Automated Induction of Sense in Context

James PUSTEJOVSKY
Patrick HANKS
Anna RUMSHISKY

Brandeis University

jamesp,patrick,arum@cs.brandeis.edu

Abstract

In this paper, we introduce a model for sense assignment which relies on assigning senses to the contexts within which words appear, rather than to the words themselves. We argue that word senses as such are not directly encoded in the lexicon of the language. Rather, each word is associated with one or more stereotypical syntagmatic patterns, which we call *selection contexts*. Each selection context is associated with a meaning, which can be expressed in any of various formal or computational manifestations. We present a formalism for encoding contexts that help to determine the semantic contribution of a word in an utterance. Further, we develop a methodology through which such stereotypical contexts for words and phrases can be identified from very large corpora, and subsequently structured in a selection context dictionary, encoding both stereotypical syntactic and semantic information. We present some preliminary results.

Introduction

This paper describes a new model for the acquisition and exploitation of selectional preferences for predicates from natural language corpora. Our goal is to apply this model in order to construct a dictionary of normal *selection contexts* for natural language; that is, a computational lexical database of rich selectional contexts, associated with procedures for assigning interpretations on a probabilistic basis to less normal contexts. Such a semi-automatically developed resource promises to have applications for a number of NLP tasks, including word-sense disambiguation, selectional preference acquisition, as well as anaphora resolution and inference in specialized domains. We apply this methodology to a selected set of verbs, including a subset of the verbs in the Senseval 3 word sense discrimination task and report our initial results.

In this paper we modify the notion of word sense, and at the same time revise the manner in which senses are encoded. The notion of word sense that has been generally adopted in the literature is an artifact of several factors in the status quo, notably the availability of lexical resources such as machine-readable dictionaries, in which fine sense distinctions are not supported by criteria for selecting one sense rather than another, and WordNet, where *synset* groupings are taken as defining word sense distinctions. The feature sets used in the supervised WSD algorithms at best use only minimal information about the typing of arguments. The approach we adopt, *Corpus Pattern Analysis (CPA)* (Pustejovsky and Hanks, 2001), incorporates semantic features of the arguments of the target word. Semantic features are expressed in terms of a restricted set of shallow types, chosen for their prevalence in selection context patterns. This type system is extended with predicate-based noun clustering, in a bootstrapping process.

CPA Methodology

The Corpus Pattern Analysis (CPA) technique uses a semi-automatic bootstrapping process to produce a dictionary of selection contexts for predicates in a language. Word senses for verbs are distinguished through corpus-derived syntagmatic patterns mapped to Generative Lexicon Theory (Pustejovsky, 1995) as a linguistic model of interpretation, which guides and constrains the induction of senses from word distributional information. Each pattern is specified in terms of lexical sets for each argument, shallow semantic typing of these sets, and other syntagmatically relevant criteria (e.g., adverbials of manner, phrasal particles, genitives, negatives).

The procedure consists of three subtasks: (1) the manual discovery of selection context patterns for specific verbs; (2) the automatic recognition of instances of the identified patterns; and (3) automatic acquisition of patterns for unanalyzed cases. Initially, a number of patterns are manually formulated by a lexicographer through corpus pattern analysis of about 500 occurrences of each verb lemma. Next, for higher frequency verbs, the remaining corpus occurrences are scrutinized to see if any low-frequency patterns have been missed. The patterns are then translated into a feature matrix used for identifying the sense of unseen instances for a particular verb. In the remainder of this section, we describe these subtasks in more detail. The following sections explain the current status of the implementation of these tasks.

Lexical Discovery

Norms of usage are captured in what we call selection context patterns. For each lemma, contexts of usage are sorted into groups, and a stereotypical CPA pattern that captures the relevant semantic and syntactic features of the group is recorded. Many patterns have alternations, recorded in satellite CPA patterns. Alternations are linked to the main CPA pattern through the same sense-modifying mechanisms as those that allow for exploitations (coercions) of the norms of usage to be understood. For example, here is the set of patterns for the verb *treat*. Note that these patterns do not capture all possible uses, and other patterns may be added, e.g. if additional evidence is found in domain-specific corpora.

(1) CPA Pattern set for **treat**:

I. [[Person 1]] treat [[Person 2]] ({at | in} [[Location]])
(for [[Event = Injury | Ailment]]); NO [Adv[Manner]]

II. [[Person 1]] treat [[Person 2]] [Adv[Manner]]

IIIa. [[Person]] treat [[TopType 1]] {{as | like} [[TopType 2]]}

IIIb. [[Person]] treat [[TopType]] {{as if | as though | like}
[CLAUSE]}

IV. [[Person 1]] treat [[Person 2]] {to [[Event]]}

V. [[Person]] treat [[PhysObj | Stuff 1]] (with [[Stuff 2]])

There may be several patterns realizing a single sense of a verb, as in (IIIa/IIIb) above. Also, there may be several equivalent alternations or there may be a stereotype. Note that alternations are different realizations of the same norm, not exploitations (i.e., not coercions).

(2) Alternations for **treat** Pattern 1:

[[Person 1]] treat [[Person 2]] ({at | in} [[Location = Hospital]])
(for [[Event = Injury | Ailment]]); NO [Adv[Manner]]

Alternation 1:
[[Person 1 <--> Medicament | Med-Procedure | Institution]]

Alternation 2:
[[Person 2 <--> Injury | Ailment | Bodypart]]

CPA Patterns

A CPA pattern extends the traditional notion of selectional context to nclude a number of other contextual features, such as minor category parsing and subphrasal cues. Accurate identification of the semantically relevant aspects of a pattern is not an obvious and straightforward procedure, as has sometimes been assumed in the literature. For example, the presence or absence of an adverbial of manner in the third valency slot around a verb can dramatically alter the verb's meaning. Simple syntactic encoding of argument structure, for instance, is insufficient to discriminate between the two major senses of the verb *treat*.

(3) a. They say their bosses *treat* them with respect.
b. Such patients are *treated* with antibiotics.

The ability to recognize the shallow semantic type of a phrase in the context of a predicate is of course crucial – for example, in (3a) recognizing the PP as (a) an adverbial, and (b) an adverbial of manner, rather than an instrumental co-agent (as in (3b)), is crucial for assigning the correct sense to the verb *treat* above.

In the CPA model, automatic identification of selection contexts not only captures the argument structure of a predicate, but also more delicate features, which may have a profound effect on the semantic interpretation of a predicate in context. There are four constraint sets that contribute to the patterns for encoding selection contexts are listed below.

The notion of a selection context pattern, as produced by a human annotator, is expressed as a BNF specification in Table 1. This specification relies on word order to specify argument position, and is easily translated to a template with slots allocated for each argument. Within this grammar, semantic roles can be specified for each argument.

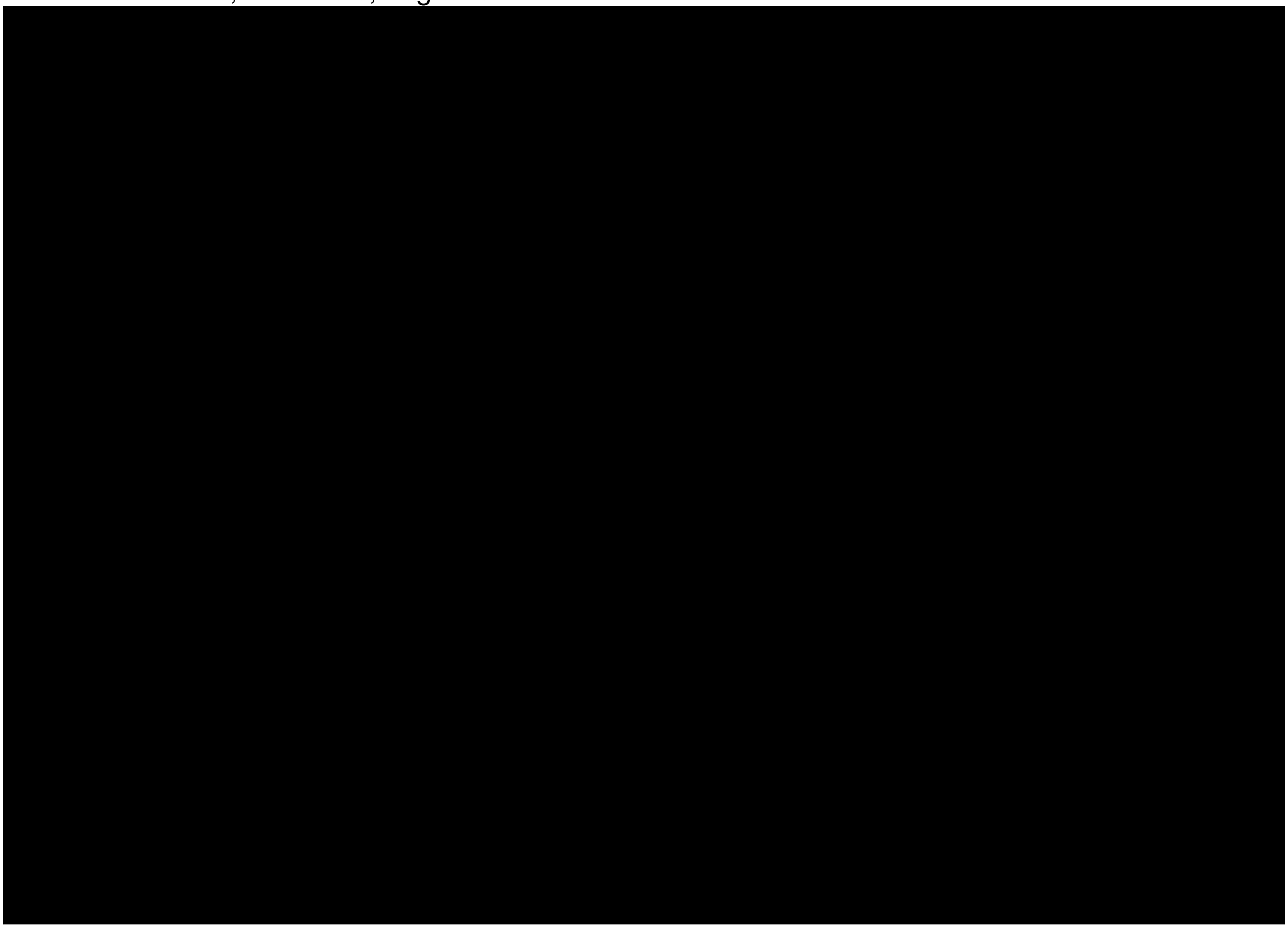
(4) a. *Shallow Syntactic Parsing*: Phrase-level recognition of major categories.

b. *Shallow Semantic Typing*: 50-100 primitive shallow types, such as Person, Institution, Event, Abstract, Artifact, Location, and so forth. These are the top types selected from the

Brandeis Shallow Ontology (BSO), and are similar to entities (and some relations) Named employed in Entity Recognition tasks, such as TREC and ACE.

c. *Minor Syntactic Category Parsing*: e.g., locatives, purpose clauses, rationale clauses, temporal adjuncts.

d. *Subphrasal Syntactic Cue Recognition*: e.g., genitives, partitives bare plural/determiner distinctions, infinitivals, negatives.



English contains only about 8,000 verbs, of which we estimate that about 30% have only one basic pattern. The rest are evenly split between verbs having 2-3 patterns and verbs having more than 4 or more patterns. About 20 light verbs have between 100 and 200 patterns each. This is less alarming than it sounds, because the majority of light verb patterns involve selection of just one specific nominal head, e.g., take account, take plunge, take photograph, with few if any alternations. The pattern sets for verbs of different frequency groups differ in terms of the number and type of features each pattern requires, the number of patterns in a set for a given verbs, the number of alternations for each pattern, and the type of selectional preferences affecting the verb's arguments.

Brandeis Shallow Ontology

The Brandeis Shallow Ontology (BSO) is a shallow hierarchy of types selected for their prevalence in manually identified selection context patterns. At the time of writing, there are just 65 types, in terms of which patterns for the first one hundred verbs have been analyzed. New types are added occasionally, but only when all possibilities of using existing types prove inadequate. Once the set of manually extracted patterns is sufficient, the type system will be re-populated and become pattern-driven.

The BSO type system allows multiple inheritance (e.g. Document PhysObj and Document Information. The types currently comprising the ontology are listed above. The BSO contains type assignments for 20,000 noun entries and 10,000 nominal collocation entries.

BSO Types

- TopType
  - Event
    - Action
      - SpeechAct
    - Activity
    - Process
    - State
  - Entity
    - PhysObj
      - Artifact
        - Machine
          - Vehicle
        - Hardware
        - Document (2)
        - Music (2)
        - Artwork (2)
        - Film (2)
        - Program (2)
        - Software (2)
        - Medium
      - Garment
      - Drug
      - Substance
        - Vapor
      - Animate
        - Bird
        - Horse
        - Person
          - HumanGroup (2)
      - Plant
        - PlantPart
    - Body
      - BodyPart

- Institution
    - HumanGroup (2)
  - Location
    - Dwelling
      - Accommodation
  - Energy
  - Abstract
    - Attitude
    - Emotion
    - Responsibility
    - Privilege
    - Rule
    - Information
      - Document (2)
      - Music (2)
      - Artwork (2)
      - Film (2)
      - Program (2)
      - Software (2)
    - Word
    - Language
    - Concept
    - Property
      - VisibleFeature
        - Color
        - Shape
    - TimePeriod
      - Holiday
    - CourseOfStudy
    - Cost
    - Asset
    - Route

The acquisition strategy for selectional preferences for predicates proceeds as follows:

(5) a. Partition the corpus occurrences of a predicate according to the selection contexts pattern grammar, distinguished by the four levels of constraints mentioned in (4). These are uninterpreted patterns for the predicate.
b. Within a given pattern, promote the statistically significant literal types from the corpus for each argument to the predicate. This induces an interpretation of the pattern, treating the promoted literal type as the specific binding of a shallow type from step (a) above.

c. Within a given pattern, coerce all lexical heads in the same shallow type for an argument, into the promoted literal type, assigned in (b) above. This is a coercion of a lexical head to the interpretation of the promoted literal type induced from step (b) above.

In a sense, (5a) can be seen as a broad multi-level partitioning of the selectional behavior for a predicate according to a richer set of syntactic and semantic discriminants. Step (5b) can be seen as capturing the norms of usage in the corpus, while step (5c) is a way of modeling the exploitation of these norms in the language (through coercion, metonymy, and other generative operations). To illustrate the way in which CPA discriminates uninterpreted patterns from the corpus, we return to the verb *treat* as it is used in the BNC. Although there are three basic senses for this verb, the two major senses, as illustrated in (1) above, emerge as correlated with two distinct context patterns, using the discriminant constraints mentioned in (4) above.

(6) a. [[Person 1]] treat [[Person 2]]; NO [Adv[Manner]]
b. [[Person 1]] treat [[Person 2]] [Adv[Manner]]

Given a distinct contextual basis for the analysis of the actual statistical distribution of the words in each argument position, we can promote statistically significant literal types for these positions. E.g., for pattern (a) above, this induces *Doctor* as Person 1, and *Patient* as bound to Person 2. This produces the interpreted context pattern for this sense:

[[doctor]] treat [[patient]]

Promoted literal types are corpus-derived and predicate-dependent, and are syntactic heads of phrases that occur with the greatest frequency in argument positions for a given sense pattern; they are subsequently assumed to be subtypes of the particular shallow type in the pattern. Step (5c) above then enables us to bind the other lexical heads in these positions as *coerced* forms of the promoted literal type. This can be seen below in the concordance sample, where *therapies* is interpreted as *Doctor*, and *people* and *girl* are interpreted as *Patient*.

(8) a. returned with a doctor who treated the girl till an ambulance arrived.
b. more than 90,000 people have been treated for cholera since the epidemic began
c. nonsurgical therapies to treat the breast cancer, which may involve

Model Bias

The assumption within GL is that semantic types in the grammar map systematically to default syntactic templates (cf. Pustejovsky, 1995). These are termed *canonical syntactic forms (CSFs)*. For example, the CSF for the type proposition is a tensed S. But there are many possible realizations (such as infinitival S and NP) for this type due to the different possibilities available from generative devices in a grammar, such as coercion and co-composition. The resulting set of syntactic forms associated with a particular semantic type is called a *phrasal paradigm* for that type. The model bias provided by GL acts to guide the interpretation of purely statistically based measures.

Automatic Recognition of Pattern Use

Essentially, this subtask is similar to the traditional supervised WSD problem. Its purpose is (1) to test the discriminatory power of CPA-derived feature-set, (2) to extend and refine the inventory of features captured by the CPA patterns, and (3) to allow for predicate-based argument groupings by classifying unseen instances. Extension and refinement of the inventory of features should involve feature induction, but at the moment this part has not been implemented. During the lexical discovery stage, lexical sets that fill some of the argument slots in the patterns are instantiated from the training examples. As more predicate-based lexical sets within shallow types are explored, the data will permit identification of the types of features that unite elements in lexical sets.

Automatic Pattern Acquisition

The algorithm involves the following steps:

(9) a. Collect all constituents in a particular argument position;
b. Identify syntactic alternations;
c. Perform clustering on all nouns that occur in a particular argument position of a given predicate;
d. For each cluster, measure its relatedness to the known lexical sets, obtained previously during the lexical discovery stage and extended through WSD of unseen instances. If none of the existing lexical sets pass the distance threshold, establish the cluster as a new lexical set, to be used in future pattern specification.

Step (9d) must include extensive filtering procedures to check for shared semantic features, looking for commonality between the members. That is, there must be some threshold overlap between subgroups of the candidate lexical set and the existing semantic classes. For instance, checking if, for a certain percentage of pairs in the candidate set, there already exists a set of which both elements are members.

Current Implementation

The CPA patterns are developed using the British National Corpus (BNC). The sorted instances are used as a training set for the supervised disambiguation. For the disambiguation task, each pattern is translated into a set of preprocessing-specific features.

The BNC is preprocessed using the Robust Accurate Statistical Parsing system (RASP) and semantically tagged with BSO types. The RASP system (Briscoe & Carroll, 2002) tokenizes, POStags, and lemmatizes text, generating a forest of full parse trees for each sentence and associating a probability with each parse. For each parse, RASP produces a set of grammatical relations, specifying the relation type, the headword, and the dependent element. All our computations are performed over the single top-ranked tree for the sentences where a full parse was successfully obtained.

We use endocentric semantic typing, i.e., the headword of each constituent is used to establish its semantic type. The semantic tagging strategy is similar to the one described in Pustejovsky et al. (2002). Currently, a subset of 24 BSO types is used for semantic tagging.

A CPA pattern is translated into a feature set, which in the current implementation uses binary features. It is further complemented with other discriminant context features which, rather than distinguishing a particular pattern, are merely *likely* to occur with a given subset of patterns; that is, the features that only partially determine or co-determine a sense. In the future, these should be learned from the training set through feature induction from the training sample, but at the moment, they are added manually. The resulting feature matrix for each pattern contains features such as those in (11) below. Each pattern is translated into a template of 15-25 features.

(11) **Selected context features**:

a. **obj_institution**: object belongs to the BSO type 'Institution'
b. **subj_human_group**: subject belongs to the BSO type 'HumanGroup'
c. **mod_adv_ly**: target verb has an adverbial modifier,
with a -ly adverb
d. **clausal_like**: target verb has a clausal argument
introduced by 'like'
e. **iobj_with**: target verb has an indirect object by 'with'
f. **obj_PRP**: direct object is a personal pronoun
g. **stem_VVG**: the target verb stem is an -ing form

Each feature may be realized by a number of RASP relations. For instance, a feature dealing with objects would take into account RASP relations 'dobj', 'obj2', and 'ncsubj' (for passives). The features such as (11a)-(11e) are typically taken directly from the pattern specification, while features such as in (11f) and (11g) would typically be added as co-determining the pattern.

Results and Discussion

The experimental trials performed to date are too preliminary to validate the methodology outlined above in general terms for the WSD task. Our results are encouraging however, and comparable to the best performing systems reported from Senseval 2. For our experiments, we implemented two machine learning algorithms, instance-based k-Nearest Neighbor, and a decision tree algorithm. For these experiments, kNN was run with the full training set. Table 2 shows the results on a subset of verbs that have been processed, also listing the number of patterns in the pattern set for each of the verbs.

[redacted]

Further experimentation is obviously needed to gauge the effectiveness of the selection context approach for WSD and other NLP tasks. But it is clear that the traditional sense enumeration approach, where senses are associated with individual lexical items, must give way to a model where senses are assigned to the contexts within which words appear. Furthermore, because the variability of the stereotypical syntagmatic patterns that are associated with words appears to be relatively small, such information can be encoded as lexically-indexed contexts. A comprehensive dictionary of such contexts could prove to be a powerful tool for a variety of NLP tasks.

References

T. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*.

J. Pustejovsky, A. Rumshisky, and J. Castano. 2002. Rerendering Semantic Ontologies: Automatic Extensions to UMLS through Corpus Analytics. *LREC 2002 Workshop on Ontologies and Lexical Knowledge Bases*.

J. Pustejovsky. 1995. Generative Lexicon. Cambridge (Mass.): MIT Press.

Appendix A. Sample CPA Patterns

Verb: toast, CPA Patterns

PATTERN 1: [[Person]] toast [[Food]]

Subj: [[Person]] ; Obj: [[Food]]

4 generous slices bread Preheat grill and toast bread lightly on both sides. Cut cheese into
an array of dainty cakes and pastries and hot toasted teacakes in silver dishes complete with lids.
a plate containing four small triangles of toasted wholemeal bread, sparsely buttered. Now eat
a Britain in which everyone wore a vest, ate toasted crumpets for tea and never went out without a
cheese and a jar of ready-made tomato sauce. Toast the muffin and put a tablespoon of the sauce
that when she returns from the kitchen with a toasted sandwich and her man barks # I said cheese
where you can buy snacks such as pizzas and toasted sandwiches. On the terrace there is
breakfast. She wanted a pint of recaff and a toasted cheese sandwich. Perhaps a bowlful of Wally
and visitors. It was always a high-tea with toasted cheese, homemade bread and butter, and bara
Add the yogurt. Garnish the chicken with toasted almonds. Serve with brown rice. DAY 6 #
with two parts brown or white, add chopped toasted hazelnuts or pine nuts, dried apricots and
in small cylindrical shapes and rolled in toasted oatmeal to give it a more distinctive

herbs freshly ground pepper and salt lightly toasted pine nuts or flaked almonds Cut the courgette
green and purple, roasted sesame seeds, toasted sunflower seeds, walnut oil and red wine

EXPLOITATIONS/COERCIONS:

Just a few more minutes of this blissful heat toasting her limbs, then she ' d rouse herself and
by day you need to do no more than gently toast your body on sandy beaches or explore the
only comfortable place was the kitchen; they toasted their raw hands and feet by the fire and the

UNCLASSIFIED:
was poured into glasses and the meal was toasted. This is America at home # Moran boasted.

PATTERN 2: [[Person 1]] toast [[Person 2]]
Subj: [[Person 1]] ; Obj: [[Person 2]]

Oliver made a speech. He said he wanted to toast a bridesmaid but there were n' t any around so
at one of his favourite Indian restaurants, toasting each other ' s health and futures in an
the whole party is gathered round a table, toasting each other noisily in champagne. The loudest
knowledge to himself. He raised his glass, toasting his host and hostess silently, his smile
the Isle of Man and Chris and family will be toasting Keith over a meal. But the man of the
the other holding a wine cup. He grinned and toasted me silently. I glared back at the bastard.
iced champagne was served for the guests to toast the emperor and the New Year. Jewelled
handsome in their grey morning coats. We toasted the couple with champagne, sampled jean-paul

ALTERNATIONS/COERCIONS:
[[Person 2]] <> {[[Event]] | memory | health | success | achievement}

Charlie JACKPOT pools winner Charlie Hill toasted his 2 million pound windfall yesterday and
duty. They had a pub lunch on his last day to toast his early retirement. The sacking cost Alex a
negotiations throughout the night. Workers toasted news of the deal with free beer at a shop
his funeral reception, members of his family toasted his memory in peach wine. Mr Hutton of
but a tooth-glass of tap water, in which she toasted my health. Byron Rogers, writing eloquently
but Morland ' s Original was being used to toast the brewer ' s successful bid defence. Greene

EXPLOITATIONS/COERCIONS:

up in a cosy bar. Soon we were once again toasting our good sense at booking Christmas week
wine, and the Declaration of Independence was toasted in it. To this day there are many collectors
meaning # railway station # So, we toast the railway station and I tell him the only
The ultimate hall of mirrors. We toast the Sears Tower, the tallest building in the

Verb: watch, CPA Patterns (fragment)

PATTERN 1: [[Person]] watch [[Event]]

REALIZATIONS for Event:

1: [N[Event]]

Other areas will probably watch the outcome of this experiment with interest.

2: [[Event]] = {[[Person]] INF/V}

two armless Lilliputian queens preside, watching a giant bathe.

3: [[Event]] = {[[Person]] ING/V}

sat on the shore and watched people splashing about in the water.

ALTERNATIONS/COERCIONS:
[[Event]] <--> [[Person]] | [[PhysObj]] (participating in Event)

He watched the ex-soldier till he was gone
Emily, following the direction of my gaze, watched the light
She could feel his eyes watching her. Then he smiled.