

Inferring Semantics from Collocation Clusters to Represent Verbs and Nouns

Bowen Hui

Department of Computer Science, University of Toronto
10 King's College Road
Toronto, Ontario, Canada
bowen@cs.utoronto.ca

Abstract

Current lexical semantic theories provide representations at a *coarse grained* level. In this paper, I will provide motivations for a *fine grained* representation for verbs and nouns. An initial case study is done to serve as evidence that a more detailed representation is needed for tasks that require high accuracy rates, such as machine translation. An automatic approach to gather fine grained information via corpus extraction is described. Lastly, issues of lexical representation and cross-lingual translation are discussed.

1 Problem

Some translational ambiguity arises when there are lexical gaps between languages. The particular problem of interest is when a general concept exists in one language but only detailed realizations of that concept exist in another language. A classic example of this problem is illustrated by the following sentences involving 'wear' in Japanese. (Notation: TOP denotes topic marker, ACC denotes accusative case, PROG denotes progressive tense.)

(1) Japanese 'wear'

Koji-wa aoi shatsu-o **kite-iru**
Koji-TOP blue shirt-ACC wear-PROG
"Koji is wearing a blue shirt"

Koji-wa kuroi buutsu-o **haite-iru**
Koji-TOP black boots-ACC wear-PROG
"Koji is wearing black boots"

Koji-wa aoi megane-o **kakete-iru**
Koji-TOP blue glasses-ACC wear-PROG
"Koji is wearing blue glasses"

Hutchins & Somers (1992) cite eight different 'wear' verbs in Japanese. The correct 'wear' verb depends on many aspects of the nominal complement. These include at least: location – where the noun is worn on the body; material – what the noun is made of; and shape – the overall shape of the noun (e.g. bendable, tube-like, donut-shape). 'wear' is just one example. In fact, gaps occur with other verbs and in other languages as well. Consider the following sentences. (Notation: NOM denotes nominative case, PAST denotes past tense, CLASS denotes classifier, DET denotes determiner.)

- (2) Japanese ‘break’
- | | | | |
|------------------------|-----------|--|-------------------|
| Rika-ga | isu-o | | kowashi-ta |
| Rika-NOM | chair-ACC | | break-PAST |
| “Rika broke the chair” | | | |
| Rika-ga | kabin-o | | wat-ta |
| Rika-NOM | vase-ACC | | break-PAST |
| “Rika broke the vase” | | | |
- (3) Cantonese ‘hit’
- | | | | |
|---------------------|------------|-------|--------|
| Tim | da | go | sai-lo |
| Tim | hit | CLASS | kid |
| “Tim hit the kid” | | | |
| Tim | pak | toi | |
| Tim | hit | table | |
| “Tim hit the table” | | | |
- (4) Vietnamese ‘wash’
- | | | | | | |
|-----------------------|------------------------|---------|------|--------|--------|
| Chau | ru⁺a | cài | chén | màu | vàng |
| Chau | wash | DET | bowl | colour | yellow |
| “Chau washed dishes” | | | | | |
| Chau | giã | t | ddò | | |
| Chau | wash | clothes | | | |
| “Chau washed clothes” | | | | | |

These examples show that different realizations have the same subcategorization in both the source and target languages. Therefore, one cannot rely on the use of syntactic surface forms to differentiate the semantic features of these verbs. Thus, Levin’s claim (1993) that syntactic constructions correlate with semantic frames, although demonstrated successfully for English, does not immediately apply to the above languages. To disambiguate these verbs in their respective languages, one needs to *extract* finer grained features. Furthermore, the translated root word is in most cases mono-morphemic in the target language. Therefore, the representation discussed by Palmer & Wu (Wu and Palmer, 1994; Palmer and Wu, 1995) for ‘break’ in Mandarin Chinese cannot be used either. There is no general Mandarin word equivalent to the English ‘break’. Mandarin has more specific verbs that are constructed serially by specifying the action taken to break something and the resulting state of the object. For example, to break a vase in Mandarin translates literally into “hit-shatter one CLAS flower pot”. To resolve translational ambiguities, Palmer & Wu devised a decision tree of the two components of ‘break’. When verbs have such a serial construction, it is possible to decompose them into their parts and use decision trees to represent them. However, this approach is not applicable to words such as ‘break’ in Japanese, ‘hit’ in Cantonese, or ‘wash’ in Vietnamese, where there is no decomposable morphology. Therefore, for languages with little morphology, a different approach is necessary.

2 Current Systems and Theories

How does a typical machine translation (MT) system deal with translational ambiguity? The translation mechanism depends on the architecture of the system. Consider Figure 1, from (Dorr et al., 2000), showing three general architectural systems: direct, transfer, and interlingua.

Systems that perform *direct* word for word translations look up each individual word in a bilingual lexicon and give the listed entry in the target language as a viable translation. If multiple target words are listed, then a (potentially incorrect) “default” verb is chosen. Systems that

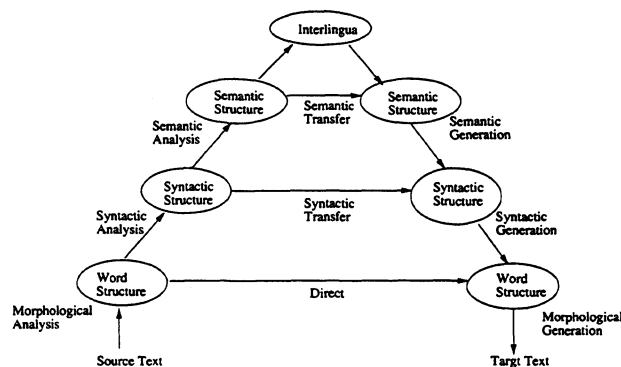


Figure 1: Types of MT Systems

apply syntactic *transfer* would not be able to handle the ambiguity illustrated above (without extra heuristics) because the sentences have the same syntactic constructions. Both semantic transfer and *interlingua* may be able to resolve the ambiguity accurately, *if* the representation of the lexical items are detailed and accurate. The purpose of the author's research is to represent lexical items in multiple Asian languages at an adequate level of granularity.

Imagine an enriched representation of verbs and nouns available to us. If MT systems use a sophisticated lexicon for organizing nouns, such as WordNet (Fellbaum, 1998), would that help resolve translational ambiguity? The problem of interest has to do with ambiguous verbs. Unless the organization of nouns reflect on their relationships with the verbs, such a lexicon would not help improve MT accuracy. Furthermore, Montemagni & Pirrelli (1998) has shown that using distribution-based semantic similarity has significant improvements over purely semantically-based similarity in tasks such as word sense disambiguation. What about using sophisticated verb representations in the lexicon, such as Pustejovsky's generative framework (1993) or Jackendoff's Lexical Conceptual Structure (LCS) (1990)? Thus far, there has been little work done in extending these theories to deal with lexical divergences in Asian languages. However, both theories have room for extensions to encode more detailed information. To understand the representations used in these theories, let us consider two lexical entries of verbs.

In Pustejovsky's framework, the FORMAL component of the qualia structure is used to distinguish the entry itself from other entries in the domain's lexicon. Taking his example of 'break' from p.80 of (Pustejovsky, 1993), the verb's qualia formal is: **broken**(e_2, y). This means that the process of breaking resulted in a state, e_2 , involving the object, y . In order to create entries for 'break' that accounts for different realizations of the verb, the qualia formal needs to be extended with more detailed predicates. The above examples show that 'waru' (infinitive of 'watta') is used when the broken object is of glass material and 'kowasu' (infinitive of 'kowashita') is used otherwise. In this case, the formal qualia of 'break' could be revised as **broken**(e_2, y) & **made_of_glass**(y).

LCS has been extended to account for lexical divergences between European languages (Dorr, 1992). In this work, Dorr et al. have extended the set of fields and primitives to account for the divergences in Spanish, German, and English. Since there are other lexical divergences between European languages and Asian languages, it would be interesting to compare the differences. Let us consider an example entry of 'wash', taken from (Dorr, 1992).

[Event DO ([Thing X], [Event WASH ([Thing X], [Thing Y])])]

The thing doing the washing, X , and the thing being washed, Y , are not encoded in detail. In Vietnamese, the verb 'ru⁺' is used for washing dishes and 'giã.t' is used for washing clothes. To account for this split, further extensions need to be made on the type Thing to specify what kind of object is undergoing the action.

From these examples, it is clear that both theories can account for lexical gaps but the specific details that are needed have not yet been developed. Therefore, we need to first identify and describe what the divergences are and then propose a way to incorporate the fine grained differences.

3 Approach

How fine grain do the representations need to be? The level of encoding depends on the range of items that need to be discriminated. For example, English 'wear' has only one surface realization while Japanese 'wear' has at least eight realizations. The number of features needed to discriminate all the Japanese 'wear' verbs will be greater than that of English. The general approach is to use verbs and nouns with different degrees of realizations from multiple languages as evidence towards fine grained features. Since languages such as Japanese, Cantonese, and Vietnamese exhibit a variety of lexical gaps, these languages are used to provide the necessary data. As mentioned above, linguistic features of verbs are well developed in lexical semantic theories (Jackendoff, 1990), (Pustejovsky, 1993), (Levin, 1993), and conceptual features of nouns are well developed in prototype and semantic theories (Rosch et al., 1976), (Lakoff, 1988), (Fellbaum, 1998). The idea presented in this paper is to combine these two kinds of features for one lexical item, whether the item is a verb or a noun. To do this, co-occurrences are used to create clusters for verbs and nouns. Each cluster reflects a distinguishing fine grained feature of the verb. The resulting clusters give insight to features that are common to the nouns in each cluster. Furthermore, these clusters give an indication of what conceptual features can be developed for the verbs. Finally, these features can be incorporated into existing lexical semantic theories.

4 A Case Study

This section describes a case study conducted on the verb 'wear' in Cantonese and Vietnamese¹. Experiments were designed to address how native speakers select amongst different 'wear' verbs when its nominal complement is given. These experiments serve as a preliminary investigation to see whether clustering nouns using verb information differ from clustering nouns using conceptual information. The former clustering approach used collocation statistics between the noun and the particular verb that it appears with in grammatical construction. The latter is based on how *similar* the nouns are to each other. If a significant difference shows up then that means a conceptual organization of nouns is not an accurate reflection of the distribution of nouns with the verbs. Furthermore, the author claims that if such a difference shows up, in MT tasks where accuracy is highly valued, a conceptually based lexicon used to resolve translational ambiguity will result in ungrammatical sentences.

The experiments concluded that there are similarities and differences between the clusters generated from the two approaches. The result of this investigation support the original hypothesis of this paper: verbs and nouns need a finer grained representation in order to achieve higher accuracy in MT tasks.

4.1 Overview of Experiments

Three experiments were conducted. The first experiment involved two native speakers of Cantonese and Vietnamese translating sentences with 'wear' using 66 nominal complements² and under different syntactic constructions. These nouns were chosen by searching for different complements of 'wear' in WordNet v1.6 (Fellbaum, 1998). These results indicated what the 'wear' verbs are in Cantonese and Vietnamese and how native speakers select among them with respect

¹Japanese data was not collected due to a lack of access to Japanese native speakers.

²These nouns are shown in Appendix A.

to its complement. Based on this elicited data, a subset of these verbs were eliminated because they do not have the meaning of 'wear'. An initial set of 10 Cantonese 'wear' verbs and 16 Vietnamese 'wear' verbs were chosen to be used in the second experiment.

The second experiment was a multiple choice task, where participants were given sentences with the verb blanked out. Each sentence has the intended sentence frame "John is wearing (ADJ) NOUN" or "Mary is wearing (ADJ) NOUN", where ADJ is an optional colour adjective and NOUN is one of the noun complements from experiment one. Sentences were written in the respective orthographies, where 'is wearing' was replaced with a blank and the intended meaning was provided in English below each sentence. Associated with each sentence was a set of verbs with a box beside each verb. The list of verbs provided were:

Cantonese:

tsun dzoek dai lam da tsa tip lou lup jau

Vietnamese:

ma(.c mang ddeo ddo[^].i ddi co' du'ng tru'm
dda('p ba(ng da'n thoa bo[^].i xu'?c so?n xa'i

In addition, participants were able to choose "other" and explicitly write a more or equally appropriate verb. In the case where multiple verbs were selected, the participants were asked to indicate their grammaticality preferences by ranking them. In total, there were 22 participants and each had to fill out 66 sentences. The results of this experiment were clustered and will henceforth be referred to as "collocation clusters".

The last experiment asked the 22 participants to group the noun complements used in the 66 sentences into what they felt were similar groups. Cue cards were given to participants to categorize. Each card had a noun complement written on it. Participants were also asked to give a short justification for each group of items they created. The results of the last experiment provided results for "conceptual clusters".

Considering the meanings of the verbs collected in experiment one and the "other" verbs suggested by various participants in experiment two, the author concluded that there were at least 13 'wear' verbs in Cantonese and 10 in Vietnamese. The verbs were:

Cantonese:

tsun dzoek dai lam tsa lou lup jap tou jau tat pei pun

Vietnamese:

ma(.c mang ddeo ddo[^].i ddi thoa bo[^].i xu'?c so?n khoa'c

4.2 Clustering Analysis

In cluster analysis (CA), no assumptions were made about the number of groups or the group structure of the data. The specific CA technique used in this study is an *agglomerative hierarchical method*. Every data point has an associated distance or similarity with respect to every other data point. Each data point begins by being in a cluster of its own. Based on a specified similarity measure, the two closest clusters merge to form a new cluster. This new cluster's distance is calculated based on the distances of the two original clusters. This process repeats at different levels until all the clusters combine into one. The data in this experiment used two standard measures for comparison purposes, *single linkage* and *average linkage* (Johnson and Wichern, 1992). Single linkage was chosen because it does not impose any constraints on the shape of the clusters and average linkage was chosen because it tends to produce clusters with small variance without constraining the shape of the clusters (Milligan, 1980).

The input to these algorithms was distance matrices. Each participant's data was represented as a two-dimensional matrix, where the columns were 'wear' verbs and the rows were noun complements. Each entry in the matrix was the ranking that the participants assigned to the

verb for that particular noun. Call this the ranking matrix. The entries in these ranking matrices were converted into scores by:

$$s_{ij} = 11 - r_{ij}, \text{ for } r_{ij} \neq 0 \\ = 0, \text{ for } r_{ij} = 0$$

where s_{ij} is the new entry and r_{ij} is the corresponding entry in the ranking matrix. These scores range from 0 to 10, where 0 denotes ungrammatical usage and values between 1 and 10 reflect the degree of verb preference, with 1 being the least preferred and 10 being the most preferred.

For each language, the ranking matrices were summed across participants. As a result, two matrices with weighted rankings from all the participants were generated. These two matrices served as input to the agglomerative hierarchical CA methods. The data was clustered with each verb operating as a dimension and each noun as a Euclidean point in space.

The data was clustered into 5, 7, 10, 13, 15, 17, and 20 clusters. Applying average linkage and single linkage to each of these seven sets, a total of 14 *collocation clusters* were generated for each language. To produce the *conceptual clusters*, the same CA techniques were applied to the data in experiment three. Each participant's data was represented as a two-dimensional incidence matrix, where the columns and rows were the nouns to be categorized. Each entry in the matrices reflected membership of that pair of nouns, i.e., an entry, e_{ij} , is recorded as:

$$e_{ij} = 1 \text{ if noun}_i \text{ and noun}_j \text{ are in the same group} \\ = 0 \text{ otherwise}$$

For each language, the incidence matrices were summed across participants. As a result, two similarity matrices with summed memberships from all the participants were generated, which served as input to the algorithms. The data was clustered with groups identified by participants operating as dimensions and each noun as a point in that space. Note that the distances generated from these similarity matrices are not true distances³. The data was clustered into 5, 7, and 10 clusters. Applying average linkage and single linkage to each of these three sets, a total of 6 *conceptual clusters* were generated for each language.

4.3 Results

Judgements differed across participants for both experiments two and three. In experiment two, there was more variance in the choice of verb and the number of possible grammatical verbs for one sentence in Vietnamese than there was in Cantonese. In experiment three, participants varied in the number of groups they created and the number of elements in each group in both Vietnamese and Cantonese. In summary, the resulting clusters are presented in Figures 2, 3, 4, and 5. Groups are drawn in solid circles unless there is only one member in the group. *Stronger* groups are drawn inside the dotted circle and *weaker* groups are drawn outside. Groups are considered strong if they consistently clustered together under different sizes of clusters or using both metrics for the same sized cluster. Groups that appeared in one cluster when there were fewer clusters and then broke off when there were more clusters are shown as subgroups of the original cluster. Groups and items that only appeared once are considered outliers and are not shown here.

There are not many strong collocation clusters in Cantonese. Many of the items showed up as individual outliers. This result correlates to the 'wear' verbs that are used with these nouns; the common 'wear' verbs are used for the items in the strong groups and the less common 'wear' verbs are used for the few remaining outliers. A similar result is found with the Vietnamese collocation clusters. The clusters that appeared depend strongly on the verbs that were used.

³Johnson & Wichern (1992) state that true distances are preferred "but most clustering algorithms will accept subjectively assigned distance numbers that may not satisfy, e.g. the triangle inequality".

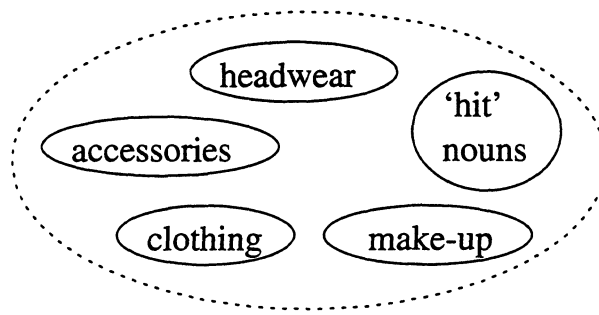


Figure 2: Cantonese Collocation Clusters

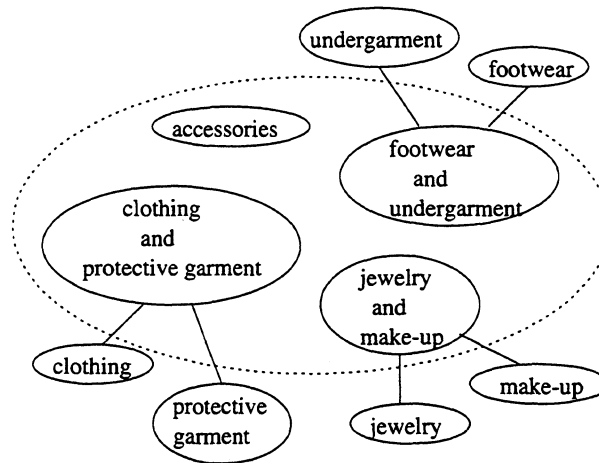


Figure 3: Cantonese Conceptual Clusters

Because there are more common realizations of ‘wear’ in Vietnamese, more clusters resulted. As expected, both of these results correlate with the realization patterns of ‘wear’ in the respective languages.

On the other hand, there are more conceptual clusters in Cantonese. The small clusters resemble those in the Vietnamese conceptual clusters, with the exception that Cantonese has a cluster for undergarments while Vietnamese has a cluster for the item hat. This result suggests that the participants’ conceptual models of these nouns are similar.

4.4 Implications

Distinctions between collocation clusters and conceptual clusters have implications on MT systems that disambiguates using a semantic lexicon (e.g., using or creating a noun subnet for a Chinese WordNet-like hierarchy). First of all, designers of such a network must decide on which categories to use and how these categories are structured among themselves. Consider the example of ‘tie’ and ‘cast’ in Cantonese that use ‘hit’ to mean ‘wear a tie’ and ‘wear a cast’. In a semantic network, the ‘hit’ group would not surface and ‘tie’ is simply a member of ‘accessories’ (as shown in Figure 3). The partial structure of what this network may look like is shown in Figure 6. This system, when asked to translate the ‘wear’ verb for ‘tie’, starts searching at the ‘wear’ node. Then it goes down to ‘accessories’, finds that ‘tie’ is a member of it, and incorrectly returns the verb corresponding to the category ‘accessories’ (in the case of Cantonese, the verb is ‘dai’). In a linguistic semantic network, the ‘hit’ group surfaces as its own cluster (as shown in Figure 2). From Figure 6, the search starts at ‘wear’, goes down to

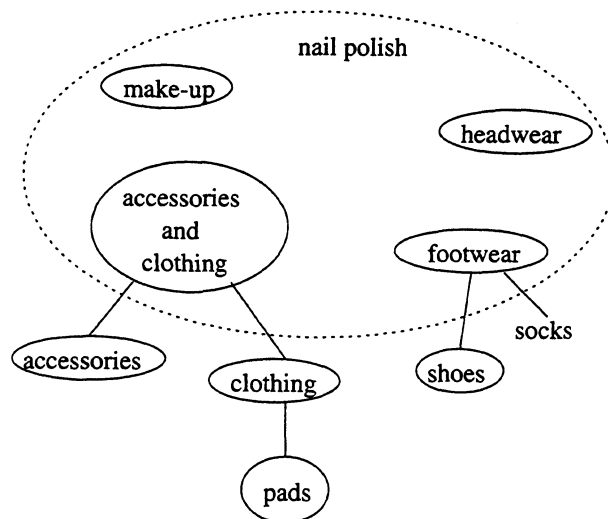


Figure 4: Vietnamese Collocation Clusters

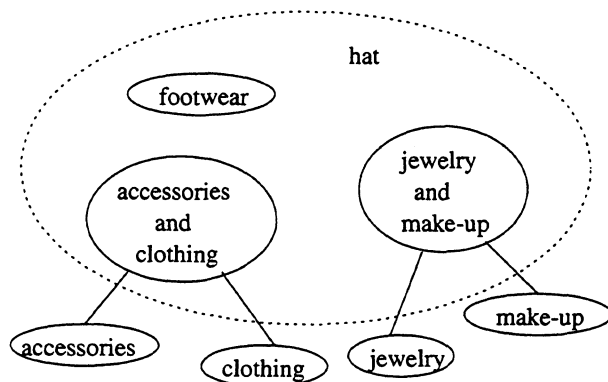


Figure 5: Vietnamese Conceptual Clusters

'accessories', goes down to 'hit', finds that 'tie' is a member of it, and then correctly returns the verb corresponding to the 'hit' cluster (i.e., 'da'). Therefore, when a cluster surfaces based on conceptual categories but not collocation statistics, inaccurate translations result. Therefore, if the accuracy of an MT system is dependent on the structure of the network, then it would be important to choose a linguistically-oriented clustering network using statistical approximations rather than a conceptually-oriented one.

On the other hand, when a cluster surfaces conceptually but not from its collocations, system error or redundancy occurs. For example, 'slippers' in Cantonese is clustered into the group 'footwear' psychologically (see Figure 7). In selecting the corresponding verb for 'slippers', the system starts at the 'wear' node, goes to 'clothing', goes to 'footwear', finds that 'slippers' is a member of 'footwear' and returns the verb associated with 'footwear'. At this point, either 'footwear' has no verb associated with it which may result in a system error or the same verb for 'clothing' (i.e., 'tsoek') is returned. Therefore, in such a network, an MT system either errors on the translation or encodes redundant categorical information which has no significance on the translation.

To recapitulate, a comparison between collocation clusters and conceptual clusters shows that MT systems that resolve lexical divergences based on conceptually organized lexicons generate

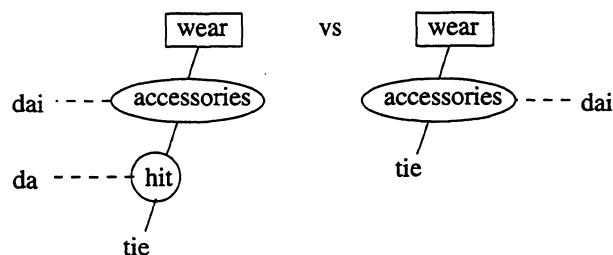


Figure 6: Category for 'tie'

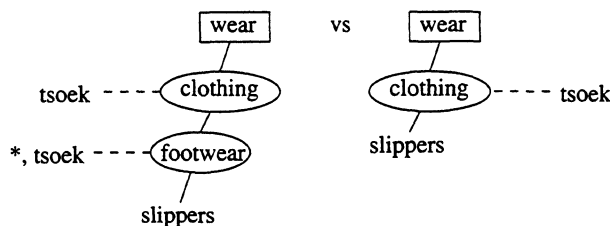


Figure 7: Category for 'slippers'

inaccurate translations. The results shown in this preliminary investigation suggest that a richer representation of verbs and nouns based on collocation statistics enhances the accuracy of MT systems. The next step is to come up with features that are common to the items in these clusters, and incorporate these features into existing lexical semantic theories. With this equipment, a robust experiment can then be devised to compare the accuracy improvements over systems with fine grained representations versus systems with coarse grained representations.

5 Conclusions and Future Work

Current lexical semantic theories (Jackendoff, 1990), (Pustejovsky, 1993), and verb disambiguation (Siegel, 1998), selection (Dorr, 1992), and classification work (Stevenson and Merlo, 2000) all represent verbs at a coarse grained level. Section 1 discussed examples from Japanese, Cantonese, and Vietnamese where current MT systems with coarse grained representations resolve translational ambiguity with *ad hoc* solutions. Fine grained lexical semantics is well motivated for high accuracy tasks, such as MT. To obtain fine grained features, the proposed method is to derive meaning from collocation statistics. To demonstrate that this kind of semantics is different from verb representations in linguistic theories and noun representations in prototype theories, a case study on 'wear' was conducted for Cantonese and Vietnamese. This investigation suggested that hierarchical dictionaries built upon collocation clusters are more accurate than those built upon conceptual clusters. In some cases, conceptual clusters may reduce translation accuracy. This idea warrants further research.

The clustering algorithms and definitions of distances need further exploration. Other techniques and defined distances such as the centroid method, complete linkage, and sums of squares should be tried and compared. Research is currently underway to gather a larger set of data via corpus extraction in multiple Asian languages. The first stage is to replicate the results in experiment two by analyzing 'wear' sentences with active voice and present tense. Then other sentence constructions and tenses will be considered, such as passives, questions, past tense, and future tense. A larger set of verbs will also be considered, including 'wear', 'wash', 'break', 'hit', and 'see'. This data can be collected by applying semi-automatic techniques on monolingual tagged corpora. First, the set of target verbs must be identified as seeds to be extracted.

If the corpus is not parsed, then the nominal complements of the seeds need to be extracted semi-automatically. These nouns and verbs can then be tallied and plugged into clustering algorithms, as in the case study, and results will follow. If bilingual corpora were available, then the target verbs need not be specified as seeds to the extraction algorithm. For instance, the English word 'wear' could be used to *discover* the seeds in an English-Chinese aligned corpus. However, gaps between English and Chinese may produce incomplete sets of seeds. Other bootstrapping methods for discovering the seeds for extraction need to be investigated.

Once a more robust set of clusters are found, the goal is to develop fine grained features needed to discriminate the data and to explore how these additions can be incorporated into different lexical semantic theories such as LCS (Jackendoff, 1990) or Pustejovsky's generative framework (Pustejovsky, 1993). Some suggestions were discussed in Section 2. Finally, evaluation will take place using an MT system to translate sentences involving the verbs and nouns in the collected data.

Appendix A

The set of nouns used in the case study were: shirt, eye mask, cloak, shoes, boots, toe ring, (finger) ring, insoles, cast, (face) mask, blanket, jewellery, belt, suspenders, military uniform, gloves, (hand) cream, nail polish, tattoo, necklace, pyjamas, wig, tie, bib, bathing suit, apron, underwear, socks, slippers, anklet, (long/short) skirt, scarf, gymsuit, jacket, jeans (pants), hat, helmet, protective gear, elbow pads, kneepads, parachute, spacesuit, armour, diaper, fur coat, sweater, vest, funeral wear, fake teeth, fake mole, wrinkles, suit, bandage, overalls, nylons, watch, glasses, bellybutton ring, hair clips, hair band, earrings, bra, bracelet, baseball glove, lipstick, perfume.

Acknowledgements

I am in debt to the support and insights provided by Nick Cercone and the statistical consultation by Dieter Ayers. Thanks also to Craig Boutilier, Pat A. Shaw, Suzanne Stevenson, and Diana Zaiu Inkpen for their helpful comments.

References

- Dorr, B.J. 1992. The use of lexical semantics in interlingua machine translation. *Journal of Machine Translation*, 7:3:135-193.
- Dorr, B.J., P.W. Jordan, and J.W. Benoit. 2000. A survey of current research in machine translation. *Advances in Computers*, 49.
- Fellbaum, C. 1998. (Ed.) *WordNet: An Electronic Lexical Database*. MIT Press.
- Hutchins, W. J. and H. L. Somers. 1992. *An Introduction to Machine Translation*. Academic Press.
- Jackendoff, R. 1990. *Semantic Structures*. MIT Press.
- Johnson, R. A. and D. W. Wichern. 1992. *Applied Multivariate Statistical Analysis 3rd ed.* Prentice Hall.
- Lakoff, G. 1988. *Women, Fire, and Dangerous Things*. The University of Chicago Press.
- Levin, B. 1993. *English Verb Classes and Alternations*. Chicago Press.
- Milligan, G. W. 1980. An experimentation of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45:325-342.
- Montemagni, S. and V. Pirrelli. 1998. Augmenting wordnet-like lexical resources with distributional evidence, an application-oriented perspective. *Proc. of the COLING/ACL Workshop on Usage of WordNet in NLP Systems*.

- Palmer, M. and Z. Wu. 1995. Verb semantics and english-chinese translation. *MT Journal*, pages 1-32.
- Pustejovsky, J. 1993. *Generative Lexicon*. MIT Press.
- Rosch, E., C. Simpson, and R.S. Miller. 1976. Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2:491-502.
- Siegel, E. V. 1998. Disambiguating verbs with the wordnet category of the direct object. *Proc. of the COLING/ACL Workshop on Usage of WordNet in NLP Systems*.
- Stevenson, S. and P. Merlo. 2000. Automatic lexical acquisition based on statistical distributions. *Proc. of the 18th International Conference on Computational Linguistics (COLING 2000)*.
- Wu, Z. and M. Palmer. 1994. Verb semantics and lexical selection. *Association for Computational Linguistics*.

