Transfer in Experience-Guided Machine Translation

Gang Zhao
Lernout & Hauspie Speech Products

Junichi Tsujii
Information Science, University of Tokyo
Language Engineering, UMIST

Abstract

Experience-Guided Machine Translation (EGMT) seeks to represent the translators' knowledge of translation as *experiences* and translates by analogy. The transfer in EGMT finds the experiences most similar to a new text and its parts, segments it into units of translation and translates them by analogy to the experiences and then assembles them into a whole. A research prototype of analogical transfer from Chinese to English is built to prove the viability of the approach in the exploration of new architecture of machine translation. The paper discusses how the experiences are represented and selected with respect to a new text. It describes how units of translation are defined, partial translation is derived and composed into a whole.

1   INTRODUCTION

Translation is 'negotiation' [10] between two language systems in specific contexts. It involves two operations: *segmentation* and *transference.* Segmentation [11, 30] breaks up the source text into *units of translation* [8, 12, 21, 29, 30]. Transference establishes and combines translation equivalence.

The knowledge of translation equivalence is not treated as it should be in dominant machine translation architectures. The transfer systems [2, 14, 17, 23, 28]. based on theories of grammar, recognise the necessity of descriptions of translation correspondences but tend to over-simplify the issue of units of translation. They implicitly assume that the source-language constructs derived according to a monolingual grammar constitute units of translation for bilingual consideration [27]. In interlingua systems [4, 9, 18, 28], the translation is seen as a process of characterising the source text in terms of intermediate expression and verbalising the intermediate expression in the target language. The underlying assumption is that the knowledge of professional translators is nothing different from that of mere bilingual speakers, hence no need for contrastive knowledge.

The difficulty in an expert-system approach to machine translation comes from the modelling and processing the translators' expertise. A formal and generic description of segmentation and transference is problematic [3, 19, 24]. Generalisation may be made but with too many exceptions for effective knowledge processing. In addition, the justification of an equivalence in a particular context of translation involves multidisciplinary factors. 'Since these factors belong to a variety of different areas of life, there is a question whether a comprehensive account of translation in the form of a coherent and homogeneous theory can ever be achieved.' [7]

This paper explores the possibility of describing the translator's knowledge in concrete terms, generalising from the specific *experience* of translation at run time and translating by analogy to the most appropriate experience. In this approach, the translation engine computes the similarity of a new text and experience texts, identifies the best analogue (the most similar experience) and transfers the new text by the proposal from the best experience. A prototype is implemented to perform analogical translation from Chinese into English.

EGMT is similar to the example-based machine translation [5, 6, 15, 16, 20, 25, 26] in that it translates according to specific cases of past translation. But its similarity computation is not confined to lexical dimensions. Multiple dimensions of similarity such as morphological, syntactic, textual, communicative are considered in selecting the best analogue, using the same algorithm of similarity computation. In addition, the partial similarity and partial solution are treated in both similarity computation and translation composition.

2   EXPERIENCE AS KNOWLEDGE FOR TRANSLATION

The expertise of translators is regarded as a collection of *experiences* about specific cases of translation. Each experience is concerned with translation equivalents and their justifications. In the current pro-

totype. equivalents are described in lexico-syntactic terms whereas the justification for the translation equivalence is captured on various levels of linguistic description, such as textual, semantic description. The decision comes from the conviction that translation is essentially a linguistic process, though various extra-linguistic factors are considered.

The experience relates a *source text* to a *target text*. They can be of any grammatical units such as morphemes, words, phrases, or discourse and do not have to be of parallel categories. The experience consists of the Description of the source text and Mappings from it to the target text.

The Description depicts the (extra-)linguistic properties of the source text, which are justifications of the current experience. It uses two formal constructs of description: constituent structure and feature collocation. The constituent structure is put in terms of phrase structures with functional annotations. The feature collocation describes (co-)occurrences of (extra-)linguistic features in or among the constituents. The illustration below is parts of the Description of the sentence, wo[I] zhengli[sort out] hao[well] ziliao[data] zai[then] huiqu[go back] (I shall sort out the data before I go back.)

```
[S (np//vp)
    [NP (pron)
        [PRON wo[I]]]
    [VP (vp₁ | vp₃)
        [VP₁ (vp₂\\np₁)
            [VP₂ (v\adj)
                [V zhengli[sort out]]
                [ADJ hao[well]]]]
            [NP₁ (n)
                [N ziliao[data]]]]
        [VP₃ (adv/vp₄)
            [ADV zai[again]]
            [VP₄ (v₁)
                [V₁ huiqu[go back]]]]]]
```

$< s, \{s.speech\text{-}act = commissive\} >$
$< pron, \{pron.semantic\text{-}type = I/02\} >$
$< adj, \{adj.semantic\text{-}type = III/110\} >$
$< n, \{n.semantic\text{-}role = patient\} >$
$< v, \{v.semantic\text{-}frame = agent\_patient,$
$v.semantic\text{-}type = XIII/75/13\} >$
...

The Description is used mainly in *experience selection*, which compares the experience texts with the new text in terms of their similarity in structure and features in order to identify the most similar experience.

The Mappings between the source and target texts holds information required in segmentation and transference of the new text. In each statement of equivalence, units of translation, their corresponding equivalents in the target text and feature constraints on the equivalence are specified.

Below are three statements of equivalence about the same sentence described above.

1. $[S*[NP][VP[VP_1][VP_3[ADV][VP_4]]]] =$
$[S_t[S_{1t}[NP_t][VP_{1t}]][S_{2t}[CONJ_t][S_{3t}[NP_t][VP_{4t}]]]$
...

6. $[S[NP][VP[VP_1][VP_3*[ADV][VP_4]]]] =$
$[S_t[S_{1t}][S_{2t}[CONJ_t][S_{3t}[NP_t][VP_{4t}]]]] |$
$< s, \{pron.semantic\text{-}role = agent,$
$v.semantic\text{-}frame = agent\_patient\} > \wedge$
$< s, \{pron.semantic\text{-}role = agent,$
$v_1.semantic\text{-}frame = agent\_\} > \wedge$
$< adv, \{adv.semantic\text{-}type = XVI/36\} > \wedge$
$< adv, \{adv.cohesion = logical/sequential\} >$
...

12. $[VP_2[V][ADJ*hao[well]]] = \{VP_{2t}[V_t]\} | <$
$adj, \{adj.semantic\text{-}role = manner/aspect/perfect\} >$
$\wedge < s, \{s.deictics = time/future\} >$

The left-hand side of the equation defines the unit of translation. The asterisked node is the top node of the unit concerned. The nodes outside its domination is the structural context. The right-hand side is the translation equivalent. The equivalence is indicated by the subscripts. The section following the bar specifies contextual constraints in features. Statement 12 states that the word hao[well] as an adjective is transferred to nothing in English if it is in $[VP[V]...]$ and the sentence is of perfect aspect and future time.

3   SIMILARITY COMPUTATION

A linguistic object is described by a set of features along (extra-)linguistic dimensions. The comparison of two discrete linguistic objects is possible when their feature descriptions are projected onto a single feature system. The similarity between two feature descriptions is thus the degree of their commonality with respect to this feature system.

The feature system, represented as a *feature graph*, is a single-rooted, directed and acyclic graph. The vertices represent descriptive categories and the edges indicate the relationship among the categories. The external vertex symbolises a feature used in a feature description. A feature projected onto the feature system is in effect a sub-graph of this general feature graph.

Each vertex is weighted to signify its geographical location in the feature graph relative to the root. It is called *positional weight*. It is relevant to a particular *weight path*, P_v, which is a list of connected vertices beginning from the root to some external vertex, v. It is assigned by the following formula, where $u \in P_v$, i

is the ordinal position of u in P_v and $|P_v|$ is the length of the weight path.

$$pWeight(u, P_v) = (i_u - 1) \div (|P_v| - 1)$$

The longer the distance is from the root, the heavier is the weight. It is expressed by a numerical value between 0 and 1 inclusive. The weight of the root is zero and that of external vertexes is one.

3.1 Computation of Feature Similarity

Given two features, u and v, their relationship is determined by their respective locations in the general feature graph. Their similarity is computed on the basis of $P_u \cap P_v$, where P_u is the path from the root to u and P_v to v. The last member of $P_u \cap P_v$ is the least generic common dominating vertex for u and v, d. The positional weight of $D_{u,v}$ is the similarity between u and v.

$$vSimilarity(u, v) = \frac{pWeight(d.P_u) + pWeight(d.P_v)}{2}$$

The algorithms discussed in this section has been used to study the behavioural tendency of 26000 Chinese words. The positive results are reported in [33].

3.2 Computation of Graph Similarity

The vertex similarity is useful for comparing primitive linguistic objects, features in a feature graph. But the comparison of two complex linguistic objects is a process of inter-graph comparison in terms of graph similarity rather than vertex similarity. Features from different graphs must be compared and their composite similarity computed in order to determine their graph similarity.

The relevant weight paths for vertex similarity computation start from one and the same vertex, the root of the graph and least generic common dominating vertex is a member of both paths. The vertex similarity required for intergraph comparison is based on the concept of vertex parallelism. First of all, the comparison involves two different graphs, F and G, where $F \neq G$ and $P_v^F \cap P_w^G = \emptyset$. Secondly, the roots of the graphs under comparison, r^F and r^G, must be of comparable descriptive category, $r^F \approx r^G$, otherwise the similarity is undefined. Thirdly, the least generic parallel vertices for v in F and w in G is the last pair $< a^F, b^G >$ of $P_v^F \cap P_w^G$

Given two comparable feature graphs (rooted with an identical descriptive category), their commonality is computed through the combinations of their respective sets of external vertexes. Each path of one graph will be compared with those of another, one by one. The graph similarity is computed by the following formula, where m is the number of external vertices of F and n that of G. $P_{v_i}^F$ is the weight path in F from its root to one of its external vertices.

$$gSimilarity(F, G) =$$
$$\sum_{i=1,j=1}^{m,n} \frac{pWeight(a.P_{v_i}^F) + pWeight(b.P_{w_j}^G)}{2}$$
$$\overline{m \times n}$$

In Figure 1, the external vertexes of Graphs I and II form 12 inter-group combinations. The feature similarity of vertexes in each vertex pair is computed on the basis of their parallel least generic dominating vertexes. For instance, Vertex 5 of Graph I and Vertex 7 of Graph II have two parallel dominating vertexes of the same category from their respective roots, namely, S and VP. The positional weight of VP in Graph I relevant to Vertex 5 is 0.5 and that in Graph II relevant to Vertex 7 is 0.3333. The feature similarity of the paired vertexes, therefore, is 0.4167. The degree of commonality is defined as the average of the feature similarities of all the combinations. The commonality between Graphs I and II, as indicated graphically in Graph III, is thus 0.2945.
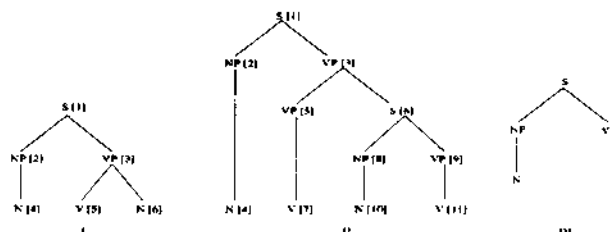

Figure 1: A Comparison of Phrase Structures

4 EXPERIENCE SELECTION

The experience selection compares the new text with the texts of the experiences and identifies the most similar experience. The process first retrieves a set of *relevant experiences*, which match with the new text on a particular dimension of description. This dimension is used as the indexing system of the experience bank. The current prototype explores the lexico-syntactic indexes of experiences. The initial selection thus returns a set of experiences having the same constituent pattern.

Similarity can be computed along multiple dimensions: lexical, syntactic, functional or semantic, as long as the knowledge model underlying the description is in a single-rooted, directed and acyclic graph. The similarity is computed in terms of lexical properties, semantic features, structural constituency and selectional restriction. The lexical similarity is computed from morphological and syntactic features (see [31] for a description of the Chinese lexicon used for the purpose). The structural similarity and functional similarity are computed on the basis of constituent structure, using the algorithms of similarity computation described above.

The similarity of feature co-occurrences takes the following steps.

- Given a feature cooccurrence in the experience. $< \sigma. \{n.P = p, v.Q = q\} >$, where P at the vertex n has the value p and Q at the vertex v is instantiated to q, σ is the vertex that spans the feature relationship, the same feature cooccurrence, $< \sigma_t, \{n_t.P = x, v_t.Q = y\} >$, is hypothesised between n_t and v_t on condition that

  - $n \equiv n_t$, $v \equiv v_t$, $\sigma \equiv \sigma_t$, in terms of structural category

  - $\sigma \equiv \sigma_t$ in terms of their functional composition

- Compute the similarities of p and x, q and y and average the sum similarity values.

In the current study, the similarity between an experience and a new text is the average of similarity values computed along all the dimensions, though *relevance weight* can be added to highlight a particular dimension.

5 EXPERIENCE APPLICATION

The experience selection finds the best experience for each constituent of the new text. The experience application follows these experiences to translate the parts of the new text and assemble the translation of the parts into a whole. The process requires two data: the decomposition of the new text and the best analogues for its constituents.

5.1 Segmentation

Though the new text is decomposed by the source language grammar, there is no assumption made that the units of translation coincide with its constituency. The first step is thus to segment the part of the new text by analogy to the best experience. This is a process of mapping top-down the constituent structure of the experience text to that of the new text. The more they are alike, the more reliable is the segmentation. Suppose that $[VP \ [VP][VP]]$ in Graph I is the part to be translated and Graph II is the analogue in Figure 2. The subgraph marked out by the dotted line in Graph II is the unit of translation explicitly stated in the experience. The new text does not match completely with the experience text. But the experience can be generalised to ignore the surface difference and make a valid proposal of segmentation marked by the dotted line in Graph I. The dotted line in Graph I encircles a proposal of segmentation.

5.2 Transferring the parts

The segmentation of the new text by analogy to the experience text is also a process of setting up mappings between the parts of the new text and the experiences. Figure 3 illustrates the relationships required
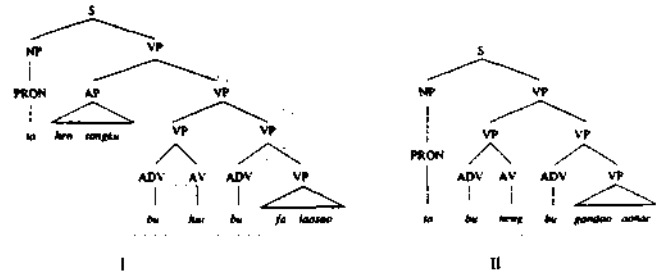

Figure 2: Segmentation

for deriving partial translation. Graph I is the part of the new text. Graph II is the corresponding part of the experience. Graph III is the transference of the part of the experience text and Graph IV the transference of the part of the new text. The curves between Graphs I and II are mappings derived during segmentation. The curves between Graphs II and III are specified in the Mappings of the experience. Graph IV is created according to the structure of Graphs I, II and III and mappings among them. The transference can be structural transformation or lexical translation or both. The transfer process derives translation, following the mappings among these data objects.

5.3 Assembling the transfer results

The constituency of the new text is an important clue to how the transfer results are combined into a whole. The decomposition of the new text is f/s (see Figure 4), where p and q are its vertices. T_p and T_q are the results of transference for p and q respectively. T_q can be grafted onto T_p since p directly dominates q.

There are three consequences of attempting to graft T_q onto T_p. The complete match occurs when T_p and T_q are equivalent or the former is a subgraph of the latter. For instance, T_q can be a substructure of T_p when $[x \ [y][z]]$ match with either $[a \ [b][c]]$ or with $[c \ [d][e]]$. The extension can be partial or total. The total extension occur where there are no overlapping children between T_p and T_q, for instance when x matches with b, d or e. In the case of a partial extension, the non-overlapping vertices will be added. Suppose x is mapped into c and y into d, the result of the extension will be $[a \ [b][c \ [d][e][z]]]$. The combination fails if x is not mapped into any vertices in T_p. If T_p and T_q are not immediately related, their combination is undefined.
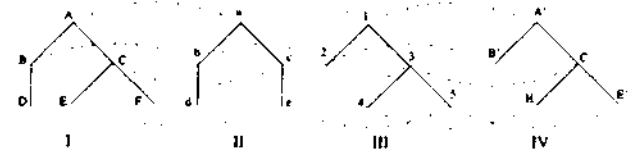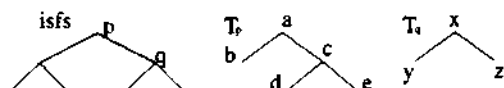

Figure 3: Resolution Derivation

Figure 4: Assembling results of resolutions

6 Analysis of examples

This section describes six examples to illustrate the robustness in the architecture of EGMT. They show that the system is capable of recognising partial similarities, defining units of translation and composing the partial translation and that it can be 'stretched' to cover unexperienced phenomena on the basis of incomplete knowledge. For better illustration, the system uses a very limited experience bank of 78 instances of annotated experiences and yet manages to handle some common phenomena in Chinese-English translation.[2]

Original: women⌈we⌋ diaocha⌈investigate⌋ hao⌈well⌋ le⌈have-been⌋ zai⌈then⌋ xie⌈write⌋ baogao⌈report⌋.
Translation: We shall investigate before we write the report.

Figure 5 sums up the description and the results of experience selection for the sentence. The sentence is decomposed into phrase structures with functional annotations and feature annotations in analysis. The result of experience selection is recorded between angled brackets at each node. The integer indicates the number of relevant experience retrieved for that subproblem. The real number is the aggregate similarity of the experience chosen for it. The translation is composed of partial translations derived according to 16 partially similar experiences.

Original: ta⌈he⌋ xiang⌈to⌋ ni⌈you⌋ baogao⌈report⌋ le⌈have-been⌋ shigu⌈accident⌋ de⌈⌋ yuanyi⌈cause⌋.
Translation: He reported the cause of the accident to you.
The correct rendering should be 'has reported' due to the time feature of *present*, but the only experience of treating baogao⌈report⌋ le⌈have-been⌋ is in a sentence with the time feature of *past*.

---

[2]The annotation uses 20 synonym types from 1750 synonymous categories in *A Concise Synonym Dictionary of Chinese* [13], 10 semantic roles and 5 semantic frames from dependency relationship description of Chinese [32] (22 roles and 19 frames), 19 phrase structure types, 10 word classes, 6 types of grammatical dependency, 6 textual features. 4 speech acts, 5 deictic features. The experience bank consists of one textual, nine sentential, thirty-six phrasal and thirty-one lexical experiences. The experiments conducted involve similarity computation along semantic, syntactic, textual and communicative dimensions and cover complex transference, such as change of grammatical structure (simple into complex sentence, word classes), lexical gaps, translation polysemy.
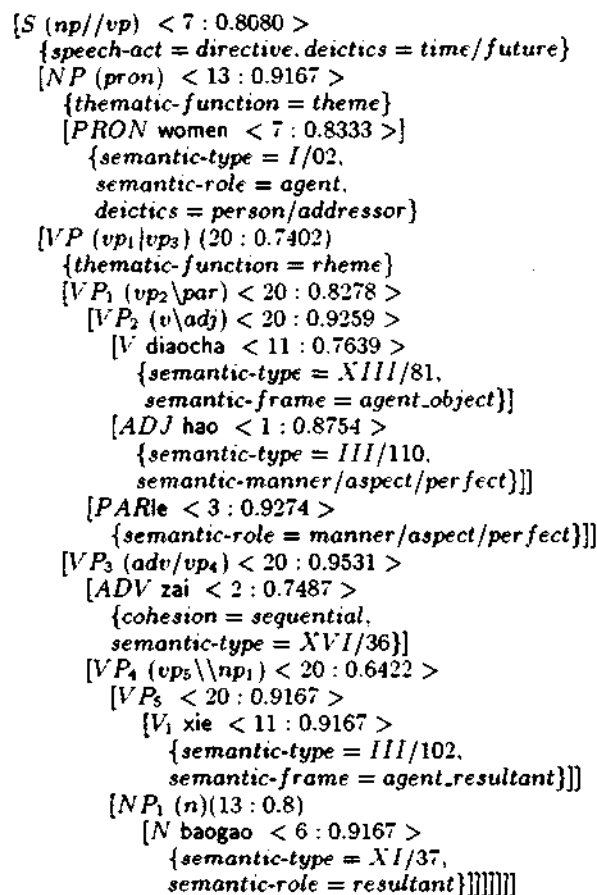
$[S\ (np//vp)\ < 7 : 0.8080 >$
  $\{speech\text{-}act = directive.deictics = time/future\}$
  $[NP\ (pron)\ < 13 : 0.9167 >$
    $\{thematic\text{-}function = theme\}$
    $[PRON\ \text{women}\ < 7 : 0.8333 >]$
      $\{semantic\text{-}type = I/02,$
      $semantic\text{-}role = agent,$
      $deictics = person/addressor\}$
  $[VP\ (vp_1|vp_3)\ (20 : 0.7402)$
    $\{thematic\text{-}function = rheme\}$
    $[VP_1\ (vp_2\backslash par)\ < 20 : 0.8278 >$
      $[VP_2\ (v\backslash adj)\ < 20 : 0.9259 >$
        $[V\ \text{diaocha}\ < 11 : 0.7639 >$
          $\{semantic\text{-}type = XIII/81,$
          $semantic\text{-}frame = agent\_object\}]$
        $[ADJ\ \text{hao}\ < 1 : 0.8754 >$
          $\{semantic\text{-}type = III/110,$
          $semantic\text{-}manner/aspect/perfect\}]]$
      $[PARle\ < 3 : 0.9274 >$
        $\{semantic\text{-}role = manner/aspect/perfect\}]]$
    $[VP_3\ (adv/vp_4)\ < 20 : 0.9531 >$
      $[ADV\ \text{zai}\ < 2 : 0.7487 >$
        $\{cohesion = sequential,$
        $semantic\text{-}type = XVI/36\}]$
      $[VP_4\ (vp_5\backslash\backslash np_1)\ < 20 : 0.6422 >$
        $[VP_5\ < 20 : 0.9167 >$
          $[V_1\ \text{xie}\ < 11 : 0.9167 >$
            $\{semantic\text{-}type = III/102,$
            $semantic\text{-}frame = agent\_resultant\}]]$
        $[NP_1\ (n)(13 : 0.8)$
          $[N\ \text{baogao}\ < 6 : 0.9167 >$
            $\{semantic\text{-}type = XI/37,$
            $semantic\text{-}role = resultant\}]]]]]]$

Figure 5: A Result of Experience Selection

Original: wo⌈I⌋ qu⌈go⌋ xie⌈write⌋ baogao⌈report⌋, ziliao⌈data⌋ ni⌈you⌋ zhengli⌈sort out⌋.
Translation: I shall go [and] write the report. As for the data you sort out the data.
The sentence is translated into a text of two sentences. No distinction is made between sentences and texts in the experience description. They are handled by the same mechanism. The Chinese sentence is a complex sentence consisting of two clauses connected by a comma. The object of the second clause is topicalised, which is a common phenomenon in Chinese. The system has experience of translating topicalisation in Chinese with the phrase 'as for'. Though the experience topicalises a constituent with different a semantic feature, the system finds overwhelming evidence to ignore the difference and translates this accordingly. The abnormality with 'go write' is due to the inadequacy of the best experience.

Original: ziliao⌈data⌋ zhunbei⌈prepare⌋ hao⌈well⌋ le⌈have-been⌋.
Translation: The data have been prepared.

The sentence has a patient subject with no explicit passive markers. There are two properties that are uncovered in the experience bank. One is the place where the word. ziliao⌈data⌋ occurs in the sentence. It has been used as grammatical object, but not as subject. The other is the two post-modifiers of the verb, zhunbei⌈prepare⌋. The words hao⌈well⌋ and le⌈have-been⌋ have occurred in the experiences as postmodifiers separately. In the present case, they occur one after another, modifying the same verb. The system manages to find the best examples for dealing with these unseen properties and piece the partial resolutions properly.

Original: ta⌈he⌋ zhengli⌈sort out⌋ le⌈have-been⌋ wu⌈five⌋ fen⌈copy⌋ fenxi⌈analyse⌋ jieguo⌈result⌋.
Translation: He sort out five copies of the result of the analysis.
The correct version would be 'He sorted out five copies of the result of the analysis.' There is experience of treating le⌈have-been⌋ in a sentence with the present time. The time of the current example, however, is characterised as past. The main predicate verb, 'sort out' could be inflected as part of morphological post-processing.

Original: jieguo⌈result⌋ wo⌈I⌋ fenxi⌈analyse⌋ le⌈have-bee⌋.
Translation: As for the result, I have ⎕ the result.
The correct version would be 'As for the result, I have analysed it.' The experience about translating fenxi⌈analyse⌋ is only concerned with its being used a modifier of a noun. In the present case, it is used as a main verb. Although the best experience is found for it (treating it similarly as zhunbei⌈prepare⌋), the string-to-string translation regardless of contextual constraints is not allowed in the current implementation, the translation stops short of putting the string equivalent in the sentence.

7 CONCLUSION

The use of similarity computation in analogical transfer has notable significance to the architecture of machine translation systems.

Firstly, the applicability of experience is expressed in approximate values bounded between 0 and 1. It indicates a gradual continuum of approximation to the perfect solution. The approach makes no assumption that the problem space can be clearly partitioned and neatly described in Boolean terms. It works effectively with overlapping coverage of experiences. The system is inherently robust with no need for ad hoc fail-soft procedures to deal with exceptional cases.

Secondly. the applicability of an experience to a particular problem is relative to the body of competing experiences in the experience bank. Its performance is improved by accumulating experiences.

Thirdly. though experiences are descriptions of spe-

cific cases of translation, the mechanism of similarity computation allows them to be 'stretched' and used on different problems. In other words they can be abstracted to a certain point where they are similar to the problem. The level of abstraction is captured by the degree of similarity. The more that has to be abstracted away, the less similar the experience is to the new problem. Moreover, the abstraction of experiences is dynamic: the applicability of an experience is undefined without a specific problem supplied at run time. The dynamic abstraction is an important feature of experience-guided machine translation in contrast to the rule-based architecture, where the level of abstraction and applicability of a rule is determined beforehand at compile time. It not only improves the reusability of experiences but also helps in knowledge coding. The knowledge expert can concentrate on specific cases of translation instead of generalising about an ill-structured knowledge domain.

A second important feature of the analogical transfer is that it is capable of recognising partial similarity and combining partial solutions. A new text with partial resemblance to various experience texts is translated by analogy to the partial translations suggested by these texts. In the case-based reasoning in EGMT, the case is not a monolithic whole. It is structured and can be seen as containing subcases. The compositionality of cases is essential for processing creativity in the use of language.

In EGMT, the unit of translation has an important place. It is explicitly specified in the experience and analogised in treating a new text.

The results support the viability of the approach and prove that the approach is worth pursuing further. The future research will concentrate on extensions of the experience bank in contents and size so that the translation processes can be examined against a wider scope of data and improved accordingly.

References

[1] ANANIADOU, S., Ed. *Proceedings of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing* (Manchester, July 1992). Centre for Computational Linguistics, UMIST.

[2] ARNOLD, D., AND SADLER, L. EUROTRA: an assessment of the current state of the EC's MT programme. *Translating and the Computer 13* (1992), 137–151.

[3] BROECK, R. V. The concept of equivalence in translation theory: some critical reflections. In *Literature and Translation*, J. S. Holmesand, J. Lambert, and R. V. Broeck, Eds. Acco, Leuven. 1978.

[4] FARWELL, D., AND WILKS. Y. ULTRA: a multilingual machine translation. In *Progress in ma-*

chine translation, S. Nirenburg, Ed. IOS Press, Amsterdam, 1993, pp. 45–52.

[5] FURUSE, O., AND IIDA, H. Transfer-driven machine translation. In Ananiadou [1], pp. 95–111.

[6] FURUSE, O., AND IIDA, H. Constituent boundary parsing for example-based machine translation. In Proceedings of the 16th International Conference on Computational Linguistics (Kyoto, August 1994), pp. 105–111.

[7] GUTT, E. A. Translation and Relevance: cognition and context,. Basil Blackwell, Oxford, 1991.

[8] HARRIS, B. Bi-text: a new concept in translation theory. Language Monthly 54 (1988), 8–9.

[9] KAJI, H. HICATS/JE: a Japanese-to-English machine translation system based on semantics. In Machine Translation Summit (Hakone, September 1987), M. Nagao, Ed., pp. 101–106.

[10] KAY, M., GAWRON, J. M., AND NORVIG, P. Verbmobil: a translation system for face-to-face dialog. Centre for the Study of Language and Information, Standford, 1994.

[11] KELLY, I. D. K., Ed. Progress in Machine Translation: natural language and personal computers. Sigma Press, Wilmslow, England, 1989.

[12] LEDERER, M. La traduction simultanée, un poste d'observation du langage. In Interpréter pour Traduire [22], pp. 245–255.

[13] LIN, X.-G., AND FEI, B. Concise Synonym Dictionary of Chinese. Commercial Press, Bejing, 1987.

[14] MAEGAARD, B., AND PERSCHKE, S. Eurotra: general system design. Machine Translation 6, 2 (1991), 73–82.

[15] NAGAO, M. A framework of a mechanical translation between Japanese and English by analogy principle. In Artificial and Human Intelligence, A. Elithorn and R. Banerji, Eds. North Holland, Amsterdam, 1984, pp. 173–180.

[16] NAGAO, M. Some rationales and methodologies for example-based approach. In Ananiadou [1], pp. 82–94.

[17] NAGAO, M., TSUJII, J.-I., AND NAKAMURA, J. The Japanese government project for machine translation. In Machine Translation Systems, J. Slocum, Ed. CUP, Cambridge, 1985, pp. 141–186.

[18] NIRENBURG, S., AND LEVIN, L. Knowledge representation support. Machine Translation 4, 1 (March 1989), 25–52.

[19] PYM, A. Translation and Text transfer: an essay on the principles of intercultural communication. Peter lang, Frankfurt am Main, 1992.

[20] SATO, S., AND NAGAO, M. Toward memory-based translation. In Proceedings of the 13th International Conference on Computational Linguistics (Helsinki, July 1984), H. Karlgren, Ed., Yliopistopaino, pp. 247–252.

[21] SELESKOVITCH, D. La compréhension d'une pensée à travers son expression. In Interpréter pour Traduire [22], pp. 294–308.

[22] SELESKOVITCH, D., AND LEDERER, M. Interpréter pour Traduire. Didier, Paris, 1975.

[23] SLOCUM, J., BENNETT, W. S., BEAR, J., MORGAN, M., AND ROOT, R. METAL: the LRC machine translation system. In Machine Translation Today: the state of the art. M. King. Ed. Edinburgh University Press, Edinburgh, 1987, pp. 319–350.

[24] SNELL-HORNBY, M. Translation Studies: an integrated approach. John Benjamins, Amsterdam, 1988.

[25] SUMITA, E., AND IIDA, H. Experiments and prospects of example-based machine translation. In Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (Berkeley, 1991), pp. 185–192.

[26] SUMITA, E., IIDA, H., AND KOHYAMA, H. Translating with examples: a new approach to machine translation. In Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation (Linguistics Research Center, University of Texas at Austin, July 1990), pp. 203–212.

[27] TSUJII, J.-I., AND ANANIADOU, S. Knowledge-based processing in MT. In Proceedings of the First International Conference on Knowledge Base and Knowledge Sharing (Tokyo, 1993).

[28] VAUQUOIS, B., AND BOITET, C. Automated translation at Grenoble University. Computational Linguistics 11, 1 (1985), 28–36.

[29] VÁZQUEZ-AYORA, G. On the notion of an analytical unit of translation. Babel 28, 2 (1982), 70–81.

[30] VINAY, J., AND DARBELNET, J. C. Stylistique Comparée du Français et de l'Anglais. Harrap, London, 1958.

[31] ZHAO, G. Computer-assisted classification of Chinese words. In Proceedings of International Conference on Chinese Computing '94: the latest technological advancement and applications (Singapore, June 1994), pp. 317–325.

[32] ZHAO, G. *Analogical Translator: experience-guided transfer in machine translation*. PhD thesis, Language Engineering, UMIST, 1997.

[33] ZHAO, G. Similarity computation for analogical nlp: applications in chinese lexical studies and translation. In *Proceedings 1998 of International Conference on Chinese Information Processing* (Beijing, China, November 1998), H. Changning, Ed., pp. 466–474.