

Figure 1: System architecture.

house implementation of a named entity recognizer. We also extract from the FN lexicon mappings of target words and the semantic frames they evoke.

Various features corresponding to constituents were extracted and passed to SVM and ME classifiers. For example, in Figure 2, the frame dis-

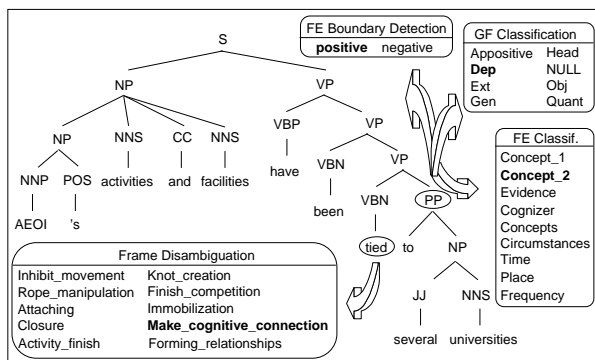


Figure 2: Classification examples for each sub-task.

ambiguation sub-task extracts features corresponding to the constituent *tied* in order to predict the right frame between the semantic frames that can be evoked by this target word. In this figure, the correct categories for each sub-task are shown in boldface.

The complete set of features extracted for all the classification sub-tasks is illustrated in Figure 3. These represent a subset of features used in previous works (Gildea and Jurafsky, 2002; Florian et al., 2002; Surdeanu et al., 2003; Xue and Palmer, 2004; Bejan et al., 2004; Pradhan et al., 2005) for automatic semantic role labeling and word sense disambiguation. Figure 3 also indicates whether or not a feature is selected for a specific classification task.

In the remaining part of this section we describe in detail each classification sub-task and the features that have the most salient effect on improving the corresponding classifiers.

## 2.1 Frame Disambiguation

In FrameNet, some target words can evoke multiple semantic frames. In order to extract the semantic structure of an ambiguous target word, the first step is to assign the correct frame to the target word in a given context. This task is similar with the word sense disambiguation task.

We select from the FN lexicon 556 target words that evoke at least two semantic frames and have at least five sentences annotated for each frame, and assemble a multi-class classifier for each ambiguous target word. As described in Figure 3, for this task we extract features used in word sense disambiguation (Florian et al., 2002), lexical features of the target word, and NAMED ENTITY FLAGS associated with the root node in a syntactic parse tree. For the rest of the ambiguous target words that have less than five sentences annotated we randomly choose a frame as being the correct frame in a given context.

## 2.2 Frame Element Identification

The idea of splitting the automatic semantic role labeling task into FE boundary detection and FE label classification was first proposed in (Gildea and Jurafsky, 2002) and then adopted by other works in this task. The problem of detecting the FE boundaries is cast as the problem of deciding whether or not a constituent is a valid candidate for a FE.

NO	TWFD	FEBD	GFLC	FEFC	Feature Description	NO	TWFD	FEBD	GFLC	FEFC	Feature Description
01	v				TW UNIGRAMS: The words, stem words and part of speech (POS) unigrams that are adjacent to target word expressions;	20				v	CW: The content word of the constituent computed as described in (Surdeanu et al., 2003);
02	v				TW BIGRAMS: The words, stem words and POS bigrams that are adjacent to target word expressions;	21				v	CW POS: The POS corresponding to the content word;
03	v				TW WORD: The target word expression;	22				v	CW STEM: Stemmed content word;
04		v	v		TW STEM: The stem word(s) of the target word expression;	23				v	GOVERNING CATEGORY: Test whether the noun phrase constituents are dominated by verbal phrases or sentence phrases;
05	v				TW POS: The POS of the target word;	24	v				SYNTACTIC DISTANCE: The length of the syntactic path;
06		v	v		TW CLASS: The lexical class of the target word, e.g. verb, noun, adjective;	25				v	PP FIRST WORD: If the constituent is a prepositional phrase, return the first word in the phrase;
07	v				NAMED ENTITY FLAGS: Set of binary features indicating whether a constituent contains, is contained or exactly identifies a named entity;	26				v	HUMAN: Test whether the constituent phrase is either a personal pronoun or a hyponym of first sense of PERSON synset in WordNet;
08	v				VERB WSD: If the target word is a verb, extract the head noun of the direct object and the prepositional object included in the verbal phrase;	27				v	CONSTITUENTS NUMBER: The number of candidate FEs;
09	v				NOUN WSD: If the target word is a noun, extract the head word of the verbal phrase that is in a verb-subject or verb-object relation with the noun;	28				v	CONSTITUENTS LIST: Constituents labels list of the candidate FEs;
10	v				ADJECTIVE WSD: If the target word is an adjective, extract the head noun that is modified by the adjective;	29				v	SAME CLAUSE: Test whether the constituent is in the same clause with the target word;
11		v	v		PHRASE TYPE: The syntactic category of the constituent;	30				v	GF: The grammatical function of a candidate frame element;
12		v	v	v	DIRECTED PATH: Path in the syntactic parse tree between the constituent and the target word preserving the movement direction;	31				v	GF LIST: The list of grammatical functions associated to the candidate FEs;
13			v		UNDIRECTED PATH: Same syntactic path as DIRECTED PATH without preserving the movement direction;	32	v	v		v	FRAME: The name of the semantic frame that is evoked by the target word;
14		v			PARTIAL PATH: Path from the constituent to the earlier common ancestor of the target word and the constituent;	33				v	NP SISTER: Determine whether the constituent has a noun phrase sister;
15		v	v	v	POSITION: Test whether the constituent contains the target word, or appears before or after the target word;	34				v	FIRST/LAST WORD: Return the first/last word of the constituent phrase;
16		v	v		VOICE: Test if the verbal target word has active or passive construction;	35	v			v	FIRST/LAST POS: Return the first/last POS in the constituent;
17		v	v		HW: The head word of the constituent;	36				v	LEFT/RIGHT SISTER LABEL: Return the left/right sibling constituent label;
18		v	v		HW POS: The syntactic head POS of the constituent;	37				v	LEFT/RIGHT SISTER HEAD: Return the left/right sibling head word;
19		v	v		HW STEM: The stem word of the constituent's head word;	38				v	LEFT/RIGHT SISTER STEM HEAD: Return the left/right sibling stemmed head word;
						39				v	LEFT/RIGHT SISTER POS HEAD: Return the left/right sibling head POS;
						40				v	TW STEM & HW STEM: Join of TW STEM and HW STEM;
						41				v	TW STEM & PHRASE TYPE: Join of TW STEM and PHRASE TYPE;
						42				v	VOICE & POSITION: Join of VOICE and POSITION.

Figure 3: Feature set for extracting frame semantic structures.

We consider a binary classifier over the entire FN data and extract features for each constituent from a syntactic parse tree. Because this experimental setup allows training the binary classifier on a large set of examples, the best feature combination consists of a restrained number of features. Most of these features are from the set proposed by (Gildea and Jurafsky, 2002). Another feature that improved the prediction of FE boundaries in every feature selection experiment is the FRAME feature. Since the frame disambiguation is executed before the FE boundary detection in the pipeline architecture, we can use the FRAME feature at this step. This feature helps the binary classifier distinguish between frame element structures from different semantic frames.

### 2.3 Grammatical Function Classification

Once we identify the candidate boundaries for frame elements, the next step is to assign the grammatical functions to these boundaries. In FrameNet, the grammatical functions represent the manner in which the frame elements satisfy grammatical constraints with respect to the target word.

For this task we train a multi-class classifier over the entire lexicon to predict seven categories of GFs that exist in FN. In addition, we assign the NULL category for those FEs that double as target words.

The features are extracted only for the constituents that are identified as FEs in the previous FE boundary identification sub-task. The best feature set in this phase includes the features proposed by (Gildea and Jurafsky, 2002) and the FRAME feature.

### 2.4 Frame Element Classification

The task of FE classification is to assign FE labels to every constituent identified as FE. In order to predict the frame elements, which are locally defined for each semantic frame, we built 489 multi-class classifiers, where each classifier corresponds to a frame in FrameNet. This partitioning of the FN lexicon has the advantage of increasing the overall classification performance and efficiently learning the frame elements labels. On the other hand, this approach suffers from the lack of annotated data in some frames and hence it requires using a large set of features.

The advantage of designing the classifiers in a pipeline architecture is best illustrated in this sub-task. Some of the most effective features for FE classification are extracted using information from previous sub-tasks: FRAME feature is made available by the TWFD sub-task, CONSTITUENTS NUMBER and CONSTITUENTS LIST are made available by the FEBD sub-task, and GF and GF LIST are made available by the GFLC sub-task.

