"Hidden semantics": what can we learn from the names in an ontology?*

Allan Third
Computing Department, Open University, UK
a.third@open.ac.uk

Abstract

Despite their flat, semantics-free structure, ontology identifiers are often given names or labels corresponding to natural language words or phrases which are very dense with information as to their intended referents. We argue that by taking advantage of this information density, NLG systems applied to ontologies can guide the choice and construction of sentences to express useful ontological information, solely through the verbalisations of identifier names, and that by doing so, they can replace the extremely fussy and repetitive texts produced by ontology verbalisers with shorter and simpler texts which are clearer and easier for human readers to understand. We specify which axioms in an ontology are "defining axioms" for linguistically-complex identifiers and analyse a large corpus of OWL ontologies to identify common patterns among all defining axioms. By generating texts from ontologies, and selectively including or omitting these defining axioms, we show by surveys that human readers are typically capable of inferring information implicitly encoded in identifier phrases, and that texts which do not make such "obvious" information explicit are preferred by readers and yet communicate the same information as the longer texts in which such information is spelled out explicitly.

1 Introduction

There has been increasing interest in recent years in the generation of natural language texts from, or using, ontologies ((Cregan et al., 2007; Kaljurand and Fuchs, 2007; Smart, 2008), for example). Such "verbalisations" – translations of the logic of, for example, OWL (W3C Consortium, 2012), into human-readable natural language – can be useful for a variety of purposes, such as communicating the results of ontology inference, generating custom texts to suit a particular application domain or assisting non-ontology-experts in authoring, reviewing and validating ontologies. This paper takes as its starting point an observation about ontology structure and use. The purpose of an ontology (specifically, the so-called "T-box"[1]) is to define the terms of a particular domain in order to allow automated inference of the semantics of that domain. Given that machines are essentially *tabulae rasae* with regard to nearly any kind of world knowledge, it is therefore necessary to spell out the meanings of most terms in what (to a human) would be excruciating detail. In most, if not all, ontology languages, and certainly in OWL, identifiers – the "names" for individual entities, classes and relations[2] – are atomic units. That is to say, every identifier is treated by the machine as simply a flat string, with no internal structure or semantics. The corresponding natural language constructions – noun and verb phrases – by contrast have a very rich internal structure which can communicate very subtle semantic distinctions. Best practice for human ontology developers recommends that for every entity in an ontology, either its identifier should be a meaningful simple or complex term, or it should have a (localised) label which is a meaningful simple or complex natural language

*Many thanks to Richard Power and Sandra Williams for their help and comments. This work was supported by Engineering and Physical Sciences Research Council Grant Ref. G033579/1.

[1] "Terminology box"

[2] "Property" is the OWL terminology for a relation between two entities

INLG 2012 Proceedings of the 7th International Natural Language Generation Conference, pages 67–75,
Utica, May 2012. ©2012 Association for Computational Linguistics

term. For example, in the domain of education, a class intended to represent the real-world class of junior schools ought to have (in English) an identifier such as `junior_school` or a label such as "junior school". Ontology developers who follow this best practice (and, according to (Power, 2010), the vast majority do) produce ontologies in which the entities are easily recognisable and understood by human readers who can parse these identifiers, to infer, for example, that "junior school" is a subclass of the class "school". As it stands, however, a machine will *not* make this inference. In order for the machine to comprehend the semantics of this example, there must additionally be an axiom equivalent to "a junior school is a school".

The motivation for this work is the desire to identify which kinds of identifier or label are "obvious" in this way. That is to say, if we treat an OWL identifier as if it were in fact a multi-word natural language expression, can we infer at least some of its semantics from its properties as a noun phrase, for example? This has two overall purposes: given an existing ontology, definitional axioms for "obvious" identifiers can be omitted when verbalising for a human user, in order to shorten the text and make it more relevant, and, conversely, during the process of ontology creation, if a human uses an obvious identifier, a reasonable guess can be made as to its definitional axiom(s), and these can be presented to the user for confirmation, thus saving the user the need to spend time and energy spelling out the obvious for the machine's purposes. This paper addresses the first of these two purposes. Note that the aim of this work is not particular to consider how best to *realise* entity names in a verbalisation, but rather, how to use the names of entities to *guide* the choice and construction of sentences.

This work was undertaken in the context of the SWAT (Semantic Web Authoring Tool) project, which is investigating the application of NLG/NLP to ontology authoring and editing (Williams et al., 2011),(Williams and Power, 2010),(Power and Third, 2010),(Stevens et al., 2011), (Power, 2010),(The SWAT Project, 2012).

2 Existing work

Other researchers have attempted to take advantage of the internal structure of ontology identifiers to infer semantics, but these have exclusively been concerned with the question of checking or improving an ontology's coverage of its domain. Examples include (Wroe et al., 2003; Fernandez-Breis et al., 2010; Egaña Aranguren et al., 2008). To the best of our knowledge, our current research is the first to take advantage of identifier structure to infer semantics in order to improve verbalisation and produce more human-focused texts.

3 Hypothesis

Informal feedback from existing work indicates that many readers are dissatisfied with the kinds of text produced by ontology verbalisers, feeling them to be somewhat fussy and unnatural. Some of this can no doubt be put down to the verbalisations themselves – it is very difficult to find a natural way to express that one property is the inverse of another without resorting to mathematical variables – but, as with other generation tasks, the problem is not necessarily just how things are said, but also in the selection of which things to say at all. Our hypothesis takes two parts:

1. linguistically-complex identifiers/labels are often defined by "obvious" axioms in the OWL ontologies containing them, and
2. ontology verbalisations which omit such "obvious" axioms lead to a better reading experience for users without sacrificing content.

A prerequisite for these is also the claim that linguistically-complex identifiers are reasonably common in ontologies. (Power, 2010) demonstrated very clearly that recommended best-practice is in fact followed very well in much ontology development, and entities do tend to be given meaningful names.

One caveat is necessary here. We are talking about what an average human reader might reasonably expect to follow from a given use of language. However, observing that a black horse is a horse, a grey horse is a horse, a brown horse is a horse, and so on, does not guarantee the truth of any inferences we might make on encountering a reference to a clothes-horse. There will always be situations in which ordinary uses of language will need to be made more precise. An interesting future direction of this work would be to investigate whether it is possible to detect exactly when such clarification is necessary, in the context of ontology verbalisation, at least.

4 Definitions

Of course, "complex", "obvious", and so on can be loaded terms, and it is necessary to make them precise before continuing.

4.1 Simple and complex identifiers

Identifiers[3] may consist of a single natural language word, in which case we call it *simple*, or multiple words, in which case it is *complex*. The words in a complex identifier may be separated by spaces ("junior school"), punctuation (`junior_school`) or capitalisation (`juniorSchool`). In any case, it is trivial to separate these words into a list automatically.

4.2 Content words

In looking for "defining" axioms, we often need to ignore words occurring in complex identifiers which have some grammatical function. For example, if comparing "has as father" to other identifiers in the same ontology, we may ignore "has" and "as" and consider only the *content word* "father". "Has" occurs far too frequently to be of any use in identifying axioms relating to the semantics of "has as father", although it is of course relevant to what those semantics actually are in any one of such axioms.

4.3 Constructed identifiers

A complex identifier is *constructed* if its component words (or just its content words) are themselves simple identifiers in the containing ontology. For example, if an ontology contains identifiers corresponding to "French", "woman" and "French woman", then "French woman" is a constructed identifier. We may wish to relax this definition slightly to consider constructed identifiers where *most* of the component (content) words are also identifiers, or where component words are morphological variants of other identifiers.

[3]Henceforth, for brevity, "identifier" may mean "OWL identifier", if that is human-meaningful, or it may mean "label", otherwise.

4.4 Defining axioms

The meaning of a constructed identifier can be *defined* in an ontology by axioms in which all, or most, of its component or content words occur as, or in, other identifiers. For example, if there is an identifier `van_driver`, there is likely to be an axiom similar to

A van driver drives a van.

So, for a complex identifier I, we take an axiom A to be a *defining axiom* if:

- A contains at least two distinct identifiers,
- I occurs in A, and either
- for each identifier $J \neq I$ in A, the content words in J are a subset of the content words in I, OR
- the content words in I are a subset of the union of the content words of at least two other identifiers in A.

The third condition is relatively straightforward – a phrase such as "white van man" can be defined in OWL using at most terms corresponding to "white", "van" and "man", but not every word in a complex phrase must appear in its defining axiom. Adjectives often work like this: we accept "a junior school is a school" as being a defining axiom of "junior school", but "junior" only appears in the definiendum. It is worth noting here that a defining axiom need not be the whole of the definition of its definiendum; a complex identifier may have more than one defining axiom associate with it, in which case its definition would be the set of all of its defining axioms.

The fourth condition perhaps seems stranger. The intention here is to capture defining axioms such as

A French woman is a woman whose nationality is French

where "nationality" is *not* a content word of "French woman", and yet there is an underlying relationship between this "extra" word/phrase and one of the content words of "French woman", namely in this case that "French" is a nationality. One goal of future work might be to look into ways to identify such underlying relationships from OWL semantics in order to use them in new contexts.

5 Corpus study

So far, we have given criteria for which identifiers we consider to be linguistically-complex and for which axioms we believe serve as definitions for such identifiers. The immediately obvious question is whether these criteria are useful. To test this, we evaluate them against a corpus of 548 OWL ontologies collected from a variety of sources, include the Tones repository (TONES, 2012), the Swoogle semantic web search engine (Finin et al., 2004) and the Ontology Design Patterns corpus (ODP, 2012). The corpus includes ontologies on a wide range of topics and featuring a variety of authoring styles.

By using the OWL API (OWL API, 2012), a Java program was written to scan through the corpus for identifiers matching the definition of "complex" above, and for each such identifier found, look for defining axioms for it. Of the logical entity types possible in OWL – Named Individuals, Classes, Data- and Object-Properties – it was decided to omit Named Individuals from the current study. Much as with proper nouns in natural language, the names of individuals typically have less internal structure than other kinds of entity or relation names, and those which do have such structure (such as, e.g., "Lake Windermere") are usually very simple. Individuals are also not really "defined" as such. One may state what are deemed to be highly-salient features about them, such as that Lake Windermere is a lake, but this is not a definition. Had we included individuals in this study, it was thought that the large number of non-defined names would artificially depress the statistics and give an inaccurate view of the phenomena being studied. Re-running the analysis including Named Individuals confirmed this hypothesis: less than 10 ontologies in the whole corpus contained any defining axioms for named individuals, with the most common pattern having only 77 occurrences in the whole corpus – a negligible frequency. It would be interesting to look at these cases in more detail, however, to examine what kinds of individual are defined in this way.

Having identified defining axioms across the corpus, the results were then abstracted, by replacing the occurrences of content words of each identifier in an axiom with alphabetic variables, so that

`SubClassOf(junior_school school)`

and

`SubClassOf(domestic_mammal mammal)`

both become

`SubClassOf(AB B).`

The occurrences of each abstracted axiom pattern could then be counted and tabulated. Table 1 shows the most frequent 10 patterns, comprising 43% of all results. Across the whole corpus, 69% of all identifiers were complex, according to our definition, and of those, 45% had at least one defining axiom. These figures indicate that the phenomenon we are investigating is in fact a very common one, and hence that any improvements to ontology verbalisation based on taking advantage of identifier semantics are likely to be significantly useful.

Of all the patterns identified, 64% involve the SubClassOf axiom type ("A junior school is a school"). A further 14% involve InverseObjectProperties ("Bill is father of John if and only if John has father Bill"), and another 14% involve ObjectPropertyDomain or ObjectPropertyRange ("If something has as segment X, then X is a segment"). Collectively, then, these axiom types cover 92% of all defining axioms. An informal glance at the results involving SubClassOf axioms shows that what appears to be the case in Table 1 is generally true – the bulk of the SubClassOf axioms involve some form of adjective construction.

It should be noted here that the intention was to use the absolute bare minimum of linguistic knowledge in identifying these axioms – almost everything is done by matching substrings – in order to avoid influencing the results with our own intuitions about how we think it ought to look. It is reassuring to see nonetheless how far it is possible to get without involving linguistic knowledge. Indeed, one of the ontologies in the test corpus has identifiers in Italian, and it was confirmed by a bilingual English/Italian speaker that the axiom patterns our software identified for that ontology were just as "obvious" in Ital-

Table 1: 10 most frequent patterns of defining axiom

No. of occurrences	Pattern	Example
1430	SubClassOf(AB B)	SubClassOf(representation-activity activity)
1179	SubClassOf(ABC BC)	SubClassOf(Quantified set builder Set builder)
455	InverseObjectProperties(hasA isAof)	InverseObjectProperties(HasInput IsInputOf)
387	SubClassOf(ABCD BCD)	SubClassOf(Continental-Statistical-Water-Area Statistical-Water-Area)
348	SubClassOf(ABCD CD)	SubClassOf(NonWikipediaWebPage WebPage)
240	SubClassOf(ABC AC)	SubClassOf(Process-Resource-Relation Process-Relation)
229	ObjectPropertyRange(hasA A)	ObjectPropertyRange(hasAnnotation Annotation)
192	ObjectPropertyRange(hasAB AB)	ObjectPropertyRange(hasTrustValue TrustValue)
188	InverseObjectProperties(AB ABof)	InverseObjectProperties(situation-place situation-place-of)
179	InverseObjectProperties(Aof hasA)	InverseObjectProperties(contentOf hasContent)

ian as they are in English. There are limitations, of course. A defining axiom such as "a pet owner is a person who owns a pet" would *not* be picked up by this software, as "owner" and "owns" do not match each other as strings. To bypass this particular limitation, the software has been modified to allow the optional use of a (language-specific) stemming algorithm before substring matching, so that both "owner" and "owns" would be matched as "own", for example. The current work, however, focuses on the non-stemmed results for reasons of simplicity and time; we intend to carry out further research using the stemmed results in future.

6 Generation study

6.1 Design

A core part of our claim for these defining axioms is that their semantics are in some sense "obvious". A human reading a phrase such as "junior school" is unlikely to need to be told explicitly that a junior school is a school. This claim needs to be tested. Furthermore, it was suggested above that ontology verbalisations would be improved in quality for human readers if such "obvious" sentences were omitted and the semantics implied by the internal structure of noun and verb phrases were used to improve verbalisation. Again, it is necessary to test whether any improvement does occur.

In order to test the first of these claims, a survey was designed, in which each question would consist of a (verbalised) identifier phrase, followed by three sentences containing that identifier phrase. Respondents were asked to indicate which of those sentences, if any, they were able to deduce from the phrase itself, without relying on any domain knowledge. The questions were based on the top 8 patterns of defining axiom from Table 1, and the containing ontology of each was verbalised using the SWAT ontology verbalisation tool ((The SWAT Project, 2012)). The choice of 8 was motivated by an intended survey size of 10 to 14 questions allowing for some duplication of patterns in order to vary, e.g., the order of elements in sentences, and to minimise the effects of possible domain knowledge on behalf of respondents.

Figure 1 shows an example of a question from the first survey. The prediction was that respondents would be more likely to select sentences based on a defining axiom pattern than sentences which are not based on any such pattern.

The second claim required a more involved test. It was decided to present respondents with two paragraphs of text, both verbalised from the same set of axioms "about" the same class or property. One paragraph of each pair contains verbalisations of every initial axiom, possibly with common subjects or objects aggregated between sentences (the "full" paragraph). The other omitted the verbalisations of any defining axioms, and allowed aggregation of common elements from *within* identifiers where that was justified by one of the omitted defining axiom. For example, the already-aggregated (in the full paragraph) sentence

The following are kinds of atrium cavity: left atrium cavity, right atrium cavity

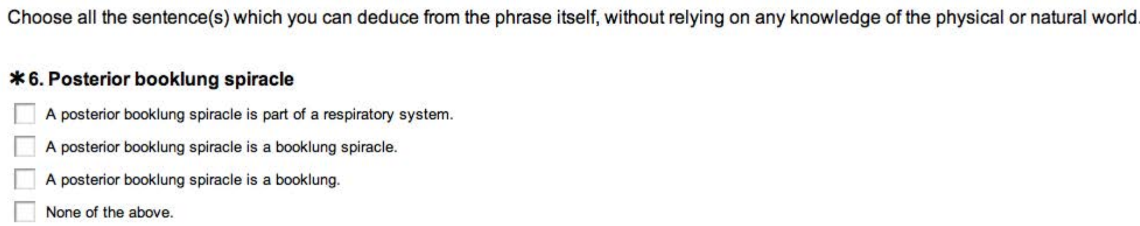
was further aggregated to

The following are kinds of atrium cavity: left, right.

because of the defining axioms

A left (right) atrium cavity is an atrium cavity.

This second paragraph is the "simplified" paragraph. Both paragraphs were checked in each case to ensure that the original set of axioms could be


Choose all the sentence(s) which you can deduce from the phrase itself, without relying on any knowledge of the physical or natural world.

*** 6. Posterior booklung spiracle**

- [ ] A posterior booklung spiracle is part of a respiratory system.
- [ ] A posterior booklung spiracle is a booklung spiracle.
- [ ] A posterior booklung spiracle is a booklung.
- [ ] None of the above.

Figure 1: Sample question from survey 1

inferred without any external information, providing an objective guarantee that both carried the same semantic content even if one only did so implicitly. Respondents were asked to judge whether each pair of paragraphs expressed the same information, to express a preference (or lack of preference) for one paragraph over the other, and to select those sentences from each paragraph which conveyed information which was not conveyed by the other paragraph. Figure 2 shows an example survey question.

Three hypotheses were tested simultaneously by this survey. The first was that respondents would be able to detect when two paragraphs contained the same information at a probability significantly greater than chance and the second that respondents would tend to prefer the simplified paragraphs. The third hypothesis was that respondents would be unlikely to label information as being "missing" from a paragraph when that information was implicitly expressed.

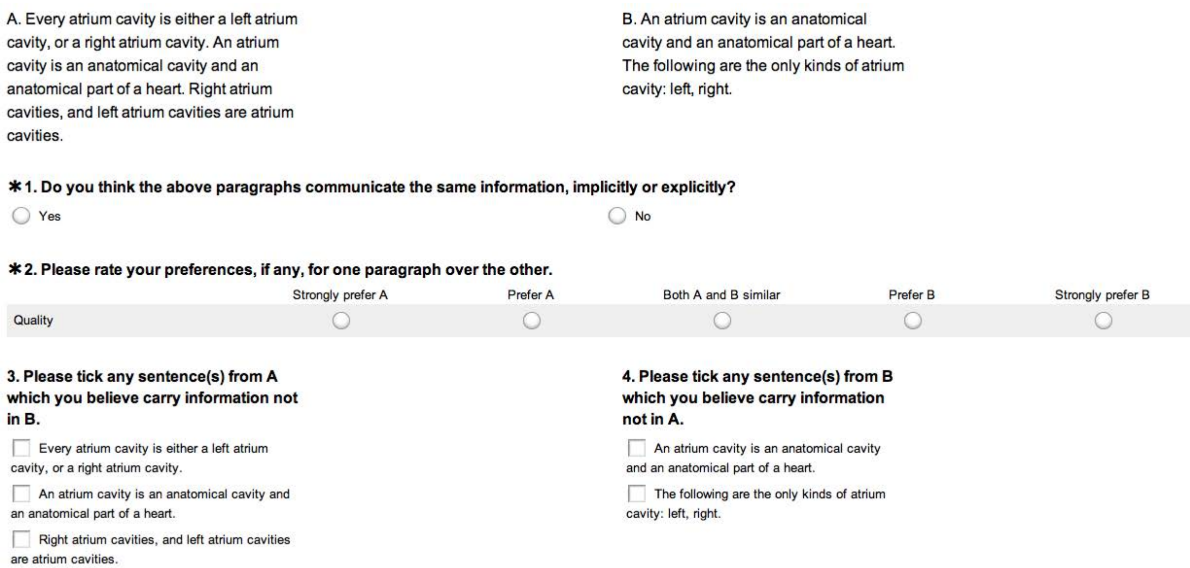
Our initial survey design also included pairs of paragraphs which genuinely did contain different information, to serve as a control, and so respondents' ability to judge pairs of paragraphs as carrying the same information would be compared to their ability to judge when the presence of different information. However, in piloting that design, nearly every respondent reported such examples as being highly confusing and distracting. This is perhaps not surprising; the task of telling when two texts express the same content is not symmetrical with the task of telling when they express different content. The latter is considerably easier, by virtue of the fact that different content will involve different words or phrases, or noticably different sentence structures. Because of this, the decision was taken only to test texts which objectively did contain the same logical content, and to compare the results to chance. Each paragraph pair was controlled to minimise the effects of ordering of information and, where possible, of length.

To maximise take-up and completion, it was decided to try to keep the length of time taken to complete each survey down to around five minutes. Consequently, survey 1 had 14 of the relatively simple identifier inference questions and survey 2 had 4 of the more complex paragraph-comparison questions. Both surveys were published via SurveyMonkey (Monkey, 2012) and were publicised via the social networking sites Twitter (Twitter, 2012), Facebook (Facebook, 2012) and Google+ (Google+, 2012).

6.2 Results and evaluation

The first survey attracted 30 respondents, the second 29. The data collected from the first survey are summarised in Table 2, where S is "sentence predicted to be obvious by a defining axiom pattern" and J is "sentence judged inferrable from the given identifier". Applying a 2×2 χ^2 contingency test results in $\chi^2 = 342.917$, $df = 1$ and $P < 0.0001$, indicating an extremely significant association between the predicted obviousness of a sentence and respondent judgement of that sentence as following from the given identifier.

It is interesting, however, to note the top row of Table 2: for sentences which are predicted to hold, human judges are ambivalent as to whether to judge it as definitely true or not. One interpretation of this result is that, while it is very clear that *non*-defining axioms can *not* be inferred from identifier phrases, people are hesitant to commit to *asserting* these axioms in an unfamiliar domain, perhaps for fear of an unknown exception to the general rule. For example,



A. Every atrium cavity is either a left atrium cavity, or a right atrium cavity. An atrium cavity is an anatomical cavity and an anatomical part of a heart. Right atrium cavities, and left atrium cavities are atrium cavities.

B. An atrium cavity is an anatomical cavity and an anatomical part of a heart. The following are the only kinds of atrium cavity: left, right.

*** 1. Do you think the above paragraphs communicate the same information, implicitly or explicitly?**

○ Yes ○ No

*** 2. Please rate your preferences, if any, for one paragraph over the other.**

	Strongly prefer A	Prefer A	Both A and B similar	Prefer B	Strongly prefer B
Quality	○	○	○	○	○

3. Please tick any sentence(s) from A which you believe carry information not in B.

☐ Every atrium cavity is either a left atrium cavity, or a right atrium cavity.

☐ An atrium cavity is an anatomical cavity and an anatomical part of a heart.

☐ Right atrium cavities, and left atrium cavities are atrium cavities.

4. Please tick any sentence(s) from B which you believe carry information not in A.

☐ An atrium cavity is an anatomical cavity and an anatomical part of a heart.

☐ The following are the only kinds of atrium cavity: left, right.

Figure 2: Sample question from survey 2

while "a Qualifier Noun is a Noun" is usually a good rule of thumb, "a clothes-horse is a horse" is a clear counterexample. So perhaps the better interpretation of these results would be to say that, presented for example with a phrase of the form "Qualifier Noun", a reader would not be surprised if it turned out that the entity referred to is also a "Noun". Either way, these statistics suggest that it could well be safe, when generating texts, to omit defining axioms and allow readers' default assumptions to apply. A simple improvement suggests itself. In the situation where a particular defining axiom pattern would be predicted, but its negation is in fact present, the said negation is automatically highly-salient. It is always likely to be worthwhile verbalising "a clothes-horse is not a horse."

Table 2: Results of the survey on identifier inference.

	J	not J
S	224	211
not S	44	739

It is also interesting to separate out the results of this survey by type of axiom. There were three general families of defining axiom type tested – SubClassOf ("A junior school is a school"), InverseObjectProperties ("Bill is father of John if and only if John has father Bill") and ObjectPropertyRange ("If something has as segment X, then X is a segment"). Table 3 shows the results broken down by these categories, where "SC" is SubClassOf, "IOP" is InverseObjectProperties" and "OPR" is ObjectPropertyRange.

Table 3: Breakdown of identifier inference results by axiom type.

	J	not J
SC	152	109
IOP	52	64
OPR	20	38

A 3×2 χ^2 test results in $\chi^2 = 13.54$, $df = 2$ and $P = 0.001148$, indicating that the judgement of a sentence as obvious or not varies to a significant degree with the type of sentence it is. This is perhaps to be expected, given that not all axiom-types can be verbalised by sentences of similar linguistic complexities. In particular, it is very difficult to see how to verbalise ObjectPropertyRange sentences without appealing to the use of variables such as X and Y, which tend to lead to rather clunky sentences. Sentences corresponding to SubClassOf axioms are most likely to be judged as obvious. Further work is necessary to determine the reasons for these differences empirically.

Table 4: Results of paragraph comparison survey (I)

	Yes	No
Same info	74	22
Prefer simplified paragraph	61	24

Table 4 summarises the results of the "same information" and "preference" questions from the paragraph-comparison survey, aggregated across questions. Comparing each of these to a random distribution of Yes/No answers gives, in turn, $\chi^2 = 15.198$, $df = 1$ and $P < 0.0001$ (same information) and $\chi^2 = 8.498$, $df = 1$ and $P = 0.0036$ (preference), indicating an extremely significant likelihood of judging two paragraphs containing the same information as in fact doing so, and a significant likelihood of preferring the more concise of such paragraphs.

More interesting are the results shown in Table 5. Here, taken across all paragraph-pairs, E denotes that the information expressed by a sentence in one paragraph is explicitly expressed in the other paragraph, and J denotes the judgement as to whether each sentence was judged to express information *not* also expressed in the other paragraph. These distributions of observations need to be compared, for explicit and implicit in turn, to the expected distributions of judgements as to whether the information is missing or not. For explicit information, the expected distribution is zero judgements of "missing" – where sentences were explicit in both paragraphs, they were in fact identical in both paragraphs and so should never have been judged missing – and 696 judgements of "not missing". It scarcely needs a statistical test to show that the actual observations of 3, and 693, respectively, do not differ significantly from these expectations. Nonetheless, Fisher's exact test (since one of the expected values is 0, ruling out χ^2) gives P=0.2495. For implicit information, the null hypothesis is that implicit information is indistinguishable from absent information, and so the expected distribution is 290 judgements of "missing" and zero judgements of "not missing", compared to observations of 33, and 257, respectively. Applying Fisher's exact test gives P less than 0.0001, indicating an extremely significant difference. In other words, implicit information is readily distinguishable from absent information, as predicted.

Table 5: Results of paragraph comparison survey (II)

	J	not J
E	3	693
not E	33	257

7 Conclusion and further work

Beginning from some observations about identifier use and semantics in ontologies, we posed two hypotheses, that linguistically-complex identifiers are often defined by "obvious" axiom patterns in terms of the content words contained in those identifiers, and that these "obvious" axiom patterns could be omitted from ontology verbalisations in order to produce clearer texts without significant information loss. By means of an ontology corpus study, and the survey evaluation of generated NL texts with human readers, we have confirmed these hypotheses. As a result, these generation strategies have already been incorporated into the SWAT ontology verbaliser and ontology authoring tool and are already being evaluated in use by ontology developers as those tools progress.

Of course, there are many avenues along which this work could be taken further. We have barely scratched the surface when it comes to using underlying logical formalisms, and the information "hidden" in identifiers to improve generated text. Further investigation of the possibilities of language-specific stemming algorithms in defining-axiom-pattern detection, the interactions between multiple defining axioms for the same entities to form whole definitions, and exploitation of the logical contents of an ontology to determine the salience of "usual" or "unusual" features in order to aid text organisation, all offer rich opportunities to improve natural-language generation from ontologies. We look forward to being able to look further into these areas, and to identifying which of these phenomena can perhaps be generalised to other NLG applications by means of ontologies.

References

Anne Cregan, Rolf Schwitter, and Thomas Meyer. 2007. Sydney owl syntax - towards a controlled natural language syntax for owl 1.1. In *OWLED*.

M. Egaña Aranguren, C. Wroe, C. Goble, and R. Stevens. 2008. In situ migration of handcrafted ontologies to reason-able forms. *Data & Knowledge Engineering*, 66(1):147–162.

Facebook. 2012. Facebook. http://www.facebook.com. Last checked: 10th February 2012.

J. Fernandez-Breis, L. Iannone, I. Palmisano, A. Rector, and R. Stevens. 2010. Enriching the gene ontology via the dissection of labels using the ontology pre-processor language. *Knowledge Engineering and Management by the Masses*, pages 59–73.

Tim Finin, Yun Peng, R. Scott, Cost Joel, Sachs Anupam Joshi, Pavan Reddivari, Rong Pan, Vishal Doshi, and Li Ding. 2004. Swoogle: A search and metadata engine for the semantic web. In *In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, pages 652–659. ACM Press.

Google+. 2012. Google+. http://plus.google.com, February. Last checked: 10th February 2012.

Kaarel Kaljurand and Norbert E. Fuchs. 2007. Verbalizing owl in attempto controlled english. In *Proceedings of Third International Workshop on OWL: Experiences and Directions, Innsbruck, Austria (6th–7th June 2007)*, volume 258.

Survey Monkey. 2012. Survey monkey. http://www.surveymonkey.com. Last checked: 10th February 2012.

ODP. 2012. Ontology design patterns. http://ontologydesignpatterns.org. Last checked: 10th February 2012.

OWL API. 2012. The OWL API. http://owlapi.sourceforge.net. Last checked: 10th February 2012.

Richard Power and Allan Third. 2010. Expressing OWL axioms by English sentences: dubious in theory, feasible in practice. In *23rd International Conference on Computational Linguistics*.

Richard Power. 2010. Complexity assumptions in ontology verbalisation. In *48th Annual Meeting of the Association for Computational Linguistics*.

Paul R Smart. 2008. Controlled natural languages and the semantic web. July.

R. Stevens, J. Malone, S. Williams, R. Power, and A. Third. 2011. Automating generation of textual class definitions from owl to english. *Journal of Biomedical Semantics*, 2(Suppl 2):S5.

The SWAT Project. 2012. Last checked: 10th February 2012.

TONES. 2012. The TONES ontology repository. http://owl.cs.manchester.ac.uk/repository/browser. Last checked: 10th February 2012.

Twitter. 2012. Twitter. http://twitter.com. Last checked: 10th February 2012.

W3C Consortium. 2012. Last checked: 10th February 2012.

Sandra Williams and Richard Power. 2010. Grouping axioms for more coherent ontology descriptions. In *6th International Natural Language Generation Conference*, pages 197–202.

Sandra Williams, Allan Third, and Richard Power. 2011. Levels of organisation in ontology verbalisation. In *Proceedings of the 13th European Workshop on Natural Language Generation (forthcoming)*.

CJ Wroe, R. Stevens, CA Goble, and M. Ashburner. 2003. A methodology to migrate the gene ontology to a description logic environment using. In *Pacific Symposium on Biocomputing*, volume 8, pages 624–635.