Two-Dimensional Visual Language Grammar ⋆

Siska Fitrianie and Leon J.M. Rothkrantz

Man-Machine-Interaction Group, Delft University of Technology
Mekelweg 4 2628CD Delft, The Netherlands
{s.fitrianie, l.j.m.rothkrantz}@ewi.tudelft.nl

Abstract. Visual language refers to the idea that communication occurs through visual symbols, as opposed to verbal symbols or words. Contrast to a sentence construction in spoken language with a linear ordering of words, a visual language has a simultaneous structure with a parallel temporal and spatial configuration. Inspired by Deikto [5], we propose a two-dimensional string or sentence construction of visual expressions, i.e. spatial arrangements of symbols, which represent concepts. A proof of concept communication interface has been developed, which enables users to create visual messages to represent concepts or ideas in their mind. By the employment of ontology, the interface constructs both the syntax and semantics of a 2D visual string using a Lexicalized Tree Adjoining Grammar (LTAG) into (natural language) text. This approach captures elegantly the interaction between pragmatic and syntactic descriptions in a 2D sentence, and the inferential interactions between multiple possible meanings generated by the sentence. From our user test results, we conclude that our developed visual language interface could serve as a communication mediator.

1 Introduction

Skills at interpreting visual symbols play an important part in humans' learning about the world and understanding of language. Words are also composed by symbols, of course. There are nonverbal symbols that can provide essential meanings with their succinct and eloquent illustrations. Humans respond to these symbols as messages, though often without realizing exactly what it is that has caused us to reach a certain conclusion. Such symbols are often visual, though they can be auditory or even tactile. The research described in this paper concentrates on visual nonverbal symbols. In particulary, it focuses on exploring such symbols to represent concepts, i.e. objects, actions, or relations.

According to [15], human communication involves the use of concepts to represent internal models of humans themselves, the outside world and of things with which the humans are interacting. In earlier work, we have investigated a languange independent communication using visual symbols, i.e. icons [8], based

⋆ The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024

on signs and symbols, which are understood universally [15]. As a proof of concept, an iconic interface has been developed in a specific domain. It is applied on a communication interface on PDAs in crisis situations. In such situations, wired communication could be disabled by the breakdown of the infrastructure or information overload and speech communication is difficult due to noisy environments. Visual language has been used successfully in human-computer interaction, visual programming, and human-human communication. Words are easily forgotten, but images stay in our minds persistently [6]. Therefore, icons can evoke a readiness to respond for a fast exchange of information and a fast action as a result [14]. Furthermore, a direct manipulation on the icons allows us to have a faster interaction [11].

A visual (communication) language uses a set of spatial arrangements of visual symbols with a semantic interpretation that is used in carrying out communication [3]. This language is based upon a vocabulary of visual symbols where each symbol represents one or more meaning, which are created according to the metaphors appropriate for the context of this type of languages. The sentence structure of a visual language is different from a sentence in spoken language [12]. The spoken language is composed by a linear ordering of words, while a visual language has a simultaneous structure with a parallel temporal and spatial configuration, e.g. the sign language syntax for deaf people [1], comic illustrations [4], diagrams. Based on this, we propose a two-dimensional syntax structure that enables a visual language sentence constructed in a 2D way.

2 Related Work

The approach described in this paper basically is inspired by Deikto, a game's interaction language [5]. The creator aimed at producing the smallest language capable of expressing the concepts used in the story world. A sentence in Deikto is represented by a connected acyclic graph, where a predecessor element explains its successor. A graph element can be a lexicon or another clause. To help the players, the game provides hints of what lexicon(s) can be selected based on its word class. Deikto follows a rigid grammar by assigning each verb the parts of the sentence in its dictionary definition. This revolving verb approach is supported by Fillmores case grammar [7] and Schanks conceptual dependency theory [17]. In case grammar, a sentence in its basic structure consists of a verb and one or more noun phrases. Each phrase is associated with the verb in a particular case relationship. The conceptual dependency defines the interrelationship of a set of primitive acts to represent a verb. The approach employs rules and a set of primitive acts, e.g. ATRANS: transfer possession of an object. Instead of using the word class (e.g. subject, verb, object), both approaches use thematic roles, e.g. agent, patient, instrument, etc, to define a sentence structure.

VIL, a one-dimensional visual language application [13], is designed to allow people to communicate with each other by constructing sentences solely relying on icons. The system is based on the notion of simplified speech by reducing a significant complexity. It is also inspired by the case concept of Fillmore and the verb classification of Schank.

In the following sections we will give an overview of the 2D visual language we propose. Further, we will concentrate on the visual sentence construction and conversion to text or speech as well as a proof of concept visual language interface we have developed. Finally, we also present our test results.

3 Two-Dimensional Visual Language

According to [1], the syntax analysis of the visual language does not reduce to classical spoken sentence syntax. There exists a set of "topic" and "comment" relations, in which a comment explains a topic. Therefore, in our two-dimensional syntax structure, a sentence is constructed in an acyclic-graph of visual symbols, i.e. icons (see fig. 1(b)). Conventional textual syntax structures are not considered 2D, since the parser processes them as 1D streams of symbols.

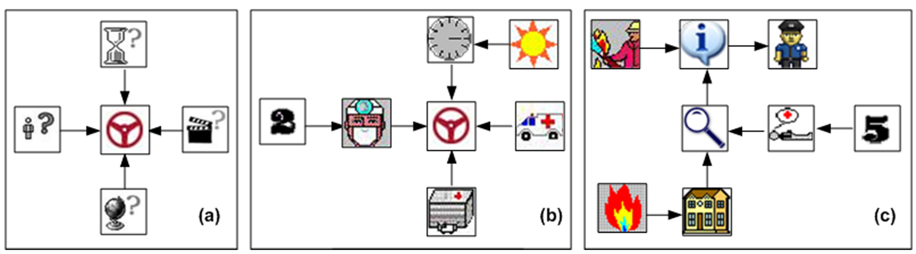

Fig. 1. Examples of visual language sentences: (a) hints for the verb "drive" and (b) a simple sentence: "Two paramedics drove an ambulance to the hospital at 15.00" and (c) a compound sentence: "The firefighter informs the police that he will search five victims in the burning building"

A visual symbol can be connected by an arrow to another symbol, in which the former explains the latter symbol. Each symbol represents concepts or ideas. The sentence may be constructed from any part, however as soon as a verb is selected, the structure of the sentence will be determined. We define a case for every symbol in our vocabulary. For example: the symbol "paramedic" contains number, location, status, name, as attributes of the case. Particulary for symbols that represent verbs, based on the theory of [7][17], we define a case of every verb and follow the frame syntactic analysis used for generating the VerbNet [10]. For example: the case of the verb "drive" contains agent, theme, location, time.

To help the user, attributes of each selected visual symbol's case on a 2D sentence are displayed (see fig. 1(a)). As the icon is deselected, the hint will disappear to reduce the complexity. A hint symbol can be selected and replaced by a visual symbol that is grammatically correct to form a sentence. The approach gives a freedom to users to fill in the parts of a sentence, but at the same time the system can restrict the choices of symbols which lead to a meaningful sentence. A user may attach symbols that are not given by the hint by inserting a new node on a specific node on the sentence. This means that the new node explains the selected node. Our developed grammar allows a compound sentence construction, which can be done by inserting another verb on a verb node of the sentence (see fig. 1(c)) or noun phrase conjunction.

4 Knowledge Representation

Visual symbols have no meaning outside of their context. A visual symbol can be interpreted by its perceivable form (syntax), by the relation between its form and what it means (semantics), and by its use (pragmatics) [2]. The relation between visual symbols and words can have an ambiguous meaning. We tackled this by creating a verbal context that can link both visual and verbal thoughts together to form a symbol that can be remembered and recalled. An ontology is employed to represent context that binds verbal and visual symbols together.

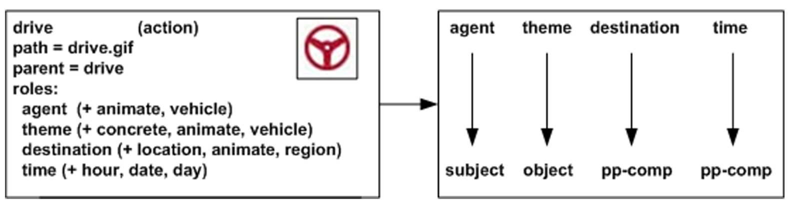

Fig. 2. Schematic vie of the two components of a verb lexical definition: semantic types and linking to syntactic arguments

We employ the ontology to store information of the vocabulary, i.e. visual symbols that represent nouns, pronouns, proper-nouns, adjectives, and adverbs, and their properties (i.e. attributes of a case). The ontology provides a natural way to group its elements based on their concepts and provides the system information about their semantic description, e.g. symbols of ”firefighter” and ”paramedic” are grouped under ”animate” and a symbol of ”ambulance” is under ”vehicle”. In particularly for verb symbols, based on [10], a lexeme has one or more sense definitions, which consist of a semantic type with associated thematic roles and semantic features, and a link between the thematic roles and syntactic arguments. The definition also defines required and optional roles. Figure 2 shows a case for the verb ”drive”.

5 Lexicalized Tree Adjoining Grammar by Visual Symbols

Each visual symbol provides only a portion of the semantics of a visual language sentence. The meaning of a 2D sentence with more than just one symbol can still be represented by the same set of symbols, but it turns out to be very difficult to determine the sentence meaning. The syntax of interactions between concepts (that are represented by visual symbols) enriches progressively the semantic of the sentence. The only thing that can be automatically derived from the semantics of the symbols in a visual language sentence is a fixed word or phrase belonging to these symbols.

Based on [16], we assign our vocabulary with Lexicalized Tree Adjoining Grammar (LTAG) [9]. Figure 3(a) shows the example of our TAG trees vocabulary. For this purpose, we exploit XTAG grammar [18] that presents a large existing grammar for English verbs. Mapping our VerbNet-based syntactic frames to the XTAG trees greatly increases the robustness of the conversion of 2D visual language sentences to natural language text/speech.

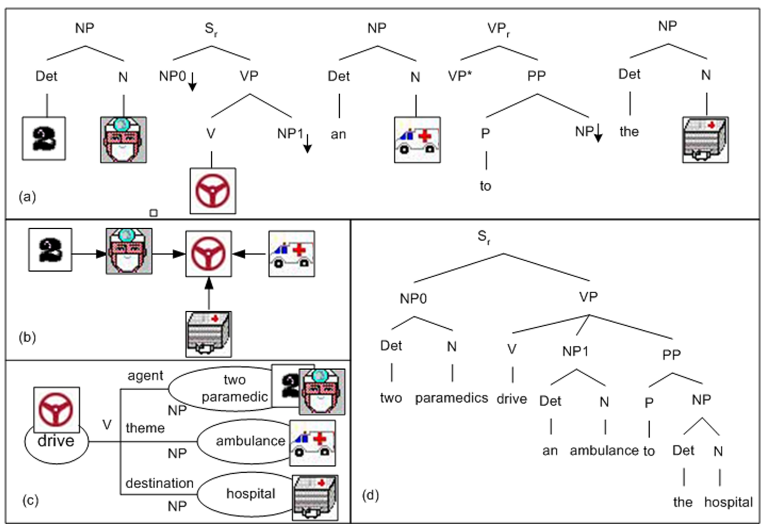

Fig. 3. Conversion to natural language text: (a) examples of the iconized TAG elementary trees, (b) example of a 2D sentence: "Two paramedics drove an ambulance to the hospital", (c) example of mapping the thematic roles to the basic syntactic tree defined by the case of the verb "drive", and (d) example of a parse tree as the results of mapping the basic syntactic tree to the TAG trees

Based on the case of every symbol in a 2D sentence, a parser processes a 2D stream of symbols and maps the thematic roles of them into the basic syntactic tree on the VerbNet-based vocabulary. Presumably, transformation of VerbNet's syntactic frames are recoverable by mapping the 2D sentence onto elementary trees of TAG tree families. For this purpose, the parser exploits the system's ontology to have the syntactic argument of every symbol in the sentence. Figure 3 shows an example of parsing a 2D sentence using LTAG. We specify the semantics of a 2D sentence in two ways. First, our developed ontology offers a simple syntax-semantics interface for every symbol. As shown in fig. 2, each verb case has restricted the choice of symbols to form the sentence, i.e. by associating thematic roles to semantics features. The meaning of a TAG tree is just the conjunction of the meanings of the elementary trees used to derive it, once appropriate case elements are filled in. Finally, the VerbNet structure provides an explicitly constructed verb lexicon with syntax and semantics. By this way, the syntax analysis and natural language construction can be done simultaneously.

6 Reporting Crisis Situations

Our visual language communication tool was designed for reporting observations in a crisis situation. A user can arrange an acyclic graph of visual symbols, i.e. icons, as a realization of his/her concepts or ideas. Besides supporting a fast interaction by converting the message into natural language, the tool also can be combined with any language application, e.g. a text to speech synthesizer, a language translator, etc, through a socket network. The current prototype provides a speech synthesizer to read aloud the resulted natural language text with correct pronunciations. Figure 4 shows the architecture of our developed tool.

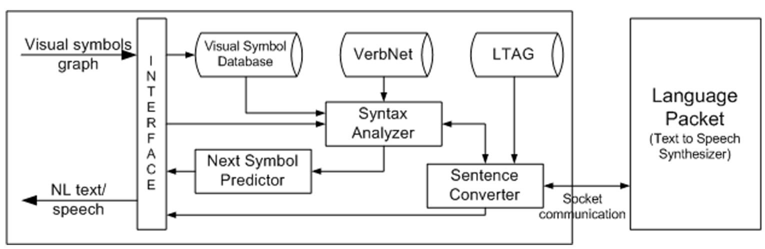

Fig. 4. The architecture of our visual language communication tool on a PDA

Figure 5 shows the interface of our visual language communication tool. On the interface, visual symbols are grouped in clusters based on their concept. The interface provides a next-symbol predictor to help users to find their intended symbols fast. It predicts which symbols are most likely to follow a given segment of a visual symbol graph based on its syntax structure. When a user selects one of the suggestions, it is automatically inserted into the graph to replace a selected hint symbol. The probability of the prediction of a visual symbol is estimated with n-grams language modelling. To compute the multi-grams model, our tool collects the data during the interaction. The interface provides a real time distinctive appearance of which visual symbols can be selected next according to syntactical rules. To construct a message, a user can select visual symbols from the menu or from the prediction window. If the user changes the input graph, the resulted text will be refreshed.

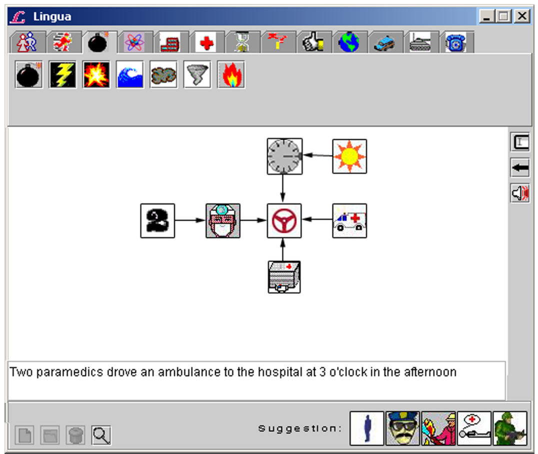

Fig. 5. The interface of our visual language communication tool on a PDA

7 Evaluation

We performed a similar user test as reported in [8]. It aimed to assess whether or not users were still capable to express their concepts in mind using the provided visual symbols in a 2D way. The test also addressed the usability issues on interacting with 2D sentence constructions on the visual language interface. Eight people took part in the test. The tasks were created using images of real crisis situations. The participants were asked to report what they might experience, by creating 2D visual language sentences on the interface. While performing the task, they were asked to think aloud. There were no incorrect answers, except if any task was not performed at all. All activities were recorded and logged for analyses purposes.

The experimental results showed that our target users were able to compose 2D visual language messages to express their concepts and ideas in mind. The users used some time to find another concept when they could not find a relevant concept from the provided visual symbols to represent their message. Although adaptation time was needed to recognize some icons, the results also indicated that the hints given while creating 2D-icon string helped the user to compose a complete report.

8 Conclusion

A two-dimensional visual language grammar, as the continuation of our research [8], has been developed. The idea is inspired by a game's language interaction, Deikto [5]. A sentence can be created using a spatial arrangement of visual symbols, i.e. icons. To support a natural visual language sentence construction, the arrangement may not be in a linear order. We combine LTAG syntax and VerbNet frames so that we can analyze the syntax and semantics of visual sentences and convert them to text/speech simultaneously and easier. The approach naturally and elegantly captures the interaction between the graph structure of the visual sentences and the tree strucuture of LTAG syntax, and the inferential interactions between multiple possible meaning generated by the sentences.

An experimental visual language interface has been developed that is applied for reporting observations in a crisis situation. Our target users could express their concepts and ideas solely using a spatial arrangement of visual symbols. However, future work should be done to gather data about how people might create visual messages in their real life and how they experience this. Therefore, the 2D visual language grammar can cover all possible message constructions.

References

1. Bevelier D., Corina D.P. and Neville H.J.: Brain and Language: a Perspective from Sign Language. Neuron, Cell Press. **21** (1998) 275–278.
2. Chandler D.: Semiotics, the Basic. Routledge (2001).
3. Chang S.K., Polese G., Orefice S., and Tucci M.: A Methodology and Interactive Environment for Iconic Language Design. International Journal of Human Computer Studies. **41** (1994) 683–716.
4. Cohn N.: Visual Syntactic Structures, Towards a Generative Grammar of Visual Language. http://www.emaki.net/essays/ (2003).
5. Crawford C.: Erasmatron. http://www.erasmatazz.com (2005).
6. Frutiger A.: Sign and Symbols, Their Design and Meaning. New York: van Nostrand Reinholt (1989).
7. Fillmore C.J.: The Case for Case. in Universals in Linguistic Theory, Ed. by Emmon Bach and Robert Harms. New York: Holt, Rinehart and Winston (1968).
8. Fitrianie S. and Rothkrantz L.J.M.: Language-Independent Communication using Icons on a PDA. Text, Speech and Dialogues. Lecture Notes in Artificial Intelligence. Springer. **3658** (2005) 404–411.
9. Joshi A.K., Levy L., and Takahashi M.: Tree Adjunct Grammars. Journal of the Computer and System Sciences. **10** (1975) 136–163.
10. Kipper-Schuler K.: VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon. Ph.D. thesis proposal, University of Pennsylvania (2003).
11. Kjeldskov J. and Kolbe N.: Interaction Design for Handheld Computers. In Proc. of APCHI02. Science Press. China. (2002).
12. Lester P.M.: Visual Communication: Images with Messages, 4th Ed. Wadsworth/Thompson, Belmont, CA (2006).
13. Leemans N.E.M.P.: VIL, A Visual Inter Lingua. Doctoral Dissertation, Worcester Poly-technic Institute, USA (2001).
14. Littlejohn S.W.: Theories of Human Communication, 5th Ed. Wadsworth (1996).
15. Perlovsky L.I.: Emotions, Learning and Control. In Proc. of Intl. Symposium: Intelligent Control, Intelligent System and Semiotics (1999).
16. Ryant N. and Kipper K.: Assigning Xtag Trees to VerbNet. 7th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+7). Vancouver, Canada (2004).
17. Schank R.: Computer Models of Thought and Language. W. H. Freeman. San Francisco (1973).
18. XTAG Research Group: A Lexicalized Tree Adjoining Grammar for English. Technical Report IRCS-01-03. IRCS, University of Pennsylvania (2001).