Toward Communicating Simple Sentences Using Pictorial Representations

Rada Mihalcea and **Ben Leong**
Department of Computer Science
University of North Texas
{rada,cl0198}@cs.unt.edu

Abstract

This paper evaluates the hypothesis that pictorial representations can be used to effectively convey simple sentences across language barriers. Comparative evaluations show that a considerable amount of understanding can be achieved using visual descriptions of information, with evaluation figures within a comparable range of those obtained with linguistic representations produced by an automatic machine translation system.

1 Introduction

Universal communication represents one of the long-standing goals of the humanity – borderless communication among people, regardless of the language they speak. According to recent studies (Eth, 2005), (Gibbs, 2002) there are about 7,000 languages spoken worldwide. From these, only about 15–20 languages can currently take advantage of the benefits provided by machine translation, and even for these languages, the automatically produced translations are not error free and their quality lags behind the human expectations.

In this paper, we investigate a new paradigm for translation: **translation through pictures**, as opposed to translation through words, as a means for producing universal representations of information that can be effectively conveyed across language barriers. Regardless of the language they speak, people share almost the same ability to understand the content of pictures. For instance, speakers of different languages have a different way of referring to the concept of *apple*, as illustrated in Figure 1(a). Instead, a picture can be understood by all people in the same way, replacing the multitude of linguistic descriptions with one, virtually universal representation (Figure 1(b)).

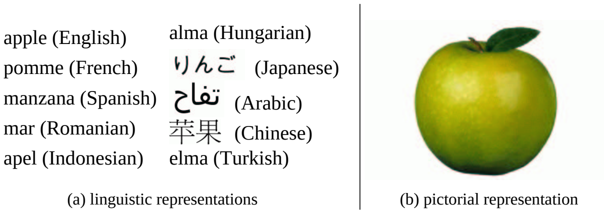


Figure 1: Linguistic and visual representations for the concept “apple”.

In addition to enabling communication across languages, the ability to encode information using pictorial representations has other benefits, such as language learning for children or for those who study a foreign language, communication to and from preliterate or non-literate people, or language understanding for people with language disorders.

This paper describes a system for the automatic generation of pictorial translations for simple sentences, and evaluates the hypothesis that such pictorial descriptions can be understood independent of language-specific representations. An example of the pictorial translations that we target is shown in Figure 2(a).

There are of course limitations inherent to the use of visual representations for the purpose of communication. First, there are complex informations that cannot be conveyed through pictures, as in e.g. *"An inhaled form of insulin won federal approval yesterday,"* which require the more advanced representations that can only be encoded in language. Second, there is a large number of concepts that have a level of abstraction that prohibits a visual representation, such as e.g. *politics* or *regenerate*. Finally, cultural differences may result in varying levels of understanding for certain concepts. For instance, the prototypical image for *house* may be different in Asian countries as compared to countries in Europe. Similarly, the concept of *coffee* may be completely missing from the vocabulary of certain Latin American tribes, and therefore images representing this concept are not easily understood by speakers of such languages.

While we acknowledge all these limitations and difficulties, we attempt to take a first cut at the problem, and evaluate the amount of understanding for simple sentences when "translated through pictures," as compared to the more traditional linguistic translations. Note that we do not attempt to represent complex states or events (e.g. emotional states, temporal markers, change) or their attributes (adjectives, adverbs), nor do we attempt to communicate linguistic structure (e.g. prepositional attachments, lexical order, certainty, negation). Instead, we focus on generating pictorial translations for simple sentences, using visual representations for basic concrete nouns and verbs, and we evaluate the amount of understanding that can be achieved with these simple visual descriptions as compared to their linguistic alternatives.

Starting with a given short sentence, we use an electronic illustrated dictionary (PicNet) and state-of-the-art natural language processing tools to generate a pictorial translation. A number of users are then asked to produce an interpretation of these visual representations, which are then compared with the interpretation generated based on a linguistic description of the same information. Results show that a considerable amount of understanding can be achieved based on visual descriptions of information, with evaluation figures within a comparable range of those obtained for automatically produced linguistic representations.

2 Understanding with Pictures

The hypothesis guiding our study is that simple sentences can be conveyed via pictorial representations with limited or no use of linguistic descriptions. While linguistic expressions are certainly irreplaceable when it comes to complex, abstract concepts such as *materialism* or *scholastics*, simple concrete concepts such as *apple* or *drink* can be effectively described through pictures, and consequently create pictorial representations of information.

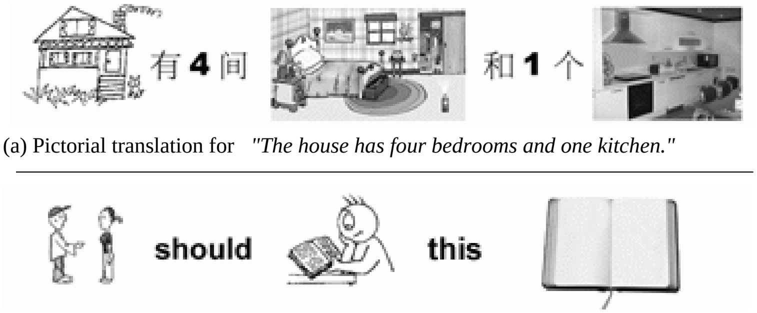
Our goal is to test the level of understanding for *entire* pieces of information represented with pictures, e.g. short sentences such as *I want to drink a glass of water*, which is different than testing the ability to grasp a single concept represented in a picture (e.g. understand that the concept shown in a picture is *apple*). We therefore perform our experiments within a translation framework, where we attempt to determine and evaluate the amount of information that can be conveyed through pictorial representations.

Specifically, we compare the level of understanding for three different ways of representing information: (1) fully conceptual, using only pictorial representations; (2) mixed linguistic and conceptual, using representations consisting of pictures placed within a linguistic context; and finally (3) fully linguistic, using only words to represent information.

2.1 Translation Scenarios

We conduct out experiments under the assumption that there is a language barrier between the two participants in an information communication process. The sender (speaker) attempts to communicate with a receiver (listener), but the only communication means available is a language known to the sender, but not to the receiver. We therefore deal with a standard translation framework, where the goal is to convey information represented in an "unknown" (source) language to a speaker of a "known" (target) language. The following three translation scenarios are evaluated:

Scenario S1. No language translation tool is available. The information is conveyed exclusively through pictures, and while linguistic representations can still be used to suggest the presence of ad-



(a) Pictorial translation for *"The house has four bedrooms and one kitchen."*

(b) Mixed pictorial and linguistic translation (automatic) for *"You should read this book."*

I eat the egg and the coffee work as breakfast.

(c) Linguistic translation (automatic) for *"I eat eggs and coffee for breakfast."*

Figure 2: Sample pictorial and linguistic translations for three input texts.

ditional concepts, they are not understood by the information recipient. In this scenario, the communication is performed entirely at conceptual level. Figure 2(a) shows an example of such a pictorial translation.

Scenario S2. An automatic language translation tool is available, which is coupled with a pictorial translation tool for a dual visual-linguistic representation. The linguistic representations in the target ("known") language are produced using an automatic translation system, and therefore not necessarily accurate. Figure 2(b) shows an example of a mixed pictorial-linguistic translation.

Scenario S3. The third case we evaluate consists of a standard language translation scenario, where the information is conveyed entirely at linguistic level. Similar with the previous case, the assumption is that a machine translation tool is available, which can produce (sometime erroneous) linguistic representations in the target "known" language. Unlike the previous scenario however, no pictorial translations are used, and therefore we evaluate the understanding of information using representations that are fully linguistic. An example of such a representation is illustrated in Figure 2(c).

In the following section, we briefly describe the construction of the PicNet illustrated dictionary, which associates pictures with word meanings as defined in an electronic dictionary. We then describe an automatic system for generating pictorial translations, and evaluate its ability to convey simple pieces of information across language barriers.

3 PicNet

PicNet (Borman et al., 2005) is a knowledge-base consisting of dual visual-linguistic representations for words and phrases – seen as cognitive units encoding the smallest units of communication. Starting with a machine readable dictionary that defines the words in the common vocabulary and their possible meanings, PicNet adds visual representations to the dictionary entries, to the end of building a resource that combines the linguistic and pictorial representations of basic concepts.

PicNet relies on a Web-based system for augmenting dictionaries with illustrative images using volunteer contributions over the Web. The assumption is that all Web users are experts when it comes to understanding the content of images and finding associations between words and pictures. Given a word and its possible meanings – as defined by a comprehensive dictionary – Web users participate in a variety of game-like activities targeting the association of pictures with words.

The primary lexical resource used in PicNet is WordNet (Miller, 1995) – a machine readable dictionary containing a large number of concepts and relations between them. While the WordNet dictionary covers English concepts, it is also linked to a large number of dictionaries covering several Euro-

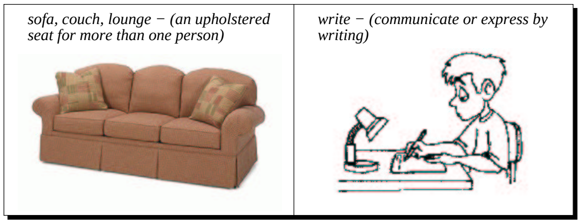

sofa, couch, lounge − (an upholstered seat for more than one person)
write − (communicate or express by writing)

Figure 3: Sample word/image associations from PicNet.

pean languages (Vossen, 1998), and to the Chinese HowNet dictionary (Carpuat et al., 2002).

Initially, PicNet was seeded with images culled from automated image searches using PicSearch (http://www.picsearch.com) and AltaVista (http://www.altavista.com/image), which resulted in 72,968 word/image associations automatically collected. Data validation is then performed by Web volunteers who can choose to participate in a variety of activities, including free association (assign a concept to a randomly selected image from the database), image upload (upload an image the user finds representative for a given concept), image validation (assign a quality vote to a randomly selected concept/image association from the PicNet dictionary), or competitive free association (a game-like activity where multiple users can simultaneously vote on a concept/image association).

3.1 Ensuring Data Quality

Collecting from the general public holds the promise of providing much data at low cost. It also makes attending to an important aspect of data collection: ensuring contribution quality. PicNet implements a scoring scheme that ranks concept/image pairs based on the total number of votes received from users of the various PicNet activities. A complete history of users' decisions is maintained and used to rank the concept/image associations. Each action provides an implicit quantified vote relating to the concept/image pair. The sum of these votes creates a score for the pair, allowing PicNet to rank images associated to a particular concept. The following list represents the possible actions that users can perform on the PicNet site, and the corresponding votes: Upload an image for a selected concept (+5); Image validation – well related to the concept (+4); Image validation – related to many concept attributes (+3); Image validation – loosely related to the concept (+1); Image validation – not related to the concept (-5); Free association (+3); Competitive free association (+n, where n is the number of users agreeing with the association).

3.2 PicNet Evaluations

Evaluations concerning the quality of the data collected through PicNet were conducted based on the concept/image associations collected up-to-date for approximately 6,200 concepts from 320 contributors. A manual inspection of 100 random concept/image pairs suggested that the scoring scheme is successful in identifying high quality associations, with about 85 associations found correct by trusted human judges. Figure 3 shows two sample concept/image associations collected with PicNet and their dictionary definitions. More details on the PicNet activities and evaluation are provided in (Borman et al., 2005).

In our picture translation experiments, PicNet is used to assign an image to basic nouns and verbs in the input sentence. Once again, no attempt is made to assign pictures to adjectives or adverbs. In addition to the image representations for nouns and verbs as collected through PicNet, we also use a set of pictorial representations for pronouns, using images from a language learning course[1].

4 A System for Automatic Pictorial Translations

The automatic translation of an input text into pictures is a non-trivial task, since the goal is to generate pictorial representations that are highly correlated with the words in the source sentence, thus effecting a level of understanding for the pictorial translations which would be comparable to that for the linguistic representations alone.

Starting with an input sentence, the text is tokenized and part-of-speech tagged (Brill, 1992), and word lemmas are identified using a WordNet-based lemmatizer. Next, we attempt to identify the most likely meaning for each open-class word using a publicly available state-of-the-art sense tagger that

[1] http://tell.fll.purdue.edu/JapanProj/FLClipart/

identifies the meaning of words in unrestricted text with respect to the WordNet sense inventory (Mihalcea and Csomai, 2005).

Once the text is pre-processed, and the open-class words are labeled with their parts-of-speech and corresponding word meanings, we use PicNet to identify pictorial representations for each noun and verb. We supply PicNet with the lemma, part-of-speech, and sense number, and retrieve the highest ranked picture from the collection of concept/image associations available in PicNet. To avoid introducing errors in the pictorial translation, we use only those concept/image associations that rank above a threshold score of 4, indicating a high quality association.

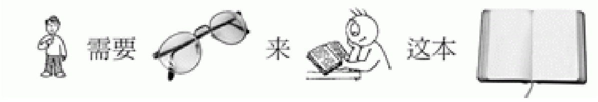
5 Experiments and Evaluation

Through our experiments, we target the evaluation of the translation quality for each of the three translation scenarios described before.

We created a testbed of 50 short sentences, consisting of 30 randomly selected examples from language learning courses, and 20 sentences from various domain-specific texts covering fields such as e.g. finance, sports, or travel. While all the sentences in our testbed are short, with an average of 15 words each, they have various levels of difficulty, ranging from simple basic vocabulary taught in beginner language classes, to more complex sentences containing domain-specific vocabulary.

Although our translation system, as described in Section 4, is designed to work with English as a source language, in order to facilitate the evaluations we have also created a Chinese version of the sentences in our data set[2]. The reason for using Chinese (rather than English) as the source “unknown” language was to ensure the fairness of the evaluation: since this research was carried out in an English-speaking country, it was difficult to find users who did not speak English and who were completely unaware of the peculiarities of the English language. Instead, by using Chinese as the source language, we were able to conduct an evaluation where the users interpreting the pictorial representations were not aware of any of the specifics of the source language (such as vocabulary, word order, or the syntactic structure specific to Chinese).

Starting with the Chinese version of each sentence in our data set, three translations were generated: (1) A pictorial translation, where verbs, nouns, and pronouns are represented with pictures, while the remaining context is represented in Chinese[3] (no pictorial translations are generated for those verbs or nouns not available in PicNet). (2) A mixed pictorial and linguistic translation, where verbs, nouns, and pronouns are still represented with pictures, but the context is represented in English. (3) A linguistic translation, as obtained from a machine translation system (Systran http://www.systransoft.com), which automatically translates the Chinese version of each sentence into English; no pictorial representations are used in this translation.



Interpretation 1: *I use glasses to read my books.*
Interpretation 2: *I need glasses to read a book.*
Interpretation 3: *I need my eye glasses to read this book.*

Figure 4: Various interpretations by different users for a sample pictorial translation.

Each of the three translations is then shown to fifteen different users, who are asked to indicate in their own words their interpretation of the visual and/or linguistic representations. For instance, Figure 4 shows a pictorial translation for the sentence *“I need glasses to read this book,”* and three interpretations by three different users[4].

[2] This represents the ”unknown” language for the translation evaluations below. The translation was generated by two native Chinese speakers, through several iterations until an agreement was reached.

[3] The pictorial translations, automatically assigned to the English version of each sentence, were manually assigned to the concepts in the Chinese sentence. It is important to note that this step was required exclusively for the purpose of conducting the evaluations. In the general case, the pictorial translations are automatically assigned to a source English sentence, and used as such in the communication process. However, since we wanted to circumvent the problem of all the users available for our study being English speakers, we chose to conduct the evaluations using a language different than English (and consequently selected Chinese as the source language).

[4] A pictorial representation was not used for the verb *“need”*, since no image association was found in PicNet for this concept.

5.1 Evaluation Metrics

To assess the quality of the interpretations generated by each of the three translation scenarios described before, we use both manual and automatic assessments of quality, based on metrics typically used in machine translation evaluations.

First, we use a human evaluation of quality, consisting of an adequacy assessment. A human judge was presented with the correct reference translation and a candidate interpretation, and was asked to indicate how much of the information in the gold standard reference translation was preserved in the candidate interpretation. The assessment is done on a scale from 1 ("none of it") to 5 ("all the information")[5].

Second, we use two automatic evaluations of quality traditionally used in machine translation evaluation. The NIST evaluation (Doddington, 2002) is based on the Bleu score (Papineni et al., 2002). It is an information-weighted measure of the precision of unigrams, bigrams, trigrams, fourgrams, and five-grams in the candidate interpretations with respect to the "gold-standard" reference translation. The other metric is the GTM score (Turian et al., 2003), which measures the similarity between texts in terms of precision, recall, and F-measure. Both measures were found to have good performance at discriminating translation quality, with high correlations to human judgments.

5.2 Results

For each sentence in our testbed and for each possible translation, we collected interpretations from fifteen different users, accounting for a total of 2,250 interpretations. No Chinese speakers were allowed to participate in the evaluations, since Chinese was the "unknown" language used in our experiments. The user group included different ethnic groups, e.g. Hispanics, Caucasians, Latin Americans, Indians, accounting for different cultural biases. While all the users were accustomed to the American culture (all of them having lived in the United States for two or more years), only a small fraction of them were English native speakers.

All the interpretations provided by the users were scored using the three evaluation measures: the GTM F-measure and the NIST scores, and the manually assessed adequacy. Table 1 shows the evaluation results, averaged across all users and all sentences.

	Evaluation		
	automatic		manual
Type of translation	NIST (Bleu)	GTM	Adequacy
S1: Pictures	41.21	32.56	3.81
S2: Pictures+linguistic	52.97	41.65	4.32
S3: Linguistic	55.97	44.67	4.40

Table 1: Results for the three translation scenarios, using automatic and manual evaluation criteria. Standard deviations were measured at: 7.80 for the NIST score, 6.30 for the GTM score, and 0.31 for the adequacy score.

The lower bound is represented by the "no communication" scenario (no language-based communication between the two speakers), corresponding to a baseline score of 0 for all the translation scores. For the human adequacy score, the upper bound consists of a score of 5, which reflects a perfect interpretation. For the NIST and the GTM scores, it is difficult to approximate an upper bound, since these automatic evaluations do not have the ability to account for paraphrases or other semantic variations, which typically get penalized in these scores. Previous evaluations of a NIST-like score on human-labeled paraphrases led to a score of 70%, which can be considered as a rough estimation of the upper bound.

5.3 Discussion

The results indicate that a significant amount of the information contained in simple sentences can be conveyed through pictorial translations. The human adequacy score of 3.81, also reflected in the automatic NIST and GTM scores, indicate that about 76%[6] of the content can be effectively communicated using pictures. This score is explained by the

[5] Traditionally, human evaluations of machine translation quality have also considered fluency as an evaluation criterion. However, since we measure the quality of the *human*-produced *interpretations* (rather than measuring directly the quality of the automatically produced translations), the interpretations are fluent, and therefore do not require an explicit evaluation of fluency.

[6] The fraction of the adequacy score for pictorial translations (3.81) divided by the maximum adequacy score (5.00).

intuitive visual descriptions that can be assigned to some of the concepts in a text, and by the humans ability to efficiently *contextualize* concepts using their background world knowledge. For instance, while the concepts *read* and *book* could also lead to a statement such as e.g. *"Read about a book,"* the most likely interpretation is *"Read a book,"* which is what most people will think of when seeing the pictorial representations of these two concepts.

5.3.1 Data Analysis

In an attempt to understand the level of difficulty associated with the understanding of pictorial translations for different sentence types, we performed a detailed analysis of the test set, and measured the correlation between various characteristics of the test sentences and the level of understanding achieved during the sentence interpretation experiments. Specifically, given a sentence feature (e.g. the number of words in a sentence), and an evaluation score for translation quality (e.g. the NIST score), we determined the Pearson correlation factor (r) between the feature considered and the quality of the interpretation. In all the correlation experiments, we report correlation measures using the NIST evaluation scores, but similar correlation scores were observed for the other evaluation metrics. As typically assumed in previous correlation studies, a Pearson factor of $0.10 - 0.29$ is associated with a *low* correlation, $0.30 - 0.59$ represents a *medium* correlation, and $0.60 - 1.00$ is considered *high* correlation.

Based on correlation analyses for a number of features, the following observations were drawn.

Sentence length. There is a high negative correlation ($r = -0.67$) between the number of words in a sentence and the level of understanding achieved for the pictorial translations. This suggests that the understanding of pictorial translations increases with decreasing sentence length. Our pictorial translation paradigm is therefore most effective for short sentences.

Ratio of words with a given part-of-speech. There is a medium positive correlation ($r = 0.44$) between the proportion of nouns in a sentence and the level of understanding, and a medium negative correlation ($r = -0.47$) between the number of function words and the quality of interpretation, indicating that sentences that are "dense" in concepts (large number of nouns, small number of function words) are easier to understand when represented through pictures.

Syntactic complexity. We modeled syntactic complexity by counting the number of different syntactic phrases (e.g. noun phrases), and by determining the high-level structure of the syntactic parse tree (e.g. subject-verb, subject-verb-indirect_object, etc.). We found that the understanding of pictorial translations decreases with increasing syntactic complexity, with a medium negative correlation observed between the number of noun-phrases ($r = -0.49$) or prepositional phrases ($r = -0.51$) in a sentence and the quality of interpretation. Although no significant correlation was found between the level of understanding of a pictorial translation and the structure of the syntactic parse tree, on average better interpretations were observed for sentences with a complete subject-verb-direct_object structure (as compared to e.g. sentences with a subject-verb structure).

Semantic classes. Using the semantic classes from WordNet (26 semantic classes for nouns and 15 semantic classes for verbs), we determined for each sentence the number of concepts belonging to each semantic class, and measured the correlation with the level of understanding for pictorial translations. We found a low positive correlation ($r = 0.20 - 0.30$) associated with the number of nouns belonging to the semantic class "animal" (e.g. *dog*) and "communication" (*e.g. letter*) and the verbs from the semantic classes of "cognition" (e.g. *read*) and "consumption" (e.g. *drink*). No significant correlations were found for the other semantic classes.

Word frequency. For each of the sentences in the test set, we determined the frequency of each constituent word (excluding stopwords) using the British National Corpus. These word frequencies were then combined into a score which, after normalization with the length of the sentence, reflects the usage frequency for the concepts described in a sentence. We found a medium positive correlation ($r = 0.38$) between the combined frequency of the words in a sentence and the level of understanding for pictorial translations, suggesting that it is easier to understand and interpret the pictorial representations associated with frequently used words.

5.3.2 Translation Score Analysis

An analysis of the translation scores listed in Table 1 reveals interesting aspects concerning the amount of understanding achieved for different translation scenarios.

The score achieved through the pictorial translations alone (S1) represents a large improvement over the score of 0 for the "no communication" baseline (which occurs when there are no means of communication between the speakers). The score achieved by this scenario indicates the role played by conceptual representations (pictures) in the overall understanding of simple sentences.

The difference between the scores achieved with scenario S1 (pictorial representations) and scenario S2 (mixed pictorial and linguistic representations) points out the role played by context that cannot be described with visual representations. Adjectives, adverbs, prepositions, abstract nouns and verbs, syntactic structure, and others constitute a linguistic context that cannot be represented with pictures, and which nonetheless has an important role in the communication process.

Finally, the gap between the second (S2) and the third (S3) scenarios indicates the advantage of words over pictures for producing accurate interpretations. Note however that this is a rather small gap, which suggests that pictorial representations placed in a linguistic context are intuitive, and can successfully convey information across speakers, with an effectiveness that is comparable to full linguistic representations.

There were also cases when the pictorial representations failed to convey the desired meaning. For instance, the illustration of the pronoun *he*, a *riverbank*, and a *torch* (for *He sees the riverbank illuminated by a torch*) received a wrong interpretation from most users, perhaps due to the unusual, not necessarily commonsensical association between the *riverbank* and the *torch*, which most likely hindered the users ability to effectively contextualize the information.

Interestingly, there were also cases where the interpretation of the pictorial translation was better than the one for the linguistic translation. For instance, the Chinese sentence for *I read email on my computer* was wrongly translated by the machine translation system to *I read electricity on my computer post.* which was misleading, and led to an interpretation that was worse than the one generated by the illustration of the concepts of *I*, *read*, *email*, and *computer*.

Overall, while pictorial translations have limitations in the amount of information they can convey, the understanding achieved based on pictorial representations for simple short sentences was found to be within a comparable range of the understanding achieved based on an automatic machine translation system, which suggests that such pictorial translations can be used for the purpose of communicating simple pieces of information.

6 Related Work

Early research efforts in cognitive science and psychology (Potter et al., 1986) have shown that a picture can successfully replace a noun in a rapidly presented sentence, without any impact on the interpretation of the sentence, nor on the speed of understanding, suggesting that the human representation of word meanings is based on a conceptual system which is not tied to a given language.

Work has also been done on the design of iconic languages for augmentative communication for people with physical limitations or speech impediments, with iconic keyboards that can be touched to produce a voice output for communication augmentation (Chang and Polese, 1992). Also related to some extent is the work done in visual programming languages (Boshernitsan and Downes, 1997), where visual representations such as graphics and icons are added to programming languages to support visual interactions and to allow for programming with visual expressions.

Another area related to our work is machine translation, which in recent years has witnessed significant advances, with large scale evaluations and well attended events organized every year. Despite this progress, current machine translation techniques are still limited to the translation across a handful of languages. In particular, statistical methods are restricted to those language pairs for which large parallel corpora exist, such as e.g. French-English, Chinese-English, or Arabic-English. Dealing with morphologically complex languages (e.g.

Finnish), languages with partial free word order (e.g. German), or languages with scarce resources (e.g. Quechua) prove to be very challenging tasks for machine translation, and there is still a long way to go until a communication means will be available among all the languages spoken worldwide.

Finally, a significant amount of research work has been done in automatic image captioning (e.g. (Barnard and Forsyth, 2001), (Pan et al., 2004)). This topic is however outside the goal of our current study, and therefore not overviewed here.

7 Conclusions

Language can sometime be an impediment in communication. Whether we are talking about people who speak different languages, students who are learning a new language, or people with language disorders, the understanding of linguistic representations in a given language require a certain amount of knowledge that not everybody has.

In this paper, we described a system that can generate pictorial representations for simple sentences, and proposed "translation through pictures" as a means for conveying simple pieces of information across language barriers. Comparative experiments conducted on visual and linguistic representations of information have shown that a considerable amount of understanding can be achieved through pictorial descriptions, with results within a comparable range of those obtained with current machine translation techniques.

Future work will consider the analysis of more complex sentences of various degrees of difficulty. Cultural differences in picture interpretation are also an interesting aspect that we plan to consider in future evaluations.

References

K. Barnard and D.A. Forsyth. 2001. Learning the semantics of words and pictures. In *Proceedings of the IEEE International Conference on Computer Vision*.

A. Borman, R. Mihalcea, and P. Tarau. 2005. Picnet: Augmenting semantic resources with pictorial representations. In *Proceedings of the AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors*, Stanford, CA.

M. Boshernitsan and M. Downes. 1997. Visual programming languages: A survey. Technical report, U.C. Berkeley.

E. Brill. 1992. A simple rule-based part of speech tagger. In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, Trento, Italy.

M. Carpuat, G. Ngai, P. Fung, and K. Church. 2002. Creating a bilingual ontology: A corpus-based approach for aligning WordNet and HowNet. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, Taipei, Taiwan, August.

S. Chang and G. Polese. 1992. A methodology and interactive environment for iconic language design. In *Proceedings of IEEE workshop on visual languages*.

G. Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of Human Language Technology HLT-2002*, San Diego.

2005. http://www.ethnologue.com.

W.W. Gibbs. 2002. Saving dying languages. *Scientific American*, pages 79–86.

R. Mihalcea and A. Csomai. 2005. Senselearner: Word sense disambiguation for all words in unrestricted text. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics*, Ann Arbor, MI.

G. Miller. 1995. Wordnet: A lexical database. *Communication of the ACM*, 38(11):39–41.

J.Y. Pan, H.J. Yang, C. Faloutsos, and P. Duygulu. 2004. Gcap: Graph-based automatic image captioning. In *Proceedings of the 4th International Workshop on Multimedia Data and Document Engineering (MDDE)*, Washington, DC.

K. Papineni, S. Roukos, T. Ward, and W. Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, Philadelphia, PA, July.

M.C. Potter, J.F. Kroll, B. Yachzel, E. Carpenter, and J. Sherman. 1986. Pictures in sentences: understanding without words. *Journal of Experimental Psychology*, 115(3).

J. Turian, L. Shen, and I. D. Melamed. 2003. Evaluation of machine translation and its evaluation. In *Proceedings of the Machine Translation Summit*, New Orleans.

P. Vossen. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht.