Natural Language Generation in the Context of the Semantic Web

Editor(s): Philipp Cimiano, Universität Bielefeld, Germany
Solicited review(s): John Bateman, Universität Bremen, Germany; Ion Androutsopoulos, Athens University of Economics and Business, Greece; Philipp Cimiano, Universität Bielefeld, Germany

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 given content input into a natural language output, given some communicative goal. Although this input can take various forms and representations, 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 (SW), 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 content input into a natural language output, given some communicative goal in a specific context [121]. This input can take various forms and representations, 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., [114,129]. 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 introduce 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 present 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 it may allow for an explicit parameterization of only one or several dimensions, as shown in Figure 1, which provides a summary of the main characteristics of an NLG system's input, output and context.

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. 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).[1] 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 [6,30,112], Schank's scripts [70,105], Sowa's conceptual graphs [110], a variety of frame representations such as KL-ONE and LOOM [69,77,119], 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 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; see, e.g., the aggregation of the syntactic representations of the two sentences *"The wind will be weak. The rain will be light"* to produce the more fluent text *"The wind will be weak and the rain light"* ;
(5) generation of referring expressions, 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* architec-

[1]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 form, template); input representation language (first-order logic, OWL-DL, etc); single or multiple input(s).
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.

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	Language only (written or spoken) or multimodal (e.g., text or speech with table or figures) and the degree of the multimodality.

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.
Physical location	Wearable devices involved in, e.g., the generation of narrations about the artifacts the user is looking/has looked at in a museum, the generation of information about people in the user's vicinity at a conference, etc., the generation of an air quality bulletin for a specific location, etc.

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

ture [101,120]. These modules are 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 planner is in charge of deciding *what to say* and organizing the chosen content into a coherent whole of a text plan. The sentence planner is in charge of mapping the text plan to the linguistic structure of a sentence plan, 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 pipeline's modules 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. Some NLG approaches have addressed these issues. For instance, Nicolov et al. [108] 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 [75] address the issue of deciding what content to include in a description for the generation of referring expressions.[2]

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., [11,88], follow various paths at decision-making points, postponing the selection of the best action until enough evidence has been gathered. Optimization approaches [45,91] model decision-making as a network of states con-

[2]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.

nected 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 by Konstas and Lapata [81] 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. [42] 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 implement simplistic shortcuts for tasks that are not in their focus. As a consequence, the theoretical justification of the approach, and the maintainability, output quality and variability of the implemented system cannot be predicted based only on the fact whether templates are used or not.

Statistical and heuristic-based realizations of different NLG tasks have also become increasingly popular in recent years (see [84] for some recent approaches), from sentence planning and realization [11,88,135], text planning [98], ordering of content units [50] to content selection [4].

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 compare those judgments to 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 [4,12,14,15,49]. For example, to evaluate their approach for ordering semantic structures, Duboue and McKeown [49] use a rule-based planner as a gold standard reference and several random orderings as baseline.

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.[3,4] Content selection and discourse structuring tend to output the same type of semantic structure as they take as input, whilst 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, our focus is on macro-level content selection.[5]

Determining what content to convey depends largely on the context dimensions detailed in Figure 1, and is traditionally achieved using rule-based and template-based approaches. 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 [43,83,111]. In this paper, we borrow Dai et al.'s [35] terminology by calling the first paradigm "closed planning" and the second one "open planning".[6]

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

[3] Although conceptual aggregation operates on the input structure, we do not discuss it here as it has traditionally been addressed in an ad-hoc way. See however Barzilay and Lapata's proposal [5] for a statistical approach to the aggregation of database entries.

[4] Referring expression generation involves both content selection as well as linguistic realization and as such can operate on the input structure [85]. However, as we point out in Section 2.2.1, our focus in this paper is on macro- rather than micro- level content selection and so we do not include this NLG task in our discussion.

[5] A few approaches to micro-level content selection are discussed in Section 2.1 above, namely Nicolov et al.'s [108] and Jordan and Walker's [75] approaches. Krahmer et al.'s [83] approach for referring expression generation is yet another notable example of micro-level content selection.

[6] The distinction between *closed* and *open* planning is used by Dai et al. [35] presumably to contrast between the fact that in the first type of approaches the "content is confined by a query with [a] condition" and the fact that in the second type of approaches, the content is not confined to a query but depends on the graph nodes and relations around a central topic node.

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. For instance, in O'Donnell et al.'s [111] ILEX system, potential rhetorical relations are established between sets of facts. In the work of Demir et al. [43], *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, e.g., in-Demir et al.'s approach [43] or be statistically inferred from a corpus of texts aligned with data as, e.g., in Barzilay and Lapata's approach [4].

During content selection, message determination can also be performed. Message determination groups content units into messages that can later facilitate linguistic expression.[7] These messages may correspond to fine-grained content units such as facts or events (e.g., [111,114]) or to topic-related groupings of fine-grained content units (e.g., [121]), and be realized in the text as constituents, clauses or sentences. Message determination has been addressed mainly using templates in a pipeline architecture [121], thus ensuring that each message can be rendered in natural language. For example, the following `RainSpellMsg` message taken from a weather reporting generator is only built if rain occurs on more than a specified number of days in a row [121]:

```
((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))))
```

Message determination can also occur after content selection and prior to discourse structuring and the Elementary Discourse Units (EDUs) thus built are used as basic units for discourse structuring (e.g., [139]).

2.2.2. *Discourse structuring*

Although often handled together with content selection as a single task using text schemas in the sense of McKeown [96] (see [121]), 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 either after [71,92,123] or interleaved with the selection of content [95,107,111]. In this latter case, during the search for the relevant content through the content graph, only those nodes are taken into account that can be connected to nodes already selected via a discourse relation—thus ensuring discourse coherence. The discourse relations between content nodes are introduced either directly into the content graph prior to NLG proper [123], via *plan operators* during text planning [71,95,107], or via a projection from semantic relations established between content nodes [24,82].

A very popular discourse structure theory in NLG is *Rhetorical Structure Theory* (RST) [90] 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 to build up coherent discourse structures; see, e.g., [72,92,107].

In some approaches, the problem of discourse structuring is reduced to the problem of determining the best order between selected facts using empirical approaches [46,50].

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 [60], sequential graph-rewriting approaches [51,109,110,130] and structure mapping approaches [80,89,113,138].

Discrimination networks amount to decision trees that select the most specific lexical unit (i.e., leaves) that subsumes the target object or event according to some context specified in the non-terminal nodes

[7] This type of content selection is sometimes referred to as content determination [121], which we think has a broader scope than content selection.

which governs path selection. For example, the concept `Ingest` with value restrictions `actor:John` and `theme: Milk027` will lead to the selection of the lexical unit `'drink'` since `theme: Milk027` is liquid.

Graph-rewriting approaches as a rule presuppose that the lexical entities in the lexicons are defined in terms of semantic/conceptual forms, such that the selection of the most appropriate lexical item for a given semantic input is essentially performed using pattern matching of the input against the semantic forms of the lexical entities, be it through unification [51,110,130] or some matching metric [109].

Structure mapping approaches map a given input structure onto a lexicalized output structure by using 1) a semantic dictionary that maps a sense onto one or more lexical entries, and 2) a lexicon that provides the lexico-syntactic constraints for each individual entry. For instance, the semantic dictionary will map the semanteme 'cause' onto one of the lexical items [*to*] *cause*, *cause*$_N$, *because*, *due to*, depending on the role of 'cause' in the input semantic structure (e.g., head of a full statement, argument of the head of a full statement, or discourse marker between statements). For [*to*] *cause*, the lexicon will contain subcategorization patterns which specify that its second actant can be an NP, a subordinate with *that*, etc. For *cause*$_N$, it will contain, e.g., the subcategorization pattern with *cause*$_N$'s second actant as a PP with *of*, and so on. See, e.g., [89] for a detailed example.

Both graph-rewriting and structure mapping approaches are especially relevant for multilingual generation since they imply a clear distinction between a semantic or conceptual (language-independent) layer of representation and a lexical (i.e., language-dependent) layer. These two approaches are also especially suitable for taking into account collocational and other realizational constraints. Thus, for the lexicalization of the predicative item 'treatment', one of the collocations listed in the entry for *treatment* can be picked (e.g., *give* [*a*] *treatment*, *provide* [*a*] *treatment*, *receive* [*a*] *treatment*, *respond* [*to a*] *treatment*), depending on the availability of the arguments of 'treatment' in the input, on the context, etc.

In addition to these three main approaches, some proposals treat the items of the semantic representation as lexical items, such that they do not change the type of the output; see e.g., [119], where no distinction is made between ontology labels and lexical items, and, e.g., the appropriate lexical reference to a dog in a given statement (*animal*, *dog*, *dachshund*, ...) is chosen by navigating in the ontology.

Finally, language-oriented ontologies that are both domain- and task- (i.e., NLG-) independent have been introduced into generation from early on in order to facilitate the mapping between domain representation- and task- (i.e., NLG) -dependent linguistic representation, whilst giving sufficient room for flexible verbalization. The most prominent of these, without doubt, is the *Penman Upper Model* (UM) [7]. Originally used in the context of the systemic-functional generators PENMAN [94] 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]. However, as pointed out by Bateman *et al.* [10], it is important to note that the GUM is not a lexical semantics, but rather a grammatical semantics repository. The node labels in the GUM are, in fact, words rather than semantemes or concepts. Thus, the semantic counterpart in the GUM of the word *courtyard* will be still *courtyard*, although typed as 'spatial'.

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 the systems can scale up to large datasets, and 3) evaluability, so that competitive evaluation of the individual 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 (e.g., SimpleNLG [59]) are the most reused NLG resources, 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 should be and how these representations should be mapped from one to the other (see, e.g., the different representations and mapping approaches for lexicalization described in Section 2.2.3). Efforts such as RAGS ("Reference Architecture for Generation Systems") 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 [101].

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 [58]. Furthermore, any new NLG task or module needs to be plugged in with existing ones so as to produce an evaluable output (i.e., text) [122].

Although it is possible for some NLG applications to use pre-existing independent domain knowledge bases and datasets—see, e.g., data-to-text systems [114] or systems that use existing medical ontologies [28]—many NLG systems require modeling and acquiring of the domain knowledge from the ground up. This means that the datasets used tend to be small and the approaches taken idiosyncratic, mainly template-based. However, as mentioned in Section 2, statistical approaches that tend to require the use of larger datasets are increasingly popular in NLG. 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 assess and interpret raw and/or basic *domain knowledge* to derive new additional knowledge which is typically found in the natural language output, i.e., the so-called *domain communication knowledge* [80]. For example, in the context of the generation of air quality bulletins, raw measurements can be interpreted to find out minima and maxima, to infer a rating from a measurement or to infer a cause relation between pollutant rating(s) and air quality index [139].

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. [19,39] distinguish between SW-oriented NLG applications that help users who are not knowledge engineers to understand and use ontologies, and NLG applications that present in a user-friendly way (e.g., reports, letters) the formal knowledge encoded and manipulated in ontologies by larger applications. Similarly, Smart [126] distinguishes between Natural Language Interfaces (NLIs) for ontology engineering and NLIs for the publication of knowledge based on a specific communicative goal using NLG technologies. Gardent et al. [57] distinguish between applications for querying, for authoring and for verbalizing ontologies, with the latter subsuming both the tasks of documenting ontologies and publishing knowledge in an end-user application.

These uses of NLG of course predate the Semantic Web. For example, Ševčenko [134] presents a template-based approach for the verbalization of the logical axioms of the SUMO language-independent upper-ontology. These verbalized axioms are used 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 tuned 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.[8] 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.[9]

[8]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 over graphical or formal language interfaces (e.g., [78]), 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., [79,137]. In some cases, some generation capabilities exist but are very limited, such as the generation of sentences or phrases to guide the user through the completion of a formal query in natural language using a BNF grammar generated (at least in part) dynamically from the source ontology [17]. Some NLI tools for authoring focus on the creation of new content and therefore do not need to render any previously existing content. These tools present content using the same text introduced by the user when authoring it, without performing any NLG, e.g., [44,86,125].

[9]One exception with respect to input size is Sun and Mellish [132], 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 [66]	OWL-DL ont.	Yes	No	LP, DE	Class description	Say all	No	+	P	Yes	Eng
Jarrar [74]	ORM	Yes	No	DE	Constraint	Say all	No	– –	S	N/A	Multi.
Sun [132]	RDF Graph	Yes	No	LP,DE	Graph	Say all	No	+	S	N/A	Eng
Kaljurand [76]	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 [99,100]	OWL-DL ont.	Yes	No	LP,DE	Class subsumers	Say most relevant facts	No	N/A	N/A	N/A	N/A
Davis [40]	OWL-DL ont.	Yes	No	DE	Classes, Properties Instance	Say all	No	–	P	No	Eng
Stevens [131]	(Subset of) OWL-DL ont.	Yes	No	DE	Class description	Say all	No	–	P	Yes	Eng
Power [115]	(restricted) OWL-DL graph	Mixed	No	DE	Class axiom Individual assertion	Say what user selects	No	–	S	N/A	Eng
Hielkema [67,68]	OWL-Lite/RDF graph	No	No	DE	Metadata description about intellectual artifact	Say what user selects	No	–	T	Yes	Eng
Dongilli [48]	OWL/RDF graph	Yes	No	DE	Query	Say what user selects	No	+	S	Yes	Eng
Wilcock [142]	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 [55]	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 [140]	Protegé frames RDF ont.	Yes	Artist bios (summary, chronology)	LP	Artist name	Say typical facts	Yes	+	T	Yes	Eng
Dai [35]	Semantic network	Yes	No	LP	Specific entity in network	Say most relevant facts	No	+	P	Yes	Chn, Eng
Dannélls [38]	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), Fluency (Flu.)={High(+),Medium(–),Low(– –)}, Output Size={S(entence), P(aragraph), T(ext)}, Coherence (Coh.), Language (Lang.) = {Eng=English, Multi=Multilingual, Gre=Greek, Spa=Spanish, Fin=Finnish, Swe=Swedish, Chn=Chinese}.[a]

[a] *User profile* refers to NLG systems that generate different texts for different users given the same input using an explicit user profile/model.
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. *Coherence* can only apply above the sentence level.

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), see, e.g., [2,21,66]; (ii) to verbalize the content interactively selected by the user, see, e.g., [48, 67,115]; (iii) to verbalize the most typical facts found in (real or virtual) target texts, see, e.g., [38,140]; or (iv) to verbalize the most relevant facts according to the context, see, e.g., [35,55].

Whereas (iii) is typically achieved by closed planning, (iv) is achieved either by closed planning or by open planning, both of which were introduced in Section 2.2.1. Some approaches 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, e.g., [23,24,27].

In what follows we describe the main approaches to address these four communicative goals, focusing on closed planning approaches for saying the most typical things, and on open planning approaches for saying the most relevant things.

3.1.1. *Say (almost) everything*

This communicative goal subsumes approaches in which there is a verbalization request (i) for the entire input graph (e.g., [20,21,132]), (ii) for the entire set of constraints or axioms (*abox* or *tbox*) in the domain model (e.g., [40,74,76]), (iii) for all the possible queries that can be built from a class or an individual associated to a term (e.g., [2]), or (iv) for all the axioms related to a class description (e.g., [66,131]). In (i) and (ii), the content selection element is minimal (if at all) and consists in eliminating redundancies [20,21]. In (iii) and (iv), the content selection consists mainly in selecting the queries or axioms to verbalize for the given term or class description. For instance, starting from the class or individual associated to the input term (e.g., `Enzyme`), Ang et al. [2] retrieve all the transitive query paths from the query term across its object properties up to 3 triples in length. These are paraphrased as natural language queries by the query formulator (e.g., `Which enzyme has been found in fungi, and acts on substrate?`) before being presented to the user for selection. Stevens et al. [131] 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). For the ordering and aggregation of the selected content, e.g., for the generation of class descriptions, a simple albeit idiosyncratic strategy for text planning often suffices.

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 coined by Hallett et al. [62]. In conceptual authoring, a supporting NLI guides the user through the authoring process, whether to formulate a query or to design or populate an ontology, e.g., [48,53,67,115]. The user selects or authors the concepts or individuals from the ontology by editing the underlying knowledge representation displayed to her via an interface editor as NL statements generated automatically from the knowledge representation.[10] 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.

For example, the editor might show the user the feedback text **`[some event]`** associated with a generic event [62]. The brackets indicate that the text acts as a place-holder which can be replaced by a specific event from a set of events presented to the user as a list of verbs on a menu (e.g., `consulted, examined, treated, etc.`). Once the user selects `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 as follows: **`[some person]`** `examined` **`[some person]`** `[`*`in some way`*`]`. The feedback text has three place-holders (two obligatory in boldface and one optional in italics), each of which can be specified by the user through further menu option selection until no more obligatory arguments are required.

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

[10]Power et al. in earlier papers (see, e.g., [117]) refer to this 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.

querying, one can be sure that the input to the system matches a valid query, thus avoiding the pitfall of "linguistic vs. conceptual failure" [54], 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 [61,62,117], the use of conceptual authoring for SW ontologies was a natural step forward [48,53,67,68,115]. Some of these approaches even propose that the authoring process can be automatically supported by OWL DL standard reasoning (i.e., class satisfiability, subsumption, consistency and instance checks [48,115]). For example, the editor would point out to the user that the feedback text *Mary owns an animal* associated with the axiom $\{\texttt{Mary}\} \sqsubseteq \exists\ \texttt{own.animal}$ is redundant as it can be deduced from the following axioms already in the ontology: $\texttt{pet} \sqsubseteq \texttt{animal}$ and $\{\texttt{Mary}\} \sqsubseteq \exists\ \texttt{own.pet}$ [115].

3.1.3. Say the most typical (using closed planning)

As already mentioned in Section 2.2.1, 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. [27] use SPARQL queries as content selection rules. Dannélls et al. [38] retrieve relevant triples from multiple datasets and ontologies using a single SPARQL endpoint 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 [21,38], simple ordering rules (as, e.g., class first, then properties) for prototypical texts [131,142], or partial order constraints as annotations on the ontology [47,55].

Weal et al. [140] use a template-based approach for generating biographies of artists, combining text fragments with sentences generated dynamically from the facts. Both text fragments and facts are harvested from the web using information extraction technologies and 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, open planning approaches introduced in Section 2.2.1 are particularly promising for NLG from large and heterogeneous SW datasets for which the creation of templates and rules becomes prohibitively expensive. In this section, we present two additional approaches to relevance-based planning with a more SW flavour to them.

Mellish et al. [99,100,102] 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, to answer a question such as *What is an A?* where A (*the target*) is an atomic concept in an ontology. They argue that simply selecting axioms for rendering them in natural language might result in a text with overly complex and repetitive sentences, inadequate focus and misleading and incomplete information [102]. They propose instead an approach called "Natural Language Directed Inference" (NLDI) in which the axiom (either original or inferred) which results in the best content plan evaluated according to NLDI constraints is selected. These constraints include 1) using a naturally expressive concept language that, similar to Natural Language, expresses constraints conjunctively rather than disjunctively, 2) limiting the selected concept complexity to match the linguistic complexity of the sentence, and 3) selecting the simplest constraint amongst logically equivalent ones.

Dai et al. [35] present an approach called "Semantic Network Language Generation", 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, each of which can be linguistically realized as a sentence. The 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 [83] method for referring expression generation. An annotated corpus of texts is used to train the patterns used by the system's components. Their approach has been tested on semantically parsed Wikipedia summaries about Chinese cities.

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 [38], SPARQL rules [26], XSLT [141,142] or XML [40] templates, and aggregation patterns based on entity-sharing between triples or axioms [21,67,131] 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 labels on the one end of the continuum, e.g., [40,74]. For example, Davis et al. [40] 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 [116], has been validated in two separate studies on collections of freely available ontologies by Mellish and Sun [103] and Power [118]. Mellish and Sun analyzed the naming patterns of properties and classes whilst Power assessed the complexity and hence linguistic expressibility of OWL axioms. According to these authors, what these direct mapping approaches lose in output fluency they gain in domain independence and simplicity of the engineering solution, thus reducing the cost of creating domain and ontology-specific lexicons; see e.g., [20]. Nonetheless, tasks that are not, strictly speaking, NLG tasks, can further contribute to improve fluency, such as final spell checking [2], removal or monitoring of repetitions/redundancies [20,21,66,140] or presentation of sentence coordinated constituents in a bulleted list [66].

In what follows, we discuss the main different approaches to mapping SW content onto linguistic representation, elaborating first on the approaches that rely on the ontology's linguistic expressibility assumption, including the ones that use Controlled Natural Languages in mapping grammars for axiom verbalization. We then introduce 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. Exploiting the ontology's linguistic expressibility assumption

In this approach, 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 [103], 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 [133] or FrameNet [37] 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 [124,131,144] 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 [66,115]. To palliate for the deficiency in naming conventions, others propose to involve (expert) users of the NLG system in the acquisition of linguistic resources for each domain, such as the creation of new entries in the lexicon for newly added ontology concepts [41,67,115].

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* (CNL), for which an unambiguous mapping from the formal languages used in the Semantic Web can be defined, e.g., [40,41,64,76,127]. 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 [41,76] *Attempto Controlled English* (ACE), a subset of which is used in the round-trip meaning preserving mapping with OWL, and Davis et al.'s [40] CLOnE, a very simple CNL with a bidirectional mapping to only a small subset of OWL.

Finally, it is important to note that these approaches exploit 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 [74] 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 [55], 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, and in the case of Cimiano et al.'s proposal [34]. In NaturalOWL, properties are furthermore 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 *#manufacturedBy* is given in Figure 2. It is basically a template in the form of an RDF annotation with a sequence of slots. Each slot can be filled by an expression referring to the owner of the property (e.g., `a laptop`), the value (or filler) of the property (e.g., `Toshiba`), or a string (e.g., "`was manufactured`"). The `owlnl:retype` element of the first and third slots, with value `re_auto`, allows the system to select automatically the owner and filler's linguistic rendering depending on context (e.g., whether to use a noun phrase like `this laptop` or a pronoun to refer to the owner).

For the purpose of annotating ontologies, the authors of NaturalOWL developed a tool [1,18,56] supported by reasoning that can be used to assign multilingual annotations to OWL ontologies.

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. Thus, if we specify in GUM terms that *build* is of type 'constructive-doing', that there is an "actor" *John* of type 'person', an "actee" *house* of type object, and that between *John* and the person *Mary* an 'inclusive accompaniment' relation holds, we can use the grammatical pattern defined for verbal lexemes that subclassify 'constructive doing' and draw upon the variety of standard realizations of the 'inclusive accompaniment' relation (e.g., *John built a house with Mary*, *John and Mary built a house*, *Mary took part when John built a house*, etc.).

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 op-

```
<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. 2. A micro-plan (or template) for the property *#manufacturedBy* for the NaturalOWL System [55]

erate on any ontology that contains the same four upper relations.

Bouayad-Agha et al. [23,24] use a linguistic generator based on a Meaning–Text Theory (MTT) model [104] that takes as input a conceptual representation in the sense of Sowa [128]. The authors suggest, for the future, to integrate the conceptual representation as part of the linguistic ontology already in use for content selection and text planning (see Section 3.1.3 above) and hence to map the text plan onto the conceptual representation using SW technology.

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, 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 heterogeneous knowledge sources, such as domain and domain communication knowledge (see Section 2.3) and conceptual and discourse representations, can be modelled 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 [35,43,55,99,100,111].

- The availability of vast numbers 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 [35] 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., [18,53,54]), to infer additional domain communication knowledge (see Section 2.3) from existing knowledge by applying domain-specific manually crafted rules (e.g., [25,26]) or to support content selection with subsumption reasoning (e.g., [99,100,102]).

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 discuss 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 (see Section 3.2). However, the codification and modeling of NLG-relevant knowledge in SW standards and formalisms can provide potential benefits for the interoperability between NLG modules and tasks as well as for 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; see, e.g., the *Generalized Upper Model* [9] and the *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) [93].[11,12] Lexical databases which are commonly used by the NLP community, WordNet and Framenet,[13,14] have also been converted to RDF/OWL. In addition to these resources, recently 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

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

data; see for instance LexOnto [32], LexInfo [29] and the *Linguistic Information Repository* (LIR) [106]. 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 [97], one of the main authors of RAGS, this is due to the complexity of the framework, lack of tools to support its implementation (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. In order to exploit the full potential of linked data, NLG will have to adapt to vast amounts of data belonging to multiple domains, described using multiple vocabularies encoded in knowledge representations of varying degrees of expressivity.[15] This will have an impact not only on the interpretation, assessment and selection of content, but also on other NLG tasks which may operate or reason on the same input representation—for instance, ordering content units 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. [38] approach multilingual generation from multiple independent datasets by adopting a unifying view based on the PROTON upper ontology.[16] This view allows them to treat domain-specific and generic datasets as a single body of knowledge with respect to querying and reasoning. For the verbalization of SPARQL queries, Ell et al. [52] exploit the mapping from DBPedia entities to types in the YAGO ontology to lexicalize entities found in the queries.[17] 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 [136]. 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 (see, e.g., [4] for the selection of content 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 together.

4.3. Summarization of large volumes of data

We have already argued that the Semantic Web is both a chance and a challenge for NLG techniques to scale up 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 [65], and NLG provides the means for presenting semantic data in an organized, coherent and accessible way. NLG approaches have started to address the generation of information from large/multiple datasets (see Sections 2.2.1 and 3.1.4). These approaches can be further developed on SW data

[15] 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 [87]. 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; see Halpin et al.'s [63] discussion of the diverse use of the `owl:sameAs` property, which does not indicate only full identity.

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

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

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 rely on what content should be 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., [31,146]. 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 [143] 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 paraphrase, 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. In this context, NLG is used for regeneration guided by SW data, and is combined with parsing, IE and TM techniques 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. [140] (Section 3.1.3) and Dai et al. [35] (Section 3.1.4) are the only approaches 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. There is a long tradition in NLG of working with 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 content (as, e.g., [55]), or a history of the content communicated to a specific user (as, e.g., [36]). 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]). Both modeling of the contextual information and its influence on the NLG process are still rather elementary. 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 (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.

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 [145] that answers questions about pop trivia, Janzen and Maas' system [73] that answers questions about products for a virtual in-store shopping environment, or Weal et al.'s ArtEquAkt system [140] 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, the increasing popularity of the Semantic Web representations among NLG researchers and the (hopefully) increasing awareness of the needs of NLG by the SW researchers are a good foundation for further progress towards this goal.

Acknowledgements

We would like to thank the reviewers for their very constructive and detailed comments on earlier versions of the paper. All mistakes and omissions are of course ours. This work was carried out within the *Personalized Environmental Service Configuration and Delivery Orchestration* (PESCaDO) project, supported by the European Commission under the contract number FP7-ICT-248594.

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* (Stroudsburg, PA, USA, 2005), G. Wilcock, K. Jokinen, C. Mellish, and E. Reiter, Eds., Association for Computational LInguistics, 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), CEUR Workshop Proceedings, Volume 435, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-435/, CEUR.

[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* (Los Alamitos, CA, 2011), D. Al-Dabass, A. Orsoni, A. Pantelous, G. Romero, and J. Felez, Eds., IEEE Explore Digital Library, pp. 366–371.

[4] 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)* (Stroudsburg, PA, USA, 2005), Association for Computational Linguistics, pp. 331–338.

[5] BARZILAY, R., AND LAPATA, M. Aggregation via set partitioning for natural language generation. In *Proceedings of the Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics* (Stroudsburg, PA, USA, 2006), Association for Computational Linguistics, pp. 359–366.

[6] BASILE, V., AND BOS, J. Towards Generating Text from Discourse Representation Structures. In *Proceedings of the 13th European Workshop on Natural Language Generation* (Stroudsburg, PA, USA, 2011), Association for Computational LInguistics, 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* (Stroudsburg, PA, USA, 1990), D. McDonald and M. Meteer, Eds., Association for Computational LInguistics, 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)* (Stroudsburg, PA, USA, 2012), Association for Computational Linguistics, 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 Linguistics* (Stroudsburg, PA, USA, 2006), Associa-

tion for Computational Linguistics, 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)* (Stroudsburg, PA, USA, 2006), Association for Computational Linguistics, 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* (Stroudsburg, PA, USA, 2011), Association for Computational Linguistics, 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 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.8562&rep=rep1&type=pdf. In *Proceedings of the 15th Workshop on Information Technologies and Systems* (2005), K. T. . J. L. Zhao, Ed., 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* (Washington DC, 2007), vol. 2, IEEE Explore, 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), http://gate.ac.uk/conferences/iswc2003/* (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* (Berlin, 2005), A. Gómez-Pérez and J. Euzenat, Eds., Lecture Notes in Computer Science, Springer Verlag, 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)* (Berlin, 2004), F. Meziane and E. Métais, Eds., vol. 3136 of *Lecture Notes in Computer Science*, Springer Verlag, 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* (Stroudsburg, PA, USA, 2012), Association for Computational LInguistics, 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).

[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)* (Berlin, 2012), Lecture Notes in Computer Science, 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)* (Stroudsburg, PA, USA, 2011), Association for Computational LInguistics, 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* (Berlin, 2011), G. Antoniou, M. Grobelnik, E. P. B. Simperl, B. Parsia, D. Plexousakis, P. D. Leenheer, and J. Pan, Eds., Lecture Notes in Computer Science, Springer Verlag, 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)* (Stroudsburg, PA, USA, 2011), Association for Computational Linguistics, 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* (Berlin, 2009), L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. HyvÃűnen, R. Mizoguchi, E. Oren, M. Sabou, and E. P. B. Simperl, Eds., Lecture notes in Computer Science, Springer Verlag, 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), 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, held in conjunction with the International Semantic Web Conference (ISWC)* (Stroudsburg, PA, USA, 2007), Association for Computational LInguistics, 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] CIMIANO, P., LÜKER, J., NAGEL, D., AND UNGER, C. Exploiting Ontology Lexica for Generating Natural Language Texts from RDF Data. In *Proceedings of the 14th European Workshop on Natural Language Generation, held in conjunction with ACL 2013* (Stroudsburg, PA, USA, 2013), Associa-

tion for Computational Linguistics, pp. 10–19.

[35] 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.

[36] DANNÉLLS, D. The Value of Weights in Automatically Generated Text Structures. In *CICLing 2009: Computational Linguistics and Intelligent Text Processing* (Berlin, 2009), Lecture Notes in Computer Science, Springer Verlag, pp. 233–244.

[37] 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), J. Kelleher, B. Mac Namee, I. van der Sluis, A. Belz, A. Gatt, and A. Koller, Eds., Association for Computational Linguistics, pp. 179–183.

[38] 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), ACM, pp. 239–242.

[39] 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 (2005), 64–84.

[40] 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)*, Lecture Notes in Computer Science. Springer Verlag, Berlin, 2008, pp. 50–65.

[41] 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, Berlin, 2009, pp. 276–308.

[42] DEEMTER, K. V., KRAHMER, E., AND THEUNE, M. Real vs. Template-based NLG: a False Opposition? *Computational Linguistics 31*, 1 (2005), 15–24.

[43] 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)* (Stroudsburg, PA, USA, 2010), J. Kelleher, B. Mac Namee, and I. van der Sluis, Eds., Association for Computational LInguistics, pp. 17–27.

[44] DENAUX, R., DIMITROVA, V., COHN, A., DOLBEAR, C., AND HART, G. Rabbit to OWL: Ontology Authoring with a CNL-Based Tool. In *Proceedings of the Workshop on Controlled Natural Language (CNL)* (Berlin, 2009), N. Fuchs, Ed., Lecture Notes in Computer Science, Springer Verlag, pp. 246–264.

[45] 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), Association for Computational Linguistics, pp. 654–659.

[46] DIMITROMANOLAKI, A., AND ANDROUTSOPOULOS, I. Learning to Order Facts for Discourse Planning in Natural Language Generation. In *Proceedings of the 9th European Workshop on Natural Language Generation, held in conjunction with the Biannual Meeting of the European Chapter of the Association for Computational Linguistics (EACL)* (Stroudsburg, PA, USA, 2003), Association for Computational Linguistics, pp. 23–30.

[47] DONGILLI, P. Natural Language Rendering of a Conjunctive Query http://web.inf.unibz.it/krdb/pub/tr/krdb08-3.pdf. Tech. rep., KRDB Research Centre Technical Report No. (KRDB08-3). Bozen, IT: Free University of Bozen-Bolzano, 2008.

[48] 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* (Palo Alto, CA, 2006), G. Sutcliffe and R. Goebel, Eds., AAAI Press, pp. 658–663.

[49] DUBOUE, P., AND MCKEOWN, K. Content Planner Construction via Evolutionary Algorithms and a Corpus-based Fitness Function. In *Proceedings of the 2nd International Natural Language Generation Conference (INLG'02)* (Stroudsburg, PA, USA, 2002), Association for Computational Linguistics, pp. 89–96.

[50] DUBOUE, P. A., AND MCKEOWN, K. R. Empirically Estimating Order Constraints for Content Planning in Generation. In *Proceedings of the 39th Annual Meeting of Association for Computational Linguistics* (Stroudsburg, PA, USA, 2001), Association for Computational Linguistics, pp. 172–179.

[51] ELHADAD, M., MCKEOWN, K., AND ROBIN, J. Floating Constraints in Lexical Choice. *Computational Linguistics 23*, 2 (1997), 195–239.

[52] ELL, B., VRANDECIC, D., AND SIMPERL, E. SPARTIQULATION: Verbalizing SPARQL Queries. In *Proceedings of the Workshop Interacting with Linked Data (ILD 2012), CEUR Workshop Proceedings Volume 913 http://ceur-ws.org/Vol-913* (2012), C. Unger, C. P., L. V., E. Motta, B. P., and C. R., Eds., CEUR.

[53] 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) http://ceur-ws.org/Vol-565/* (2010), S. Handschuh, T. Heath, V. T. Thai, I. Dickinson, L. Aroyo, and V. Presutti, Eds., CEUR Workshop Proceedings, Volume 565, CEUR.

[54] FRANCONI, E., GUAGLIARDO, P., AND TREVISAN, M. Quelo: a NL-based Intelligent Query Interface. In *Pre-Proceedings of the Second Workshop on Controlled Natural Languages http://ceur-ws.org/Vol-622/* (2010), M. Rosner and N. Fuchs, Eds., CEUR Workshop Proceedings, Volume 622, CEUR.

[55] 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* (Stroudsburg, PA, USA, 2007), Association for Computational Linguistics, pp. 143–146.

[56] 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), Association for Computational Linguistics, pp. 17–20.

[57] 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).

[58] GATT, A., AND BELZ, A. Introducing Shared Tasks to NLG: The TUNA Shared Task Evaluation Challenges. In *Empirical Methods in Natural Language Generation: Data-oriented Methods and Empirical Evaluation*, E. Krahmer and M. Theune, Eds., Lecture Notes in Computer Science. Springer Verlag, Berlin, 2010, pp. 264–293.

[59] GATT, A., AND REITER, E. SimpleNLG: a Realisation Engine for Practical Applications. In *Proceedings of the 12th European Workshop on Natural Language Generation, held in conjunction with the Binnual Meeting of the European Chapter of the Association for Computational Linguistics* (Stroudsburg, PA, USA, 2009), Association for Computational Linguistics, pp. 90–93.

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

[61] HALLETT, C., SCOTT, D., AND POWER, R. Intuitive Querying of e-Health Data Repositories. In *Proceedings of the UK E-Science All-Hands Meeting http://www.allhands.org.uk/2005/proceedings/proceedings/proceedings.pdf* (2005), S. Cox and D. Walker, Eds., EPSRRC.

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

[63] HALPIN, H., AND HAYES, P. J. When OWL:sameAs isn't the Same: An Analysis of Identity Links on the Semantic Web. In *Proceedings of the Linked Data on the Web Workshop (LDOW2010), CEUR Workshop Proceedings Volume 628* (2010), C. Bizer, T. Heath, T. Berners-Lee, and M. Hausenblas, Eds., CEUR.

[64] HART, G., JOHNSON, M., AND DOLBEAR, C. Rabbit: Developing a Control Natural Language for Authoring Ontologies. In *The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008, Proceedings.* (Berlin, 2008), S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, Eds., Lecture Notes in Computer Science, Springer Verlag, pp. 348–360.

[65] 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), held in conjunction with ISWC 2006* (2006), L. Rutledge, Ed.

[66] HEWLETT, D., KALYANPUR, A., KOLOVSKI, V., AND HALASCHEK-WIENER, C. Effective NL Paraphrasing of Ontologies on the Semantic Web. In *Proceedings of the Workshop on End-User Semantic Web Interaction, held in conjunction with the 4th Internatioanl Semantic Web conference (ISWC), CEUR Workshop Proceedings, Volume 172 http://ceur-ws.org/Vol-172/* (2005), A. Bernstein, I. Androutsopoulos, D. Degler, and B. McBride, Eds., CEUR.

[67] 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), Association for Computational Linguistics, pp. 69–72.

[68] HIELKEMA, F., MELLISH, C., AND EDWARDS, P. Evaluating an Ontology-driven WYSIWYM Interface. In *Proceedings of the Fifth International Natural Language Generation Conference* (Stroudsburg, PA, USA, 2008), Association for Computational Linguistics, pp. 138–146.

[69] 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, London, etc., 1990, pp. 193–227.

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

[71] 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.

[72] HOVY, E. Automated Discourse Generation Using Discourse Structure Relations. *Artificial Intelligence 63*, 1-2 (1993), 341–386.

[73] 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* (Los Alamitos, CA, 2009), IEEE Computing Society, pp. 361–366.

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

[75] 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.

[76] KALJURAND, K., AND FUCHS, N. E. Verbalizing OWL in Attempto Controlled English. In *Proceedings of Third International Workshop on OWL: Experiences and Directions http://ceur-ws.org/Vol-258/* (Washington DC, 2007), C. Golbreich, A. Kalyanpur, and B. Parsia, Eds., CEUR Workshop Proceedings, Volume 258, CEUR.

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

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

[79] KAUFMANN, E., BERNSTEIN, A., AND ZUMSTEIN, R. Querix: a Natural Language Interface to Query Ontologies based on Clarification Dialogs. In *The Semantic Web—ISWC 2006* (Berlin, 2006), I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, Eds., Lecture Notes in Computer Science, Springer Verlag, pp. 980–981.

[80] KITTREDGE, R., KORELSKY, T., AND RAMBOW, O. On

the Need for Domain Communication Knowledge. *Computational Intelligence 7*, 4 (1991), 305–314.

[81] 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)* (Stroudsburg, PA, USA, 2012), Association for Computational Linguistics, pp. 752–761.

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

[83] KRAHMER, E., ERK, S., AND VERLEG, A. Graph-Based Generation of Referring Expressions. *Computational Linguistics 29*, 1 (2003), 53–72.

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

[85] KRAHMER, E., AND VAN DEEMTER, K. Computational Generation of Referring Expressions: A Survey. *Computational Linguistics 38*, 1 (2012), 173–218.

[86] KUHN, T. Acewiki: A Natural and Expressive Semantic Wiki. In *Proceedings of the Fifth International Workshop on Semantic Web User Interaction (SWUI 2008), CEUR Workshop Proceedings, Volume 543* (2008), D. Degler, M. Schraefel, J. Golbeck, A. Bernstein, and L. Rutledge, Eds., CEUR.

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

[88] LANGKILDE, I., AND KNIGHT, K. Generation that Exploits Corpus-based Statistical Knowledge. In *Proceedings of the 17th international conference on Computational linguistics (COLING) and the 36th Annual Meeting of the Association for Computational Linguistics* (Stroudsburg, PA, USA, 1998), Association for Computational Linguistics, pp. 704–710.

[89] 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.

[90] 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.

[91] 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)* (Stroudsburg, PA, USA, 2005), G. Wilcock, K. Jokinen, C. Mellish, and E. Reiter, Eds., Association for Computational Linguistics, pp. 101–108.

[92] 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)* (Palo Alto, CA, 1997), AAAI Press, pp. 629–635.

[93] MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., AND OLTRAMARI, A. IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web. Deliverable D18, Ontology Library (final) http://www.loa.istc.cnr.it/papers/d18.pdf, 2009.

[94] 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.

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

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

[97] 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), J. Kelleher, B. Mac Namee, I. van der Sluis, A. Belz, A. Gatt, and A. Koller, Eds., Association for Computational Linguistics, pp. 85–93.

[98] 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), E. Hovy, C. DiMarco, and G. Hirst, Eds., pp. 98–107.

[99] MELLISH, C., AND PAN, J. Finding Subsumers for Natural Language Presentation. In *Proceedings of the International Workshop on Description Logics, CEUR Workshop Proceedings, Volume 189, 2006* (2006), B. Parsia, U. Sattler, and D. Toman, Eds., CEUR, pp. 127–134.

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

[101] 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.

[102] 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* (Stroudsburg, PA, USA, 2005), Association for Computational Linguistics, pp. 118–124.

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

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

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

[106] 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.

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

[108] NICOLOV, N., MELLISH, C., AND RITCHIE, G. Sentence Generation from Conceptual Graphs. In *Proceedings of the Third International Conference on Conceptual Structures, ICCS '95. Conceptual Structures: Applications, Implementation and Theory* (Berlin, 1995), Lecture Notes in Computer Science, Springer Verlag, pp. 74–88.

[109] NIRENBURG, S., MCCARDELL, R., NYBERG, E., HUFFMAN, S., KENSCHAFT, E., AND NIRENBURG, I. Lexical

Realization in Natural Language Generation. In *Proceedings of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages* (1988), pp. 18–26.

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

[111] 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.

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

[113] 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.

[114] 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.

[115] 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), Association for Computational Linguistics, pp. 9–15.

[116] 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), Association for Computational Linguistics, pp. 132–136.

[117] 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)* (Chichester/London/New York, 1998), H. Prade, Ed., Wiley, pp. 675–681.

[118] 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), Association for Computational Linguistics, pp. 1006–1013.

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

[120] 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)* (Stroudsburg, PA, USA, 1994), D. McDonald and M. Meteer, Eds., Association for Computational LInguistics, pp. 163–170.

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

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

[123] 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.

[124] SCHÜTTE, N. Generating Natural Language Descriptions of Ontology Concepts. In *Proceedings of the 12th European Workshop on Natural Language Generation (EWNLG)* (Stroudsburg, PA, USA, 2009), Association for Computational Linguistics, pp. 106–109.

[125] SCHWITTER, R., AND TILBROOK, M. Controlled Natural Language Meets the Semantic Web. In *Proceedings of the Australasian Language Technology Workshop* (Canberra, 2004), A. Asudeh, C. Paris, and S. Wan, Eds., The Australian Speech Science and Technology Association, pp. 55–62.

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

[127] SMART, P., BAO, J., , 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)* (Berlin, 2009), N. Fuchs, Ed., Lecture Notes in Computer Science, Springer Verlag, pp. 206—225.

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

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

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

[131] 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).

[132] 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 at the European Conference on AI (ECAI)* (Riva del Garda, Italy, 2006), C. Callaway, A. Corradini, J. Kreutel, J. Moore, and M. Stede, Eds.

[133] 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* (Stroudsburg, PA, USA, 2007), Association for Computational Linguistics, pp. 105–108.

[134] ŠEVČENKO, M. Online Presentation of an Upper Ontology. In *Proceedings of Znalosti 2003, ISBN 80-248-0229-5* (2003), V. Svátek, Ed., VŠB, Technical University of Ostrava.

[135] 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.

[136] WALTER, S., UNGER, C., AND CIMIANO, P. A Corpus-Based Approach for the Induction of Ontology Lexica. In *Proceedings of the 18th International Conference on Applications of Natural Language to Information Systems (NLDB)* (Berlin, 2013), E. Métais, F. Meziane, M. Saraee, V. Sugumaran, and S. Vadera, Eds., Lecture Notes in Computer Science, Springer Verlag, pp. 102–113.

[137] 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 Verlag, Berlin, 2007, pp. 473–487.

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

[139] 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.

[140] 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 (2007), 537–562.

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

[142] WILCOCK, G., AND JOKINEN, K. Generating Responses and Explanations from RDF/XML and DAML+OIL. In *Proceedings of the Workshop Knowledge and Reasoning in Practical Dialogue Systems, held in conjunction with the International Joint Conference on Artificial Intelligence IJCAI* (2003), A. Jönsson, J. Alexandersson, T. Becker, K. Jokinen, and M. Merkel, Eds., vol. 2003, pp. 58–63.

[143] 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.

[144] 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), Association for Computational Linguistics, pp. 197–201.

[145] 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) http://www.informatik.uni-trier.de/ ley/db/conf/icaart/icaart2009.html* (2009), J. Filipe, A. Fred, and B. Sharp, Eds., INSTICC Press, pp. 115–122.

[146] 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), ACM, pp. 707–716.