S-SSTC: A SYNCHRONOUS ANNOTATION SCHEMA OF STRUCTURED STRING-TREE CORRESPONDENCE (SSTC)

MOSLEH H. AL-ADHAILEH
Computer Aided Translation Unit
School of Computer Sciences
Universiti Sains Malaysia
11800 PENANG, MALAYSIA
mosleh@cs.usm.my, mosleh@hotmail.com

TANG ENYA KONG
Computer Aided Translation Unit
School of Computer Sciences
Universiti Sains Malaysia
11800 PENANG, MALAYSIA
enyakong@cs.usm.my

ABSTRACT

In this paper, we propose a synchronous annotation schema called Synchronous Structured String-Tree Correspondence (S-SSTC). This synchronous variant of SSTC structure is used to describe not only the correspondence between a text in a language and its abstract representation tree but also, the correspondence between different languages. We will also describe how S-SSTC provides the flexibility to treat some of the non-standard cases, which are problematic to other synchronous formalisms. The proposed S-SSTC schema is well suited to describe the correspondence between different languages, in particular, relating a language with its translation in another language (i.e. in Machine Translation). Also it can be used as annotation for translation systems that automatically extract transfer mappings (rules or examples) from bilingual corpora. The S-SSTC is very well suited for the construction of a Bilingual Knowledge Bank (BKB), where the examples are kept in form of S-SSTCs.

KEYWORDS: Parallel text, Structured String-Tree Correspondence (SSTC), Synchronous SSTC, Bilingual Knowledge Bank (BKB), Tree Bank Annotation Schema.

1. INTRODUCTION

In machine translation (MT), one major new line of investigation that has emerged in recent years is the corpus-based method which uses corpora of parallel texts in the construction of translation and computer-aided translation systems to improve the translation output. To exploit such parallel texts, particularly for example-based MT (EBMT) and transfer rules extraction for MT systems, the description of the lexical and structural correspondences is required at a finer granularity between the various corresponding parts of such parallel texts. This requirement is due to lexical and structural differences between different natural languages. To achieve such correspondence, a formal and flexible annotation structure is needed to serve as a means for representing the translation examples extracted from parallel texts. To this end, the main issues in this paper is the definition of a generic annotation schema called Synchronous Structured String-Tree Correspondence (S-SSTC).

This annotation schema consists of two SSTCs and synchronous correspondences as a result of applying the idea of synchronization as introduced on grammar formalisms, e.g. S-TAG (Abeillé et al., 1990; Shieber & Schabes, 1990). Each SSTC structure (Boitet & Zaharin, 1988) is a text in a language, its abstract representation tree and the correspondence between the text and its representation tree. The synchronous correspondences describe the relation between these two SSTCs at different levels of their representation tree structure. The S-SSTC is fairly flexible and can handle different cases of correspondences between different natural languages. In the case of EBMT and transfer rules extraction systems, the S-SSTC is used to represent translation relations between a source language and its target language. The synchronous correspondences describe such relation between different levels, i.e. lexical and structural level, of their representation structure. Also, we have defined a set of constraints to ensure the correct establishment of the synchronous correspondence in the S-SSTC in a natural manner.

In this paper, we will present the proposed S-SSTC – a schema well suited to describe the correspondence between two languages. S-SSTC is flexible and able to handle the non-standard correspondence cases exist between different languages. It can also be used to facilitate automatic extraction of transfer mappings (rules or examples) from bilingual corpora.

2. STRUCTURED STRING-TREE CORRESPONDENCE (SSTC)

SSTC is a general structure that can associate an arbitrary tree structure to string in a language as desired by the annotator to be the interpretation structure of the string, and more importantly is the facility to specify the correspondence between the string and the associated tree which can be non-projective (Boitet & Zaharin, 1988). These features are very much desired in the design of an annotation scheme, in particular for the treatment of linguistic phenomena, which are non-standard, e.g. crossed dependencies (Tang & Zaharin, 1995).

***Definitions*[1]:**

- *An **SSTC** is a general structure, which is a string in a language associated with an arbitrary tree structure; i.e. its interpretation structure, and the correspondence between the string and its associated tree, which can be non-projective; i.e. SSTC is a triple (**st, tr, co**), where **st** is a string in one language, **tr** is its associated representation tree structure and **co** is the correspondence between **st** and **tr**.*
- *The correspondence **co** between a string and its representation tree is made of two interrelated correspondences:*
  - *a) Between nodes and substrings (possibly discontinuous).*
  - *b) Between (possibly incomplete) subtrees and (possibly discontinuous) substrings.*
- *The correspondence can be encoded on the tree by attaching to each node **N** in the representation tree two sequences of **INTERVALS** called **SNODE(N)** and **STREE(N)**.*
- ***SNODE(N)**: An interval of the substring in the string that corresponds to the node **N** in the tree.*
  ***STREE(N)**: An interval of the substring in the string that corresponds to the subtree having the node **N** as root.*

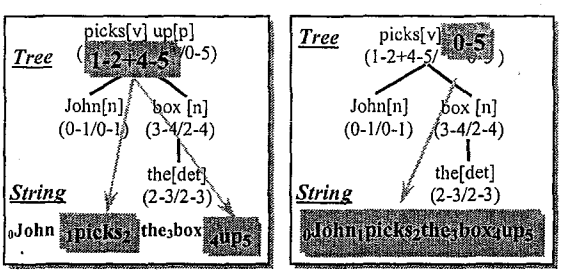

Figure 1: An SSTC recording the sentence "**John picks the box up**" and its dependency tree together with the correspondences between substrings of the sentence and subtrees of the tree.

Figure 1 illustrates the sentence **"John picks the box up"** with its corresponding SSTC. It contains a non-projective correspondence. An interval is assigned to each word in the sentence, i.e. (0-1) for "**John**", (1-2) for "**picks**", (2-3) for "**the**", (3-4) for "**box**" and (4-5) for "**up**". A substring in the sentence that corresponds to a node in the representation tree is denoted by assigning the interval of the substring to SNODE of the node, e.g. the node "**picks up**" with SNODE intervals (1-2+4-5) corresponds to the words "**picks**" and "**up**" in the string with the similar intervals.The correspondence between subtrees and substrings are denoted by the interval assigned to the STREE of each node, e.g. the subtree rooted at node "**picks up**" with STREE interval (0-5) corresponds to the whole sentence "**John picks the box up**".

The case depicted in Figure 1, describes how the SSTC structure treats some non-standard linguistic phenomena. The particle **"up"** is featurised into the verb **"pick"** and in discontinuous manner (e.g. **"up" (4-5)** in **"pick-up" (1-2+4-5)**) in the sentence **"He picks the box up"**. For more details on the proprieties of SSTC, see Boitet and Zaharin (1988).

3. SYNCHRONOUS SSTC STRUCTURE

Much of theoretical linguistics can be formulated in a very natural manner as stating correspondences (translations) between layers of representation structures (Rambow & Satta, 1996), such as the relation between syntax and semantic. An analogous problem is the definition of correspondences between a language and its translations in other languages. Therefore, the synchronization of two adequate structures seems to be an appropriate representation for that.

We propose a flexible annotation schema called Synchronous Structured String-Tree Correspondence (S-SSTC)) to realize additional power and flexibility in expressing structural correspondences at the level of language sentence pairs. For example, such schema can serve as a mean to represent translation examples, or find structural correspondences for the purpose of transfer grammar learning (Menezes & Richardson, 2001), (Watanabe et al., 2000), (Meyers et al., 2000), (Matsumoto et al., 1993), (kaji et al., 1992), and example-base machine translation EBMT[2] (Sato & Nagao, 1990), (Sato, 1991), (Richardson et al., 2001), (Al-Adhaileh & Tang, 1999).

3.1 Synchronous SSTC Annotation Schema

In this section, we will discuss the definition and the formal properties of S-SSTC. S-SSTC consists of a pair of SSTCs with an additional synchronization relation between them. The use of S-SSTC is motivated by the desire to describe not only the correspondence between the text and its representation structure in one language (i.e. SSTC) but also the correspondence between two languages (synchronous correspondence).

Definitions:

- *Let each of S and T be SSTC which consists of a triple (st, tr, co), where st is a string in one language, tr is its associated representation tree structure and co is the correspondence between st and tr, as defined in Section 2.*
- *A synchronous SSTC S_{syn} is defined as a triple (S, T, $\varphi_{(S,T)}$), where $\varphi_{(S,T)}$ is a set of links defining the synchronization correspondence between S and T at different internal levels of the two SSTC structures.*

[1] These definitions are based on the discussion in (Tang, 1994) and Boitet and Zaharin (1988).

[2] for a comprehensive overview about EBMT, see Somers(1999)

- *A link $\ell \in \varphi_{(S,T)}$ can be either of type ℓ_{sn} or ℓ_{st} which defines the synchronous correspondences between nodes of tr in S, and nodes of tr in T.*
  - *ℓ_{sn} records the synchronous correspondences at level of nodes in S and T (i.e. lexical correspondences between specified nodes), and normally $\ell_{sn} = (X_1, X_2)$, where X_1 and X_2 are sequences of SNODE correspondences in co, which may be empty.*
  - *ℓ_{st} records the synchronous correspondences at level of subtrees in S and T (i.e. structural correspondences between subtrees), and normally $\ell_{st} = (Y_1, Y_2)$, where Y_1 and Y_2 are sequences of STREE correspondences in co, which may be empty.*
- *A synchronous correspondence link $\ell \in \varphi_{(S,T)}$ can be of type ℓ_{sn} or ℓ_{st}.*
- *ℓ_{sn} is a pair$(\ell_{sn_s}, \ell_{sn_t})$, where ℓ_{sn_s} is from the first SSTC and ℓ_{sn_t} is from the second SSTC.*
- *ℓ_{sn} is represented by sets of intervals such that:*
  - *$\ell_{sn_s} = \{ i_1\_j_1 +\ldots+ i_k\_j_k +\ldots+ i_p\_j_p \} \mid i_k\_j_k \in$ **X:SNODE** correspondence in co of the first SSTC.*
  - *$\ell_{sn_t} = \{ i_1\_j_1 +\ldots+ i_k\_j_k +\ldots+ i_p\_j_p \} \mid i_k\_j_k \in$ **X:SNODE** correspondence in co of the second SSTC.*
- *ℓ_{st} is a pair$(\ell_{st_s}, \ell_{st_t})$, where ℓ_{st_s} from the first SSTC and ℓ_{st_t} from the second SSTC as defined below:*
  - *$\ell_{st_s} = \{ i_1\_j_1 +\ldots+ i_k\_j_k +\ldots+ i_p\_j_p \} \mid i_k\_j_k \in$ **Y:STREE** correspondence in co of the first SSTC or $(i_k\_j_k) = (i_k\_j_k) - (i_u\_j_v) / i_u \geq i_k \wedge j_v \leq j_h$: i.e. $(i_u\_j_v) \subseteq (i_k\_j_k)$ which corresponds to an incomplete subtree.*
  - *$\ell_{st_t} = \{ i_1\_j_1 +\ldots+ i_k\_j_k +\ldots+ i_p\_j_p \} \mid i_k\_j_k \in$ **Y:STREE** correspondence in co of the second SSTC or $(i_k\_j_k) = (i_k\_j_k) - (i_u\_j_v) / i_u \geq i_k \wedge j_v \leq j_h$: i.e. $(i_u\_j_v) \subseteq (i_k\_j_k)$ which corresponds to an incomplete subtree.*
- *The synchronous correspondence between terminal nodes with **X:SNODE** = **Y:STREE** will be of both ℓ_{sn} and ℓ_{st} correspondence such that $\ell_{sn} = \ell_{st}$.*

Note: The synchronous correspondences can be between SSTCs that contain non-standard phenomena; i.e. featursiation and discontinuity (crossed dependency). In these cases the synchronous correspondence is strait forward (following the above definitions); e.g. see Figure 2 and Figure 4.

The S-SSTC will be used to relate expressions of a natural language to its associated translation in another language. For convenience, we will call the two languages *source* and *target* languages, although S-SSTC is non-directional. S-SSTC is defined to make such relation explicit. Figure 2 depicts a S-SSTC for the English *source* sentence ***"John picks the heavy box up"*** and its translation in the Malay *target* sentence ***"John kutip kotak berat itu"***. The gray arrows (⟶) indicate the correspondence between the string and it representation tree within each of the SSTCs, and the dot-gray arrows (⇢) indicate the relations (i.e. synchronous correspondence) of synchronization between linguistic units of the *source* SSTC and the *target* SSTC.

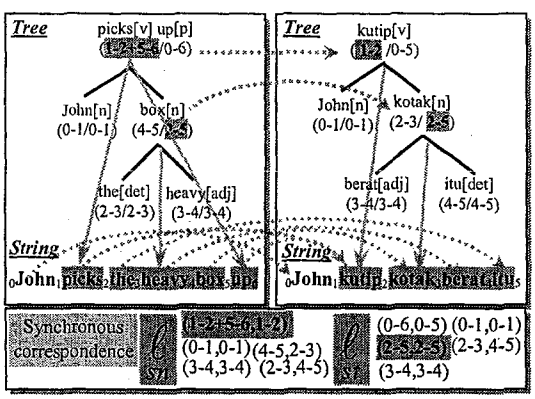

Figure 2: Synchronous SSTC for the sentence **"John picks the heavy box up"** and its Malay translation **"John kutip kotak berat itu"**, together with the synchronous correspondence between them.

Based on the notation used in S-SSTC, Figure 2 illustrates the S-SSTC for the English sentence ***"John picks the heavy box up"*** and its translation in the Malay language ***"John kutip kotak berat itu"***, with the synchronous correspondence between them. The synchronous correspondence is denoted in terms of SNODE pairs for ℓ_{sn} and STREE pairs for ℓ_{st}. For ℓ_{sn} each pair is of $(\ell_{sn_s}, \ell_{sn_t})$, where ℓ_{sn_s} is SNODE interval/s from the *source* SSTC and ℓ_{sn_t} is SNODE interval/s from the *target* SSTC. As for ℓ_{st} each pair is of $(\ell_{st_s}, \ell_{st_t})$, where ℓ_{st_s} is STREE interval/s from the *source* SSTC and ℓ_{st_t} is STREE interval/s from the *target* SSTC. For instance, as depicted in Figure 2, the fact that **"picks up"** in the *source* corresponds to

"kutip" in the *target* is expressed by the pair $(\ell_{sn_s}, \ell_{sn_t}) \Leftrightarrow (1\text{-}2{+}5\text{-}6,1\text{-}2)$ under the ℓ_{sn} synchronous correspondence. Whereas, the fact that ***"John picks the heavy box up"*** is corresponds to ***"John kutip kotak berat itu"*** is expressed by $(\ell_{st_s}, \ell_{st_t}) \Leftrightarrow (0\text{-}6,0\text{-}5)$ under the ℓ_{st} synchronous correspondence. Also the fact that **"box"** in the *source* corresponds to **"kotak"** in the *target* under the pair $(\ell_{sn_s}, \ell_{sn_t}) \Leftrightarrow (4\text{-}5,2\text{-}3)$ in the ℓ_{sn} synchronous correspondence. Whereas, the phrase **"the heavy box"** is corresponds to the phrase **"kotak berat itu"** in the *target* is expressed by $(\ell_{st_s}, \ell_{st_t}) \Leftrightarrow (2\text{-}5,2\text{-}5)$ under the ℓ_{st} synchronous correspondence.

4. HANDLING NON-STANDARD CASES WITH S-SSTC

As mentioned earlier, there are some non-standard phenomena exist between different languages, that cause challenges for synchronized formalisms. In this Section, we will describe some example cases, which are drawn from the problem of using synchronous formalisms to define translations between languages (e.g. Shieber (1994) cases). Due to lack of space we will only brief on some of these non-standard cases without going into the details.

Figure 2 illustrates a case where the English sentence has non-standard cases of featurisation, crossed dependency and a many-to-one synchronous correspondence in **"picks up"**.

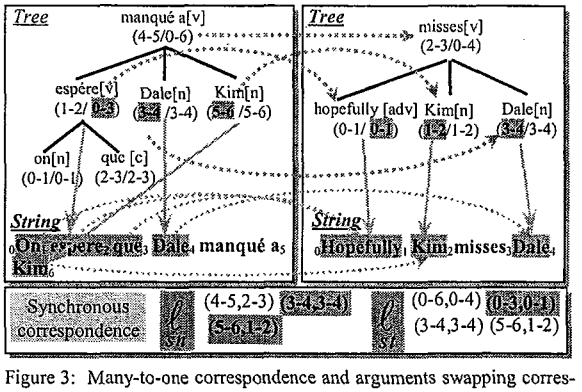

Figure 3: Many-to-one correspondence and arguments swapping correspondence in the *French* sentence **"On espére que Dale manqué a Kim"** and its corresponding *English* sentence **"Hopefully Kim misses Dale"**.

Figure 3, shows two non-standard cases between languages; e.g. French and English. First, the case of many-to-one correspondence, where a word (single node) in one language corresponds to a phrase (subtree) in the other, namely, the adverbial **"hopefully"** is translated into the French phrase **"On espére que"**. Second, a case of argument swap (reordering of subtrees) in the *English* **"Kim misses Dale"** and its corresponding translation **"Dale manqué a Kim"** in *French*.

Figure 4 describes the cases of clitic climbing in French and the non-projective correspondence (i.e. crossed dependency). It shows the flexibility of SSTC and the proposed S-SSTC in handling such popular cases.

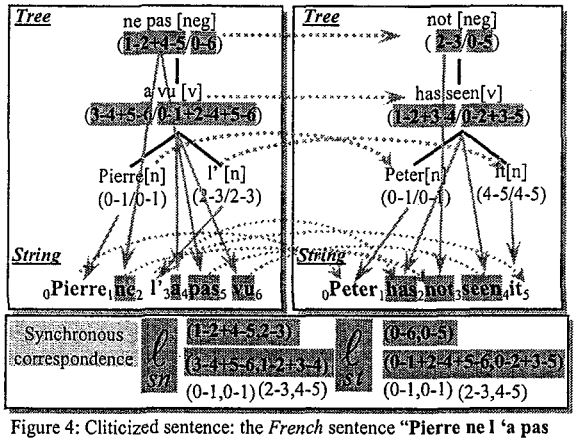

Figure 4: Cliticized sentence: the *French* sentence **"Pierre ne l 'a pas vu"** and its corresponding *English* sentence **"Peter has not seen it"**.

Figure 5 exemplifies a case where the number of nodes in the synchronized SSTCs or subSSTCs is the same, but they exhibit different structures. Nodes participating in the domination relationship in one SSTC may be mapped to nodes neither of which dominates the other (i.e. elimination of dominance).

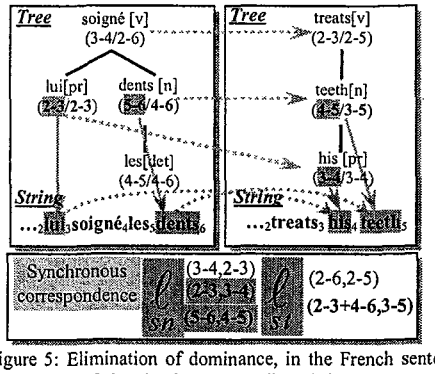

Figure 5: Elimination of dominance, in the French sentence **"le docteur lui soigné les dents"** and its corresponding English sentence **"the doctor treats his teeth"**.

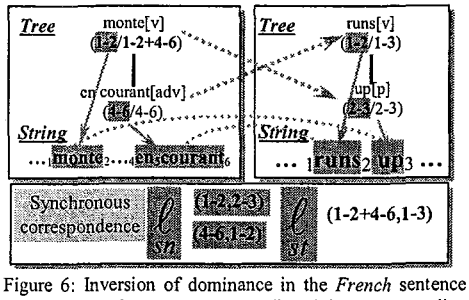

Figure 6: Inversion of dominance in the *French* sentence **"Jean monte la rue en courant"** and its corresponding *English* sentence **"John runs up the street"**.

Another even more extreme relationship between the synchronized pair involving inverted correspondences is exemplified in Figure 6.

Figure 7, depicts the case when partial subtree/s from the first SSTC has/ve a synchronous correspondence with partial subtree/s in the second SSTC. The *German* word "**beschenkte**" corresponds to the *English* phrase "**give present**" which is a partial subtree from the tree rooted by the word "**give**" in the *English* SSTC. This synchronous correspondence is recorded under the ℓ_{st} where the operation (-: minus) is used to calculate the Y:STREE interval/s for the partial subtree/s.

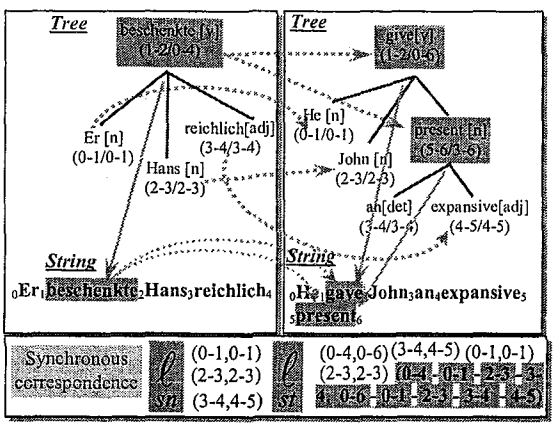

Figure 7: Partial subtree/s correspondence: the *German* sentence "**Er beschenkte Hans reichlich**" and its corresponding *English* sentence "**He gave John an expensive present**"; i.e. the use of (-) operation to calculate the Y:STREE interval.

5. SYNCHRONOUS CORRESPONDENCE CONSTRAINTS BETWEEN NATURAL LANGUAGES (NLs)

As we mentioned in Section 2, in the SSTC the correspondences between the surface text and the associated representation tree structure are ensured by means of intervals; i.e. (X:SNODE, Y:STREE). This explicitly indicates which word/s of the text correspond/s to which node in the tree. For describing a NL using SSTC, a set of constraints were defined to govern such correspondences (Lepage, 1994):

*- **X:SNODE** and **Y:STREE** intervals are governed by the following constraints:*

*i) **Global correspondence:** an entire tree corresponds to an entire sentence.*

*ii) **Inclusion:** a subtree which is part of another subtree T, must correspond to a substring in the substring corresponding to T.*

*iii) **Membership:** a node in a subtree T, must correspond to a word which is member of the substring corresponding to T.*

In a similar manner, in order to describe the synchronous correspondences between NLs using S-SSTC, we define a set of constraints to govern the synchronous correspondences between the different NLs. These constraints will be used to make explicitly the synchronous correspondences in a natural manner.

*- ℓ_{sn} and ℓ_{st} are governed by the following **constraints:***

- ***Singleness:** A node N which has a synchronization correspondence, can participate in one and only one $\ell \in \ell_{sn}$, and one and only one $\ell \in \ell_{st}$. This means allowing one-to-one, one-to-many and many-to-many, but the mappings do not overlap.*
- ***Inclusion:** Given two ℓ_{st} correspondence pairs $\ell_{st_1} = (\ell_{st_{s_1}}, \ell_{st_{t_1}})$ and $\ell_{st_2} = (\ell_{st_{s_2}}, \ell_{st_{t_2}})$, ℓ_{st_1} and ℓ_{st_2} satisfy the inclusion constraint if and only if $\ell_{st_{s_1}} \subseteq \ell_{st_{s_2}}$ and $\ell_{st_{t_1}} \subseteq \ell_{st_{t_2}}$.*
- ***Membership:** Given two correspondence pairs $(\ell_{st_s}, \ell_{st_t}) \in \ell_{st}$ and $(\ell_{sn_s}, \ell_{sn_t}) \in \ell_{sn}$, ℓ_{sn} and ℓ_{st} satisfy the membership constraints if and only if $\ell_{sn_s} \subseteq \ell_{st_s}$ and $\ell_{sn_t} \subseteq \ell_{st_t}$. This means the lexical correspondences are always members in the structural correspondences.*
- ***Dominance:** Given two subtrees S and T, there is a correspondence $\ell \in \ell_{st}$ between S and T satisfy the dominance constraints if and only if $\forall\ \ell \subseteq STREE(S)$ correspond to $\forall\ \ell \subseteq STREE(T)$.*
- ***Globality:** Given a S-SSTC, there must be $\ell \in \ell_{st}$ satisfies the globality constraints between the the root node R_s of the entire tree in the first SSTC and the root node R_t of the entire tree in the second SSTC, if and only if $(\ell_{st_s}, \ell_{st_t}) \in \ell_{st}$ such that $\ell_{st_s} = STREE(R_s)$: INT(String) in the first SSTC, and $\ell_{st_t} = STREE(R_t)$: INT(String) in the second SSTC. This mean the whole tree in the first SSTC corresponds to the whole tree in the second SSTC, and the whole string in the first SSTC corresponds to the whole string in the second SSTC).*

Note that these constraints can be used to license only the linguistically meaningful synchronous correspondences between the two SSTCs of the S-SSTC (i.e. between the two languages). For instance, when building translation units in EBMT approaches (Richardson et al., 2001), (Al-Adhaileh &Tang, 1999), (Sato & Nagao, 1990), (Sato, 1991), (Sadler & Vendelmans, 1990), etc., where S-SSTC can be used

to represent the entries of the BKB or when S-SSTC used as an annotation schema to find the translation correspondences (lexical and structural correspondences) for transfer-rules' extraction from parallel parsed corpus (Menezes & Richardson, 2001), (Watanabe et al., 2000), (Meyers et al., 2000), (Matsumoto et al., 1993) and (kaji et al., 1992). Note that the grammar alignment rules used in (Menezes & Richardson, 2001) can be reformulated using these constraints to construct the transfer mappings from a synchronous *source-target* example.

Figure 8 shows an example from Menezes and Richardson (2001), the logical form for the Spanish-English pair: ("***En Información del hipervínculo, haga clic en la dirección del hipervínculo***", "***Under Hyperlink Information, click the hyperlink address***").

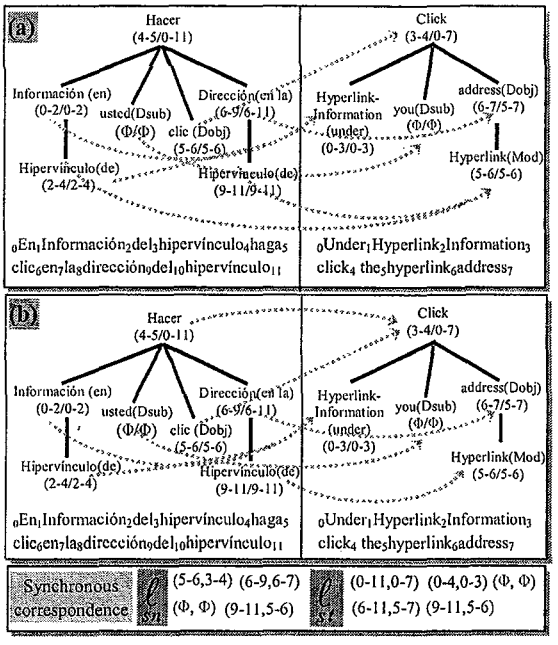

Figure 8: (a) the lexical correspondences, (b) the structural correspondences after applying the constraints.

Recently, the development of machine translation systems requires a substantial amount of translation knowledge typically embodied in the bilingual corpora. For instance, the development of translation systems based on transfer mappings (rules or examples) that automatically extracted from these bilingual corpora. All these systems typically first obtain a tree structures (normally a predicate-argument or a dependency structure) for both the source and target sentences. From the resulting structures, lexical and structural correspondences between the two structures are extracted, which are then presented as a set of examples in a bilingual knowledge bank (BKB) or transfer rules for translation process.

However, what has so far been lacking is a schema or a framework to annotate and express such extracted lexical and structural correspondences in a flexible and powerful manner. The proposed S-SSTC annotation schema can fulfill this need, and it is flexible enough to handle different type of relations that may happen between different languages' structures. S-SSTC very well suited for the construction of a BKB, which is needed for the EBMT applications. Al-Adhaileh and Tang (2001) presented an approach for constructing a BKB based on the S-SSTC.

6. DISCUSSION AND CONCLUSION

In this paper, we have presented the concept of S-SSTC which provides a flexible way of representing both the lexical and structural correspondences between two languages, namely:

1. ***Node–to–node lexical correspondence***: This correspondence is recorded in terms of pair of intervals (X_s,X_t) where X_s and X_t is SNODE interval(s) for the source and the target SSTC respectively.
2. ***Subtree–to–subtree structural correspondence***: This correspondence is very much needed for relating the two different languages at a level higher than the lexical level, i.e. at the level of phrases. This correspondence is recorded in terms of pair of intervals (Y_s,Y_t) where Y_s and Y_t is STREE interval(s) for the source and the target SSTC respectively.

S-SSTC is a flexible schema, which is able to handle non-standard phenomena that may occur between different languages. We conclude this paper with some observations on the S-SSTC:

i. A natural way to express correspondences between two languages (e.g. a text and its translation) in a very fine-grained manner which is very much nedded in handling some non-standard cases.

ii. A natural way to specify bi-directional structural transfer, as SSTC is used to specify both the structural analyzers and generators (i.e. bi-directional).

iii. Synchronous SSTC inherits from the SSTC the independence from the choice of the tree structure. Such property is needed to ensure that the proposed S-SSTC can be used as a generic annotation schema which is capable of annotating a string with representation structures proposed by different linguistic theories.

iv. The transfer between two languages, such as source and target languages in machine translation, can be done by putting directly into correspondence large elementary units without going through some interlingual representation. In particular the flexibility of using the proposed S-SSTC as annotation for translation systems

that automatically extract transfer mappings (rules or examples) from bilingual corpora.

v. Synchronous SSTC can be easily extended to record the correspondences between more than two languages, hopefully with transitive property, especially in constructing multilingual knowledge banks (MKB) (i.e. synchronization between multiple languages)

Furthermore, the SSTC structure can easily be extended to keep multiple levels of linguistic information, if they are considered important to enhance the performance of the machine translation system (i.e. ***Features transfer***). For instance, each node representing a word in the annotated tree structure can be tagged with part of speech (POS), semantic features and morphological features.

A GUI editor has been implemented for view, edit, create and correct the S-SSTC components, as illustrated in Figure 9.

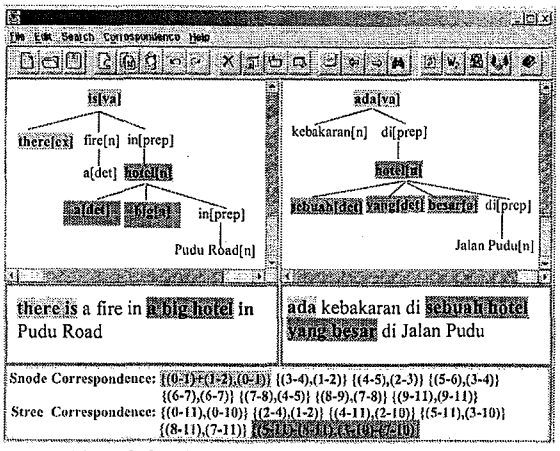

Figure 9: Synchronous SSTC Editor.

REFERENCES

Abeillé, A., Schabes, Y. and Joshi, A. (1990). Using lexicalized TAGs for machine translation. In *Proceedings of the 13th International Conference on Computational Linguistics (COLINGS'90)*, Helsinki, Finland, pp 1-6.

Al-Adhaileh, M.H. and Tang, E.K. (1999). Example-Based Machine Translation Based on the Synchronous SSTC Annotation Schema. In *Proceedings of Machine Translation Summit VII*, Singapore, pp 244-249.

Al-Adhaileh, M.H. and Tang, E.K. (2001). Converting a Bilingual Dictionary into a Bilingual Knowledge Bank Based on the Synchronous SSTC Annotation Schema. In *Proceedings of Machine Translation Summit VIII*. Spain, pp 351-356.

Boitet, C. and Zaharin, Y. (1988). Representation trees and string-tree correspondences. In *Proceedings of the 12th International Conference on Computational Linguistics (COLINGS-88)*, Budapest. Hungary, August, pp 59-64.

Kaji H., Kida Y., and Morimoto Y. (1992), Learning Translation Templates from Bilingual Text. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, France, pp 672-678.

Lepage, Y. (1994). Texts and Structures – Pattern-matching and Distances, ATR report TR-IT-0049, Kyoto, Japan.

Matsumoto, Y., Ishimoto H., and Utsuro, T. (1993). Structural Matching of Parallel Texts. In *Proceedings of the 31th annual meeting of Association for Computational Linguistics (ACL-93)*, pp 23-30.

Menezes, A. and Richardson, S. (2001). A Best-first Algorithm for Automatic Extraction of Transfer Mappings from Bilingual Corpora. In *the workshop on Data-Driven Machine Translation, at the 38th Annual Meeting of the Association for Computational Linguistic (ACL 2001)*, Toulouse, France.

Meyers A., Kosaka M., and Grishman R. (2000). Chart-based Transfer Rule Application in Machine Translation. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*. Saarbrücken, Germany / Luxembourg.

Rambow, O. and Satta, G. (1996). Synchronous Models of language. In *Proceedings of the 34th Meeting of the Association for Computational Linguistics (ACL-96)*, Santa Cruz, CA, USA.

Richardson S., Dolan W., Menezes A. and Pinkham J. (2001). Achieving commercial-quality translation with example-based methods. In *Proceedings of Machine Translation SUMMIT VIII*, Spain, pp 293-297.

Sadler V. and Vendelmans, R. (1990). Pilot Implementation of a Bilingual Knowledge Bank. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, Vol. 3, Helsinki, Finland, pp 449-451.

Sato S. and Nagao M. (1990). Towards Memory-based Translation. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, Vol. 3, Helsinki, Fenland, pp 247-252.

Sato, S. (1991). *Example-Based Machine Translation.* Ph.D. thesis, Kyoto University, Japan.

Shieber, S. (1994). Restricting the Weak Generative Capacity of Synchronous Tree Adjoining Grammar. *Computational Intelligence*, 10(4): 371-385.

Shieber, S. and Schabes, Y. (1990). Synchronous Tree Adjoining Grammars. In *Proceedings of the 13th International Conference on Computational Linguistics (COLINGS-90)*, Helsinki, Finland, pp 253-258.

Somers, H. (1999). Review article: Example-based Machine Translation, *Machine Translation*, 14: 113-157.

Tang E. K. (1994). *Natural Language Analysis in Machine Translation (MT) Based on the String-Tree Correspondence Grammar (STCG)*. PhD. thesis, Universiti Sains Malaysia, Penang, Malaysia.

Tang, E. K. and Zaharin, Y. (1995). Handling Crossed Dependencies with the STCG. *In Proceedings of NLPRS'95*, Seoul, Korea.

Watanabe H., Kurohashi S., and Aramaki E. (2000). Finding Structural Correspondences from Bilingual Parsed Corpus for Corpus-based Translation. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, Luxembourg/Saarbruecken, Germany.