User adaptation in a hybrid MT system: Feeding user corrections into synchronous grammars and system dictionaries

Susanne Preuß[1], Hajo Keffer[1], Paul Schmidt[1], Georgios Goumas[2] Athanasia Asiki[2], and Ioannis Konstantinou[2]

[1] GFAI – Gesellschaft zur Förderung der Angewandten Informationsforschung
http://www.iai-sb.de/iai/
[2] National Technical University of Athens
School of Electrical and Computer Engineering
Computing Systems Laboratory
http://www.cslab.ece.ntua.gr

Abstract. In this paper we present the User Adaptation (UA) module implemented as part of a novel Hybrid MT translation system. The proposed UA module allows the user to enhance core system components such as synchronous grammars and system dictionaries at run-time. It is well-known that allowing users to modify system behavior raises the willingness to work with MT systems. However, in statistical MT systems user feedback is only 'a drop in the ocean' in the statistical resources. The hybrid MT system proposed here uses rule-based synchronous grammars that are automatically extracted out of small parallel annotated bilingual corpora. They account for structural mappings from source language to target language. Subsequent monolingual statistical components further disambiguate the target language structure. This approach provides a suitable substrate to incorporate a lightweight and effective UA module. User corrections are collected from a post-editing engine and added to the bilingual corpus, whereas the resulting additional structural mappings are provided to the system at run-time. Users can also enhance the system dictionary. User adaptation is organized in a user-specific commit-and-review cycle that allows the user to revise user adaptation input. Preliminary experimental evaluation shows promising results on the capability of the system to adapt to user structural preferences.

1 Introduction

It is well-known that allowing users modify system behavior raises their willingness to work with MT systems. Thus, User Adaptation (UA) has received remarkable attention in recent years. The PACO report [1] provides a survey on post editing tools and argues that UA is an important feature for increasing the acceptance of MT systems. Most MT systems, however, have conceptual problems to provide UA. Statistical MT (SMT) systems face the problem that small-scale user feedback has insignificant effect on system resources that are based on large data sets. Therefore, SMT systems restrict their UA functionality to shallow string manipulation.On the other hand, rule-based MT (RBMT) systems face the problem that the rule components are often manually written which requires sophisticated linguistic skills that cannot be expected by the end-users. Therefore, classical RBMT also provides only shallow string manipulation operations as UA or restricted facilities for bilingual lexicon extension.

The FAUST project explicitly addresses the problem of user feedback. The project's goal is to develop interactive MT systems which adapt rapidly (!) and intelligently to user feedback. A major issue is to identify noise that users tend to produce when giving feedback to MT systems. The major achievement planned is to "develop mechanisms for instantaneously incorporating user feedback into the machine translation engines" [2]. For SMT this amounts to specially weighting user feedback. Whether this is different to string replacement remains to be seen.

In this paper we present a UA mechanism built on top of a novel, Hybrid MT system that makes use of already available tools and lexica, automatically built large monolingual corpora and synchronous grammas, and easily assembled small bilingual corpora. A key goal of the MT system is to provide a full translation path for new language pairs with significantly reduced implementation effort. Adhering to this philosophy, the UA module is capable of providing a lightweight and effective mechanism that allows users modify system components towards a system adapted to their needs.

2 A hybrid MT system

Design philosophy: The key idea of the hybrid MT system is to use publicly available tools and resources so that development time and costs are kept low and the system can be easily extended to new language pairs without trading translation quality. Such tools are statistical taggers and lemmatizers and statistical chunkers. Large monolingual corpora for target languages (TL) are automatically built out by properly mining the world wide web. The only bilingual resources needed are small parallel bilingual corpora consisting of a few hundred sentences and bilingual dictionaries which are often available from publishers, at least if the system is not used commercially. Any other linguistic resources are derived from the resources listed above, mainly including SL models for statistical chunkers, synchronous grammars, statistical models for TL lemma disambiguation and TL token generation tables. For a detailed description see [3].

Translation process: The translation process consists of three steps:

1. *SL annotation*: The incoming SL sentences are annotated with the help of the statistical taggers, chunkers and clause chunkers[3].

[3] A pattern-based clause chunker has been developed that recognizes SVO and SOV patterns and can be easily adapted to new languages by specifying tag classes such as the set of finite verb tags and the set of nominal tags.

2. *Bilingual structure generation*: The synchronous grammar takes the annotated SL structures as input and generates a set of TL structures, including TL lemmas and tags.
3. *TL disambiguation*: Monolingual statistical models disambiguate the TL structures. For the best TL structure, tokens are generated.

Thus the synchronous grammar component is sandwiched between two statistical components that do SL disambiguation and TL disambiguation. This architecture makes it viable to work with relatively small bilingual corpora for synchronous-grammar generation. The resulting grammar need not be specific enough to do SL or TL disambiguation.

Synchronous grammar component: The limited size of the bilingual corpus does not allow statistically derived synchronous grammars as in [4]. Also, unlike other tree-to-tree translation approaches [5], [6], [7], the system proposed does not use deep syntactic processing of the corpus data. Instead, shallow parsers are used to annotate the small bilingual corpus. Thus, the challenge is to extract the richness of cross-linguistic rules and generalizations out of a very limited set of flatly structured and monolithically annotated bilingual sentence pairs.

To extend the coverage of the synchronous grammar beyond the patterns found in the bilingual corpus, the sentential chunk and tag alignments are broken down into the smallest self-contained alignments[4] and then converted into productions[5]. As an example for structural divergences between languages, consider the case of separable prefix verbs in German. The German verb 'annehmen' ('accept') is split into 'nehmen' and 'an' in certain constellations. Thus there is a structural change in the translation from English to German:

[4] Definition of minimal self-contained alignment:
If $SL_k \ldots SL_i$ is a sequence of source language elements and $TL_x \ldots TL_y$ is a sequence of target language elements, with k, i, x, y, n and $m \in Na$ and $k \leq i$ and $x \leq y$, then $SL_k \ldots SL_i$, $TL_x \ldots TL_y$ is a minimal self-contained alignment for the elements SL_k and TL_x if i and y are the smallest natural numbers for which the following holds:

1. each element in $SL_k \ldots SL_i$ is aligned to an element in $TL_x \ldots TL_y$ and
2. each element in $TL_x \ldots TL_y$ is aligned to an element in $SL_k \ldots SL_i$ and
3. no element SL_{k-n} is aligned to an element TL_{y+m} and
4. no element SL_{i+n} is aligned to an element TL_{x-m}

[5] The system employs two more strategies to extend the coverage of the synchonous grammar beyond the patterns found in the bilingual corpus:

Tag reduction: The complex tags of morphologically rich languages such as German are reduced to forms that contain only the information that is relevant for the translation process. Thus otherwise derivable agreement information is deleted.

Equivalence classes of tags: The productions are multiplied out according to equivalency classes of tags. I.e. finite verb tags form an equivalence class. Thus, if a production exists for a 3. person singular present tense tag, the corresponding productions for all other persons, numbers and tenses are generated.

SL: John accepted the offer from Mary.
TL: John nahm das Angebot von Mary an.
Gloss: 'John took the offer of Mary on.'

The diverging structures can be represented schematically as shown in Fig. 1 where B represents the verbal constituents.

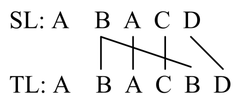

Fig. 1. Alignments

$$MO_1 \Rightarrow A_2 MO_3 \Leftrightarrow MO_1 \Rightarrow A_2 MO_3$$
$$MO_1 \Rightarrow B_2 A_3 C_4 MO_5 \Leftrightarrow MO_1 \Rightarrow B_2 A_3 C_4 B_2 MO_5$$

Fig. 2. Productions for isomorphic and complex alignments

The minimal self-contained alignments are converted into productions. Productions have the following format: 'SL rule(s)' $\Leftrightarrow$ 'TL rule(s)'. Subscripts indicate alignments. MO is the mother node. If the sequences consist of chunks, then MO is a clause node, if the sequences consist of tags, then MO is a chunk node. For each production a recursive and non-recursive variant is generated. For example, the isomorphic alignment of A – A and the complex alignment of BAC – BACB are converted into the recursive productions shown in Fig. 2[6]. Thus, structural identity between SL and TL results in hierarchical trees, whereas structural divergence results in flat trees.

In order to process the synchronous grammars, a simple Earley chart parser is adopted. The Earley chart parser takes the SL side of the bilingual productions to build up SL tree structures. The Earley chart parser also does a lexicon lookup in the bilingual dictionary. The TL side of the productions and the TL lemmas found in the lexicon are stored in specific features in the SL trees. Then the information on TL structure and TL lemmas is unpacked and a set of TL structures including TL lemmas is generated.

3 User adaptation in the Hybrid MT system

To address the challenges of UA, we first need to set a solid basis of key decisions, then provide a system design, and finally implement an effective and easy-to-use to system.

User adaptation philosophy: When it comes to designing a module with a goal to adjust a complex system to the user needs, a lot of subtle issues arise.

[6] In order to cut down the number of generated SL and TL tree structures, tag and lemma disjunctions are represented locally. Tag disjunctions contain frequencies of occurrences which are converted into probabilities that are taken into account by the TL disambiguation components. The frequencies account for minor tagging and alignment errors and foreground the most frequent cases.

Thus, a number of high-level decisions, forming a design philosophy, need to be taken. In the following paragraphs we discuss a number of them.

Make the frequent case fast: As mentioned above the ultimate goal of the UA module is to save users time. This is achieved by exploiting previous corrections to avoid future ones.

Involve the user: Ideally, the system should work transparently to the user. However, this cannot be the case for a complex, multi-user system. Wrong, funny, or malevolent corrections may contaminate the system and greatly degrade performance. Thus, it is advisable to involve the user in the adaptation process.

Keep the core system intact: The proposed system is implemented utilising a number of critical linguistic resources, such as language models, monolingual corpora and resources derived thereof, parallel corpora and bilingual lexica. A key decision in the UA approach is to keep these core system resources unaffected by user corrections in order to maintain the systems initial functionality.

User adaptation design: The proposed UA system targets adaptation regarding both word selection and general structural aspects of the translation. Users may correct single word translations or freely post-edit the sentence. The module design is based on the following key ideas:

- Freely post-edited translations affecting the parallel corpora and the 2nd translation step (bilingual structure generation)
- Word-level adaptation affecting the lexica and the 3rd translation step (TL disambiguation)

The information collected from users post-editing is stored in two types of resources, which are separate for each user:

- *User lexica*: When the user changes the translation of a single word, a new lexicon entry for the specific pair of SL and TL words will be added to the system per user. If the user decides that the new translation should always be used, then the existing entry will be replaced by the new one.
- *User parallel corpora*: The changes made by the user in the structure of a sentence are stored as a new pair of SL and TL sentence in a user parallel corpus. The UA module comprises components that interact with the systems post-processing module (and GUI in general), and components that are incorporated in the main translation engine of the system. Figure 3 shows the general design of the UA module and how user corrections are collected from the post-editing module and enhance user lexica and parallel corpora.

Supported functionality:

Word replacement: In the post-editing module users are able to replace a single word either by an alternative word provided by the system lexicon, or enter a new word. Users can freely modify the lemma form provided by the post-editing module into any inflected form. This action is logged by the UA module as a candidate for user adaptation in subsequent translations.

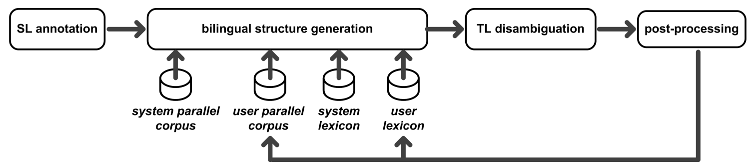

Fig. 3. The proposed UA design

Free text editing: Post-editing offers the functionality of free text editing. These changes are also logged by the UA module as a triplet: SL sentence, system TL sentence, and user TL sentence. The original system TL sentence is stored in order to facilitate quality control of the system translations after UA has been applied. Pre- and post-UA system translations can be compared.

Enable/disable UA: User adaptation is offered to registered users that can enable/disable the UA feature at will.

Review/commit changes: With UA enabled, and after a series of corrections, users are able to review corrections made since UA was enabled. In the case of replacing a word with an existing translation alternative, they are provided with the option to inform the system that the correction made was temporary (it affected only one translation) or that this replacement should replace subsequent occurrences of the SL word from now on and make it the preferred translation.

Versioning: Users are provided with the functionality to create snapshots of their adapted environment. Starting from the initial, pure system state (called vanilla) committed changes start to build the adapted environment, called working. After a period of satisfactory system behavior, the users may decide to backup the working adaptation environment into a version called stable.

Feeding user input back to the system: The additional sentence pairs are added to a user-specific copy of the bilingual corpus. The corresponding synchronous grammar productions and the dictionary update are done at runtime. The lemmas of the new lexical entries are determined by lemmatizing them in the sentence they occur in. Further linguistic annotation such as tags is not needed in the lexicon.[7].

4 Experimental evaluation scores

For an initial experiment, 265 sentences German - English have been taken from a multi-lingual EU web site. The sentences have been minimally adapted in order to get a sentence-aligned parallel corpus. Then the corpus has been split into a training corpus of 225 sentence pairs and a test corpus of 40 sentence pairs. The

[7] If the TL lemmas in the new lexical entries are not part of existing lexical entries, the token generation table and the TL lemma disambiguation models have to be updated as well since for efficiency reasons both are restricted to the TL lemmas found in the dictionary. The update of the statistical models is not done at run-time.

training corpus has been further split into randomly selected subsets consisting of 25 sentence pairs which have been subsequently added to the bilingual corpus of the MT-system thus simulating structural user adaptation. Figure 4 visualizes the evaluation scores in relation to corpus size.

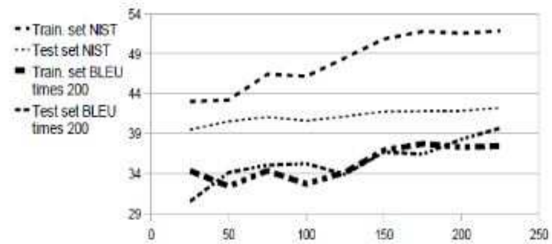

Fig. 4. Results for SL=German and TL=English translations

The evaluation scores confirm the contention that increasing the bilingual corpus improves the evaluation scores. However, the different development of BLEU and NIST scores needs further explanation. The development of the BLEU scores is expected, namely that the scores of the training set and the independent test set increase in the same order of magnitude, since in the present concept of structural user adaptation the training corpus is mainly used for mining grammatical structures which are expected to be roughly the same in the test set and the training set. The fact that the NIST scores of the training set increase to a much larger extent than the NIST scores of the test set is unexpected at first. However, considering that the NIST scores rank matches of content words higher than matches of grammatical function words, the hypothesis is that the content words of the training set are better translated because test and training set differ on the lexico-syntactic level. The German SL sentences contain many compounds and other complex lexico-syntactic expressions which translate into complex syntactic expressions in English, which have to be instantiated by bilingual productions. Thus, the difference in NIST scores is due to a difference in lexico-syntactic structures. This hypothesis is supported by the fact that allmost all of the corpus sentences provide new productions, namely 231 out of the 265. Further experiments are needed to examine whether the difference in NIST scores levels out at some point. In an additional experiment, 200 sentences randomly taken from large web corpora and translated by professional translators have been added to the corpus. They yield the same curves. While the additional experiment confirms that the domain of the corpus is irrelevant, the convergence point has not been reached yet.

In order to provide not only scores but also real system translations, consider the following example from the test set: The sentence *"Diese Übereinkunft könnte implizit von der Partei angenommen werden."* has as reference translation: *"This agreement could be implicitly adopted by the party."*. The system translations with 25, 50-100, 125 and 150-225 training sentences are: 25: *"This instrument* *knows implicitly of the party taken will."* (BLEU: 0.0000 NIST: 1.6610), 50-100: *"This instrument know implicitly of the party will be accepted."* (BLEU: 0.0000 NIST: 1.8120), 125: *"This implicitly arrangement may be taken from the party."* (BLEU: 0.0000 NIST: 1.9932), 150-225: *"Implicitly this agreement could be taken from the party."* (BLEU: 0.3928 NIST: 2.6575). Structurally the final translation is correct but the verb and preposition disambiguation does not provide the intended reading. Also, the position of the adverb is different than in the reference translation. Thus adding more reference translations will improve the evaluation scores.

Human evaluation of the translations has revealed that the translations of some German compounds are missing and that the word order of verbs in long sentences and sentences with complex sentential embeddings tends to be incorrect. The two phenomena have an interesting connection. Both can be improved by lexical user adaptation. Adding new lexical entries improves not only the translation of words, it also improves the alignment of the bilingual parallel corpus which has effects on the productions. In particular long-distance phenomena such as the differing word order of verbs in German and English can only be accounted for if the chunks intervening between the verbs are aligned. Thus adding new lexical entries also improves the treatment of word order phenomena.

5 Conclusions

First experiments show that the proposed user adaptation concept successfully allows the user to improve system behavior and that the system is robust enough to take 'naturally occuring', freely translated sentences as input. [8]

References

1. Paco-Project: Deliverable 5.1 Report on State of the art in Post-Editing Interfaces. http://www.ccl.kuleuven.be/ vincent/ccl/papers/paco_d5.1.pdf
2. Byrne, Bill, B. 2010: FAUST - Feedback Analysis for User Adaptive Statistical Translation. In Proceedings EAMT 2010. Also: http://www.faust-fp7.eu/faust/Main/FaustPapers.
3. PRESEMT Deliverable D2.1: System Specifications
4. Chiang, David. 2007. Hierarchical phrase-based translation. In Computational Linguistics 33(2). 201228.
5. Eisner, Jason 2003. Learning non-isomorphic tree mappings for machine translation. In Proc. of the 41st Annual Meeting of the Assoc. for Computational Linguistics.
6. Cowan, Brooke, Ivona Kucerova, and Michael Collins. 2006. A discriminative model for tree-to-tree translation. In Proc. of the 2006 Conference on Empirical Methods in Natural Language Processing. 232-241.
7. Zhang, Min, Hongfei Jiang, Ai Ti Aw, Jun Sun, Sheng Li, and Chew Lim Tan. 2007. A Tree-to-Tree Alignment-based Model for Statistical Machine Translation. In MT-Summit-07. 535-542.

[8] The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^o 248307 (PRESEMT project).