Developing a Japanese Corpus Annotated for Semantic Roles (JCASR)*

—Introducing a NICT Project—

Kow Kuroda
kuroda@nict.go.jp

1 What is the JCASR Project?

Development of a Japanese Corpus Annotated for Semantic Roles (JCASR) is being attempted as one of the research projects at National Institute of Information and Communications Technology (NICT), Japan. Its goal is to develop a (relatively small) corpus of Japanese texts annotated for semantic roles comprising (semantic) frames, adopting the insights from Berkeley FrameNet project [2, 7].

1.1 Members

A team of four, Kow Kuroda (head), Jae-Ho Lee, Hajime Nozawa, and Yoshikata Shibuya (all at NICT, Japan), in collaboration with Keiko Nakamoto (Bunkyo University, Japan) are working for this project now. We are working with graduate students at Kyoto University as “external” annotators.

Note that we are working independently of Japanese FrameNet project [24, 23], the official Japanese agency for Berkeley FrameNet.

1.2 Status of the Project

The JCASR project officially began two years ago. It is (still) at a preliminary, “exploratory” stage, in that we are trying to see what kinds of frames are needed at what granularity levels, without assuming a pre-existing, “ready-to-use” database of semantic frames and frame elements. Serious development of a semantically tagged corpus has not started yet, but annotation samples are available freely or privately at web sites (contact me for more details).

Some preliminary results are reported in English in [16, 12] (there are a bunch of works written in Japanese).

Tentatively, procedures to identify (a) frames for event conceptualizations (e.g., ROBBERY, PREDATION) and (b) frames for social interactions (e.g., speech acts like CLAIMING, CRITICIZING, DOUBTING, PROTESTING, WARNING) are separated. This is mainly because the latter class of frames is more complex, more selective for data, and harder to specify. Currently, Kuroda, Lee and Shibuya work for the former class; and Nozawa for the latter class.

1.3 Motivations and Goals

Needs for semantic processing have become more and more demanding. But we (still) lack resources that can be used for this purpose. Why is this so?

The reason would be that some fundamental questions remain unanswered. The most serious problem, I presume, is that it is not clear what people understand when they hear or read a sentence, let alone a text, i.e., a collection of sentences. Actually, there is little agreement what people’s understanding is and how it should be represented. This clearly has slowed, if not impeded, the progress of theories for semantic annotation/analysis. So, something needs to be done if we want to go further, even if it might look risky — research into anything interesting.

*I’m grateful to helpful comments and corrections to earlier drafts of this article by Yoshikata Shibuya (NICT) and to comments and suggestions by Keiko Nakamoto (Bunkyo University).
is always risky, isn’t it?

The goal is to establish a set of (ontological) links from “pieces of world knowledge” to text segments in terms of **semantic role tagging** in the sense specified below.

### 1.4 Development Cycle

Currently, we are following the “incremental” development like the following:

1. Select a Japanese text $T$ from a text database.
2. Segment each sentence of $T$ into text segments by the staff at NICT. Every result of segmentation always needs to be checked, because the standard outputs of a so-called “morphological analyzer” like “KNP” and “ChaSen” are sometimes inappropriate for our purposes.
3. Ask “external” annotators to annotate the segmented texts by making reference to databases $D_1$ and $D_2$ of “sample annotations” hosted at web sites, both public and private.
4. Collect the annotations by annotators as “drafts,” check and edit the results if necessary (which is very often the case) by the staff at NICT.
5. Add the edited results to the databases $D_1$ and $D_2$.
6. “Sanitize” the databases when needed.

$T$ is always chosen from Japanese texts aligned with English texts, expecting that future comparisons against other annotations (using Berkeley FrameNet database, for example) can be facilitated.

So far, all texts have been taken from the following text bases:

1. **English-Japanese Translation Alignment Data** (a collection of Japanese-English alignments of copyright-free texts like *Fables* by Aesope)
   
   http://www2.nict.go.jp/x/x161/members/mutiyama/align/index.html

2. **Japanese-English Newspaper Article Alignment Data** (JENAAD) [31]
   
   http://www2.nict.go.jp/x/x161/members/mutiyama/jea/index.html

3. **Kyoto University Corpus**

### 1.5 Statistics

Table 1 shows some statistics of the current semantic role tagging.

Target texts for $D_1$ are chosen from (2a) and (2b), which mainly consists of proses. Target texts for $D_2$ are chosen from (2c), which consists of newspaper articles. $D_1$ and $D_2$ are hosted at different web sites, with different availabilities.

- $D_1$ is hosted at:
  
  http://www.kotonoba.net/~mutiyama/cgi-bin/hiki/hiki.cgi?FrontPage

  without access restriction.

- $D_2$ is hosted at:
  
  http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?FrontPage

  with access restriction (user account is required)

### 2 Outline: What to Annotate, and How?

What I call **semantic role tagging** is a special case of **semantic tagging**. Any tagging is a semantic tagging if it annotates pieces of a text with **semantic tags**.

What tags are **semantic** tags, however? There is no straightforward answer to what they are: unlike **part-of-speech tags** (POS tags) like “N,” “NP,” “V,” “VP,” there is no generally agreed, general purpose scheme for semantic tags, but let me give you the basic idea, by taking simple examples like the following:

(3) A group of masked men attacked a bank branch in New York yesterday.

First, you segment a sentence $S$ (of a text $T$) into a set of **text segments** such as \{ a group of masked men, attacked, ... \}. Then, you choose an appropriate **semantic label or marker** (i.e., a “semantic tag”) for each of those segments. Labels of this kind are sometimes referred to as “sense tags.”
<table>
<thead>
<tr>
<th></th>
<th>(D_1) (open)</th>
<th>(D_2) (semi-closed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. of sentences</td>
<td>67</td>
<td>64</td>
</tr>
<tr>
<td>N. of text segments (token)</td>
<td>1474</td>
<td>1719</td>
</tr>
<tr>
<td>N. of text segments (type)</td>
<td>442</td>
<td>539</td>
</tr>
<tr>
<td>Freq. (average)</td>
<td>2.19</td>
<td>2.26</td>
</tr>
<tr>
<td>N. of frames (token)</td>
<td>927</td>
<td>1227</td>
</tr>
<tr>
<td>N. of frames (hapax)</td>
<td>686</td>
<td>990</td>
</tr>
<tr>
<td>Hapax ratio</td>
<td>74%</td>
<td>80%</td>
</tr>
<tr>
<td>N. of frame elements (token)</td>
<td>3031</td>
<td>3815</td>
</tr>
<tr>
<td>N. of frame elements (hapax)</td>
<td>2393</td>
<td>3149</td>
</tr>
<tr>
<td>Hapax ratio</td>
<td>78%</td>
<td>83%</td>
</tr>
</tbody>
</table>

### 2.1 Nature of the Task/Problem

What makes our task/problem very complicated (and challenging) is the fact that there is no guarantee that we have a single appropriate label.marker for any of text segments. This means that we need to deal with the inherent multidimensionality in semantic labeling/annotation.

For illustration, consider the correspondence matrix in Figure 1, where the correspondence of multiple semantic analyses, \(L_0, L_1, L_2, L_3, L_4\), against a target text \(T\) is specified.

#### 2.1.1 Labels for \(L_0\)

There is a level of semantic specification on which text segments are assigned labels like \(\{\text{HUMAN, ACT, INSTITUTION, PLACE, TIME, \ldots}\}\). The correspondence between the elements of \(T\) and those of \(L_1\) is probably what comes to your mind when you hear semantic annotation. But the specification of correspondence between \(L_1\) and \(T\) is not what we mean by semantic role annotation. This is what we call semantic type annotation/analysis. Relevant details on this layer will be briefly discussed in Section 3.1.

#### 2.1.2 Labels for \(L_1, \ldots, L_4\)

By semantic role annotation/analysis, we mean multi-level specifications of correspondences between \(T\) and \(L_1, T\) and \(L_2, T\) and \(L_3\). For this, we do not assume, or rather avoid assuming, that there is a single level \(L_i\) from which every other level \(L_j\) \((j \neq i)\) is “derived,” which many theories for semantic/pragmatic analysis tend to do without any guarantee.

#### 2.1.3 Defining the Annotation/Analysis Procedure

Under this setting, the goal of the semantic role annotation/analysis is the following:

(4) Procedure of semantic role annotation (informal): Given a sentence \(s\) segmented into segments \(W = \{w_1, \ldots, w_n\}\), to identify and specify

a. “situations” (specified in terms of “frames” in the sense of Frame Semantics and FrameNet) “evoked” by specific segments in \(W\), and

b. “semantic roles” (or “frame elements” in the sense of FrameNet) that comprise the situations identified this way.

In what follows, I want to provide some background to this approach.

### 2.2 Relation to Frame Semantics and Berkeley FrameNet

Building on the insight of Fillmore’s Frame Semantics [4, 5, 6], Berkeley FrameNet approach to semantic annotation [11, 17] (and also M. Minsky’s theory of
2.2.1 Remark 1

There are two somewhat different senses of the term “semantic role annotation.” The first one has a broader sense, in that it refers to any semantic annotation in which semantic roles are specified. In this broader sense, specifying labels at $L_1$, $L_2$, $L_3$ or $L_4$ are all semantic role annotations. The second one has a narrower sense, in that it refers to annotation of concrete semantic roles comprising concrete situations specified by labels at $L_1$ and $L_2$.

Abstract roles at $L_3$ or $L_4$ can be identified with “deep cases” in Fillmore’s Case Grammar [3] in 70’s and “thematic roles” widely exploited in linguistic analysis in 80’s and 90’s. The usefulness of such labels is limited, however: they are too general a specification, and just like semantic types, they are ineffective to link text segments to our world knowledge, against which people understand an overall text.

2.2.2 Remark 2

A strong emphasis is placed on the description of the frame/situation evocation by nouns. This is related to the first remark.

Previous research has revealed that certain nouns (like robber(s), victim(s), scene of a crime, doctor, patient, medicine, hospital) are not just “names for things” but “names for situation-specific (semantic) roles” that evoke situations/frames without help of explicit “governors” (i.e., namers) of frames/situations.

This has something common with the theory of “relational nouns” proposed by Gentner and her colleagues [1, 8, 9]: “semantic roles” referred to as “semantic role names” in our terms can be equated with “relational role categories” referred to as “relational role nouns” in Gentner’s theory, and “situations” or “frames” in our terms with “relational schema categories” in Gentner’s theory. For relevant details, see [15].

Interestingly, role names and objet names seem to have different potentials for metaphorical uses. Other things being equal, role names are more ready for metaphor,
whereas object names are more ready for simile. This was confirmed by psychological experiments on Japanese nouns by [21].

Taking these things into consideration, it would be useful to map out semantic roles to senses/concepts in an appropriate thesaurus. This would increase the usefulness of a thesaurus when it is used as a “substitute” for an ontology.

### 2.2.3 Remark 3

A hierarchical, “instantiational” relationship can be defined among $L_1$, $L_2$, $L_3$, $L_4$, in the following way:

$$L_1 \text{ is-a } L_2 \text{ is-a } L_3 \text{ is-a } L_4$$

This results in so-called “inheritance hierarchies.” It predicts “event hierarchy” like (6) and “role hierarchies” like (7a, b):

$$\text{(6)} \quad \text{ROBBERY is-a HARM-CAUSING ACTIVITY is-a ACTIVITY is-a EVENT}$$

$$\text{(7)} \quad \text{a. ROBBER is-a HARM-CAUSER is-a ACTOR is-a PARTICIPANT}$$

$$\text{b. [OWNER part-of STORE OF VALUABLES] is-a VICTIM is-a PATIENT is-a PARTICIPANT}$$

A metonymic adjustment takes place to give (7b).

### 2.2.4 What makes your understandings “better” understandings

Note, incidentally, that the granularity levels of those hierarchies need to be accommodated; otherwise, event and role hierarchies alone would make a lot of “wrong” predictions, because it allows us to “conceive” such wrong role sets as *PREDATOR, STORE OF VALUABLES, PLACE OF HUNTING, . . .*, *ROBBER, PREY, SCENE OF CRIME, . . .*.

In our approach, a strong emphasis is given to the identification and specification of “finer-grained,” “concrete,” “situation-specific” roles at levels $L_1$ and $L_2$, rather than “coarse-grained,” “abstract,” “general-purpose” roles at $L_3$ or $L_4$.

Why? We do this because we hypothesize that **better understandings are achieved at more concrete levels, rather than at more abstract levels**. This is one of the points that make our approach different from other (usually more “formally oriented”) approaches to semantic annotation/analysis which tend to assume that the deepest semantic analysis is the most abstract semantic analysis.

More formally, we assume the following:

$$\text{(8) Concreteness Bias on Semantic Interpretation (Hypothesis):}$$

the more “specific” and “concrete” your understanding is, the better it is (as long as it is not obviously wrong).

This is the hypothesis that motivates very concrete specifications like ROBBERY, PREDATION at $L_1$.

The principle stated in (8) clearly favors “overinterpretations” over “underinterpretations,” other things being equal. We are aware that this is a controversial hypothesis and will invites challenges, but it is an interesting hypothesis that deserves an exploration. No matter how controversial, this hypothesis has a clear merit: it would explain **why people makes guess, even risking misunderstandings**. This is an interesting property of human understanding that deserves a dedicated explanation.

We have a good motivation for the hypothesis. In our view, the “deepest” analysis, if any, is the most detailed analysis, acknowledging that what makes human mind alive is not its power to do abstract reasoning, but its power to counterbalance powerful reasoning by general rules and principles with messy details of the world which cannot be predicted by general principles. For this reason, human understanding needs to be “adaptive,” rather than just powerful.

“Better” understanding means “more adaptive” understanding, at least in actual life. Precise, presumption-free understandings are not always adaptive, simply because the world is essentially uncertain. This makes performers of good guesses more adaptive agents. At least, “adaptive thinking,” in the sense of Gigerenzer [10], is not expected to be error-free.
Hierarchical Frame Network (inside the Database)

F9: <Commiting a Crime>
F7b: <Losing>
F7a: <Gaining>

Indicates that morpheme M corresponds to role R.

Indicates that a role R elaborates a role R* at more abstract level.

Figure 2: SFNA of (3). Blue arrows from text segments to roles or frames indicate “lexical realization” relations, including “evocation” relations (Difference in thickness indicates difference in “strength” of evocation); Black arrows indicate “is-a” relations. Pink arrows between frames indicate “frame-to-frame” relations, some of them (e.g., “parallel”) are bidirectional.
2.2.5 Representing the activation pattern of situations with SFNA

What happens (in our brain) to the entire network of situations/frames when some of them are evoked by (combinations of) lexical items (called “lexical units” in Berkeley FrameNet) and activated by inheritances after interpretation? To illustrate this, I give the hierarchical network of situations/frames evoked or activated during the interpretation of (3) in Figure 2. This structure, called Semantic Frame Network Analysis (SFNA) of (3), is selection of situations over the entire lattice of situations (presumably stored in the brain). Diagrams like this one should tell us more about the interaction among pieces of semantic/pragmatic encodings of (3) at different layers in Figure 1. The semantic specifications at \( L_1, L_2, L_3, \) and \( L_4 \) in Figure 1 correspond to situations F4, F13, F15, and F19 in Figure 2, respectively, which are distinguished by different base color.

We posit more kinds of frame-to-frame relations (e.g., “realizes” relation, “motivates” relation, “facilitates” relation, many of them characterize causal, conditional, or logical relations) than Berkeley FrameNet, simply because it turns out that we needed them in effective semantic annotation/analysis.

It needs to be mentioned that SNFA does not assume deep syntactic parsing. We presume that surface-true, string-based “pattern matchings” will suffice to associate text segments with semantic roles, though this idea is not implemented yet. (Parallel) Pattern Matching Analysis (PMA) proposed in [13, 14] would help in implementing this idea.

2.2.6 Dealing with selectional restrictions

An important research question to this hypothesis is **if there are lower limits on “concreteness” of understanding**. We admit that this is an open question, and a dedicated research to it is reported in [22]. One important heuristics that we came up with after the research is that (i) “selectional restrictions” reflect event conceptualizations/classifications at lower levels, rather than higher levels, and (ii) you can specify as many lower-level, concrete situations as you need, as long as selectional restrictions can be specified in a realistic way, even if there are no ultimate, lowermost levels of conceptualization.

2.3 Managing “depth” of readings

Some may wonder if interpreting (3) as referring to a bank robbery is not an “overinterpretation.” The answer is both yes and no.

Most people interpret in different modes. When they are careful, they refrain from overinterpretations, seemingly prefer “underinterpretations.” But this is true only when they are in a “cautious” and “careful” mode: they are not so in a “normal” mode. Most people seem to prefer overinterpretations in a normal mode.

By normal, I mean that they are not unaware of obvious “penalties” on misunderstandings. When they are made aware of them, they become careful and try to avoid overinterpretations, being afraid of penalties. The careful mode would be more compatible with truths, but this does not reflect what people do under usual circumstances. First of all, overinterpretation is not always a bad thing. Human tendency for overinterpretation looks even “adaptive” in usual circumstances where we are encouraged to look ahead. Actually, overinterpretation seems rather “harmless” as far as it is cancelled easily.

This suggests that people can deal with the “depth” of their interpretations: they just pick up an interpretation at the most appropriate granularity/confidence level out of several “candidates” at various granularity/confidence levels, depending on external conditions on their interpretations.

The problem is, of course, how to define a set of those “candidates”? Inspired by the FrameNet approach, we hypothesize that a certain “hierarchy of situations” can define a set of such candidates. For the cases like (3), the hierarchy of (Harm-causation) events/situations, such as illustrated in Figure 3, called “hierarchical frame network analysis” (HFNA), seems to define the set of candidate interpretations. (Note: the HFNA in Figure 3 was constructed to account for the range of interpretations for Japanese sentences in which \( osou \) (meaning *attack, assault, hit in English) is used as the main verb, whether in active or passive. So, it can be the case that it does not properly characterize the interpretational range of English sentences in which *attack, assault and hit are used as the main verb. This needs to be said as a caveat).

Interpretations at multiple granularity/confidence levels can be attributed to appropriate “nodes” of HFNA in
a. The most abstract situation/frame against which *attack-* and *hit-*sentences are interpreted is at the “top” of the lattice of situations/frames in Figure 3. In other words, this node is the “root” of the situation/frame hierarchy.

b. The most concrete situations are at “leaves” of the lattice marked by thick profiles (the “bottom” of the lattice is not indicated).

c. The root node corresponds to the semantic specification at Layer/Level $L_2$ in Table 1. All other nodes in this HFNA are candidates for the specification at $L_1$. In other words, there are many “intermediate” levels for semantic specification between $L_1$ and $L_2$. This is exactly what we need to deal with the ramification of semantic interpretations.

d. When a “greedy” interpretation is attempted, (3) is interpreted against F03b: *(Bank Robbery)*. This is likely to be an overinterpretation. Note also, however, that even a greedy interpretation of (3) does not match F03a: *(Personal Robbery)*, which characterizes a personal scale harm-causing activity.

e. When a more “modest” interpretation is attempted, (3) is interpreted against B1: *(Victimization of Human by Human, Crime 2)*, which licenses situations of G: *(Power Conflict)*.

2.3.1 What underspecification means to interpretation

An attempt to make interpretations “more modest” and “less greedy” has the same effect as using (semantic) underspecification. It is equivalent to going a few steps up the lattice towards the root.

2.3.2 Filtering out many “inappropriate” interpretations

Most importantly, however, adequate interpretations of (3) need to be within the “domain” of B1: *(Victimization of Human by Human, Crime 2)*, all in oranges, in that all attempts to take interpretations “outside this domain” fail or force metaphoric or metonymic “adjustments” on the meanings of some lexical items of the sentence. It is possible to interpret (3) to mean, or “allude to,” a situation of F12b: *(Social Disaster on Smaller Scale)* but this forces a *group of masked men* to be interpreted as a nickname for a *(Down Turn)*, an expected *(Red Figures)*, or a similar kind of *(Accident)*.

This is another kind of greedy interpretative process in which lexical meaning of a *group of masked men* is “sacrificed” over the interpretive selection of F12b, which is very likely to be an overinterpretation for (3).

3 Benefits of Semantic Role Annotation

We expect that semantic role annotation along the proposed line would make a good resource of “lexically based” inferences. In what follows, let me specify very briefly why this is the case.

3.1 Limits of semantic type annotation

The most common way of annotating text segments with semantic tags is to use labels like HUMAN, THING, i.e., semantic types differentiated from semantic roles. Why is this common? It is probably because (i) it is relatively easy, in that the tagset seems to be closed (this is important indeed); and (ii) the obtained results are relatively stable and reliable, and easy to validate.

But we need to go beyond mere reliability if we want to reach people’s actual understanding of texts.

3.1.1 Dealing with guesses and “lexically based” inferences

Actually, specifying a *group of masked men* and a *bank branch* in (3) as HUMAN and INSTITUTION will not make you well-informed. For one, it does not tell you what (people understand) happened.

It should be noted that people do not avoid making “guesses” when they (try to) understand, and most guesses they make are very good ones.

What guesses do people (tend to) make for (3), for example? You can say that an average reader/hearer of (3)
Figure 3: A “lattice” of the situations against which attack- and hit-sentences are interpreted. Black arrows indicate “is-a” relations; Green arrows from sentences to situations indicate “is-interpreted-against” relations.
would presume the following unless they are “overridden” by explicit lexical specifications:

(10) a. “a group of masked men” are ROBBERs,
    b. “the bank branch” refers to a STORE OF VALUABLES (e.g., “money,” or valuable things like “jewels”),
    c. and the ROBBERs used certain WEAPONs (like “guns”, “army knives,” or even “bombs”) for THREATENING, to achieve their PURPOSEs of ROBBERY.
    d. The reason the group of men “masked” themselves was to HIDE their IDENTITIES.
    e. The reason the ROBBERs made “a group” was to FORM A TEAM to PERFORM BETTER in COLLABORATION.

Some of these are explicitly encoded in the diagram in Figure 2.

In (11), the value for WEAPON for ROBBERY is overridden by explicit lexical specification with molotov cocktails, and the evocation to ROBBERY is “cancelled” in the following case:

(11) A group of masked men attacked a bank branch in New York with molotov cocktails yesterday.

Indeed, the situation evoked in (11) is not the same as the one evoked in (3): molotov cocktails evokes a different situation of POWER CONFLICT, where EXTREMISTS (is-a ANTI-SOCIALISTS) used them as WEAPON, though somewhat in an extended, metaphorical sense.

Unlike for (3), the interpretation for (11) can hardly fall outside G: (Power Conflict between Human Groups). The reason why FO1: (Combat between Human Groups) and FO2: (Military Invasion) are dispreferred is probably that the conflict under question is not a territorial conflict but a power conflict.

In this case, the semantic role assigned to a bank branch is not STORE OF VALUABLES, but it is just EXAMPLE OF WARNING. It should be noted, however, that a kind of “presupposition preservation” takes place: both STORE OF VALUABLES and EXAMPLE OF WARNING are special cases of VICTIM.

Again, people make guesses and “adjustments” like these, and they are very good at doing it. So, it wouldn’t be an exaggeration to say that (good) guesses are part of human understanding (I personally think this is rather adaptive: linguistic communication will be very ineffective if people are disallowed to make guesses, and are forced to stick to “facts,” “truths,” or “what is really said”). For this very specific reason, we can say that people’s understanding is biased for something beyond truths. This is an aspect that semantic type specification cannot deal with.

If this is true, it implies that semantic type labeling (done at L0) will not be so useful unless they are provided with inferences that lead you to specifications at L1, L2; otherwise, you cannot deal with what people understand (including “guesses”) when they read or hear sentences like (3), (11).

3.1.2 What All This Means to “Word Sense Disambiguation”

These aspects need to be specified somehow, and we believe that Frame Semantics/FrameNet [4, 5, 6, 11, 17] approach to semantic analysis/annotation is the most promising way to go if it provides, or at least helps to discover, sets of semantic roles like { ROBBER, STORE OF VALUABLES, WEAPON . . . } for ROBBERY, { PREDATOR, PREY, . . . } for PREDATION.

The situation of PRADATION, evoked in (12), is different from the situation of ROBBERY, evoked in (3), even if the same verb attack is used, on the one hand, and (13) refers to the situation of ROBBERY, too, even though different verbs, attack and hold up, are used, on the other:

(12) A group of lions attacked impalas.
(13) A group of masked men held up a bank branch in New York yesterday.

Clearly, this has interesting implications to word sense disambiguation, on the one hand, and to characterization of selectional restrictions/preferences, on the other.

It is hard, or at least “costs” a lot, to interpret (12) as referring to situations other than FO6: (Predatory Victimization). Likewise, it is hard, or at least costs a lot, to interpret (13) as referring to situations other than FO3b: (Bank Robbery). This seems to be true, but the question is, why is this so?
The model/theory of (word) sense disambiguation that we assume to deal with this problem is like this:

(14) a. Potential senses \( \{s_1, \ldots, s_n\} \) of a verb \( v \) of a sentence \( s \) are disambiguated to \( s_i \) if and only if a certain concrete situation or “frame” is selected from candidate situations such as ROBBERY, PREDATION, each of which is evoked by a combination of words of \( s \).

b. More generally, the same thing happens to every word of \( s \), in a “parallel, distributed” way.

This characterizes roughly how selectional restrictions are met for \( s \). This means that word sense disambiguation is co-selectional process, in addition to its co-compositional nature in the sense of Generative Lexicon Theory [25, 26, 27]

3.1.3 No sharp distinction of “semantics” from “pragmatics”

Phenomena mentioned above mean that “deep” semantic analysis of a text demands effective specifications of what guesses people make, as well as of semantic types of text segments. Put differently, it does not really matter whether people’s understandings are semantically based or pragmatically based as far as our goal is to illustrate people’s text understanding: specify what people understand, but how they do so is not. The semantics/pragmatics distinction makes sense as far as how people understand is at issue, but how they do so is not. The semantics/pragmatics distinction makes sense as far as how people understand is at issue after what they understand is made clarified.

This would be both good news and bad news, depending on your perspective. This would be good news if you feel that routes to deeper semantics are promised. This would be bad news if you feel that you cannot excuse by saying “Leave it all to pragmatics” any more, because what is at issue now is what pragmatics does and how it works out: you need to specify it.

3.2 Things to Do

There are a lot of things to do. Among others, we’ll definitely need to:

(15) a. develop a theory that enables us to find the most appropriate granularity levels,

b. develop an effective annotation model that can be put into practice realistically,

c. establish a mapping model from semantic roles to “concepts” in a thesaurus

After doing these, we then need to determine how to develop a database of frames/situations.

4 Concluding Remarks: Back to Basics

So, if our approach is valid, the ultimate questions to semantic annotation/analysis would take the following form:

(16) a. How many situations/frames like ROBBERY, PREDATION, do exist (in the human mind)?

b. How do we identify them?

c. How do we validate or evaluate the allegedly “identified” situations/frames?

All of these are open questions, somehow related to the “foundations” of ontologies, to none of which easy answers can be expected. We hope we can make some contribution to this large-scale problem from linguistic analysis.

References


