasing this as “projection of the discourse and sentence specifications obtained from the preceding tasks onto the final (surface) form of the text”, since the meaning of “surface” may otherwise be unclear to some readers.	In order to clarify that this task is not just about linearization but also about morpho-syntactic realization. We have paraphrased as follows: “ mapping the discourse or sentence specifications obtained from the preceding tasks onto a syntactically, morphologically and orthographically correct text.”
15	Section 2, paragraph 2: Aggregation is also (at least partially) concerned with semantics. Hence, it seems strange not to include aggregation in the following discussion.	We added a footnote that states this point and justifies why we do not include as one of the semantically oriented tasks: “Conceptual aggregation also operates on the input structure. However, as it tends to be addressed in an ad-hoc way we did not think it deserved to be discussed in this section.”
16	Section 2.1, paragraphs 1-2: It might be worth including the discourse history, physical location and surroundings in the	We have changed the first paragraph of 2.1 to mention discourse history and physical location and surroundings, as follows: “the context can [..]allow for an explicit parameterization of only one or

	discussion of “context”.	several dimensions, such as [..], discourse history (e.g., what the user already knows) or physical location and surroundings (e.g., people in the vicinity of the user at a conference or artifacts the user is looking at in a museum for an advice or narration generation wearable device).”
17	Section 2.1, paragraph 3, “application such as Dialogue”: Calling Dialogue an information processing application sounds strange. Maybe refer to “dialogue systems” instead.	OK, we fixed this.
18	At several points throughout the article the authors use dashes at points where they do not seem necessary (e.g., “NLG-research”, “NLP-application”, “SW-standards”, “OWL-ontologies”).	Indeed. These have been removed when not necessary throughout the paper.
19	Table 1, input, Type: Isn't a table or a template also structured information?	Yes indeed, we’ve merged these.
20	Same table, “Domain independence”: Cannot a “conceptual representation” be domain dependent?	Indeed, we removed the illustration.
21	Same table, “Coherence”: I suggest replacing “coherent paragraph” by “coherent text”, since coherence is not restricted to paragraphs.	OK fixed.
22	Same table, “Fluency: fluent NL, controlled NL, or telegraphic style”: Cannot controlled NL be also fluent? Cannot a controlled NL be also telegraphic?	We’ve removed CNL. We just say “stilted NL” and do not mention how this can be achieved.
23	Page 3, first paragraph under Table 1, reference [122]: It might be worth including a pointer to E. Krahmer, M. Theune (eds.), Empirical Methods in Natural Language Generation, Springer, 2010.	We’ve included this reference, in addition to others.
24	Same page, paragraph 2: The part from “For instance” to the end of the paragraph is bit unclear. Some examples are needed to illustrate the problems of a typical pipeline architecture. For example, content selection may select information that is difficult for the discourse planner to structure coherently.	As an example we discuss 2 approaches that address the relation between content selection and GRE (Jordan and Walker), and content selection and lexico-syntactic realization (Nicolov).
25	“Syntacticization” and other uncommon terms or terms that have not been defined should be avoided.	Ok we now say “Syntactic realization”.
26	Same page, paragraph 3, “document or text planning (sometimes also referred to as content determination or macro-	We’ve removed the mention of “Content determination”.

	planning)": Content determination is usually considered part of document planning, it is not an alternative name for document planning.	
27	Last paragraph of page 3 and first paragraph of page 4: The discussion here seems to assume that any evaluation that relies on human judges is qualitative. An evaluation with human judges, however, can be quantitative (e.g., if it collects scores from the judges, and applies a statistical analysis on the scores).	Fair point. We've rewritten that paragraph and we don't talk about quantative vs qualitative but rather of human- vs corpus-based approaches.
28	Section 2.2.1, paragraph 1, last sentence: "Say everything there is to say" is also an explicit communicative goal. Maybe say "where content selection is guided by communicative goals that require particular subsets of the available information about an object".	Yes indeed. We've chosen to just mention bottom-up approaches and removing mention of top-down which wasn't very informative anyway (readers can infer/know that top-down approaches have different communicative goals than "say everything").
29	Section 2.2.1, paragraph 2: It might be worth mentioning that some systems use the RDF graph directly, i.e., a graph whose nodes correspond to classes and individuals, instead of using a graph whose are facts and edges represent selection constraints.	In the intro to section 2 "A brief overview of NLG", we now say: "To avoid any overlap with discussions later in the article, we deliberately avoid in this brief overview of NLG any mention of approaches that are specific to SW data". This should explain why we don't mention the use of RDF graphs in open planning.
30	Page 4, paragraph 2, "It has been argued ": This paragraph is unclear to me. First, the distinction between local and global coherence should be defined more clearly. Secondly, if the selected semantic units (e.g., facts) are already linked with discourse relations, hasn't global coherence already been achieved? Should "to achieve global coherence" be replaced by "to obtain a total ordering of the semantic units (e.g., facts) selected"?	We've changed the last sentence in order to omit those details that would require further explanations to be understandable. We think that those details are not necessary for what the paragraph is trying to convey. So now we say instead: "This can be exploited to achieve a total ordering of contents grounded on a coherent discourse structure -- as done by Marcu [..] in his constraint-based approach."
31	Same page, paragraph 3, "Another issue ": Please explain more clearly what EDUs are, and how they can be formed using templates or via topological search.	EDU/message determination is now discussed in section 2.2.1 on content selection as it is typically handled together with content selection. We explain more clearly what they are and give an example.
32	Section 2.3.3: Simply mentioning terms like "discrimination networks" or "semantemelexeme association with additional variation" along with pointers to articles is not particularly useful. Please explain the main ideas.	We have rephrased this paragraph, simplifying and/or explaining some of the concepts/approaches presented.
33	Same section, paragraph 2: Exactly which	We rephrased that paragraph as follows: "In order to facilitate the mapping from a language-

	gap did Penman UM bridge? Please explain more clearly how a Generalized Upper Model helps.	independent domain representation onto a task-dependent linguistic representation whilst giving sufficient room for flexible verbalization, domain- and task-independent language-oriented ontologies have been introduced into generation. The most prominent of these is,without doubt, the Penman Upper Model…” We hope that with this rephrasing it becomes clear that upper models bridge the gap between domain reps and task-dependent linguistic reps.
34	Section 3.1, paragraph 1, “enriched by the information how to”: I couldn’t parse this sentence.	N/A anymore.
35	Page 5, paragraph 2, “The layers can produce”: I could not follow this sentence. Please break it into shorter, simpler sentences.	N/A anymore.
36	Same page, paragraph 5, “A complete ontological model”: Complete in what sense?	Yes, some other word rather than “complete “ was actually intended: we now say “an explicit ontological model” instead of “a complete ontological model”.
37	Page 5, column 2, paragraph 4, “[93, 105] argue”: At several points the authors use reference numbers as subjects of sentences, which does not look right.	Ok, this has been fixed.
38	Section 3.2, paragraph 2: An example of a concept that arises in human communication and is not modeled by domain ontologies would help.	This section has now disappeared. However the text still pops up elsewhere, in section 3.1 on Text Planning from SW data, and we now provide an example, as follows: “Most approaches to text planning from SW data exploit the content provided by some input domain ontology. However, in some cases, this domain knowledge is enriched with domain communication knowledge (in the sense of Kittredge et al. [..]). The basic idea behind this distinction is that domain ontologies do not model, as a rule, all the concepts that arise in human communication. Thus, Bouayad-Agha et al [..] populate a domain communication ontology given a domain ontology using DL and/or rule-based reasoning (hence the "Mixed" task- and domain- independence of their input for both their systems reported in Table 2). To take but a couple of examples from the European football domain used by these researchers in one application [..], the fact that a team is going up or down the Football league's classification can be inferred from its current and previous classification, or the fact that it is drawing or winning/losing can be inferred from the final numerical score of the match.”
39	Same section, paragraph 4: I could not understand how NLG can be used to “verbalize the query potential of an ontology in a non-standard query language”.	N/A anymore.
40	Section 3.4, paragraphs 1 and 2: These paragraphs are unclear to me. What does “task” mean here? Why is task-specific knowledge domain-independent? Also,	This section is no more. We now paraphrase elsewhere task-independence as NLG task-independence. We hope that the slight reformulation of (G)UM in 2.2.3 and the new section 3.2.3 on use of upper models for sentence planning for NLG

	how is the mapping facilitated by the UM?	approaches to SW clarify those issues…
40	Section 3.4, last sentence: What does "meta-model" mean here?	N/A anymore.
41	Section 3.5, paragraph 1, sentence 2: What is the "Problem Description Language"? Do we need to know its name?	N/A anymore.
41	Same paragraph, "However, most NLG applications developed so far and exploited": I couldn't parse this sentence.	N/A anymore.
42	Page 8, last paragraph of Section 3: What exactly is a "reason-able view"? Do we need to know this term?	This is no more.
43	Same paragraph, last sentence: This sentence contains nested brackets, which make it difficult to follow the sentence.	N/A anymore.
44	Page 9, paragraph 2, "round trip authoring": It is worth explaining that the CNL is edited (e.g., by an ontology engineer) before being translated back into ontology axioms.	The text now appears in section 3.2.1 on the consensus model and we now explain what round-trip authoring means: "where ``round trip'' authoring is desirable, whereby the ontology is verbalized into a CNL and can be edited in this way by an ontology engineer before being translated back into ontology axioms."
45	Same page, paragraph 3, "in the ROO-authoring tool for the Rabbit-CNL": Is something missing here? Should the sentence say something like "as, for example, in the ROO-authoring tool"?	N/A anymore.
46	Same paragraph, last sentence, "Unfortunately, only a few": I couldn't understand this sentence. Isn't comparing systems that do or do not use NLG a way to compare the "underlying technology" (NLG vs. not-NLG)?	Yes, indeed we have shortened and rephrased that paragraph, which now appears in section 3, and begins as follows: Although some usability studies have shown the users preferences for NLIs rather than graphical or formal language interfaces [..] not all NLIs use NLG.
47	Page 9, paragraph 10, "interactive guided verbalization, i.e., conceptual authoring": What is "interactive guided verbalization"? Is it a synonym of "conceptual authoring", as the phrasing here seems to suggest?	N/A anymore
48	Page 10, paragraph 2, "a the Attempto Parsing Engine": Delete "a"	N/A anymore
49	Page 11, paragraph 1, "In order to achieve a more balanced linearized into a sentence": I couldn't follow this sentence. Please break it into shorter, simpler sentences. An example might also help.	N/A anymore

50	Page 11, paragraph 2, “if they can be inferred from previously selected axioms”: What are the “previously selected axioms”? Why should the (new) selected axioms be inferable from the previously selected axioms? Also, in “depth-first search of refined concepts (i.e., subsumers)”, should “refined” be “more general”?	Now in section 3.1.4, “previously selected axioms” -> axioms already selected by the approach. The new axioms are not inferred, their selection is inferred. We replaced “refined” by “more general”.
51	Page 11, column 2, paragraph 2: What is a “top-level argument”?	Section 3.1.1: We now provide an example of a class that is a top-level argument of an axiom and one that is not.
52	Same paragraph, “However the result is dampened and asked them which is”: The English of this sentence may need to be re-checked.	N/A anymore
53	Page 11, last paragraph, “The paths are translated into NL statements”: Should “NL statements” be “NL questions”?	N/A anymore
54	Page 12, first paragraph: I couldn’t follow this paragraph. Please explain in a clearer manner. An example might help.	N/A anymore
55	Section 4.1.2, paragraph 1, “the user authors the concepts of the ontology schema to formulate a query ”: Does the user author the concepts of the ontology even when formulating a query?	Now section 3.1.2: we now say: “The user selects or authors the concepts or individuals from the ontology to formulate a query or to design or populate the ontology”.
56	Same paragraph: I am familiar with WYSIWYM, but I doubt that a reader unfamiliar with this approach would be able to understand how WYSIWYM works without an example.	Now in Section 3.1.2: we’ve added an illustration (figure+explanations) taken from Hallett et al’s paper on conceptual authoring.
57	Footnote 13: This URL has already been provided.	N/A anymore
58	Page 13, last paragraph, “Argüello et al. [3]’s generator”: This should probably be “Argüello et al.’s [3] generator”.	N/A anymore
59	Page 14, paragraph 2, “Galanis and Androustsopoulos [55]’s”: This should probably be “Galanis and Androutsopoulos’s [55]”. Also, delete the extra “s” in the second surname here and in footnote 15.	N/A anymore
60	Same paragraph, “interest to the use”: This should be “interest to the user”	N/A anymore
61	Page 14, column 2, paragraph 2, (2) and (3): An article (“a”) is probably missing at the beginning of (2) and (3), i.e., “a model”, not “model”.	N/A anymore
62	Footnote 21: This URL has already been provided.	N/A anymore

63	Page 15, last paragraph: The discussion here is very difficult to follow. An example might help.	We have rephrased the explanation of Dai et al.'s approach (in Section 3.1.4) in simple terms.
64	Page 16, column 2, last paragraph: Is it necessary to introduce the SQASM acronym? Is it even necessary to mention the name of the model (Semantic QA Structure Model)?	N/A anymore
65	Page 17, column 2, paragraph 3: What are "pondered nodes"?	Now section 3.1: We have rephrased in a way that we hope makes the meaning of pondered nodes clearer, as follows: "with or without pondering the traversed nodes based on their inherent importance value and/or a user model".
66	Footnote 24: I couldn't follow the argument of this footnote.	Now as a footnote in Table 1. We have simplified it to just say: "Our assessment of the level of fluency relies not only on our knowledge of the performed NLG tasks and their scope, but also on the sample texts provided in the reference articles."
67	Reference 56: "Protégé" needs fixing.	Fixed