Automated Identification of Frame Semantic Relational Structures

Abstract: Preliminary attempts to identify semantic frames and their internal structure automatically have met with a degree of success. In a first stage, clustering is used to detect previously identified semantic frames (COMMERCIAL TRANSACTION, HIT, JUDGING, RISK) from verb definitions in Longman’s Dictionary of Contemporary English. In a second stage, nouns used in the definitions of frame-invoking verbs or in whose definitions the frame-invoking verbs occur in certain forms are searched in WordNet to identify frame elements. Suggestions for refinement of the processes are discussed.

1. Rationale

This paper reports on research addressing a particular set of semantic relationships within the lexical inventory of a language, namely frame semantic relationships. The basic notion underlying frame semantics is that the propositional content of language (and thought) is structured by conventional (recurrent, stereotypical) patterns of experience. Such patterns, or frames, establish a basic configuration of relationships between frame elements.

For example, the verbs buy, purchase, sell, pay, spend, and cost invoke a COMMERCIAL TRANSACTION frame. Buy and purchase profile the BUYER’s acquisition of some MERCHANDISE, while sell profiles the SELLER’s de-acquisition of the MERCHANDISE; pay and spend profile the BUYER’s use of MONEY for the acquisition, while cost profiles the amount of MONEY needed on the BUYER’s part for the transaction. Despite their differences, all these verbs presuppose a common set of relationships between a BUYER, a SELLER, a sum of MONEY, and some MERCHANDISE, the basic elements of the frame. The upshot of this situation is that a single conceptual proposition may be expressed by rather different linguistic means. While the variety in such lexical realizations typically express meaningful differences (as in the profiling distinctions noted above), it also obscures meaningful semantic similarities.

Text retrieval strategies should be able to produce better results operating on frame semantic expressions, in which relational similarities are made manifest, than on natural language text, where the configuration of relationships underlying the predicate-argument structure of propositions is often much more obscure. On the one hand, frame semantics addresses the recall issue by mapping sets of semantically similar, but textually divergent, expressions to common (or systematically related) frame semantic representations. On the other hand, frame semantics addresses the often-more-pressing precision issue by doing implicit lexical and syntactic disambiguation. For example, homonyms will probably map to different frames, while syntactically ambiguous expressions, even if they map to the same frame, would probably instantiate the frame in different ways.

The identification of frames and their internal structures is a labor-intensive task, which heretofore has been undertaken largely by introspection, on the one hand, and manual analysis, on the other hand. Moreover, even the large-scale FrameNet project (Lowe, Baker, and Fillmore, 1997) is limited in the range of frames whose internal structure it is seeking to identify. Thus, the development of a systematic and automated (or semi-automated) approach to the tasks of (1) frame detection and (2) identification of frame internal structure is called
for, both to extend our knowledge to a wider range of frames and to ensure consistency in the treatment of frames.

The preliminary research reported here addresses both of these tasks. The first task of identifying frames is approached by identifying sets of verb senses that invoke a common frame. The second task builds on the first by seeking to automate the identification and characterization of elements in the frame, given a set of verbs/verb senses that invoke the same frame.

2. Frame Detection

The definition of a verb sense usually reflects the frame(s) it invokes and commonly indicates (at least in part) the internal structure of the frame(s). For example, figure 1 gives the definitions of several verb senses that invoke the COMMERCIAL TRANSACTION frame. Two generalizations are worth noting: (1) Several lexemes recur in the definitions of the verb senses: obtain, give, money, value, exchange, goods, sale. (2) Words corresponding to the GOODS, the MONEY, and the BUYER are present in the definitions (no words corresponding specifically to the SELLER are present, however, in these definitions). From these observations, we hypothesize a correlational association between the sets of terms used in the definitions of verb senses and the frames they invoke. Thus an appropriate analysis of verb sense definitions may identify those invoking a common semantic frame.

For this study definitions were taken from the machine-readable version of Longman's Dictionary of Contemporary English (LDOCE). Since LDOCE uses a restricted definition vocabulary of about 2000 words (Alshawi 1989), words with closely related meanings should tend to use the same words in their definitions and thus support the pattern of discovery envisioned.

An input data set was generated covering all non-phrasal verbs, distinguishing between both homonyms and the different senses of a single verb. The definitions were processed to exclude words from a customized stop list, sample sentences, and typographic codes; the definition terms were also stemmed using the Porter (1980) stemmer. Further, verbs were added as terms to their own sets, while terms appearing no more than twice in the set of verb sense definitions were deleted. The data set consists of 53,448 verb sense/term pairs, representing 12,348 unique verb senses and 2,656 unique terms. A second version of the database, which ignores verb sense distinctions, includes 42,704 entries, representing 5,444 unique verbs and 2,008 unique terms. (The number of unique terms has shrunk because of the frequency of occurrence criterion.)

<table>
<thead>
<tr>
<th>Verb Sense</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>buy 1</td>
<td>To obtain (something) by giving money (or something else of value)</td>
</tr>
<tr>
<td>buy 2</td>
<td>To obtain in exchange for something, often something of great value</td>
</tr>
<tr>
<td>buy 3</td>
<td>To be exchangeable for</td>
</tr>
<tr>
<td>Purchase 1</td>
<td>To gain (something) at the cost of effort, suffering, or loss of something of value</td>
</tr>
<tr>
<td>sell 1</td>
<td>To give up (property or goods) to another for money or other value</td>
</tr>
<tr>
<td>sell 2</td>
<td>To offer (goods) for sale</td>
</tr>
<tr>
<td>sell 3</td>
<td>To be bought; get a buyer or buyers; gain a sale</td>
</tr>
</tbody>
</table>

Figure 1. Definitions from verbs invoking the COMMERCIAL TRANSACTION frame

Finding groups of verbs that invoke the same semantic frame is an appropriate task for cluster analysis. Here the items to be grouped are verb senses, and the terms in their definitions are the attributes used for forming clusters.

Of the hierarchical agglomerative clustering algorithms commonly applied to the clustering of text, the group average link method, which forms new clusters based on the average values of pairwise similarity measures, is one of the two more successful methods.
(Rasmussen 1992). The most efficient algorithm for this method (Voorhees 1986, p. 471) runs in \( O(N^2) \) time while using \( O(N) \) space and was implemented in this study. It relies on the observation that when between-document similarity (in this case, between-verb-sense similarity) is calculated as the inner product of appropriately weighted vectors, cluster centroids, as the mean of all vectors in their spaces, can be used for computing between-cluster similarities (Voorhees 1986, p. 469).

Limited frame semantic analysis has been undertaken manually. Based on this analysis we can identify several groups of verbs that invoke the same frame, which will serve as the benchmark against which to compare the automated process. Accordingly, the results of clustering are examined with respect to a set of 50 verbs, set out in figure 2, that invoke 4 particular frames (Fillmore 1970, 1971; Fillmore and Atkins 1992; Lowe, Baker, and Fillmore 1997). Verbs in this set are referred to hereafter as “focus verbs.”

<table>
<thead>
<tr>
<th>COMMERCIAL TRANSACTION</th>
<th>buy, charge, cost, forgive, pay, price, purchase, sell, spend</th>
</tr>
</thead>
<tbody>
<tr>
<td>HITTING</td>
<td>Bump, clobber, conk, hit, jab, punch, slap, smack, smash, smite, strike</td>
</tr>
<tr>
<td>JUDGING</td>
<td>Absolve, accuse, apologize, blame, charge, condemn, convict, credit, criticize, defend, denounce, excuse, exonerate, forgive, justify, praise, scold, vindicate</td>
</tr>
<tr>
<td>RISK</td>
<td>Bet, chance, dare, endanger, expose, gamble, hazard, imperil, risk, threaten, venture, wager, warn</td>
</tr>
</tbody>
</table>

**Figure 2. Frame-invoking verbs**

There are two hurdles, however, in basing an evaluation on such evidence. First, human-generated frame semantic analysis is usually given at the word level, but the groupings sought here are at the word sense level. Second, hierarchical clustering ultimately groups all data items into a single cluster; meaningful intermediate clusters—which are hoped to correspond to frame semantic groupings—are identified on the basis of threshold values for which there is no objective basis (1.0 was used here). Given these hurdles, a formative evaluation is appropriate.

When Voorhees’ algorithm was run against the verb sense database, substantial grouping of verb senses invoking the same frame occurred. Of 29 COMMERCIAL TRANSACTION verb senses, 17 were grouped within 2 clusters; 15 of 23 HIT verb senses were gathered into a single cluster; 17 of 27 JUDGING verb senses were collected into 5 clusters; and 18 of 29 RISK verb senses fell into two clusters.

The major problem here is that each cluster averages approximately 100 verb senses. While some of the additional verb senses in these 10 clusters also invoke the 4 semantic frames under examination, many do not. To some extent the depressed precision ratios could be dealt with by using a higher similarity threshold value for dividing between clusters. (In this context, precision is a measure of the degree to which verb senses that are clustered together invoke the same frame; recall is a measure of the degree to which verb senses that invoke the same frame are grouped together into a single cluster or into a set of clusters that can be shown to be related). However, verb senses now grouped into the same cluster might then fall into different clusters.

When Voorhees’ algorithm was run against the verb database, 21 of the 50 verbs occurred in the near vicinity of a same-frame-invoking verb, 18 in pairs and 3 in a triple. Half of these small groupings echoed groupings evident in the verb sense clusterings, but half
made connections not evident among the verb senses. This suggests that verb sense clustering and verb clustering might profitably be used in tandem to establish the final groupings.

Clustering using Voorhees’ algorithm for the group average link method has demonstrated some degree of success in grouping verb senses that invoke the same frame into a small number of clusters. Three problem areas need to be addressed, two that concern recall and one that concerns precision.

On the recall side, the first problem is how to group together an even higher percentage of verb senses that invoke the same frame. It is unreasonable to think that a monohierarchical arrangement of verb senses will emerge, since verb senses may simultaneously invoke multiple frames. It may then be advantageous, if possible, to restrict the data set to ‘pure’ verb senses, i.e., those that invoke only a single frame.

A second recall problem is how to identify clusters that invoke the same frame. The clustering process has frequently grouped frame-related verb senses into, not just a single cluster, but two or more clusters. If such clusters emphasize different aspects of the frame, the drawing together of related clusters is a crucial step in the eventual success of the task.

The third problem area concerns precision. Here the issue is how to filter out verb senses that have been grouped with a set of frame-related verb senses, but that do not invoke the same frame. It is probably not necessary to filter out all ‘interlopers,’ since frame internal structure discovery is based on finding those participant types that occur most often in the input set. So long as the non-frame-invoking verb senses are not themselves systematically related, it is unlikely they will pick out a participant type at a frequency level to rival those of the true frame participants. Still the next step will presumably work better with less extraneous material.

The first and third problem areas, which address cluster membership—what’s in and what’s out of the cluster, respectively—can be addressed simultaneously. Two different approaches apply here. The first approach is to vary the clustering process, which can be achieved by changing the input to the process (e.g., modifying the stemming procedure, varying the stopword list, initializing term weights in various ways) or by changing the clustering algorithm itself. The second approach to varying cluster membership concerns how to define clusters. So far a simple threshold value has been used, but other criteria, for example, some number of monotonically decreasing similarity values, could be used instead of or in conjunction with an absolute threshold for setting off clusters.

The most significant problem area is the second one, how to group multiple clusters together that contain frame-related verb senses. The approach to its probable solution lies in combining multiple sources of information. As noted, clustering at the verb level made some desired connections not made in clustering at the verb sense level. Combining data from both clustering processes may be useful in identifying clusters invoking the same frame. Semantic relationships in sources such as WordNet might also aid in ferreting out cross-cluster connections.

3. Identification of Frame-Internal Structure

Identifying the internal structure of the frames started with definitions for the verbs known to invoke the COMMERCIAL TRANSACTION, JUDGING, HITTING, and RISK frames. Nouns within their definitions were searched in WordNet, a hierarchically-structured lexical resource, to identify nodes of the noun tree which recur most often in those definitions. The hypothesis underlying this approach is that words referring to the various elements of the frame structure would tend to recur in the definitions of verbs that invoke the frame. For example, part of the definition for the verb sell, as taken from the Shorter Oxford English Dictionary (Shorter OED), reads:
Make over or dispose of (a thing) to another in exchange for money etc.; esp. dispose of (merchandise, possessions, etc.) to a buyer for or at a specified price.

While the first part of the definition refers generally to the GOODS as a thing and to BUYER as another, naming only the MONEY frame element specifically, the more specific restatement in the second part gives a near synonym for GOODS in its use of merchandise, specifically refers to the BUYER, and refers to an aspect of the MONEY role in its use of the term price.

A verb's definition, however, often lacks explicit reference to the frame element that would commonly appear as its subject in a natural language text. It is also common for verb definitions to omit explicit reference to their surface objects, as when the Shorter OED defines purchase as:

Acquire by payment; be an equivalent price for.

Since definitions of frame-invoking verbs commonly lack explicit mention of those frame elements which fill Agent (or Experiencer) and Theme thematic roles, it is desirable to supplement the data set being passed through WordNet in such a way as to better ensure that all frame elements will be mentioned. Examination of the definitions of the words which best identify these frame elements (e.g., seller, buyer) or their synonyms (e.g., merchandise) —

A person who sells
A person who buys
The commodities of commerce; goods to be bought and sold

— suggests that the set of nouns in whose definitions frame-invoking verbs occur either as a third person singular present indicative and or as a past participle may constitute a profitable supplemental source.

Accordingly, two sets of nouns were generated for each frame, one based on the occurrence of the nouns in the definitions of the focus verbs (the verb data set) and the other based on the appearance of the specified forms of the focus verbs in the definitions of the nouns (the noun data set). Both sets were drawn from the Shorter OED. This source was used because its definitions are known to conform with the patterns described above (while LDOCE's restricted vocabulary aids the clustering process, it seems not to work as well for the frame element identification task).

No mention has yet been made of polysemy, a phenomenon that often creates problems for natural language processing. The focus verbs averaged four senses in the Shorter OED, many of which are divided into subsenses. No word sense disambiguation was done with regard to the generation of the two sets of nouns, it being assumed that the commonalities inherent in the frame-invoking senses of the verbs would, when ranked by frequency, outweigh presumably random commonalities. Thus, the only nouns excluded from the data gathering process were (1) those that appeared only in archaic, obsolete, or rare senses; (2) those that named grammatical categories (the Shorter OED definitions occasionally incorporate a limited amount of syntactic information); and (3) those appearing on a minimal stopword list. The verb data set included 368 unique nouns, while the noun data set included 693 unique nouns.

Three sets of nouns (the verb data set, the noun data set, and the verb and noun data sets combined) were searched against the WordNet noun hierarchy to identify the subtrees in which they occur. Again, no word sense disambiguation was performed, so multiple subtrees were often identified for each noun. Each subtree consists of a hierarchical series of synsets,
concepts represented by words and phrases which in some specific sense are synonyms or near-synonyms of each other. The first synset in each subtree is one of a small number of unique beginners in the WordNet noun hierarchy; all subsequent synsets represent narrower concepts in WordNet's extensive IS-A noun hierarchy, which may involve as many as a dozen levels.

Trees consisting of the union of the separate subtrees were generated for each frame. At each node, these trees recorded how many nouns passed through the node (including those terminating at that node), how many nouns terminated at that node, and how many focus verbs contributed to the presence of the node. Eight different weights were computed for each node. The manner in which the weights were computed varied along two dimensions: (1) which of two weighting schemes was used, and (2) which of four accumulation schemes was used. One of the weighting schemes (WGT 1) accorded more importance to the number of nouns whose subtrees terminated at the node, the other (WGT 2) to the number of verbs motivating those nouns. The accumulation schemes differed by how many levels of basic weights were summed to form a node's total weight. In one scheme (ACCUM 0), there was no accumulation across levels; the node's basic weight was also its total weight. In the other three schemes (ACCUM 1, ACCUM 2, ACCUM 3), a node's total weight resulted from summing its basic weight with the basic weights of all its descendants up to one, two, or three levels down, respectively. The accumulation strategies are based on the observation that definitions refer to frame elements at different levels of specificity. The nodes were then sorted by weight.

Figure 3 sets forth the top ten synsets from the best results for COMMERCIAL TRANSACTION, which came from the noun data set, using WGT 2 and ACCUM 0. On the one hand, the preponderance of frame-related synsets is an encouraging sign and suggests that lexical data can be used to determine the internal structure of a frame. On the other hand, a number of interpretational issues are raised that require resolution prior to large-scale implementation.

| Weight | Syntactic/semantic
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>Monetary value, price cost</td>
</tr>
<tr>
<td>62</td>
<td>price, cost</td>
</tr>
<tr>
<td>33</td>
<td>grocery store, grocery, market</td>
</tr>
<tr>
<td>33</td>
<td>buyer, purchaser, emptor, vendee</td>
</tr>
<tr>
<td>20</td>
<td>Payment, paying, defrayal, defrayment</td>
</tr>
<tr>
<td>11</td>
<td>Cost</td>
</tr>
<tr>
<td>11</td>
<td>Payment</td>
</tr>
<tr>
<td>11</td>
<td>Merchandise, wares, product</td>
</tr>
<tr>
<td>11</td>
<td>Distributor</td>
</tr>
<tr>
<td>11</td>
<td>Price</td>
</tr>
</tbody>
</table>

Figure 3. Top nodes for COMMERCIAL TRANSACTION (noun data set, WGT2, ACCUM 0)

In the best of all worlds, synsets representing all (but only) frame element concepts would appear at the top of the sorted node list and a large gap would occur between the weights of these nodes and the weights of other nodes. In truth, the first mention of synsets equivalent to some frame elements occur further down in the lists; some synsets that have nothing to do with the frames appear high in the lists; and the weights follow a Zipf-like distribution pattern, with no exclusive clustering of frame element synsets at weights far exceeding the weights of other nodes.

These complications arise from the nature of standard lexical relationships. Homonymy and polysemy probably account for the presence of those synsets that have little relevance to the frame in question. While the more relevant synsets have tended to rise to the top, this has
not always been the case. If each frame were represented by a more comprehensive set of lexical items at the outset, this problem might conceivably be resolved or at least reduced.

Synonymy and hyponymy lead to the presence of multiple synsets on the list that are more-or-less closely related in meaning, and it is not immediately obvious how to unify them. Indeed, this semantic closeness arises at several turns. For instance, seemingly identical synsets that occur in multiple places in the WordNet (poly)hierarchy represent one kind of closeness; the difference in ancestry precludes the concepts they represent from being strictly the same. Strictly hierarchical relationships between synsets creates another kind of closeness. This is the relationship between cost and price in figure 3. Hierarchy also creates the co-hyponymous relationship between price, cost and monetary value, price, cost.

Although preliminary in nature, the current investigation nevertheless provides cautiously optimistic evidence that lexical resource data can be fruitfully used for discovering the internal structure of frames. Improvements of the procedures used here are anticipated along two fronts: (1) Set more stringent filters in place in collecting the original data sets/noun lists. This might involve a larger stop word list, the exclusion of nouns that appear only parenthetically in a definition or the exclusion of nouns based on some syntactic criteria. Particularly if a fuller set of verbs for each frame is included in the study, the set of nouns drawn for each of those verbs needs to be more precisely limited from the beginning. (2) Determine how to cluster WordNet synsets that are closely related in meaning into a single concept. The granularity of word senses in WordNet is much finer than is needed for IR purposes. Consequently, there is the need to collapse semantically related synsets into single broader concepts. The accumulation schemes were a beginning attempt at this, but are not the full answer to the problem.

On a broader level, it would also be useful to delve into whether the WordNet hierarchy can be used to help identify the existence of frame-related words in the first place. Further studies are contemplated in which mid-to-lower level synsets in WordNet constitute the starting point for searching dictionary definitions to find those words with semantic components in common, which are then compared for common configurations of components and passed back through the process defined in this study, as refined.

References


