On the Semantic Labeling of Hand-drawn Sketches

Gennaro Costagliola[1] and Alberto Greco[2]
[1]*Dipartimento di Matematica ed Informatica – Università di Salerno*
[2]*Dipartimento di Scienze Antropologiche – Università di Genova*
E-mail: gcostagliola@unisa.it, greco@unige.it

Abstract

We propose a preliminary study for the application of Computational Linguistics techniques such as the Semantic Role Labeling of natural languages (SRL) to the case of sketch understanding. The idea of our approach originates from a psychological experiment where two groups of students were asked to produce and then recognize hand-drawn sketches, respectively. The paper presents a procedure to apply Semantic Role Labeling to the case of hand-drawn sketches and provides an example of application. This study seeks to enrich the body of research related to sketch understanding and reasoning.

1. Introduction

In any application based on hand drawn sketches it is very important to evaluate how much concepts and tasks are appropriately described in a pictorial form.

Knowing the semantic roles played by the components that appear in a hand-drawn sketch is of major importance for understanding its underlying meaning and for further reasoning. The inherent picture ambiguity and the way its constituents are spatially related can lead to very different interpretations of the same sketch.

Research on sketch understanding has taken many directions in the past decade and mostly it has tended to concentrate on the interpretations of the strokes forming the sketches [4]. In this paper we present our main ideas on how computational linguistics techniques can be used to address the problem of semantic sketch understanding. This is achieved by manually tagging constituents of sketch sentences with corresponding constituents of natural language sentences. We believe that the proposed approach may be a good starting point to be used when considering verbal descriptions in multimodal interactive interfaces, especially when meanings of speech and drawing have to be merged [1].

In particular we are interested in the use of semantic role labeling [7], one of the most promising approaches lately developed for natural language processing with interesting applications in the field of information extraction, question-answering, natural language database querying, spoken dialogue systems, machine translation, story merging, and others. While some of these applications are common for hand-drawn sketches, such as for query-by-sketch, [9], the others can provide new interests for the use of sketches.

In the following we recall the main concepts of Semantic Role Labelling as used in Computational Linguistics, describe the psychological experiment that gave start to this research, give a procedure for semantically labeling a hand drawn sketch from a verbal description of it and give some definitions of *consistent* hand-drawn sketches with an example showing the use of semantics hand-drawn sketch labeling.

2. Semantic Role Labeling

Let us recall the main characteristics of SRL. “Semantic Role Labeling involves the determination of domain-independent semantic relations among the entities and the events they participate in. Given a sentence, one formulation of the task consists of detecting basic event structures such as "who" did "what" to "whom", "when" and "where". Recently, the compilation and manual annotation with semantic roles of medium-large corpora – the PropBank [12, 13], NomBank, and FrameNet [5, 6] initiatives – has enabled the development of statistical approaches specifically for the task of semantic role labeling. SRL, especially focused on the labeling of verbal arguments and adjuncts, has become a well-defined task with a substantial body of work and comparative evaluation (e.g., see [7], CoNLL Shared Task in 2004 and 2005, Senseval-3)”, [15].

As an example of SRL let us consider the following sentences referring to the frames of Trading and Judgement, respectively:

[$_{\text{temporal}}$ At the end of the day] , [$_{\text{things being traded}}$ 251.2 million shares] **were traded** . (TRADING)

[$_{Judge}$ She] **blames** [$_{Evaluee}$ the Government] [$_{Reason}$ for failing to do enough to help]. (JUDGEMENT)

The identification of such event frames will have an important impact in many Natural Language Processing applications such as Information Extraction [14], Question Answering [11], Machine Translation [2], as well as Story Merging [10]. “Although the use of SRL systems in real-world applications has so far been limited, the outlook is promising over the next several years for a spread of this type of analysis to a range of applications requiring some level of semantic interpretation.” [15]. In our opinion these ones will naturally include sketch understanding.

There are many algorithms for semantic shallow parsing to assign semantic roles. This paper will specifically refer to the online UIUC parser [17]. The parser not only provides semantic role labeling of the sentence parts but also outputs a syntactic parse tree according to the Charniak parser, [3], allowing us to recover as much information as possible on each part of the sentence.

3. Experimental setup

The main goal in designing the experiment was to empirically identify correspondences between semantic features and pictorial elements and to explore a method for analyzing efficacy: operationally defined as the difficulty in matching pictures with original sentences, [8].

The experiment was divided into two stages and utilized two separate groups of participants – one for each stage of the experiment. In the first stage, the participants acted as “subjects” whilst in the second stage the participants acted as “judges”.

In the first stage, participants were asked to represent graphically a series of sentences describing a number of different types of situations. Sentences could describe either mathematical or non-mathematical situations. There were spatial situations (relations such as "above", "nearby", etc.), set situations like in arithmetic situations (where a set is divided into parts or subsets), and time situations (where events occur at different stages in time). Examples of sentences are:

- The house where Alan, Burt, and Chris live.
- Burt had 15 books. He bought 8 more, now he has 23 books.
- In Alan's garden there are 50 trees. Burt has more trees than Alan.

In the second stage, other participants acted as judges; their task was to match sentences (in random order) with the pictures drawn by the subjects in stage 1. The aim of this procedure was to evaluate the appropriateness of the pictures to their intended purpose, i.e. communicating information to other people. Participants in stage 2 were not taught explicit procedures, in order to encourage them to develop their own implicit procedures as they encountered more and more complex and abstract situations.

From the experiment, it has been observed that in many cases it is possible to associate parts of the drawings with parts of the originating verbal descriptions. In particular, there is evidence that the expressiveness of the drawings depends on how “well” the parts of the drawing represent the corresponding parts of the verbal descriptions and how the spatial relations (implicit ones such as nearby, above, etc. and explicit ones such as arrow, link, etc.) between the sketched parts recall the properties described in the verbal description such as possession, location, etc.. Furthermore, the appropriateness of the pictures to their intended purpose appears to be enhanced when consistent use of visual and spatial relation descriptions are used. These considerations allow us to state that it is appropriate to label parts and relations in a hand-drawn sketch with semantic roles in order to enhance sketch understanding and reasoning.

4. Semantic Sketch Role Labeling

In order to define the semantics of sketches we need to start creating a corpus of annotated sketches similarly to what it has been done with PropBank and FrameNet [12, 13, 6]. To do so we propose a way of semantically labeling sketches as shown by the procedure in Figure 1. Here each action is represented by a circled number indicating the temporal sequence of execution. The input sentence is partitioned and its parts are semantically annotated through the SRL technique (action 1 in Figure 1). Parts of the hand-drawn sketch are then labeled with the resulting sub sentences (action 2). Finally, the semantically labeled sketch is derived by merging results from the previous actions (action 3).

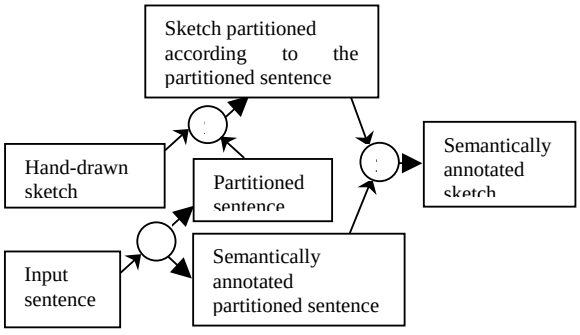


Figure 1. The Sketch Semantic Role Labeling procedure

Note that the procedure guarantees that the corresponding parts of the sketch and of the sentence are properly aligned.

The parts of the hand-drawn sketch to be labeled must be chosen accordingly. In fact, a hand-drawn sketch is a visual sentence and as such it is composed of a set of (spatially) related visual symbols or constituents [16]. An important task in the user sketch-sentence association is then the identification of the visual symbols and of the (spatial) relations to be labeled. It must be noted that, as in [16], relations in a picture can be either *implicit* such as the spatial relations nearby, below, above, etc. or *explicit* such as arrows, links, etc., i.e., relations with a visual representation.

As an example let us consider one of the sentences of the experiment in Section 3:

In Alan's garden there are 50 trees.
Burt has more trees than Alan

Figure 2 shows the hand-drawn sketch produced by a participant of stage 1 graphically depicting the sentence. In this figure, Alan and Burt are depicted as letters A and B, respectively, the garden as a waved line with flowers underneath, the two sets of trees as two tree shaped figures with the multiplication sign and a number to indicate quantity.

Table 1 provides a potential manual labeling of semantic roles for the picture in Figure 2 resulting from the procedure proposed in Figure 1.

In particular, each row contains a part of the hand-drawn sketch (column 2) drawn by the participant, the substring of the sentence with which the sketch has been manually annotated (column 1) and the annotations on the sub-sentence resulting from the Charniak parser and the Prop-Bank corpus (columns 3 and 4, respectively). While the association between the substrings and the sketch parts have been done manually, the Charniak parser and Prop-Bank annotations are done automatically by the UIUC parser, [17], based on the SRL techniques recalled in Section 2.

5. An SSRL Application

In this application we are interested in checking the semantic consistency of a hand-drawn sketch with respect to a pre-built corpus based on information coded as in Table 1. In the following, by "checking the consistency" we mean checking if the drawing style is compatible with one of the drawing styles coded in the corpus.

We distinguish three types of drawing style. Type 1 indicates the same use of spatial relations to describe

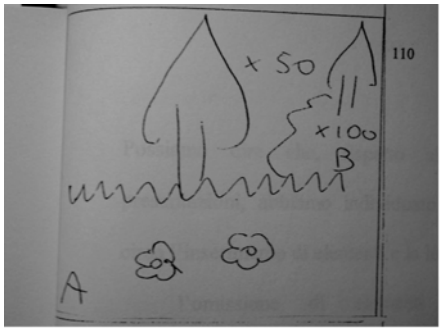
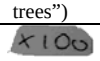


Figure 2. Participant 2 hand-drawn sketch

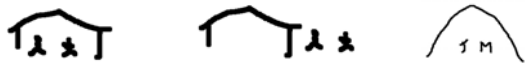
Table 1. Sketch Semantic Role Labeling for Fig. 2

Sentence	**Participant sketch**	**Charniak parser annotations**	**PROP-BANK annot.**
In			location [AM-LOC]
Alan		NNP (Proper noun, singular)	
's	nearby("Alan", "garden")	POS (Possessive ending)	
garden		NN (Noun, singular)	
there	nearby("garden", "50 trees")		
are			V: be
50		CD (cardinal number)	patient [A1]
trees.		NNS (Noun, plural)	
Burt			owner [A0]
has	nearby("Burt", "more trees")		V: have
more		JJR (Adjective, comparative)	possession [A1]
trees		NNS (Noun, plural)	

semantic connections between constituents of a sketch; type 2 indicates that semantic constituents (visual symbols) have a similar appearance; type 3 indicates that both type 1 and 2 occur. An example is shown in Figure 3: sketches (A) and (B) use the same style only with respect to type 2, while sketches (A) and (C) use the same style only with respect to type 1.

In the sequel we will only consider type 1 drawing style; in other words, two hand-drawn sketches are similar if they have a consistent use of spatial relations.

John and Mary live in a house:



(A) (B) (C)

Figure 3. The semantic concept "live" is expressed through the spatial relations "contain" in (A) and (C), and "nearby" in (B)

As an example let us consider the following sentence: **Fred owns a house** and its semantic and syntactic labeling as produced by [17], and shown in Figure 4.

Fred	owner [A0]	
owns	V: own	
a	possession [A1]	
house		

Charniak's Parse Tree

```
(S1 (S (NP (NNP Fred))
       (VP (VBZ owns)
           (NP (DT a)
               (NN house)))))
```

Figure 4. Semantic role labeling and Charniak annotation

Two possible hand-drawn sketches corresponding to the given sentence are shown in Figure 5 below.

In both cases *(i)* and *(ii)* the graphical sketch  is assigned with "Fred" and  with "house". In case *(i)* the verb "owns" is associated to the relational fact nearby("Fred", "house") or, more specifically, nearby(owner, possession). In case *(ii)*, "owns" will be associated to the relational fact arrow(owner, possession).

(i) *(ii)*

Figure 5. Hand-drawn sketches for the sentence **Fred owns a house.**

By properly matching the semantic role labeling of the sentence "Fred owns a house" and considering its visual associations against the labeling and associations of Table 1 it becomes possible to state that the hand-drawn sketch of Figure 5(i) has the same type 1 drawing style of Figure 2, while the hand-drawn sketch of Figure 5(ii) has a different drawing style.

6. Conclusions

In this paper we have presented a way for annotating hand-drawn sketches semantically. We derived its validity from the results of a psychological experiment and based its implementation on the computational linguistics technique of Semantic Role Labeling. We finally gave an example of application showing how to check for consistency between hand-drawn sketches.

7. References

[1] A. Adler and R. Davis, Speech and Sketching: An Empirical Study of Multimodal Interaction, EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling, M. van de Panne, E. Saund (Editors) (2007)

[2] H. C. Boas. Bilingual FrameNet dictionaries for machine translation. In Proceedings of LREC 2002, 1364--1371, Las Palmas, Canary Islands. (2002).

[3] E. Charniak, Statistical parsing with a context-free grammar and word statistics, Procs of the 14th National Conference on Artificial Intelligence AAAI Press/MIT Press, Menlo Park (1997).

[4] R. Davis, Magic Paper: Sketch-Understanding Research, Computer (Vol. 40, No. 9) pp. 34-41, Sept 2007.

[5] C. J. Fillmore, C. Wooters, and C. F. Baker. Building a large lexical databank which provides deep semantics. InProceedings of the Pacific Asian Conference on Language, Information and Computation (2001).

[6] FrameNet - http://framenet.icsi.berkeley.edu

[7] Gildea, D., Jurafsky, D.: Automatic labeling of semantic roles. Computational Linguistics 28(3) pp. 245-288 (2002).

[8] Greco A. A method for a cognitive analysis of graphical presentations. Tech. Rep. 0801, Lab. of Psychol. and Cog. Sci. Univ. of Genoa, (2008).

[9] T. Hisamori, G. Ohashi, Query-by-sketch interactive image retrieval using rough sets, ISIC. IEEE International Conference on Systems, Man and Cybernetics, pp.1223-1229, Oct. 2007.

[10] F. McNeill, H. Halpin, E. Klein, and A. Bundy Merging Stories with Shallow Semantics. Knowledge Representation and Reasoning for Language Processing Workshop at the EACL conference, Genoa, Italy, (2006).

[11] S. Narayanan, S. Harabagiu. Question Answering based on Semantic Structures, International Conference on Computational Linguistics (COLING 2004), Geneva, Switzerland, August 22-29, (2004).

[12] M. Palmer, D. Gildea, and P. Kingsbury. The Proposition Bank: An annotated corpus of semantic roles. In Computational Linguistics, 31(1), (2005).

[13] PropBank - http://www.cis.upenn.edu/~mpalmer/project_pages/ACE.ht

[14] M. Surdeanu, S. Harabagiu, J. Williams, and P. Aarseth. Using predicate-argument structures for information extraction. In Procs of the ACL, Sapporo, Japan, (2003).

[15] Special Issue on Semantic Role Labeling http://www.lsi.upc.edu/~carreras/srlcl.html

[16] G. Costagliola, A. De Lucia, S. Orefice, G. Tortora, A Framework of Syntactic Models for the Implementation of Visual Languages. Procs. 13th IEEE Symposium on Visual Languages, Capri, Italia, 58-65, 1997.

[17] UIUC parser - http://l2r.cs.uiuc.edu/~cogcomp/srl-demo.php