

Automatic Acquisition of Qualia Structure from Corpus Data

Ichiro YAMADA^{†a)}, Timothy BALDWIN^{††}, Nonmembers, Hideki SUMIYOSHI[†], Masahiro SHIBATA[†],
and Nobuyuki YAGI[†], Members

SUMMARY This paper presents a method to automatically acquire a given noun's telic and agentive roles from corpus data. These relations form part of the qualia structure assumed in the generative lexicon, where the telic role represents a typical purpose of the entity and the agentive role represents the origin of the entity. Our proposed method employs a supervised machine-learning technique which makes use of template-based contextual features derived from token instances of each noun. The output of our method is a ranked list of verbs for each noun, across the different qualia roles. We also propose a variant of Spearman's rank correlation to evaluate the correlation of two top-N ranked lists. Using this correlation method, we represent the ability of the proposed method to identify qualia structure relative to a conventional template-based method.

key words: *qualia structure, maximum entropy learning, ranking of word relevance, spearman's rank correlation, lexical knowledge acquisition*

1. Introduction

Recent developments in digital technology allows us to reliably use large-scale text data for a bewildering range of tasks. These tasks include knowledge extraction tasks encompassing domain, social conditions and common knowledge. One way to encode knowledge is by using a lexical knowledge base composed of lexical relations. Lexical knowledge bases have been shown to be useful for tasks including information extraction [1], question answering [2] and machine translation [3]. Manually maintaining and updating lexical knowledge bases is difficult because of the continual emergence/specialization of words and need for an ever-growing range of lexical relations. In this paper, we present a method for automatically discovering deep semantic lexical knowledge from corpus data.

Our approach incorporates the telic and agentive roles of nouns. These relations form part of the qualia structure assumed in generative lexicon theory [4]. 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.

There is a considerable body of research on extracting the formal and constitutive roles of nouns. Notably, Hearst [5], Widdows and Dorow [6], and Snow et al. [7] developed methods for automatically acquiring noun hyponyms—corresponding to the formal role—by (explicitly or implicitly) identifying a set of frequently-used unambiguous lexico-syntactic patterns. Girju et al. [8] proposed a method for learning part-whole relations, which correspond to the constitutive role. Formal and constitutive role data also exists in large-scale lexical resources such as WordNet [9], in the form of hypernym and meronym links, respectively. Telic and agentive roles, on the other hand, have received relatively little attention in terms of automatic acquisition and are not available in any large-scale lexical resources. The only research we are aware of which directly targets the task of learning telic and agentive qualia data is that of Bouillon et al. [10] and Cimiano and Wenderoth [11]. Bouillon et al. use inductive logic programming to identify “qualia pairs”—token instances of noun-verb pairs which correspond to a qualia role—from 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. Cimiano and Wenderoth use POS-tagged web data and Google counts to determine the plausibility of qualia parts. This presents an interesting alternate view on this task in giving priority to the quantity rather than quality of data, whereas in our case we prefer quality of analysis (we require that all corpus data has been parsed) over raw quantity of data (although we are still able to process 90 m words from the BNC).

An example application of telic and agentive roles is the interpretation of logical metonymy [12], 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

Manuscript received January 16, 2007.

[†]The authors are with Science & Technical Research Laboratories, NHK, Tokyo, 157-8510 Japan.

^{††}The author is with the Department of Computer Science and Software Engineering, University of Melbourne, Victoria 3010, Australia.

a) E-mail: yamada.i-hy@nhk.or.jp

DOI: 10.1093/ietisy/e90-d.10.1534

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 [13] 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.

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 a method for automatically extracting the telic and agentive roles of nouns from corpus data. Our method is based on the identification of 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-bar 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. Using these features, our proposed method employs a maximum entropy-based supervised learning technique which dynamically learns constructional and lexical preferences from the dependency data. We evaluate over the British National Corpus (BNC: [14]), as parsed with a broad-coverage parser (RASP: [15]). In our evaluation, we took a sample of 30 nouns, independently selected 50 verbs for each noun, and generated a ranked list of verbs for a given noun. We then evaluated the results using a variant of Spearman’s rank correlation which can evaluate the correlation of two ranked lists over a range of top-N items.

In the remainder of this paper, we first introduce the resources used in this research (§2). We then present our proposed method for extracting qualia structure (§3). Finally, we provide an evaluation method which is based on Spearman’s rank correlation and evaluate our method (§4), before concluding the paper (§5).

2. Resources

The method we propose makes use of a number of resources, namely: corpus data, a parser, test data for evaluating the methods, and gold-standard data that is judged by two annotators to determine the “adequacy” of each noun–verb pair

for each of the telic and agentive roles.

2.1 Corpus and Pre-Processing

The corpus is taken from the written component of the British National Corpus (BNC), composed of around 90 million words. We dependency-parsed the BNC using RASP [15], 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.

We identify instances of each target noun in the parsed BNC data, and from this generate a ranked list of verbs for each of the telic and agentive roles, as outlined in the following section.

2.2 Targeted Data

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

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. 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:

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

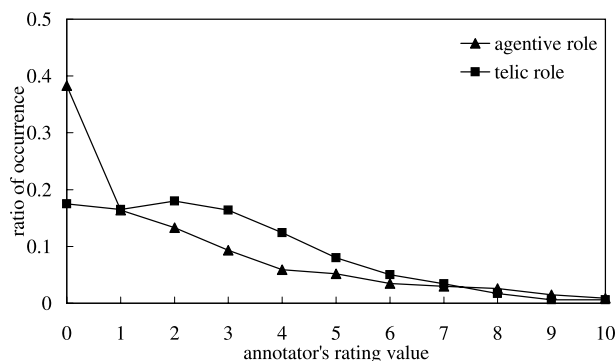


Fig. 1 Distribution of the relative occurrence of each discrete rating in the gold-standard data.

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 Gold-Standard Data

Two native English speakers were asked to judge the “adequacy” of each noun–verb pair for the telic and agentive roles, based on a 0-10 discrete numeric scale. A value of 10 means that the verb is regarded as prototypically filling the given qualia relation for the noun in question, while a value of 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 for a given (noun, verb, relation) triple are regarded as gold-standard data, based on which we evaluate the performance of our methods.

Figure 1 is a graph of the distribution of gold-standard numeric values assigned by the two annotators. The x -axis is the value of the annotator ratings and the y -axis is the ratio 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. The variance for the agentive role is larger than for the telic role, suggesting that the annotators’ judgments are more clear-cut for the agentive role than the telic role.

3. Proposed Ranking Methods for Extracting Telic and Agentive Roles

To extract the telic and agentive roles of a given noun, we propose a statistical method which generates a ranked list of verbs which express the relative “adequacy” of each noun–verb pair relative to a particular qualia role. A highly-ranked verb can be considered as a candidate for inclusion in the qualia structure of that noun.

Below, we propose two methods for extracting telic and agentive roles. The first method employs a supervised learning technique to dynamically identify templates predictive

Table 1 Positive and negative instances for the agentive and telic roles of *book*.

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, ithdraw
Telic	read, browse	call, end, appear, suit

of the different role types. The second method is a traditional template-based method which was used to benchmark our results.

3.1 Ranking Method Using Maximum Entropy Learning

Verbs which form part of the qualia structure of a noun tend to occur frequently in particular constructions [10], [16]. Our proposed method dynamically learns which of a fixed set of constructions are indicative of the telic and agentive roles for a given noun–verb pairing, 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 1 lists examples of positive and negative instances for each role of the noun *book*.

We next extracted all sentences from the parsed BNC data which incorporated either positive or negative noun–verb training pairs, resulting in 7810 positive and 13780 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 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 adequacy 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 [17] for each of the 30 nouns. To get a sense for the effectiveness of this classifier architecture at identifying agentive and telic role data, the values for $P(\text{rel} = \text{qualia role}|f_{n,v})$ in the two sample sentences *I always had books to read* and *Complete books have been written on this subject* are:

Sample sentence $f_{1, \text{book}, \text{read}}$: *I always had books to read.*

$$P(\text{rel} = \text{agentive role}|f_{1, \text{book}, \text{read}}) = 0.396$$

$$P(\text{rel} = \text{telic role}|f_{1, \text{book}, \text{read}}) = 0.943$$

Sample sentence $f_{2, \text{book}, \text{read}}$: *Complete books have been written on this subject.*

$$P(\text{rel} = \text{agentive role}|f_{2, \text{book}, \text{write}}) = 0.652$$

$$P(\text{rel} = \text{telic role}|f_{2, \text{book}, \text{write}}) = 0.455$$

$P(\text{rel} = \text{qualia role}|f_{n,v})$ is a normalized probability, in the range 0 to 1. As such, if the value of $P(\text{rel} = \text{qualia role}|f_{n,v})$ is greater than 0.5 for the given context, the noun–verb pair is considered 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*.

In order to normalize for the effects of the independent noun and verb word probabilities in calculating the frequen-

cies of occurrence of a given noun–verb pair, we use point-wise mutual information [18]. 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 calculate the probability of a verb being the telic or agentive role of a given noun by combining this mutual information with the probabilities calculated by the maximum entropy classifier, according to:

$$\text{score_ME}_{ag}(n, v) = MI(n, v)$$

$$\times \sum_{n,v} (2 \times P'(\text{rel} = \text{agentive}|f_{n,v}) - 1) / DF(n, v)$$

$$\text{score_ME}_{te}(n, v) = MI(n, v)$$

$$\times \sum_{n,v} (2 \times P'(\text{rel} = \text{telic}|f_{n,v}) - 1) / DF(n, v)$$

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

3.2 Ranking Method Using Hand-Generated Templates

Verbs which readily allow passivization of a given noun tend to be good candidates for the agentive role of that noun, i.e. frequent occurrences of sentences such as *This book was written (by him)* are 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 empirically 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 hand-generated templates are described in Tables 2 and 3, 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*.

Using the frequency of token occurrences for a given noun–verb pair within each template, we derive a single score:

$$\text{score_template}_{agentive}(n, v) = \frac{\text{Temp}F_{agentive}(n, v)}{DF(n, v)}$$

$$\text{score_template}_{telic}(n, v) = \frac{\text{Temp}F_{telic}(n, v)}{DF(n, v)}$$

Table 2 Templates for the telic role.

Template	Example
N (be ϕ) (worth deserving meriting)(V[+ing] V[+nom])	(a) <i>book worth reading</i>
N BE worthy of V[+nom]	(the) <i>book is worthy of reading</i>
N (deserves merits) V[+nom]	(the) <i>book merits reading</i>
Adverb-V[+en] N	(a) <i>well-read book</i>
Adverb V[+en] N	(a) <i>well read book</i>
N BE Adverb-V[ed]	(the) <i>book is well-read</i>
V[+ing] Noun	(I enjoy) <i>reading books</i>
N to V	(a) <i>book to read</i>

Table 3 Template for the agentive role.

Template	Example
N BE V[+en]	(the) <i>book was written (by Kim)</i>

where $TempF_{agentive}(n, v)$ and $TempF_{telic}(n, v)$ are the frequencies 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.

4. Experiment and Evaluation

To evaluate the ability of the proposed maximum entropy-based method, we carried out an experiment whereby we calculated scores for all verbs for a given noun and qualia role, and ranked the verbs accordingly. This list of verbs was then compared with the result of the traditional template-based method using a variant of Spearman's rank correlation which can evaluate a top-N item list.

4.1 Experimental Results

We applied both methods over BNC data for the 30 selected nouns to rank the 50 verbs according to each of the telic and agentive roles. The maximum entropy-based method was evaluated via 30-fold cross validation, whereby, for each noun, we took 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.

Tables 4 and 5 list the top eight verbs for the agentive and telic roles of *book*, according to the two 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 highly by both methods. In Table 5, the verb *read*—which is ranked top in the gold-standard data—ranks highest. Below, we detail a method for quantitatively evaluating the relative agreement between the different verb rankings.

4.2 Variant of Spearman's Rank Correlation

To evaluate the results of the verb rankings, we propose 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 of the two

methods for all nouns 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 the same effect on the final correlation value as 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. We therefore calculate the rank correlation over a range of values of N.

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:

$$Rs = 1 - \frac{\sum_{x=1}^m d_x^2}{E\left(\sum_{x=1}^m d_x^2\right)}$$

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

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, 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.

4.3 Evaluation of the Results of Ranking

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 Figs. 2 and 3. In each case, the y-axis represents the average value of Rs of all nouns tested 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. That the gold-standard correlation for higher ranks is high indicates that the gold-standard data has high agreement at higher ranks. The gold-standard correlation for the telic role, on the other hand, is lower than that for the agentive role, suggesting that there is greater variation in the interpretation of the concept of "purpose" than the process of creation for any given noun.

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

Rank	ME-based method (Value of score_ME)	Template-based method (Value of score_template)	Gold-standard data (Mean value of rating)
1	dedicate (1.084)	publish (0.157)	write (10.0)
2	publish (0.898)	write (0.102)	publish (8.0)
3	compile (0.651)	read (0.019)	compile (8.0)
4	dispose (0.605)	call (0.015)	print (7.5)
5	write (0.438)	dedicate (0.011)	make (7.5)
6	browse (0.408)	print (0.008)	start (7.0)
7	borrow (0.399)	keep (0.007)	design (7.0)
8	print (0.386)	compile (0.006)	translate (6.0)

Table 5 Top-8 verbs for the telic role of *book*.

Rank	ME-based method (Value of score_ME)	Template-based method (Value of score_template)	Gold-standard data (Mean value of rating)
1	read (2.814)	read (0.316)	read (10.0)
2	write (2.221)	write (0.112)	browse (9.0)
3	compile (2.115)	publish (0.079)	think (6.5)
4	dedicate (1.982)	buy (0.036)	buy (6.0)
5	buy (1.775)	keep (0.016)	provide (6.0)
6	borrow (1.695)	appear (0.015)	borrow (5.5)
7	throw (1.682)	make (0.014)	return (5.5)
8	publish (1.656)	provide (0.014)	start (5.5)

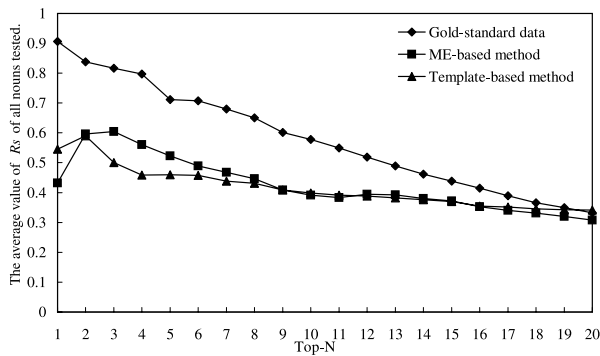


Fig. 2 Rank correlation for the agentive role.

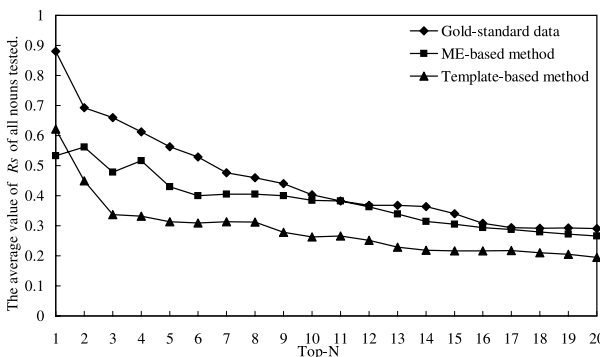


Fig. 3 Rank correlation for the telic role.

Because we expect to include only the high-ranking verbs in the qualia structure in actual system applications, our main interest in evaluation is how well the two methods perform for smaller values of N. The average number of positive instances was 3.15 in the learning process, so we compared the correlation for the top-3 items. The

experimental results for the agentive role indicate that the correlation for the top-3 items was 0.605 for the maximum entropy-based method, 0.500 for the template-based method, and 0.816 for the gold standard data. For the telic role, the correlation was 0.479 for the maximum entropy-based method, 0.337 for the template-based method, and 0.659 for the gold standard data. This result shows that the maximum entropy-based method tends to outperform the template-based method for smaller values of N, except in top-1 evaluation where the template method comes out on top for both the agentive and telic roles. Based on this, 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.

5. Conclusion

We have proposed a supervised learning technique based on sentence structure for acquiring the noun-verb relations for telic and agentive roles from corpus data. In evaluation, we developed a modified version of Spearman’s rank correlation method to evaluate the correlation of the top-N items of two ranked lists. The results of this evaluation showed that our proposed method is relatively successful at identifying qualia structure compared to a traditional template-based method.

In future research, we intend to generate a large-scale lexical knowledge base 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

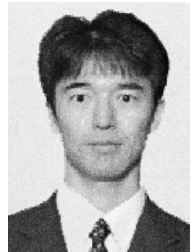
for the Study of Language and Information, Stanford University. The second author's involvement in this research was supported by the Australian Research Council through Discovery Project grant number DP0663879. The authors would like to thank Dan Flickinger, Pat Langley, Kentaro Torisawa and the anonymous reviewers for their valuable comments.

References

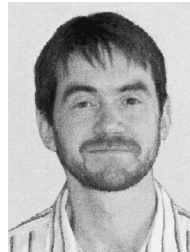
- [1] M. Stevenson and M.A. Greenwood, "Learning information extraction patterns using WordNet," Proc. 3rd International Conference of the Global WordNet Association (GWA'06), pp.95-102, Jeju Island, Republic of Korea, 2006.
- [2] K.-S. Choi, J.-H. Kim, M. Miyazaki, J. Goto, and Y.-B. Kim, "Question-answering based on virtually integrated lexical knowledge base," Proc. The Sixth International Workshop on Information Retrieval with Asian Languages (IRAL2003), pp.168-175, 2003.
- [3] M. Carpuat and D. Wu, "Word sense disambiguation vs. statistical machine translation," Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005), pp.387-394, Ann Arbor, USA, 2005.
- [4] J. Pustejovsky, *The Generative Lexicon*, MIT Press, Cambridge, USA, 1995.
- [5] M. Hearst, "Automatic acquisition of hyponyms from large text corpora," Proc. 14th International Conference on Computational Linguistics (COLING '92), pp.539-545, Nantes, France, 1992.
- [6] D. Widdows and B. Dorow, "A graph model for unsupervised lexical acquisition," Proc. 19th International Conference on Computational Linguistics (COLING 2002), pp.1093-1099, Taipei, Taiwan, 2002.
- [7] R. Snow, D. Jurafsky, and A.Y. Ng, "Semantic taxonomy induction from heterogenous evidence," Proc. COLING/ACL 2006, pp.801-808, Sydney, Australia, 2006.
- [8] R. Girju, A. Badulescu, and D. Moldovan, "Learning semantic constraints for the automatic discovery of part-whole relations," Proc. Human Language Technology Conference - Annual Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL), pp.1-8, Edmonton, Canada, 2003.
- [9] C. Fellbaum, *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, USA, 1998.
- [10] P. Bouillon, V. Claveau, C. Fabre, and P. Sébillot, "Acquisition of qualia elements from corpora - evaluation of a symbolic learning method," Proc. 3rd International Conference on Language Resources and Evaluation (LREC 2002), pp.208-215, Las Palmas, Spain, 2002.
- [11] P. Cimiano and J. Wenderoth, "Automatically learning qualia structures from the Web," Proc. ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition, pp.28-37, Ann Arbor, USA, 2005.
- [12] M. Lapata and A. Lascarides, "Detecting novel compounds: The role of distributional evidence," Proc. 10th Conference of the EACL (EACL 2003), pp.235-242, Budapest, Hungary, 2003.
- [13] F. Busa and M. Johnston, "Cross-linguistic semantics for complex nominals in the generative lexicon," AISB Workshop on Multilinguality in the Lexicon, 1996.
- [14] L. Burnard, "User reference guide for the British national corpus," Technical Report, Oxford University Computing Services, 2000.
- [15] T. Briscoe and J. Carroll, "Robust accurate statistical annotation of general text," Proc. 3rd International Conference on Language Resources and Evaluation (LREC 2002), pp.1499-1504, Las Palmas, Spain, 2002.
- [16] J. Pustejovsky, P. Anick, and S. Bergler, "Lexical semantic techniques for corpus analysis," *Computational Linguistics*, vol.19, no.2, pp.331-358, 1993.
- [17] A.L. Berger, P.F. Brown, and V.J.D. Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*,

vol.22, no.1, pp.39-71, 1996.

- [18] K.W. Church and P. Hanks, "Word association norms, mutual information and lexicography," Proc. 27th Annual Meeting of the ACL, pp.76-83, 1989.



Ichiro Yamada received his B.S. and M.S. degrees in Information Engineering from Nagoya University in 1991 and 1993. He joined NHK (Japan Broadcasting Corporation) in 1993 and has worked for NHK Science and Technical Research Laboratories since 1996. He was a visiting scholar at Stanford University from 2003 to 2004. His research interests include information extraction, contents handling and knowledge acquisition using natural language processing.



Timothy Baldwin received an M.E. and Ph.D. in Computer Science from the Tokyo Institute of Technology in 1998 and 2001, respectively. From 2001 to 2004 he worked as a senior researcher at Stanford University, including a 6 month visit to the NTT Communication Science Laboratories. Since 2004, he has been employed as a Senior Lecturer in the Department of Computer Science and Software Engineering, University of Melbourne. His research interests include information extraction, deep processing, lexical acquisition, and computational lexical semantics.



Hideki Sumiyoshi graduated in 1980 from the Electrical Engineering Department of Hiroshima Technical High School. He received his Ph.D. in Electrical Engineering from Tokyo University in 2004. He joined NHK (Japan Broadcasting Corporation) in 1980 and since 1984 has worked for NHK Science and Technical Research Laboratories. His main research interest is in developing TV program production systems using computer and computer network technologies.



Masahiro Shibata graduated in 1979 from the Electronics Department, Faculty of Engineering, Kyoto University, from which he also received his Masters degree in 1981 and Ph.D. in 2003. He joined NHK (Japan Broadcasting Corporation) in 1981 and has worked at the Science and Technical Research Laboratories and the Broadcast Engineering Department. Since 2007, he has been a senior research engineer at the Science and Technical Research Laboratories. His research interests include information retrieval technology, image database and contents handling systems.



Nobuyuki Yagi received his B.E., M.E., and Ph.D. degrees in electronic engineering from Kyoto University in 1978, 1980, and 1992, respectively. He joined NHK in 1980 and has worked at the Kofu Broadcasting Station, the Science and Technical Research Laboratories, the Engineering Administration Department, and the Programming Department. Since 2006, he has been a Senior Research Engineer in the Human & Information Science Division of the Science and Technical Research Labora-

tories of NHK (Japan Broadcasting Corporation). His research interests include image and video signal processing, multimedia processing, computer architecture, and digital broadcasting.