BTL: a Hybrid Model for English –Vietnamese Machine Translation

Dien Dinh, Hoang Kiem

IT Faculty, Vietnam National University of HCM City
HCM City, Vietnam
ddien@saigonnet.vn

Eduard Hovy

Information Science Institute, USC
Marina Del Rey, CA, USA
hovy@isi.edu

Abstract

Machine Translation (MT) is the most interesting and difficult task which has been posed since the beginning of computer history. The highest difficulty which computers had to face with, is the built-in ambiguity of Natural Languages. Formerly, a lot of human-devised rules have been used to disambiguate those ambiguities. Building such a complete rule-set is time-consuming and labor-intensive task whilst it doesn't cover all the cases. Besides, when the scale of system increases, it is very difficult to control that rule-set. In this paper, we present a new model of learning-based MT (entitled BTL: Bitext-Transfer Learning) that learns from bilingual corpus to extract disambiguating rules. This model has been experimented in English-to-Vietnamese MT system (EVT) and it gave encouraging results.

1. Introduction

The main task of every MT systems is to disambiguate the built-in ambiguities of Natural Languages in every level (word, phrase, sentence) and aspect (morphology, grammar, semantics and pragmatics). These ambiguities may be considered different possible tags (labels) of a linguistics unit (word, phrase, sentence) in its different contexts. For example, in the *morphological* aspect the word "can", it has 3 different possible POS-tags (Part-Of-Speech), but in a certain context, it must be assigned only one correct POS-tag, e.g.: "*I can can a can*", the POS-tagger must be able to classify as follows: "I_{PRO} can_{AUX} can_{V} a_{DET} can_{N}". Similarly the noun "bank" semantically has many different possible SENSE-tags (financial building; river side; etc.), but in a certain context, it must be assigned only one correct tag (N. Zinovjeva, 2000). In the following sentence: "*I enter the bank*", the SENSE-tagger must be able to identify the correct sense tag (financial building). Similar to other kinds of ambiguities (e.g. boundary of phrase, transpositions of words between the source language and target language, etc.), we may use proper taggers (e.g. Chunker, Word-order Transfer, etc.) to assign correct linguistic tags to those ambiguous linguistic units.

Formerly, human-devised rules (e.g. IF... THEN ...) have been used to assign correct tags to ambiguous units. Nevertheless, building such a complete rule-set is labor-intensive and costly whilst it does not cover all the cases. Besides, when the scale of system increases, it is very difficult to control that rule-set. In this paper, we present a new hybrid model of MT consisting of linguistic taggers. Due to the limitation of space in this 8-page paper, for more details of each linguistic tagger, please refer to the following papers: POS-tagger (Dien, 2003b), SENSE-tagger (Dien, 2002b), Word-order Transfer (Dien, 2003c), and so on.

The remains of this paper is organized as follows:

- Our Bitext Transfer Learning (BTL) model for MT: introduction to BTL model; its operation.
- The Training corpus for BTL: the English-Vietnamese bilingual Corpus (called EVC); word alignment for EVC; linguistic annotation for EVC.
- The Training algorithm for BTL: the fast Transformation-Based Learning (fTBL); apply fTBL in annotating EVC and EVT (e.g. POS-tagger, chunker, shallow parser, sense tagger).
- Evaluation and Results: compare translation results to human translations in a golden bilingual corpus via BLEU tool.
- Conclusion: limitations and future developments.

2. Our Bitext-Transfer-Learning Model for MT

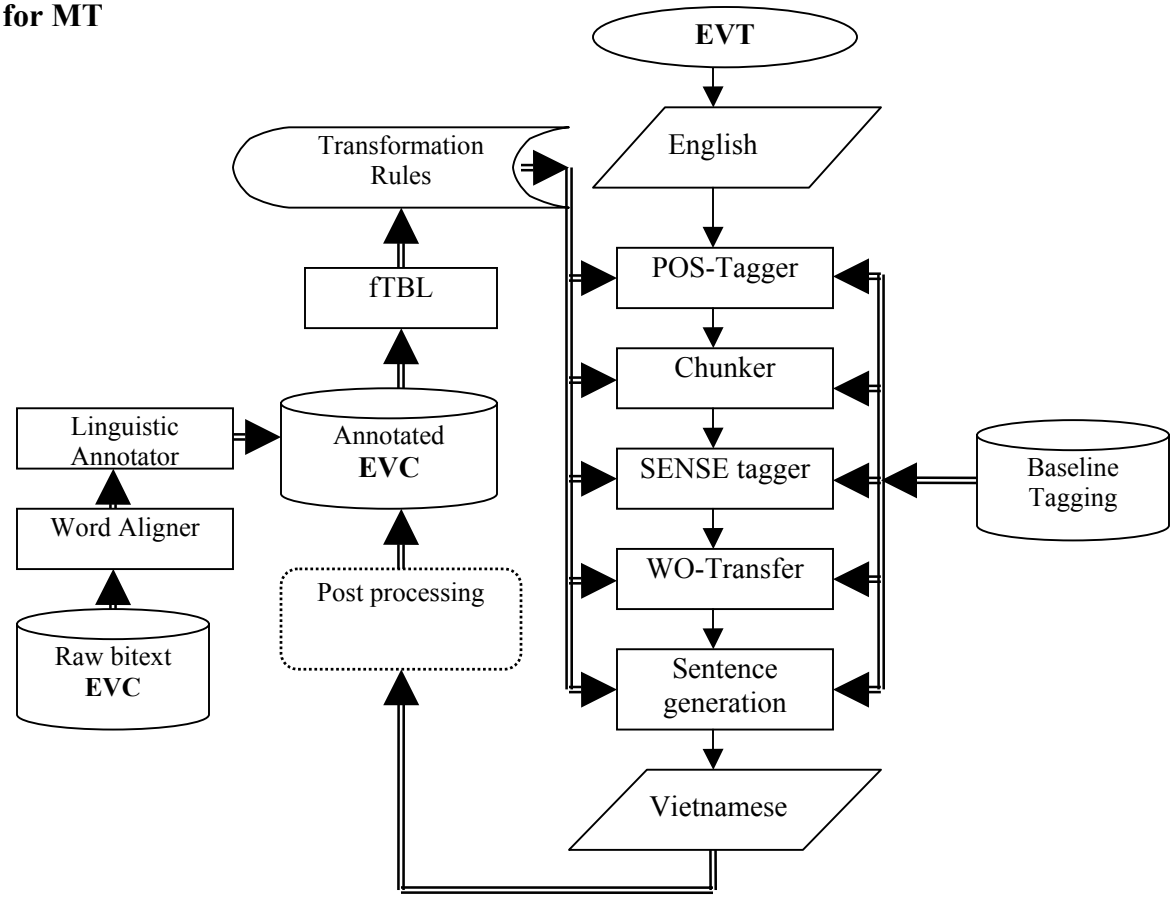

Figure 1. BTL translation model.

At first, data from raw EVC is input into a word-aligner module (figure 1) in order to align English words with corresponding Vietnamese words (figure 2). Next, this word-aligned corpus is input into a linguistic-annotator to annotate linguistic tags (e.g. Part Of Speech, Phrase Chunk, grammatical relation tags and sense tags) for EVC. (please refer to section 3 for more details of EVC)

This annotated EVC is the training data (golden corpus) for BTL through the learning algorithm fTBL. During the training period of fTBL, the system will automatically extract the transformation rules from the training corpus. These transformation rules will be used by morphological analyzer, chunker, sense tagger and word-order transfer of the English-Vietnamese Translation system (called EVT) to assign linguistic tags for new English texts. (please refer to section 4 for more details of the training algorithm fTBL)

In the training algorithm fTBL, the first step is the baseline annotation which assigns the most probably linguistic tags to linguistic units. This initial annotation will speed-up the tagging task and improve the accuracy of fTBL. So in our BTL translation model, we take advantages of the output of available powerful linguistic taggers by using them as the "baseline annotation" in order to increase the overall efficiency of our MT system.

Following the BTL translation model, after analyzing new English texts, our EVT system will produce the correct or incorrect target Vietnamese sentence. If it is incorrect, it will be post-edited by manual and combined with its source English sentence to put back to the EVC in order to enrich the training corpus. As a result, our training corpus will be larger and more covering. This enables our MT system to draw more effective transformation rules which help the system avoid previous mistakes (Dien, 2003a).

3. The Training Corpus for BTL

3.1. The raw English-Vietnamese bilingual Corpus (EVC)

The training corpus for BTL comes from the English – Vietnamese bilingual Corpus (named EVC). This 5,000,000-word corpus is collected from many different resources of bilingual texts (such as books, dictionaries, corpora, etc.) in selected fields such as Science, Technology, daily conversation (see table 1). After collecting bilingual texts from different resources, this parallel corpus has been normalized in their form (text-only), tone marks (diacritics), character code of Vietnam (TCVN-3), character font (VN-Times), etc. Next, this corpus has been sentence aligned and spell-checked semi-automatically. An example of unannotated EVC is as the following:

D02:01323: Jet planes fly about nine miles high*.

+D02:01323: *Các phi cơ phản lực bay cao khoảng chín dặm*.

The codes at the beginning of each line above refer to the corresponding sentence in the EVC corpus. For full details of building this EVC corpus (e.g. collecting, normalizing, sentence alignment, spelling checker, etc.), please refer to (Dien, 2001b).

Remarkably, this EVC includes the SUSANNE corpus (G. Sampson, 1995) – a golden corpus has been manually annotated such necessary English linguistic annotations as lemma, POS tags, chunking tags, syntactic trees, etc. This English corpus has been translated into Vietnamese by English teachers of the Foreign Language Department of Vietnam University of HCM City. In this paper, this valuable annotated corpus is used as the kernel training corpus for annotating whole our EVC.

Due to the hetegenous corpus with texts in different domains and genres, we had to classify our EVC into different smaller corpora for training different domains, such as: computer, electronics, daily conversation, etc.

3.2. Word Alignment for EVC

Next, this bilingual corpus has been automatically word aligned by a hybrid model combining the semantic class-based model (S.K.Chang and J.S.Chang, 1997) with the GIZA++. In this model, the semantic classification of LLOCE (M.Arthur, 1997) is used. Besides, the Vietnamese word segmentation was also solved in this word-alignment (D.Dien et al., 2001a). An example of the word-alignment result is as in figure 2 below. For full details of word alignment for this EVC, please refer to (Dien et al., 2002a).

Table 1. Resource of EVC Corpus

No.	Resources	The number of pairs of sentences	Number of English words	Number of Vietnamese morpho-words	Length (English words)	Percent (words/ EVC)
1.	Computer books	9,475	165,042	239,984	17.42	7.67
2.	LLOCE dictionary	33,078	312,655	410,760	9.45	14.53
3.	EV bilingual dictionaries	174,906	1,110,003	1,460,010	6.35	51.58
4.	SUSANNE corpus	6,269	131,500	181,781	20.98	6.11
5.	Electronics books	12,120	226,953	297,920	18.73	10.55
6.	Children's Encyclopedia	4,953	79,927	101,023	16.14	3.71
7.	Other books	9,210	126,060	160,585	13.69	5.86
	Total	250,011	2,152,140	2,852,063	8.59	100%

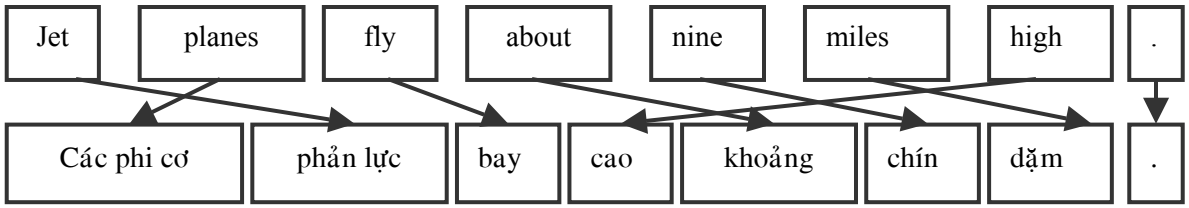

Figure 2. An example of a word-aligned pair of sentences in EVC corpus.

3.3. Linguistic Annotation for EVC

After word-aligning the EVC, linguistic units in EVC will be annotated with linguistic tags. Nevertheless, hand-annotation of even reasonably well-determined features such as part-of-speech (POS) tags has proved to be labor intensive and costly. In our work, we suggest a solution to avoid hand-annotations for word-aligned EVC by building linguistic-taggers (POS-tagger, Chunker, SENSE-tagger, etc.) using fTBL algorithm and linguistic information of corresponding Vietnamese via its word-alignment.

Our solution is motivated by I. Dagan, I.Alon, and S.Ulrike. (1991); W.Gale, K.Church and D.Yarowsky (1992). They proposed the use of bilingual corpora to avoid hand-tagging of training data. Their premise is that "different senses of a given word often translate differently in another language (for example, *pen* in English is *stylo* in French for its *writing implement* sense, and *enclos* for its *enclosure* sense). By using a parallel aligned corpus, the translation of each occurrence of a word such as *pen* can be used to automatically determine its sense". This remark is not only true for word sense but also for POS-tag and it is more exact in such typologically different languages as English vs. Vietnamese.

In fact, POS-tag annotations of English words as well as Vietnamese words are often ambiguous but they are not often exactly the same. For example (table 3), "can" in English may be "Aux" for *ability* sense, "V" for *to make a container* sense, and "N" for *a container* sense and there is hardly existing POS-tagger which can exactly POS-tag for that word "can" in all different contexts. Nevertheless, if that "can" in English is already word-aligned with a corresponding Vietnamese word, it will be easily POS-disambiguated by Vietnamese word' s POS-tags. For example, according to POS-tagset of PennTreeBank, if "can" is aligned with "có thể", it must be *Auxiliary* (MD) ; if it is aligned with "đóng hộp" it must be a *Verb*(VB), and if it is aligned with "cái hộp" it must be a *Noun* (NN). Based on this reason, we have made a POS-tagger using fTBL algorithm to bootstrap the POS-annotation results of the English POS-tagger by exploiting the POS-information of the corresponding Vietnamese words via their word-alignments in EVC. Then, we directly project POS-annotations from English side to Vietnamese via available word alignments under the model of D.Yarowsky and G.Ngai (2001). For more details of POS-tagger for EVC, please refer to (D.Dien, H.Kiem, 2003b). Similarly, because we have made use of the class-based word alignment, after aligning words, we determine the semantic class of each word. For example: according to the SENSE-tagset of LLOCE, the word "letter" has 2 senses, one is "message" (if it belongs to class G155) and the other is "alphabet" (class G148). Similarly, the word "bank" has 3 senses, one is "money" (class J104), one is "river" (class L99) and one is "line" (class J41). After aligning words, the result of semantic annotation is as table 2 and 3 below (*i* and *j* are positions of English and Vietnamese words). If the output of automatic-annotations above is still ambiguous, it will be manually corrected to become an annotated training data for our BTL.

Table 2. Result of sense tagging for "letter"

i	0	1	2	3	4	5	6
S	I	write	a	*letter*	to	my	friend
T	Tôi	viết	một	bức thư	cho	của tôi	bạn
j	0	1	2	3	5	7	6
	G 280	G 190		**G 155**		G 281	C 40

Table 3. Result of POS-tagging for "bank"

i	0	1	2	3	4
S	I	*can*	*can*	a	*can*
POS	PP	**MD**	**VB**	DT	**NN**
T	Tôi	có thể	đóng hộp	một	cái hộp
j	0	1	3	5	6

Table 4. An example of English POS-tagging and SENSE-tagging in EVC

English	Jet	*planes*	*fly*	about	nine	miles	high
E-POS-tag	NN	NNS	VBP	IN	CD	NNS	RB
Vietnamese	phản lực	(các) phi cơ	bay	khoảng	chín	dặm	cao
V-POS-tag	N	N	V	IN	CD	N	R
Sense-tag	M181	**M180**	**M28**		J4	J68	N305

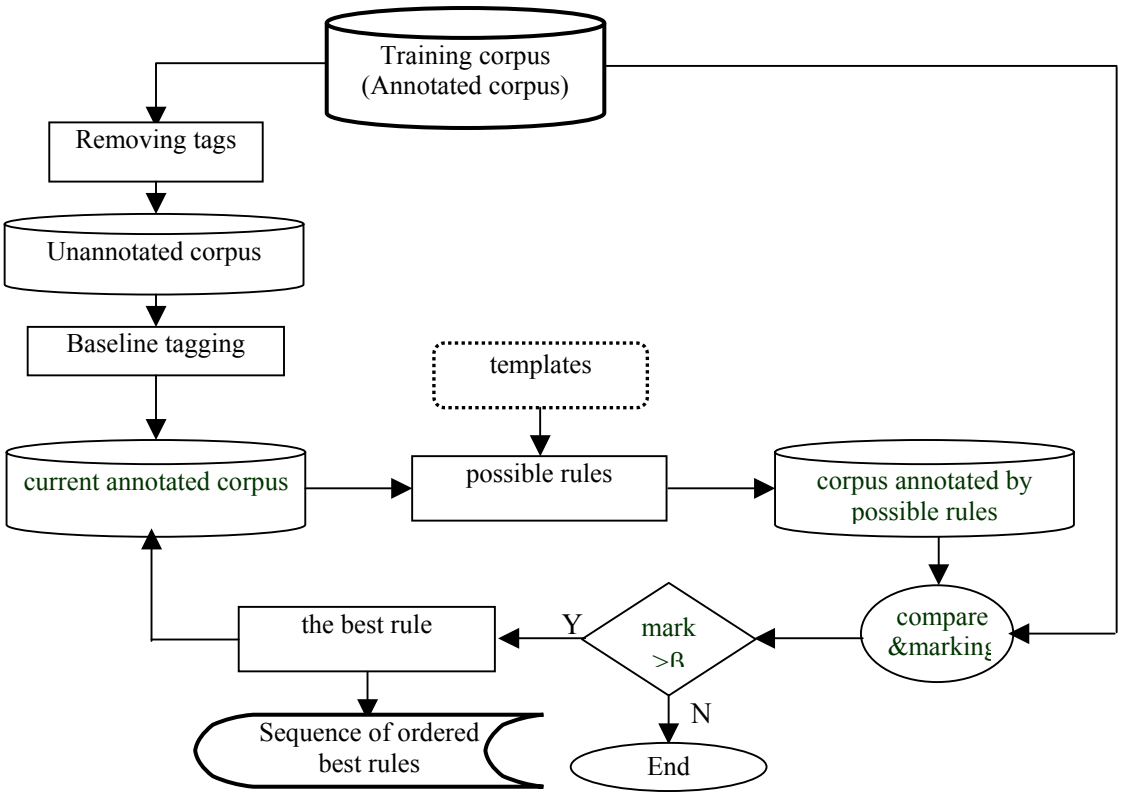

Figure 3. The flowchart of the training period in the algorithm fTBL.

4. The Training Algorithm for BTL

The main training algorithm used in our BTL is the fast Transformation-based learning (or fTBL). This algorithm has been used in annotating POS-tags and Chunker-tags for EVC, extracting the transformation-rules from annotated-EVC in order to tag for new English texts in our English-Vietnamese Translation system (or EVT).

4.1. The fast Transformation-Based Learning Algorithm (fTBL)

In 1993, Eric Brill (1993) promoted the Transformation-Based Learning (TBL) in his doctor thesis on the base of structural linguistics of Z.S.Harris. Since its birth, heretofore, TBL algorithm has been successfully applied into most of language problems. A remarkable characteristic of TBL in comparison with other learning algorithms is intuitiveness and simplicity. Linguists can fully observe and intervene during learning and tagging process as well as its intermediate and final results. In 2001, Radu Florian and Grace Ngai (2001) promoted fast-TBL to improve the speed of training stage of TBL noticeably without reducing its accuracy. For full details of TBL and fTBL, please refer to (E.Brill, 1993) and (R.Florian and G.Ngai, 2001).

4.2. The fTBL algorithm for linguistic taggers

The fTBL algorithm for linguistic-tagger can be formalized as below:

- χ : sample space, the set of language units (word/phrase). In English, it is simple to recognize the word boundary, but in Vietnamese (an isolate language), it is rather complicated and we have solved in another work (D.Dien, 2001a).

- C : set of language's tags c (classification). For example: N,V,A,.. in POS-tagset; HUM,ANI,NAT,... in sense-tagset, NP_B, NP_I, NP_O, ... in chunker-tagset, etc.
- $S = \chi x C$: the cross-product between the sample space (word/phrase) and the classification space (tagset). It is the state space where each point is a couple (word, tag) or (phrase,tag).
- π : predicate defined on S^+ space, which is on a sequence of states. This predicate π follows the human-specified templates of transformation rules. Depending on the specified linguistic-taggers, we will have different templates. For example, in the POS-tagger for English, this predicate only consists of English factors which affect the POS-tagging process, e.g.: $\bigcup_{\exists i \in [-m,+n]} Word_i$ or $\bigcup_{\exists i \in [-m,+n]} Tag_i$ or $\bigcup_{\exists i \in [-m,+n]} Word_i \wedge Tag_j$. Where, $Word_i$ and Tag_i are the word-form and the word-tag of the i^{th} word from the current word. Positive values of i mean the preceding (its left side), and negative ones mean the following (its right side). The value of i ranges within the window from $-m$ to $+n$.
- A rule r defined as a couple (π, c) which consists of predicate π and tag c. Rule r is written in the form $\pi \Rightarrow c$. This means that the rule $r = (\pi, c)$ will be applied on the sample x if the predicate π is satisfied on it, whereat, x will be assigned a new tag c.
- Giving a state s = (x,c) and rule r = (π, c), then the result state r(s), which is gained by applying rule r on s, is defined as:

$$r(s) = \begin{cases} s & \text{if } \pi(s)=\text{False} \\ (x, c') & \text{if } \pi(s)=\text{True} \end{cases}$$

- T : set of training samples (or called *golden corpus*), which were assigned correct tags. Depending on the specified linguistic-taggers, we will have different golden corpora. In the POS-tagger and Chunker for EVC, we made use of the golden corpus SUSANNE (Sampson, 1995). In the linguistic-taggers for EVT, T is the annotated and revised EVC.
- The score of each rule r = (π, c) is the difference between the result processed on the sample s of rule r and the initial state, in conformity with the following formula:

$$Score(r) = \sum_{s \in T} score(r(s)) - \sum_{s \in T} score(s)$$

$$score((x,c)) = \begin{cases} 1 & \text{if c = True(x)} \\ 0 & \text{if c} \neq \text{True(x)} \end{cases}$$

*** The training period of algorithm fTBL:**

Step 1: Initiating for each sample x in training set with the most suitable tag c (called *baseline tagging*). For instance, the word "can" in English has the highest part-of-speech probability as an *Auxiliary*. We call the first time corpus T_0. For English, we may make use of available powerful linguistic-taggers for English, e.g. Minipar of Dekang Lin (1993).

Step 2: Examining all transformation rules r influencing corpus T_k in time k^{th} and choosing a rule that has the highest Score(r) and applying this rule for corpus T_k to get new corpus T_{k+1}. We have : $T_{k+1} = r(T_k) = \{ r(s) \mid s \in T_k \}$. If there is no rule which satisfies Score(r) > β, the algorithm is stopped. β is the threshold, which is preset, and adjusted according to real demand. These rules change the linguistic-tags of words based upon the contexts they appear in. fTBL evaluates the result of applying that candidate rule by comparing the current result of linguistic-annotations with that of the golden corpus in order to choose the best one which has highest mark. These optimal rules create an ordered sequence.

Step 3: k = k+1.

Step 4: Repeat from step 2.

*** The executing period of algorithm fTBL:**

- Starting with the new unannotated text, fTBL assigns an inital linguistic-tag to each word/phrase in text in a way similar to that of the training period (baseline tagging).
- The sequence of optimal rules (extracted from the training period) are applied, which change the linguistic-tags based upon the contexts they appear in. These rules are applied deterministically in the order they appear in the sequence.

4.3. The Result of Extracted Transformation Rules

These extracted rules are intuitive rules and easy to understand by human beings. For examples:

- In the POS-tagger:

1. $((\exists i \in [-3,-1] \mid Tag_i = MD) \wedge (tag_0 = VPB)) \Rightarrow tag_0 \leftarrow VB$
2. $((tag_{-1} = TO) \wedge (tag_0 = NN)) \Rightarrow tag_0 \leftarrow VB$
3. $((\exists i \in [-2,-1] \mid Word_i = "have") \wedge (tag_0 = VBD)) \Rightarrow tag_0 \leftarrow VBN$
4. $((Word_0 = "can") \wedge (VTag_0 = MD) \wedge (tag_0 = VB)) \Rightarrow tag_0 \leftarrow MD$

The 4th rule will be understood as follows: "*if the POS-tag of current word is VB (Verb) and its word-form is "can" and its corresponding Vietnamese word-tag is MD (Modal), then the POS-tag of current word will be changed into MD*".

- In the Sense-tagger:

1. $((\exists i \in [+1,+3] \mid Word_i = "river") \wedge (Word_0 = "bank) \wedge (POS_0 = NN)) \Rightarrow tag_0 \leftarrow NAT$
2. $((SUB_0 \in HUM) \wedge (Word_0 = "enter") \wedge (POS_0 = VB) \wedge (Word_0 \in MOV) \Rightarrow OBJ_0 \leftarrow HOU$

The 1st will be understood "*if there exist a word-form is "river" within 3 positions right after the word form "bank", the SENSE-tag of current word is changed into NAT (L99: Natural)"*.

Similarly, the 2nd rule will be understood "*if sense-tag of SUBject is HUMan and the current word-form is "enter" and its POS-tag is Verb and its is a MOTION, then its OBJect will be assign to sense-tag HOU*". For example: in the sentence "*I enter the bank*", the object "bank" will be assign to "financial building" (J104: money).

- In the Word-Order-transfer:

$$(POS_{N_a} = Qwh) \cap (POS_{N_{a1}} = Aux) \cap (POS_{N_{a2}} = SP) \cap (POS_{N_{a3}} = VP) \Rightarrow N_{a2} - N_{a1} - N_{a3} - \text{"được không"}$$

This rules means that: "*if the interrogative sentence (Qwh) has the source syntax tree including: auxiliary verb (Aux) – subject (SP) – predicate (VP), it will be transferred into Vietnamese sentence as the following: subject – auxiliary verb – predicate and the inserted expletive "không" at the end of the sentence*". For example: "Can you speak English ?" ⇒ "Anh có thể nói tiếng Anh được không ?". For more details of Word-Order Transfer for EVT, please refer to (D.Dien et al., 2003c).

5. Evaluation and Results

For evaluating our EVT system, we made use of 01 file (in Computer textbook) of EVC (which is held-back for evaluating, it hasn't been used for training). This bilingual file has 866 English sentences (14,634 words). We made use of the BLEU (BiLingual Evaluation Understudy) tool (K.Papineni, 2002) of NIST (National Institute of Standard Technology) version 1.03. This evaluation is based on the comparison of the translation of machine and the translation of human-beings using N-gram co-occurrence statistics. In fact, this MT-evaluation tool is more suitable for MT systems which their target language are English, whilst in our EVT system, its target language is Vietnamese (because word order and function words are two most often used grammatical facilities in Vietnamese), i.e:

- Source: "What are you doing ?"

- Target (machine): "*Cái* gì bạn đang làm ?"

- Human translation: "Bạn đang làm gì ?"

Our experiment result is as table 5 below:

Table 5. The experiment results of EVT system

Measurements	Results
1-gram	73.28 %
2-gram	53.32 %
3-gram	43.03 %
4-gram	34.10 %
Precision:	**48.94 %**

6. Conclusion

In this paper, we have presented a hybrid model for MT which combines rule-based MT and corpus-based MT. In this model, disambiguation rules have been automatically learned from a training bilingual corpus (EVC) by fast TBL algorithm. This bilingual corpus EVC has been automatically word-aligned and annotated with linguistic-tags (e.g. POS, Chunk, SENSE, etc.) by fast TBL and linguistic information of corresponding Vietnamese words via their word-alignment links. It means that this model exploits the advantage of the approach based on rules and overcomes its defects in building these rules (replacing hand-crafted rules by automatically-extracted ones).

Currently, the initial results of BTL model is not high but in the future, its results will be improved when the training corpus EVC is completely revised and enriched more.

With this translation model, however, we can let the computer exploit automatically transfer rules for languages while observing, intervening these rules easily, because these are explicit and intuitive rules of language which are written in symbolic of normal language form. Translating while updating data, bilingual English-Vietnamese corpus becomes larger and larger, and covers almost common cases. With a more coverage training corpus, the quality of MT system will be certainly better and better.

References

M. Arthur. 1997. *Longman Lexicon Of Contemporary English (Vietnamese version by Tran Tat Thang)*. VN Education Publisher.

E. Brill. 1993. *A Corpus-based approach to Language Learning.* PhD-dissertation, Pennsylvania Uni., USA.

I. Dagan, I.Alon, and S.Ulrike. 1991. *Two languages are more informative than one*. In Proceedings of the 29th Annual ACL, Berkeley, CA, pp.130-137.

D. Dien, H.Kiem, and N.V.Toan. 2001a. *Vietnamese Word Segmentation.* Proceedings of NLPRS'01 (The 6th Natural Language Processing Pacific Rim Symposium), Tokyo, Japan, 11/2001, pp. 749-756.

D.Dien. 2001b. *Building English-Vietnamese bilingual corpus.* Master thesis in Comparative Linguistics, Literature and Linguistics Faculty of University of Social Sciences and Humanity, Vietnam National Uni. of HCM City.

D.Dien, H.Kiem, T.Ngan, X.Quang, V.Toan, Q.Hung, P.Hoi. 2002a. *Word alignment in English – Vietnamese bilingual corpus.* Proceedings of EALPIIT'02, HaNoi, Vietnam, pg 3-11.

D.Dien. 2002b. *Building a training corpus for word sense disambiguation in the English-to-Vietnamese Machine Translation.* Proceedings of Workshop on Machine Translation in Asia, COLING-02, Taiwan, 9/2002, pg.26-32.

D.Dien. 2003a. *BTL Model for English-Vietnamese Machine Translation.* Phd-dissertation in Computer Science, Information Technology Faculty of University of Natural Sciences, Vietnam National University of HCM City.

D.Dien and H.Kiem. 2003b. *POS-Tagger for English-Vietnamese Bilingual Corpus.* paper accepted at *HLT-NAACL Workshop on Parallel Text*, Edmonton, Canada.

D.Dien, T.Ngan, X.Quang, C.Nam. 2003c. *A hybrid approach to Word Order Transfer for the English-Vietnamese MT system.* paper accepted at *MT-Summit-IX*, Louisiana, USA.

R. Florian, and G.Ngai. 2001. *Transformation-Based Learning in the fast lane.* Proceedings of North America ACL-2001.

W. Gale, K.W.Church, and D. Yarowsky. 1992. *Using bilingual materials to develop word sense disambiguation methods*. In Proceedings of the Int. Conf. on Theoretical and Methodological Issues in MT, pp.101-112.

S. K. Jang and J.S. Chang. 1997. *A Class-based Approach to Word Alignment.* Computational Linguistics, 23/2, pp. 313-343.

D. Lin. 1993. *Parsing without overgeneration*, Proceedings of ACL-03, pp. 112-120, Columbus, OH.

K. Papineni, S.Roukos, T.Ward, and W.J. Zhu. 2002. *BLEU: a Method for Automatic Evaluation of Machine Translation.* Proceedings of the 40th annual ACL, PA, USA, pp.311-318.

G. Sampson. 1995. *English for the Computer: The SUSANNE Corpus and Analytic Scheme.* Clarendon Press (Oxford University Press).

D.Yarowsky and G.Ngai. 2001. *Induce, Multilingual POS Tagger and NP bracketer via projection on aligned corpora.* Proceedings of NAACL-01.

N. Zinovjeva. 2000. *Learning sense disambiguation rules for MT.* MSc-thesis, Uppsala Uni.