Automatic Discovery of Telic and Agentive Roles from Corpus Data

Ichiro Yamada
Science & Technical Research Laboratories
Japan Broadcasting Corporation
1-10-11 Kinuta, Setagaya-ku,
Tokyo 157-8510, Japan
`yamada.i-hy@nhk.or.jp`

Timothy Baldwin
CSSE
University of Melbourne
Victoria 3010, Australia
`tim@csse.unimelb.edu.au`

Abstract

We present two methods for automatically discovering the telic and agentive roles of nouns from corpus data. These relations form part of the qualia structure assumed in generative lexicon theory, where the telic role represents a typical purpose of the entity and the agentive role represents the origin of the entity. The first discovery method uses hand-generated templates for each role type, and the second employs a supervised machine-learning technique. To evaluate the effectiveness of the two methods, we took a sample of 30 nouns, selected 50 verbs for each, and then generated a ranked list of verbs for a given noun. Using a variant of Spearman's rank correlation, we demonstrate the ability of the methods to identify qualia structure.

1 Introduction

We present a study of methods for automatically discovering the telic and agentive roles of nouns based on corpus data. These relations form part of the **qualia structure** assumed in generative lexicon theory (Pustejovsky, 1995). The qualia structure of a given noun incorporates (at most) the following four roles:

- **Formal role:** the conceptual superclass of the noun.
  e.g., orientation, magnitude, shape, dimensionality, color, or position.
- **Constitutive role:** the internal constitution of the entity.
  e.g., material, weight, parts, or component elements.
- **Telic role:** the typical function of the entity.
  i.e. what the entity is for.
- **Agentive role:** the origin of the entity, or its coming into being
  e.g., creator, artifact, natural kind, or casual chain

For example, for the noun *book*, *publication* is a formal role noun, *text* is a constitutive role noun, *read* is a telic role verb, and *write* is an agentive role verb.

Research has been done on extracting the formal and constitutive roles of nouns. Hearst (1992), Widdows and Dorow (2002), and others developed methods of automatically acquiring noun hyponyms—corresponding to the formal role—by identifying a set of frequently used and unambiguous lexico-syntactic patterns. Girju et al. (2003) proposed a method of learning part–whole relations, which correspond to the constitutive role. It is also possible to use lexical resources such as WordNet (Fellbaum, 1998) to determine formal (through hypernym links) and constitutive role data (through meronym links). Telic and agentive roles, on the other hand, have received relatively little attention in terms of automatic acquisition and are not available from any large-scale lexical resources. The only work we are aware of which directly targets the task of learning telic and agentive qualia data is that of Bouillon et al. (2002),

In Proceedings of the The 18th Pacific Asia Conference on Language, Information and Computation (PACLIC 18), Tokyo, Japan.

who use symbolic learning to identify "qualia pairs"—token instances of noun–verb pairs which correspond to a some qualia role—in corpus data. Our work differs in that we can identify the qualia roles of an arbitrary noun, as suitable for the development of a lexical resource, and sub-classify noun–verb pairs according to the specific qualia role they constitute.

An example application of telic and agentive roles is the interpretation of logical metonymy (Lapata and Lascarides, 2003), such as in *Mary finished her beer*. Under the standard interpretation of logical metonymy, *finish* here predicates over an unexpressed verb, which takes *her beer* as object. By accessing the qualia structure of *beer*, it is possible to resolve the unexpressed verb by way of telic and agentive roles, resulting, e.g., in the interpretation *finished drinking her beer* (from the telic role – although in other cases the agentive role may be more appropriate). Busa and Johnston (1996) proposed an interpretation-based method of translating complex nominals from English to Italian, interpreting the relation between the nouns based primarily on telic and agentive role data. Qualia structure is also useful for QA tasks. Choi et al. (2003) proposed a QA system that uses rich lexical semantic knowledge incorporating relations between nouns and verbs of the type manifested in qualia structure.

In the qualia structure of a given noun, telic and agentive roles are described by a set of predicates (potentially specified for argument structure). For example, the prototypical telic role for *book* is normally considered to be *read*, and the prototypical agentive role is *write*. However, alternate predicates such as *study* and *publish* can also be considered to be telic and agentive roles, respectively. In line with this observation, we treat the telic and agentive roles of a given noun as a (partially) ranked list rather than a closed set of predicates. The purpose of this research is thus to generate a ranked list of verbs for a given noun for each of the telic and agentive roles, with the ranking encoding the relative prototypicality of the verb fulfilling the given role of the target noun. Verbs that rank high in this list can then be considered as the telic and agentive roles of the noun in question.

In this paper, we propose two basic methods for extracting the telic and agentive roles of nouns from corpus data. These are in the same vein as Hearst's template-based strategy (Hearst, 1992), whereby we identify highly precise (generally low-recall) syntactic constructions that are indicative of a verb constituting the agentive or telic role of a given noun. An example of such a template is an N' modified by an infinitival relative clause, such as *(a) book to read*, wherein *read* represents the purpose of *book* and is thus a candidate for the telic role. We estimate the occurrence of different verbs with a given noun in these constructional templates by running a dependency parser (RASP Briscoe and Carroll (2002)) over the British National Corpus (BNC: Burnard (2000)). The first method uses hand-generated templates for each role type. The second employs a maximum entropy-based supervised learning technique which dynamically learns constructional and lexical preferences from the dependency data. In evaluation, we took a sample of 30 nouns, independently selected 50 verbs for each, and generated a ranked list of verbs for a given noun. We then evaluated the results using a variant of Spearman's rank correlation.

In the remainder of this paper, we first introduce the resources used in this research (§2). We then present the two methods we propose for extracting qualia structure (§3). Finally, we provide details on the methodologies used to evaluate these methods (§4), before concluding the paper (§5).

2 Resources

The methods we propose make use of a number of resources, namely: corpus data, a parser, test data for evaluating the methods, and gold-standard data that are judged by two annotators to determine the "goodness" of each noun–verb pair for each of the telic and agentive roles.

2.1 Corpus and pre-processing

The corpus data is taken from the written component of the British National Corpus (BNC: Burnard (2000)), composed of around 90 million words. We dependency-parsed the BNC using RASP (Briscoe and Carroll, 2002), based on the existing BNC sentence tokenization. RASP first tags each input string based on the CLAWS-2 part-of-speech (POS) tagset, and then runs a tag sequence grammar over the word-level tags to derive a structural analysis of the sentence. This is, in turn, translated into dependency tuples specifying a head and its dependent(s), marked-up according to 23 grammatical relations (GRs). For example, the `ncmod` relation is used to capture noun–noun dependencies, and the `dobj` relation to capture direct object-type noun–verb dependencies. The following is the output of RASP for inputs *airplane tickets* and *read books*.

`ncmod(_,ticket_NN1,airplane_NN2)`	*airplane tickets*
`dobj(read_VV0,book_NN2,_)`	*read books*

Here, `NN1`, `NN2`, and `VV0` are the CLAWS-2 POS tags for singular common noun, plural common noun, and base form verb, respectively.

2.2 Test data

The primary test data used in this research is comprised of 30 nouns. 10 of these nouns were selected from the literature on qualia structure. The remaining 20 were selected randomly, and include both concrete and abstract nouns. The targeted nouns are given below.

Targeted nouns

> book, car, knife, speech, food, table, door, prisoner, juice, novel, executive, delegation, phone, clinic, cash, beef, review, letter, counter, county, sunshine, accounting, register, complexity, gaze, profession, investigation, imagination, estimate, maturity

We selected 50 verbs for each noun independently of the BNC data sentences. A small number of verbs were hand-chosen as being representative of a telic or agentive relation with the corresponding noun, while the remainder were selected randomly. For example, the verbs selected for the noun *book* are given below.

Targeted verbs for the noun *book*

> abandon, add, appear, believe, borrow, bring, browse, buy, call, compile, dedicate, design, destroy, dispose, end, expect, fill, find, follow, get, hand, hold, introduce, keep, lay, make, move, need, pack, plan, prepare, print, provide, publish, read, remove, return, show, snatch, start, steal, suit, think, throw, thrust, translate, treat, want, withdraw, write

The total number of noun–verb combinations is thus 1,500. Note that, while there is considerable variation in the verbs associated with each noun, the same 50 verbs are used in annotation/evaluation of both the agentive and telic relations.

2.3 Annotation

Two symbolic-system-major undergraduates where asked to judge the "goodness" of each noun–verb pair for the telic and agentive roles based on a 0-10 discrete numeric scale. The value 10 means that the verb is regarded as prototypically filling the given qualia relation for the noun in question, and the value 0 means that there is no way in which the verb could be construed as being related to the noun by the given qualia relation. The mean human ratings are regarded as gold-standard data, based on which we evaluate the performance of our methods.

Figure 1 is a plot of the distribution of gold-standard numeric values returned by the two annotators. The x-axis is the mean value of the annotator ratings and the y-axis is the frequency of occurrence of that rating. This shows that most of the verbs are regarded as being unrelated to the noun under the given qualia relation. The mean and variance are 2.06 and 5.97 for the agentive role, and 2.70 and 4.56 for the telic role, respectively. Because variance for the agentive role is larger than for the telic role, it would appear that the annotators' judgements on goodness are more clear-cut for the agentive role than the telic role.

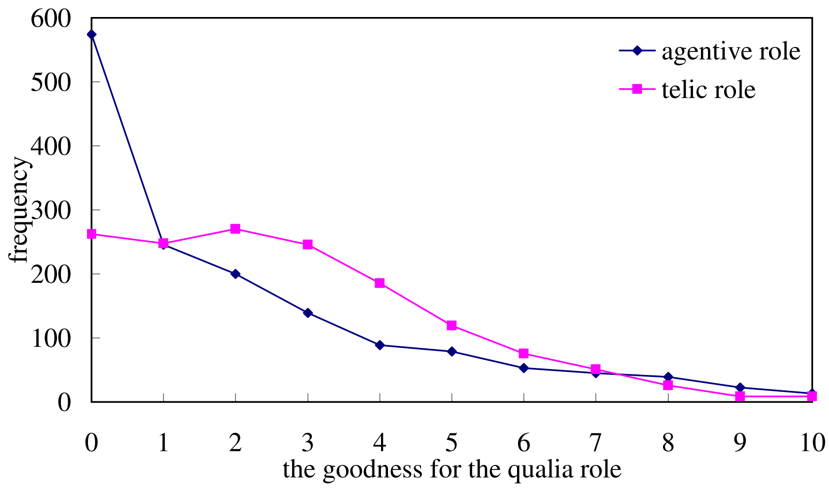

Figure 1: Distribution of total frequency for each numeric scale in gold-standard data

3 Proposed ranking methods for extracting telic and agentive roles

To extract the telic and agentive roles of a given noun, we propose two methods, each of which generates a ranked list of verbs which expresses the relative "goodness" of each noun–verb pair for the qualia role in question. A highly-ranked verb can then be considered as a candidate for inclusion in the qualia structure of that noun.

Below, we outline the two methods, the first of which uses hand-generated templates to identify candidates for each role type, while the second employs a supervised learning technique to dynamically identify templates predictive of the different role types.

3.1 Template-based ranking method

Verbs which form part of the qualia structure of a noun tend to occur frequently in particular constructions (Pustejovsky et al., 1993; Bouillon et al., 2002). For example, verbs which readily allow passivization of a given noun tend to be good candidates for the agentive role of that noun, i.e. frequent occurrence of sentences such as *This book was written (by him)* is evidence for *write* being a candidate for the agentive role of *book*. Similarly, the *a N worth V+ing* construction, e.g. *a book worth reading*, indicates that *V* (e.g. *read*) is a candidate for the telic role of *N* (e.g. *book*). We identified several constructional templates for the telic role and one for the agentive role, and counted the raw frequency of occurrence for each verb with a given noun in the two template sets. The templates used in our method are described in Tables 1 and 2, where N and V refer to the target noun and verb, respectively, V[+ing] refers to the present participle of V, V[+en] refers to the past participle of V, and V[+nom] refers to the nominalization of V.

All templates assume that the noun will occur as the deep object (ARG1) of a transitive verb. We recognize that this is an oversimplification, as evidenced by *knife* and its telic role *cut*, which we would not expect to occur as *cut (the) knife*. The principal motivation for making this simplifying assumption is empirical, in that our primary concern is with high precision, potentially at the cost of low recall.

In order to normalize for the effects of the independent noun and verb word probabilities in calculating the frequencies of occurrence of a given noun–verb pair over the set of templates, we use point-wise mutual information (Church and Hanks, 1989). If $P(x)$ is the probability of occurrence of word x, the mutual information between noun n and verb v is defined as follows.

$$MI(n,v) = log\frac{P(n,v)}{P(n)P(v)}$$

We combine this with the corpus frequency for a given noun–verb pair within each template to derive a single score:

Template	Example
N (be $\|\phi$) (worth |deserving |meriting) (V[+ing] |V[+nom])	*(a) book worth reading*
N BE worthy of V[+nom]	*(the) book is worthy of reading*
N (deserves |merits) V[+nom]	*(the) book merits reading*
Adverb-V[+en] N	*(a) well-read book*
Adverb V[+en] N	*(a) well read book*
N BE Adverb-V[ed]	*(the) book is well-read*
V[+ing] Noun	*(I enjoy) reading books*
N to V	*(a) book to read*

Table 1: Templates for the telic role

Template	Example
N BE V[+en]	*(the) book was written (by Kim)*

Table 2: Template for the agentive role

$$\begin{aligned} score\_template_{agentive}(n,v) &= MI(n,v) \times \frac{TemplateF_{agentive}(n,v)}{DF(n,v)} \\ score\_template_{telic}(n,v) &= MI(n,v) \times \frac{TemplateF_{telic}(n,v)}{DF(n,v)} \end{aligned}$$

where $TemplateF(n, v)$ is the frequency of occurrence of noun n and verb v within each template, and $DF(n, v)$ is the number of sentences in which n and v co-occur. We use this score to rank all verbs for the given noun and qualia role.

3.2 Maximum entropy learning-based ranking method

The second method employs the same basic intuition as the template-based method, but rather than relying on a fixed set of templates, dynamically learns the constructional and lexical preferences of telic and agentive noun–verb pairs based on token-level occurrences. To achieve this, we use a maximum entropy-based supervised classifier.

As training data, we treat all noun–verb pairs with an average human rating between 7 and 10 as positive instances, and all noun–verb pairs with an average rating of 0 as negative instances. Table 3 lists examples of positive and negative instances for each role of the noun *book*.

We next extracted all sentences incorporating the positive and negative noun–verb training pairs, resulting in 7,810 positive and 13,780 negative sentence exemplars for the agentive role, and 9925 positive and 5148 negative sentence exemplars for the telic role. For each sentence exemplar, we extracted all dependency noun–verb tuples involving a training noun–verb pair or test noun, as well as any noun or verb modification data, based on the output of RASP. We also extracted the local POS context of each target

Role	Positive instances	Negative instances
Agentive	print, publish, write, make, compile, design, start	abandon, appear, destroy, dispose, follow, hand, hold, keep, lay, pack, remove, return, snatch, suit, throw, thrust, withdraw
Telic	read, browse	call, end, appear, suit

Table 3: Positive and negative instances for the agentive and telic roles of *book*

noun, based on the first two characters of the CLAWS-2 POS tag (reducing the tagset from 170 to 49 tags in the process). From this, we generated a feature vector of the following form for each noun–verb pair in the sentence token:

- The grammatical relation of the noun–verb dependency tuple
- The grammatical relation of any other dependency tuples the noun occurs in, and the POS tag of other words in the dependency tuple
- The grammatical relation of any other dependency tuples the verb occurs in, and the POS tag of other words in the dependency tuple

For example, the RASP output for the question *Can I have a book to read?*, which incorporates the noun *book* and verb *read*, is as follows (with the target words in **boldface** for expository purposes):

```
(|Can:1_VM| |I:2_PPIS1| |have:3_VH0| |a:4_AT1| |book:5_NN1|
                                      |to:6_TO|  |read:7_VV0| |?:8_?|)
(|ncsubj| |have:3_VH0| |I:2_PPIS1| _)
(|dobj| |have:3_VH0| |book:5_NN1| _)
(|ncsubj| |read:7_VV0| |I:2_PPIS1| _)
(|xcomp| |to:6_TO| |book:5_NN1| |read:7_VV0|)
(|detmod| _ |book:5_NN1| |a:4_AT1|)
(|aux| _ |have:3_VH0| |Can:1_VM|)
```

Elements of the feature vector derived from this sentence are:

- The grammatical relation between *book* and *read*:
  `xcomp.to`
- The grammatical relation of other dependency tuples *book* occurs in, and the POS tag of other words in the tuple:
  `detmod, AT` (from *a* and *book*)
  `dobj, VH` (from *have* and *book*)
- The grammatical relation of other dependency tuples *read* occurs in, and the POS tag of other words in the tuple:
  `ncsubj, PP` (from *I* and *read*)

Here, `xcomp` denotes a grammatical relation between a subjectless predicate and clausal complement, `detmod` denotes a grammatical relation between a noun and determiner, `dobj` denotes a grammatical relation between a predicate and its direct object, and `ncsubj` denotes the grammatical relation between a predicate and its subject; the POS tag `AT` denotes an article, `VH` denotes the verb *have*, and `PP` denotes a pronoun. Combined, this feature vector is a positive training exemplar used in estimating the relative goodness of other noun–verb pairs with respect to the telic role, as *read* is a positive telic role instance for *book*.

We used training exemplars such as this to learn a token-level maximum entropy classifier (Berger et al., 1996) for each of the 30 nouns. This performed in the manner of cross validation, whereby, for each noun, we take exemplars from the remaining 29 nouns as training data. Additionally, for the test nouns, we use only sentences which do not include a training exemplar.

To get a sense for the effectiveness of this classifier architecture at identifying agentive and telic role data, the values for $P(rel = qualia\ role|f_{n,v})$ in the two sample sentences *I always had <u>books</u> to <u>read</u>* and *Complete <u>books</u> have been <u>written</u> on this subject* are:

Sample sentence $f1_{book,read}$: *I always had books to read.*
$P(rel = agentive\,role|f1_{book,read}) = 0.396$
$P(rel = telic\,role|f1_{book,read}) = 0.943$
Sample sentence $f2_{book,read}$: *Complete books have been written on this subject .*
$P(rel = agentive\,role|f2_{book,write}) = 0.652$
$P(rel = telic\,role|f2_{book,write}) = 0.455$

If $P(rel = qualia\,role|f_{n,v})$ is greater than 0.5 for the given context, the noun–verb pair is deemed to constitute the given qualia role. In the examples above, the classifier predicts that *read* is a telic role for *book,* and *write* is an agentive role for *book.*

We calculate the probability of a verb being the telic or agentive role of a given noun by aggregating across all positively-classified sentence-level instances of the noun–verb pair, according to:

$$\begin{aligned} score\_ME_{agentive}(n, v) &= MI(n, v) \times \frac{\sum_{n,v}(2 \times P'(rel = agentive|f_{n,v}) - 1)}{DF(n, v)} \\ score\_ME_{telic}(n, v) &= MI(n, v) \times \frac{\sum_{n,v}(2 \times P'(rel = telic|f_{n,v}) - 1)}{DF(n, v)} \end{aligned}$$

where $P'(rel = qualia\,role|f_{n,v})$ is the positive probability of $P(rel = qualia\,role|f_{n,v})$ and $DF(n, v)$ is the number of sentences in which noun n and verb v appear. Essentially, we are calculating the aggregate "goodness" of putatively positive test exemplars, normalized according to point-wise mutual information. The larger this value is, the better the "goodness" and the higher the verb is ranked for the given role type and target noun.

The token-level classification accuracy of our classifiers, aggregated across all nouns, was a modest 63.9% for the agentive role and 62.9% for the telic role. That is, the probability of our classifiers at correctly tagging a noun–verb token instance according to a given qualia role is only slightly better than random. The primary reason for the low accuracy is that many sentence tokens incorporating a given positive noun–verb instance are not indicative of the true qualia status of that pair, such that the level of false negatives tends to be high. In calculating the mean probability of positively-classified exemplars, we are able to identify the relative confidence level of the classifier in its positive judgments, and produce a type-level classification which is truly representative of the qualia status of a given noun–verb pair.

4 Evaluation

In this section, we evaluate the results of the ranking obtained from the methods discussed in the preceding section. We applied both methods over BNC data for the 30 selected nouns to rank the 50 verbs according to the telic and agentive roles. Tables 4 and 5 list the top eight verbs for the agentive and telic roles of *book*, according to the two proposed methods and the gold-standard data. In Table 4, the verbs *write, publish, compile* and *print*—all of which are positive instances in the gold-standard data—are ranked high by both methods. In Table 5, the verb *read*—which is ranked top in the gold-standard data—ranks highest, but there is little agreement between the three datasets beyond this. Intuitively, therefore, the two methods would appear to have been more successful at producing agentive role data than telic role data. Below, we detail a method for quantitatively evaluating the relative agreement between different verb rankings.

4.1 Variant of Spearman's rank correlation

To evaluate the verb rankings, we use a variant of Spearman's rank correlation to calculate the ranking over the top-N items in the two ranked lists. This is applied to the mean human ratings and the output

Template		MaxEnt		Gold-standard	
publish	0.157	dedicate	1.084	write	10.0
write	0.102	publish	0.898	publish	8.0
read	0.019	compile	0.651	compile	8.0
call	0.015	dispose	0.605	print	7.5
dedicate	0.011	write	0.438	make	7.5
print	0.008	browse	0.408	start	7.0
keep	0.007	borrow	0.399	design	7.0
compile	0.006	print	0.386	translate	6.0

Table 4: Top-8 verbs for the agentive role of *book*

Template		MaxEnt		Gold-standard	
read	0.316	read	2.814	read	10.0
write	0.112	write	2.221	browse	9.0
publish	0.079	compile	2.115	think	6.5
buy	0.036	dedicate	1.982	buy	6.0
keep	0.016	buy	1.775	provide	6.0
appear	0.015	borrow	1.695	borrow	5.5
make	0.014	throw	1.682	return	5.5
provide	0.014	publish	1.656	start	5.5

Table 5: Top-8 verbs for the agentive role of *book*

of the two methods for all noun and role types. The reason for us not wanting to use Spearman's rank correlation in its original form is that most verbs cannot be construed as fulfilling the telic or agentive roles of a given noun. Thus, if we calculate Spearman's rank correlation over the entire ranking, the high concentration of low-relevance items at the tail of the ranking will have a greater effect on the final correlation value than the items at the top of the ranking, which are the focus of interest in terms of generating a qualia structure. As a case in point, the averaged Spearman's rank correlation between the data generated by our two human annotators was a remarkably low 0.448 for the telic and 0.369 for the agentive role. Having said this, we at present have no empirical or theoretical criterion for determining the appropriate N for a given noun and qualia role, i.e., we have no way of determining whether a given noun has 1 or 3 telic roles. We therefore calculate rank correlation over a range of values of N, and leave the question of what portion of the ranked data represents true qualia data for manual post-editing.

Our reformulation of Spearman's rank correlation, Rs, analyses the ratio between the squared difference of the top N ranked items and their expected values:

$$Rs = 1 - \frac{\sum_{x=1}^{m} d_x^2}{E(\sum_{x=1}^{m} d_x^2)} = 1 - 6 \times \frac{\sum_{x=1}^{m} d_x^2}{m(2m^2 - 3nm + 2n^2 - 1)}$$

where n is the number of data items, m is the number of items at the top of the ranking, d_x is the difference between the ranks of datasets 1 and 2, and $E(x)$ is the expected value of x. If the two datasets share the same N items, the value of Rs is 1, and if they have no correlation, the value of Rs is 0. However, if the two datasets have a completely negative correlation, the value of Rs is less than -1. This is a problem when using this reformulation to evaluate negative correlations. Here, as we are only interested in evaluating positive correlations, it provides a sound evaluation metric and is faithful to the empirical nature of the conventional Spearman's rank correlation for our purposes.

4.2 Results

We used our variant of Spearman's rank correlation to evaluate the ranking of 50 verbs for the 30 nouns over the agentive and telic roles, the results of which are presented in Figures 2 and 3. In each case, the y-axis represents the value of Rs and the x-axis represents the number of top-ranking items evaluated, e.g., a value of 5 means we are evaluating the top-5 items in the ranking. In these figures, the values for the "gold-standard data" are based on the individual rankings returned by our two annotators. We can consider the values for the gold-standard correlation as an upper bound estimate for the task. The gold-standard correlation for the telic role is lower than that for the agentive role, suggesting that there is greater variation in the interpretation of the concept of "purpose" than process of creation for any given noun.

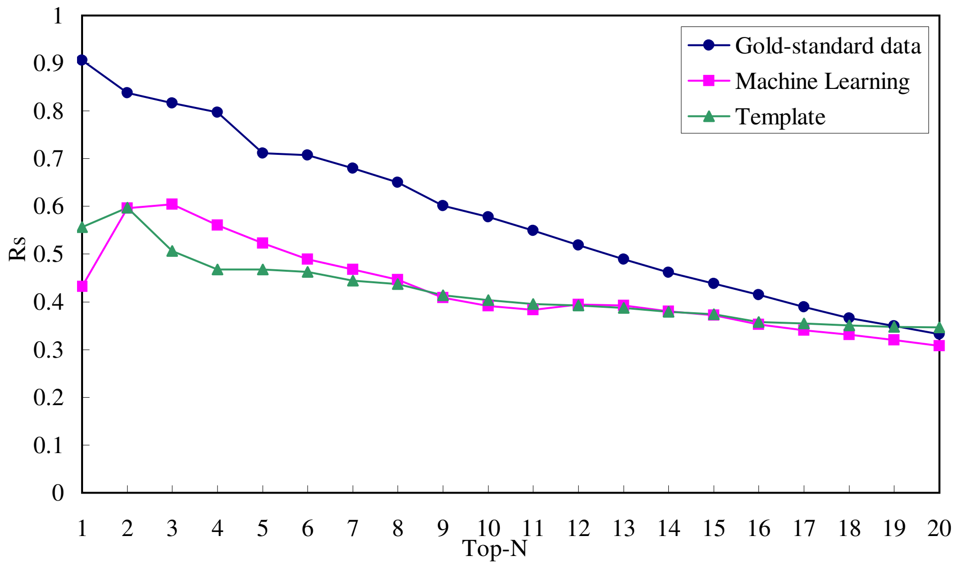

Figure 2: Rank correlation for the agentive role

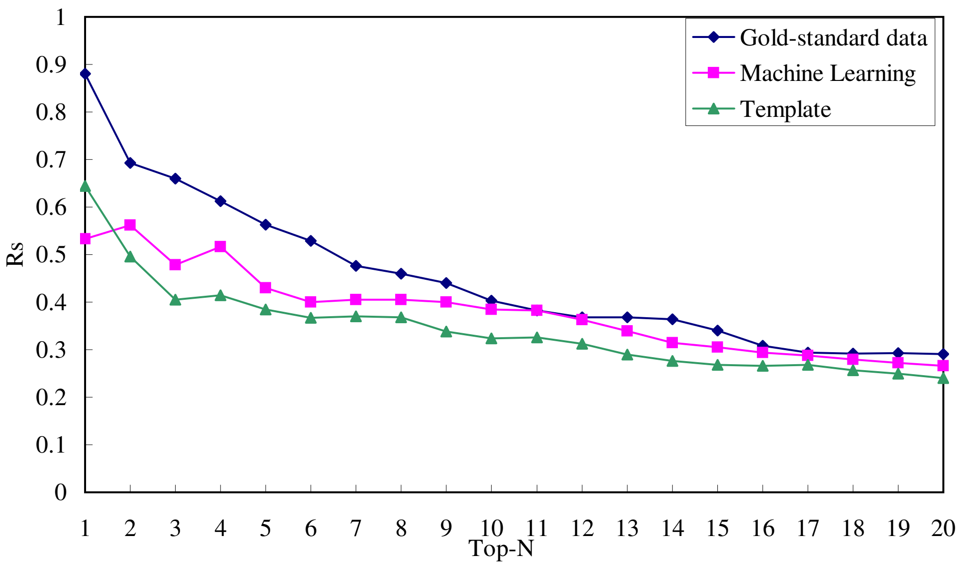

Figure 3: Rank correlation for the telic role

Because we expect to include only the high-ranking verbs in the qualia structure to be used in actual system applications, our main interest in evaluation is how well the two methods perform for smaller values of N. The experimental results for the agentive role indicate that the correlation for the top-3 items was 0.605 with the maximum entropy model, 0.507 with the template method, and 0.817 for the gold standard data. For the telic role, the correlation was 0.479 with the maximum entropy model, 0.405 with the template method, and 0.659 for the gold standard data. There was little difference in performance between the maximum entropy method and the template method, but the maximum entropy method tended to outperform the template method for smaller values of N, except in top-1 evaluation where the template method came out on top for both the agentive and telic roles. It would appear that there is grounds for hybridization, in analyzing occurrences of fixed templates but also dynamically learning the more subtle preferences of each qualia role type. We leave this as an item for future research.

Clearly, there is still room for improvement. To this end, we are planning to use more sophisticated features and filters on the training data. In this experiment, we used all sentences that incorporated positive instance noun–verb pairs to learn constructional and lexical preferences. However, not all of these had constructional features indicative of the qualia role. If we had filtered out sentences not corresponding to constructions of the type used in the template-based method, we would undoubtedly have obtained better results. Also, making more aggressive use of lexical features may aid extraction, particularly for the agentive role.

5 Conclusion

We proposed two methods for discovering the telic and agentive roles of nouns from corpus data. The first method used hand-generated templates, and the second employed a supervised learning technique based on sentence structure. In evaluation using a variant of Spearman's rank correlation, we found that these methods were moderately successful at extracting key information, but also that there was room for improvement. We are planning to use more sophisticated features and learning methods to extract qualia role data.

In future research, we intend to generate a lexical resource that incorporates qualia structure, and use this in a range of applications.

Acknowledgements

This research was carried out in part while the first author was a visitor to the CSLI LinGO Laboratory at the Center for the Study of Language and Information, Stanford University, and while the second author was employed in the CSLI LinGO Laboratory. The authors would like to thank Dan Flickinger, Pat Langley, Kentaro Torisawa, Hideki Sumiyoshi, Nobuyuki Yagi, Masahiro Shibata and the anonymous reviewers for their valuable comments.

References

Berger, A.L., P. F. Brown, and V. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

Bouillon, Pierrette, Vincent Claveau, Cécile Fabre, and Pascale Sébillot. 2002. Acquisition of qualia elements from corpora – evaluation of a symbolic learning method. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Canary Islands.

Briscoe, Ted and John Carroll. 2002. Robust accurate statistical annotation of general text. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1499–1504, Las Palmas, Canary Islands.

Burnard, Lou. 2000. *User Reference Guide for the British National Corpus*. Technical report, Oxford University Computing Services.

Busa, Frederica and Michael Johnston. 1996. Cross-linguistic semantics for complex nominals in the generative lexicon. In *AISB Workshop on Multilinguality in the Lexicon*, Sussex, UK.

Choi, Key-Sun, Jae-Ho Kim, Masaru Miyazaki, Jun Goto, and Yeun-Bae Kim. 2003. Question-answering based on virtually integrated lexical knowledge base. In *Proc. of the The Sixth International Workshop on Information Retrieval with Asian Languages (IRAL2003)*, Sapporo, Japan.

Church, Kenneth W. and Patrick Hanks. 1989. Word association norms, mutual information and lexicography. In *Proc. of the 27th Annual Meeting of the ACL*, pages 76–83, Vancouver, Canada.

Fellbaum, Christiane, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.

Girju, Roxanne, Adriana Badulescu, and Dan Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proc. of the 3rd International Conference on Human Language Technology Research and 4th Annual Meeting of the NAACL (HLT-NAACL 2003)*, Edmonton, Canada.

Hearst, Marti. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of the 14th International Conference on Computational Linguistics (COLING '92)*, Nantes, France.

Lapata, Mirella and Alex Lascarides. 2003. Detecting novel compounds: The role of distributional evidence. In *Proc. of the 10th Conference of the EACL (EACL 2003)*, Budapest, Hungary.

Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press, Cambridge, USA.

Pustejovsky, James, Peter Anick, and Sabine Bergler. 1993. Lexical semantic techniques for corpus analysis. *Computational Linguistics*, 19(2):331–58.

Widdows, Dominic and Beate Dorow. 2002. A graph model for unsupervised lexical acquisition. In *Proc. of the 19th International Conference on Computational Linguistics (COLING 2002)*, pages 1093–9, Taipei, Taiwan.