picoTrans: A User Interface Using Content Word-based Source Sentence Generation for Machine Translation

Wei SONG
University of Tokyo
Graduate School of Science and Technology
song@cl.ci.i.u-tokyo.ac.jp

Andrew FINCH
NICT
Language Translation Group
andrew.finch@nict.go.jp

Kumiko TANAKA-ISHII
University of Tokyo
Graduate School of Science and Technology
kumiko@i.u-tokyo.ac.jp

Eiichiro SUMITA
NICT
Language Translation Group
eiichiro.sumita@nict.go.jp

1 Introduction

This is the abstract of the IUI paper (Song et al., to appear in 2011). Our proposal combines statistical machine translation with the idea of a picture-based translation-aid (Graf, 2009; Meader, 1995; Warrink, 2007; Stillman, 2007; Flanagan, 2008). Picture-based translation-aids have been used in paper book forms and are currently integrated into hand-held devices but remain uncombined with machine translation systems. Briefly, in our proposed system picoTrans, the user taps picture icons appearing on the touch-screen, just like in a picture-based translation-aid. The system automatically generates the possible sentences from those selected icons, and feeds them to the machine translation in order that it can display the translated result. For example, suppose a user wished to translate the expression 'I want to go to the restaurant. ', with the picture book, the user might point at 2 pictures: 'I want to to go to ∼', and 'restaurant'. A similar scenario for our system is that the user points to a sequence of icons, however, in our case the sequence of icons is maintained on the display for the users to see, and interact with if necessary. When the input is complete, the system generates the full sentence in the source language automatically, which is then translated by the machine translation software and displayed on the screen together with the icon sequence.

The combination of the two approaches is advantageous from the perspective of both the user interface and machine translation. Firstly, from the user interface viewpoint, the major bottleneck of hand-held devices is the difficulty of text entry (MacKenzie and Tanaka-Ishii, 2007). There have been many text entry systems proposed for small devices (Sirisena, 2002), but still the entry of full sentences is a cumbersome process. Entry by the tapping of icons allows entry of words in only at most a few taps, decreasing the number of actions needed to enter a sentence, thus increasing the efficiency of entry.

Secondly, from the machine translation viewpoint, such entry by icon taps serves as a means of 'standardizing', or 'normalizing' the sentences to be translated. This is advantageous for the translation system, since a major cause of translation error arises from use of rare words or infrequent sentence forms. Most importantly, related to the entry inefficiency, users will tend to enter sentences in an abbreviated form that often includes spelling mistakes on hand-held devices. Such entries are especially difficult for a machine translation system to process. On the other hand, an icon-only approach provides conventional wordings/phrases with no typographical errors and therefore the machine translation will not suffer from errors arising from surface form variation.

2 USER INTERFACE

A diagram of the user interface in full is shown in Figure 1. In brief, we allow the user to input what they wish to express as a sequence of bi-lingually annotated icons. This is in essence the same idea as the picture-book. The form of the translation process proceeds as follows:

(1) The user selects a category for the concept they wish to express

(2) The user selects a sub-category

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

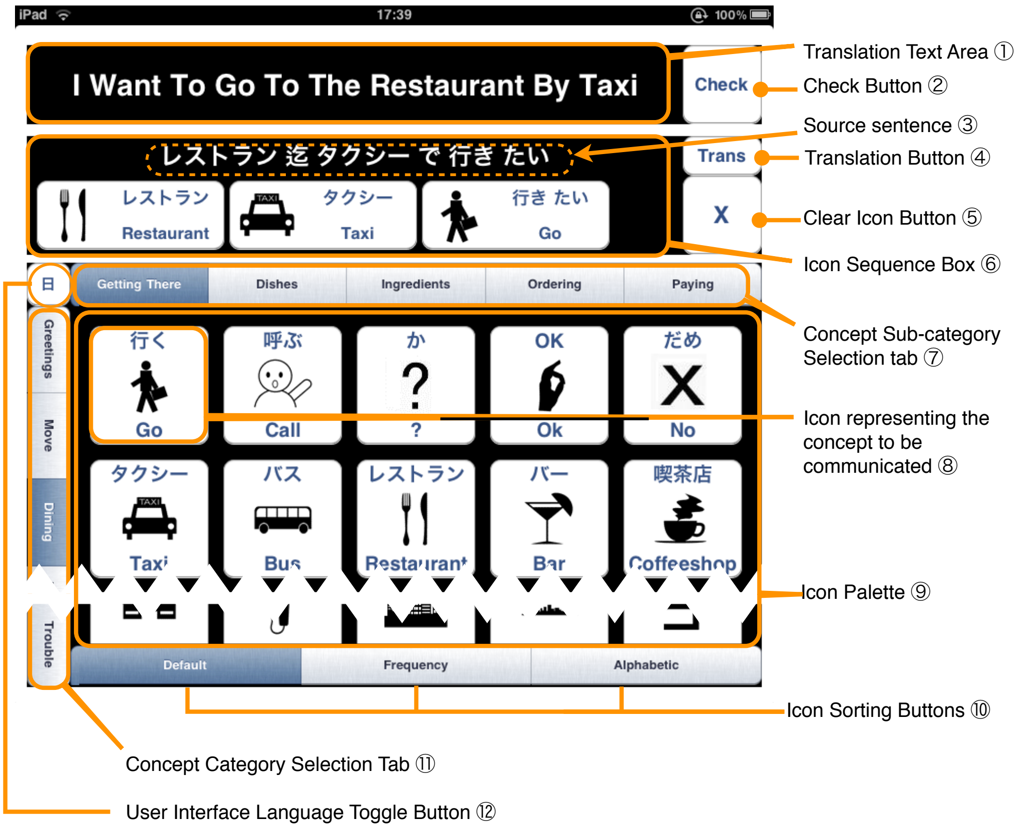

Figure 1: The annotated user interface for the picoTrans system.

(3) The user chooses the first icon in the sequence

a) Go to (1) to select another icon for the sequence

b) The icon sequence is complete, and the corresponding source sentence is acceptable. Continue to step (4) or refine the source sentence

(4) The user clicks the 'Trans' button

(5) The translation appears in the translation text area

3 System Architecture

In this section we explain the 3 components of the system, and the process of interaction between them: the user interface client, the source sentence generation server, and the machine translation system servers.

3.1 User Interface Client

The user interaction is made through an interface which is currently implemented as a prototype working on the Apple iPad mobile tablet.

3.2 Source Sentence Generation

In order to provide an input word sequence for the machine translation system, we generate a source word sequence from the icon sequence. Intuitively we build the source word sequence from pieces (contiguous word sequences) of sentences taken from a large corpus of source language sentences, we will call these pieces 'phrases'. We fit these phrases together using a statistical language model that we will describe later in this section. First we will explain how we extract the phrases from the source language side of the corpus.

3.2.1 Source Phrase Extraction

A phrase-based statistical machine translation system, generates its translations using a set of bi-

Copyright(C) 2011 The Association for Natural Language Processing. All Rights Reserved.

lingual phrase pairs, known as a *phrase table*. The source and target word sequences can be thought of as translations of each other, at the phrasal level. In our system we chose to use the source side of the phrase-pairs in the phrase-table as the basis for the phrases we associate with our icons, and we restricted the phrases extracted to be uni-grams and bi-grams. For each icon, a content word that represents this icon was chosen and a set consisting of the most frequent source phrases containing this content word was selected from the corpus. In a final step, this set of possible phrases was cleaned by a human annotator, as the statistical process used to build the machine translation system's phrase table can often extract incorrect pairs, or correct pairs that would be of no use in our system. Note that since we use the source parts of the phrase-pairs to form the source sentence to be given to the machine translation system, we are in effect guaranteeing that the machine translation system will have phrase pairs for these segments to use in the translation process; we are simplifying the translation task by forcing the input to conform closely with the translation model of the machine translation system.

3.2.2 Language Modeling

Given a sequence of source phrases representing the sequence of icons chosen by the user, we need to generate a well-formed sentence in the source language using them. In the prototype system reported in this paper, we use Japanese as the source language. Japanese is peculiar in that it has a relatively free word order, and uses particles adjacent to the main content words in the sentence to indicate their function. This motivates our choice of bigram units as the basic phrasal unit, as these were expected to include a single content word together with a particle that modifies it. In our experiments we used the SRI language modeling toolkit (Stolcke, 1999) to implement our language model.

In the natural language processing field, language models are commonly used to ensure that generated text is well-formed. In our system we generate all possible combinations of phrases for each icon in the icon sequence, and score the resulting word sequence hypotheses with a language model. We present the highest scoring hypothesis to the user as the source sentence to be translated. In the icon refinement process we use the language model in a similar manner; the user is presented with a list of the top-5 (partial) hypotheses representing the start of the source sentence up to and including the phrase associated with the icon that was tapped by the user.

3.3 Machine Translation

For our experiments we use CleopATRa (Finch et al., 2007), an in-house machine translation decoder that is based on the phrase-based machine translation techniques.

4 Evaluation

One of the main concerns about the technique proposed in our system is its expressive power within the domain, since sentences need to be expressed by only using icons that are available on the device. We therefore conducted an evaluation of the user interface to determine the proportion of in-domain sentences it was capable of representing. To do this we took a sample of 100 sentences from a set of held-out data drawn from the same sample as the training corpus, and determined whether it was possible to generate a semantically equivalent form of each sentence using the icon-driven interface and its source sentence generation process. The current version of the prototype has not been developed sufficiently to include sets of icons to deal with numerical expressions (prices, phone numbers, dates and times etc.), so we excluded sentences containing numerical expressions from our evaluation set (the evaluation set size was 100 sentences after the exclusion of sentences containing numerical expressions). Handling numerical expressions is relatively straightforward however, and we do not foresee any difficulty in adding this functionality into our system in the future. The set of icons used in the evaluation corresponded to the most 2010 frequent content words in the English side of the training corpus, that is content words that occurred more than 28 times in the corpus. Thus value was chosen such that the number of icons in the user interface was around 2000, a rough estimate of the number of icons necessary to build a useful real-world application. We found that we were able to generate semantically equivalent sentences for 74% of the sentences in our evaluation data, this is shown in Figure 2 together with statistics (based on a 30-sentence random sample from the 100 evaluation sentences) for cases where fewer icons were used. We feel this is an high level of cov-

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.

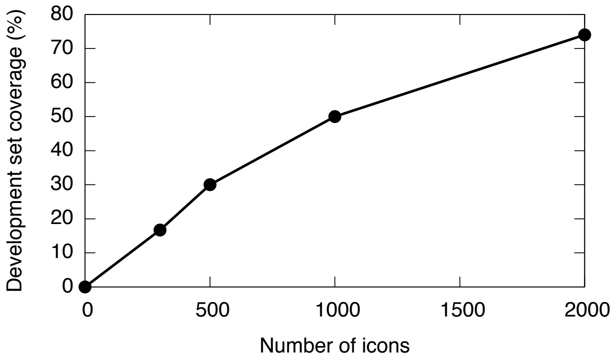

Figure 2: The coverage of unseen data with icon set size, with human interaction.

erage given the simplifications that have been made to the user interface. For 49 of the 74 sentences that we were able to cover with our system (66% of them), the system proposed the correct source sentence to the user, and no icon refinement was necessary.

We also studied the number of key-press actions needed to enter these sentences using icons relative to the number that would have been needed to input them using the device's text-entry interface. Here we assumed that each icon would require 3 key presses to select, but often the icons from the same icon subcategory can be used, and these icons would only require 1 key press, so our estimate represents an upper-bound for the number of key press actions necessary. The time required for one key press isn't equal for icon input and text input, and we did not measure this in our experiments. We also made no attempt to measure of effect of user input errors on the input process. Measuring these factors remains future work. Our measurements include the additional key presses needed to select the semantics of ambiguous icons, and also the key presses necessary to modify the source sentence to have the intended meaning.

In our experiments we found that the icon entry system required only 57% of the number of key press actions of the text entry method: 941 key presses for the icon-driven input method as opposed to 1650 for text entry.

5 Conclusion

In this paper we have presented a novel user interface that integrates ideas from two different paradigms of translation for travelers: picture-books and statistical machine translation. Our approach offers all of the advantages of the simplistic but powerful representation of the picture-books, and at the same time is able to produce natural language in the target language able to unambiguously express the source language user's meaning. The resulting system is both more expressive than the picture-book approach, and at the same time mitigates the problems due to errors in the machine translation system by facilitating more accurate translation.

References

Andrew Finch, Etienne Denoual, Hideo Okuma, Michael Paul, Hirofumi Yamamoto, Keiji Yasuda, Ruiqiang Zhang, and Eiichiro Sumita. 2007. The nict/atr speech translation system for iwslt 2007. In *In Proceedings of the IWSLT*, Trento, Italy.

Cheryn Flanagan. 2008. *Me No Speak: China*. Me No Speak.

Dieter Graf. 2009. *Point it: Traveller's Language Kit - The Original Picture Dictionary - Bigger and Better (English, Spanish, French, Italian, German and Russian Edition)*. Graf Editions.

Scott MacKenzie and Kumiko Tanaka-Ishii, editors. 2007. *Text Entry Systems —Accessibility, Mobility, Universality—*. Morgan Kaufmann.

Jonathan Meader. 1995. *The Wordless Travel Book: Point at These Pictures to Communicate with Anyone*. Ten Speed Press.

Amal Sirisena. 2002. Mobile text entry. *Retrieved July*, 27.

Wei Song, Andrew Finch, Kumiko Tanaka-Ishii, and Eiichiro Sumita. to appear in 2011. picotrans: An icon-driven user interface for machine translation on mobile devices. In *IUI '11: Proceeding of the 15th international conference on Intelligent user interfaces*.

Stillman. 2007. *Kwikpoint International Translator (English, Spanish, French, Italian, German, Japanese, Russian, Ukrainian, Chinese, Hindi, Tamil, Telug, Kannada, Malayalam, Gujarati, Bengali and Korean Edition)*. Kwikpoint.

Andreas Stolcke. 1999. Srilm - an extensible language model toolkit.

Gosia Warrink. 2007. *ICOON Global Picture Dictionary (English, Spanish, French, Italian, German, Japanese, Russian, Chinese and Hindi Edition)*. Amberpress.

Copyright(C) 2011 The Association for Natural Language Processing.
All Rights Reserved.