Towards automatic semantic role 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

In this paper we propose the use of Computational Linguistics techniques such as the Semantic Role Labeling of natural languages (SRL) to the case of semantic sketch understanding. SRL has many interesting applications in the field of information extraction, question-answering, natural language database querying, spoken dialogue system, machine translation, text summarization, story merging, etc. We propose a way of manually labeling hand drawn sketch constituents with semantic roles based on the natural language sentence to be depicted. Once this is done, applications of SRL can be extended to the case of hand drawn sketches.

1. Introduction

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

The problem of representation of abstract properties in a pictorial format has many practical implications; e.g. it is specially relevant when dealing with instructional material (cfr. graphic organizers: [19, 20]).

The research on sketch understanding has taken many directions in the last years and mostly they have been involved in the interpretations of the strokes forming the sketches [4, 12]. In this paper we propose the use of computational linguistics techniques to tackle the problem of semantic sketch understanding. This is done by manually tagging constituents of sketch sentences with the corresponding constituents of natural language sentences. The proposed approach can be a good start point also 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 system, machine translation, text summarization, story merging, and other. While some of these applications are common for hand drawn sketches, see [10] for query-by-sketch, the others can provide new interests for the use of sketches.

The main contribution of this paper is a proposal for the application of Semantic Role Labeling of natural languages (SRL) to the case of hand drawn sketches for obtaining automatic Sketch Semantic Role Labeling (SSRL). Once this is done, applications of SRL can be extended to the case of hand drawn sketches by using SSRL.

Long-term purposes of our project include:

- automatic sketch analysis, also in order to evaluate their efficacy in communication, problem solving, etc.
- creating a corpus for semantic analysis of sketches, consistent with a more general conceptual representational format

In order to be able to consistently identify correspondences between conceptual and semantic features and their actual pictorial implementation, we choose to start from empirical data, resulting from an experimental survey made with human subjects [8].

In the next Sections we will show the empirical experiment that gave start to this research, recall the main features of Semantic Role Labeling of natural languages as described in the current literature, describe Sketch Semantic Role Labeling and then illustrate one of the possible applications of SSRL.

2. Experimental setup

The main goal in designing the experiment [8] was to empirically identify correspondences between some semantic features (e.g. location, possession, etc.) and pictorial elements. The ultimate aim will be to get a method for a descriptive analysis of logical properties translation in pictorial data by filling conceptual slots. A further aim was to explore a method for analyzing efficacy: operationally defined as the difficulty in matching pictures with original sentences.

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

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

Stage 1 - Method

Participants

40 adults participated in the experiment. All of them were volunteers, students at the University of Genoa.

Materials

Descriptions referred to situations of different kinds. Table 1 shows general characteristics of situations,

Table 1 - General properties of situations

case 1 - Relationships expressible by pictorial elements
a) No relationships
b) Topological or ordinal spatial location e.g.: above / between / adjacent
c) Metric spatial location e.g.: near to…, far from…
case 2 - Properties expressible only by numbers Non concretely countable objects e.g.: one hundred houses
case 3 - Properties that require conventional symbols
a) Qualification e.g.: is red / is green
b) Single belonging or possession e.g.: Alan's
c) Multiple belonging or possession e.g.: Alan & Burt's
case 4 - Mixed (can be expressed as pictures, symbols, or both)
a) Concretely countable objects e.g.: two houses
b) Comparison, difference, correspondence e.g.: more than / less than / how many
case 5 - Transformation Reduction, addition (in this case there are two boxes, representing first-after times) e.g.: Alan had …, then he gave … / , then he bought …

Table 2 - Sentences

1. A house
2. Two houses
3. One hundred houses
4. Alan's house
5. The house where Alan, Burt, and Chris live
6. The house on the mountain
7. Alan's house on the mountain
8. Two houses on the mountain
9. One hundred houses on the mountain
10. A house with 100 windows
11. A house has 50 windows at the 1st floor and 50 windows at the 2nd floor
12. A house has 50 windows at the 1st floor and 50 windows at the 2nd floor. The 1st floor belongs to Alan; Burt and Chris live at the 2nd floor
13. A house near the mountain
14. In Alan's garden there are 50 trees. Burt has more trees than Alan
15. Alan's house is in Park Street between the town council and the chemist
16. There are red books and green books
17. There are red books and green books. There are 34 red books and 85 books in all
18. Alan has 34 books. He read 12 of them, now has 22 books to read
19. Burt had 15 books. He bought 8 more, now he has 23 books
* Sentences have been adapted from Italian.

with examples. There were four cases: 1) no pictorial-symbolic integration was required (relations could be expressed only by pictorial elements); 2) use of numbers or conventional symbols was implicitly required; 3) information could be expressed indifferently by pictorial or symbolic elements; 4) a double representation was required in order to account for transformations that implied a first-after sequence

Table 2 shows how characteristics were embodied into different sentences, in various combinations. Situations described by the sentences were arranged in increasing order of complexity and abstraction level. (Some picture examples with respect to sentence n. 14 are shown in Section 4)

Procedure

The participants were given a 19-page booklet; each page contained a verbal description of a situation printed on top of an empty box. Their task was "to represent without words" each situation. Drawings, single letters, numbers, and symbols were allowed; only a monochrome (black) pen was available. Subjects were told also that the representation should have been "clear enough in order that another person could reconstruct the sentence from the picture only". Sentences were presented in booklet pages in the same order as in table 2, by increasing complexity.

Stage 2

Participants

8 subjects, students at the University of Genoa, participated in the experiment as volunteers.

Materials

Eight items were selected from descriptions produced in stage 1.

In order to avoid that numbers acted as a possible cue for correct picture-sentence matching, all numbers in sentences were replaced by x,y,z letters.

Results

In Table 3 scores from the most difficult to the easiest sentence are shown.

Based on the analysis of the results of this experiment we have designed a procedure to start automating the process of semantic sketch understanding from a computational linguistics point of view.

The main idea is to use the knowledge gained so far in the Semantic Role Labeling of natural languages (SRL) for obtaining automatic Sketch Semantic Role Labeling (SSRL). Once this is done, applications of SRL can be extended to the case of hand drawn sketches by using SSRL.

Table 3

Sentence	Sum	Mean	Mean st.dev.
17. There are red books and green books. There are 34 red books and 85 books in all	25	.39	.49
11. A house has 50 windows at the 1st floor and 50 windows at the 2nd floor	28	.44	.34
12. A house has 50 windows at the 1st floor and 50 windows at the 2nd floor. The 1st floor belongs to Alan; Burt and Chris live at the 2nd floor	31	.48	.38
7. Alan's house on the mountain	33	.52	.36
18. Alan has 34 books. He read 12 of them, now has 22 books to read	34	.53	.32
4. Alan's house	41	.64	.39
19. Burt had 15 books. He bought 8 more, now he has 23 books	41	.64	.40
5. The house where Alan, Burt, and Chris live	42	.66	.24
10. A house with 100 windows	44	.69	.42
3. One hundred houses	46	.72	.23
8. Two houses on the mountain	46	.72	.31
9. One hundred houses on the mountain	47	.73	.40
6. The house on the mountain	47	.73	.44
1. A house	53	.83	.36
15. Alan's house is in Park Street between the town council and the chemist	54	.84	.22
16. There are red books and green books	55	.86	.27
2. Two houses	61	.95	.13
13. A house near the mountain	61	.95	.13
14. In Alan's garden there are 50 trees. Burt has more trees than Alan	62	.97	.09

In the next Sections we will recall the main features of Semantic Role Labeling for the natural languages as described in the current literature, describe our approach and then illustrate a possible application of SSRL.

3. Semantic Role Labeling

Let us recall the main characteristics of SRL as presented in [22].

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". From a linguistic point of view, a key component of the task corresponds to identifying the semantic arguments filling the roles of the sentence predicates. These predicates are mainly lexicalized by verbs but also by some verb nominalizations and adjectives. Typical predicate semantic arguments include Agent, Patient, and Instrument; semantic roles may also be found as adjuncts (e.g., Locative, Temporal, Manner, and Cause). The related tasks of determining the semantic relations among nouns and their modifiers, as well as prepositions and their arguments, are clearly important for text interpretation as well, and indeed often draw on similar role labels.

Recently, the compilation and manual annotation with semantic roles of medium-large corpora – the PropBank [16, 17], 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). As an example of SRL let us consider the following sentences referring to the frames of Trading, Judgement and Statement respectively:

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

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

[$_{\text{Message}}$ “I’ll knock on your door at quarter to six”] [$_{\text{Speaker}}$ Susan] **said**. (STATEMENT)

The identification of such event frames holds potential for significant impact in many Natural Language Processing applications , as suggested by the following works on Information Extraction [21], Question Answering [15], Summarization [14], Machine Translation [2], Story Merging [13].

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 including, in our opinion, semantic sketch understanding.

Among the many algorithms for semantic shallow parsing to assign semantic roles, in this paper we will refer to the online UIUC parser [23, 18]. The system not only provides semantic role labelling 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.

4. Semantic Sketch Role Labeling

In order to define the semantics of sketches we need to start creating corpora of annotated sketches similarly to what has been done with PropBank and FrameNet. To do so we propose a way of semantically labelling sketches according to the following two stes:

1. *Sketch-sentence association*
   By starting from a natural language sentence, the user draws a sketch and annotates (verbally or textually) parts of the sketch with parts of the original sentence
2. *Sketch semantic role labelling*
   By applying a shallow semantic parser the sentence is analyzed and labelled, together with the annotated sketch, with syntactic constructs (Charniak parser) and semantic roles based on the PropBank and FrameNet corpus

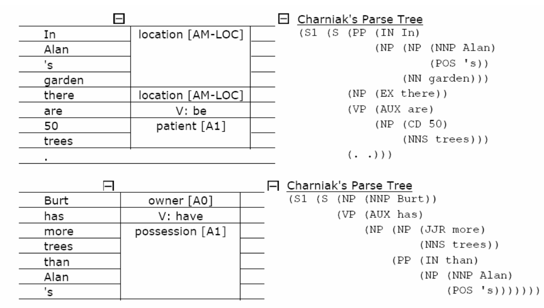


Figure 1. Semantic role labelling and Charniak annotation

As an example let us consider the sentence n. 14 from Table 4 :

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

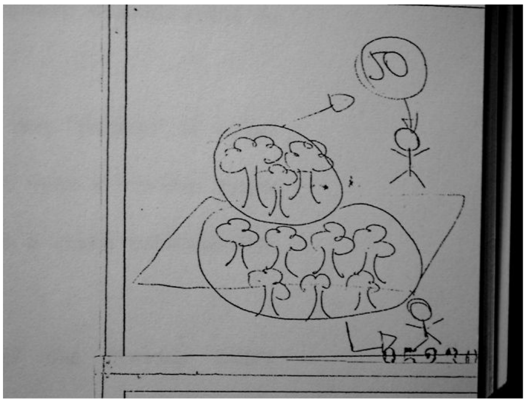


Figure 2. Partecipant 1 hand drawn sketch

Table 4. Sketch Semantic Role Labeling for Fig. 2

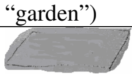
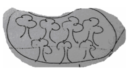
Sentence	Partecipant 1 sketch	Charniak parser annotations	PROP-BANK annot.
In			location [AM-LOC]
Alan		NNP (Proper noun, singular)	
's	nearby ("Alan", "garden")	POS (Possessive ending)	
garden		NN (Noun, singular)	
there	nearby ("garden", "50 trees");  arrow ("50", "Alan")		
are			V: be
50		CD (cardinal number)	patient [A1]
trees.		NNS (Noun, plural)	
Burt			owner [A0]
has	 arrow ("more trees", "Burt")		V: have
more		JJR (Adjective, comparative	possession [A1]
trees		NNS	

Figure 1. shows the results of the analysis obtained by applying the shallow semantic parser in [UIUC], while Figure 2 and 3 show two hand drawn sketches produced by two participants to graphically depict the sentence.

Finally, Table 2 and 3 give the manual labeling of semantic roles for each of the sketch constituents based on SRL and Charniak annotation of Figure 1..

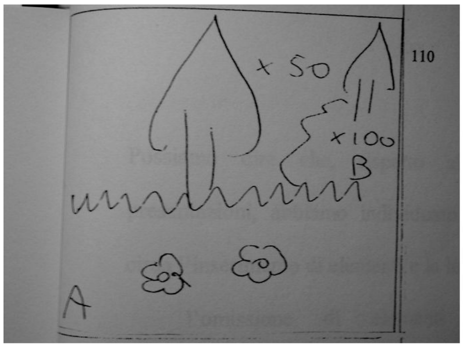
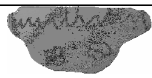
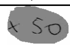
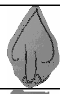
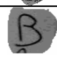
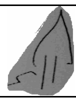


Figure 3. Partecipant 2 hand drawn sketch

Table 5. Sketch Semantic Role Labeling for Fig. 3

Sentence	Partecipant 2 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)	

Note that the gray text in the sixth row ("are") of Table 4 corresponds to the arrow between 50 and Alan. This arrow does not actually have a direct correspondence in the original text sentence: it stresses the fact that Alan has 50 trees. As a matter of fact, the arrow is to be bound to the concept of possession to be consistent with the rest of the sketch.

5. An SSRL Application

In this application we are interested in checking the consistency of an hand drawn sketch with respect to a pre-built corpus. We will use Tables 4 and 5 of the Section 4 as corpus. 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 levels of drawing style: at level 1, "the same style" indicates the same use of spatial relations to describe semantic connections between constituents of a sketch; at level 2 "the same style" indicates that semantic constituents (concepts) have a similar appearance; at level 3 both levels 1 and 2 are respected. An example is shown in Figure 4: sketches (A) and (B) use the same style only with respect to level 2, while sketches (A) and (C) use the same style only with respect to level 1

John and Mary live in a house:

(A) (B) (C)

Figure 4. The semantic concept "live" is expressed through the spatial relations "contain" in (A) and (C), and "nearby" in (B)

In this application we only consider level 1 of drawing style, in other words, two hand drawn sketches are similar if they have a consistent use of spatial relations.

As an example let us consider the following sentence:

Fred owns a house.

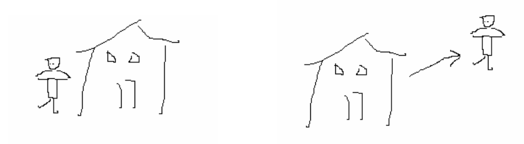
and its semantic and syntactic labelling as produced by [23]:

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 5 shows two possible hand drawn sketches corresponding to the given sentence.



(i) (ii)

Figure 5: hand drawn sketches for the sentence **Fred owns a house.**

In both cases (i) and (ii) a user will associate  with "Fred" and  with "house". In case (i) the verb "owns" will be associated to the relational fact nearby("Fred", "house") or, more specifically, nearby(owner, possession). In case 2, "owns" will be associated to the relational fact arrow(owner, possession).

By properly matching the semantic role labelling of the sentence Fred owns a house and considering its visual associations against the labelling and associations of tables 4 and 5 it becomes possible to state that the hand drawn sketch of Figure 5.i is in the drawing style level 1 of Partecipant 2, while the hand drawn sketch of Figure 5.ii is in the drawing style level 1 of Partecipant 1.

9. Conclusions

In this paper we have presented a project (as a proof of concept) showing some directions on how to annotate semantically hand drawn sketches.

We would stress the relevance of this research on multiple and multidisciplinary grounds: for the implementation of effective techniques of automatic analysis, for a more comprehensive understanding of the relationships of sketch representation, considered as a language, with natural language, and for understanding of cognitive processes implied in sketch representation.

In doing so we relied on the results of a preliminary psychological experiment and on some available computational techniques such as Semantic Role Labeling.

Long-term purposes of our project include:

- automatic sketch analysis, also in order to evaluate their efficacy in communication, problem solving, etc.

- creating a corpus for semantic analysis of sketches, consistent with a more general conceptual representational format

Another important issue to be explored for practical purposes, is that effectiveness is somehow dependent from the task at hand and pragmatic aspects should be taken into account, [9].

As following steps, it would be convenient to develop a system for systematically collecting explicit links between semantic features and pictorial components. A specific software should be designed to do so.

Adopted representational formalism might be inadequate for current purposes (e.g. how to represent parts only emerging from an inference [11]), a further project could be developing a new special representation system.

10. References

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

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

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

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

[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) (2002) 245--288

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

[9] Gurr C.A. (1999) Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. Journal of Visual Languages and Computing, 10(4).

[10] T. Hisamori, G. Ohashi, Query-by-sketch interactive image retrieval using rough sets, Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on, 7-10 Oct. 2007, pp.1223-1229

[11] Lindsay R.K. (1988) Images and inference. Cognition, 29, 229-250.

[12] A. Lovett, M. Dehghani, K. Forbus, Incremental Learning of Perceptual Categories for Open-Domain Sketch Recognition, IJCAI 2007, Hidarab, India

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

[14] G. Melli, Y.Wang, Y.g Liu, Mehdi M. Kashani, Z. Shi, B. Gu,Anoop Sarkar, and F. Popowich. 2005. Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2005 Summarization Task. In Proceedings of Document Understanding Conferences.

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

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

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

[18] V. Punyakanok, D. Roth, and W. Yih. Generalized inference with multiple semantic role labeling systems. In Proceedings of CoNLL-2005.

[19] Robinson D.H. & Kievra K.A. (1995) Visual argument: graphic organizers are superior to outlines in improving learning from text, Joumal of Educational Psychology, 87, 3, 455-467.

[20] Robinson D.H. (1998) Graphic organizers as aids to text learning, Reading Research and Instruction, 37, 2, 85-105.

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

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

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