Filtering Machine Translation Results with Automatically Constructed Concept Lattices

Yılmaz Kılıçaslan[1] and Edip Serdar Güner[1],

[1] Trakya University, Department of Computer Engineering,
22100 Edirne, Turkey
{yilmazk, eserdarguner}@trakya.edu.tr

Abstract. Concept lattices can significantly improve machine translation systems when applied as filters to their results. We have developed a rule-based machine translator from Turkish to English in a unification-based programming paradigm and supplemented it with an automatically constructed concept lattice. The test results achieved by applying this translation system to a Turkish child story reveals that lattices used as filters to translation results have a promising potential to improve machine translation. We have compared our system with Google Translate on the data. The comparison suggests that a rule-based system can even compete with this statistical machine translation system that stands out with its wide range of users.

Keywords: Concept Lattices, Rule-based Machine Translation, Evaluation of MT systems.

1 Introduction

Paradigms of Machine translation (MT) can be classified into two major categories depending on their focus: result-oriented paradigms and process-oriented ones. Statistical MT focuses on the result of the translation, not the translation process itself. In this paradigm, translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. Rule-based MT, a more classical paradigm, focuses on the selection of representations to be used and steps to be performed during the translation process.

It is the rule-based paradigm that will be the concern of this paper. We argue for the viability of a rule-based translation model where a concept lattice functions as a filter for its results.

In what follows, we first introduce the classical models for doing rule-based MT, illustrating particular problematic cases with translation pairs between Turkish and English (cf. Section 2). Then, we briefly introduce the basic notions of Formal Concept Analysis (FCA) and touch upon the question of how lattices built using FCA can serve as a bridge between two languages (cf. Section 3). This is followed by the presentation of our translation system (cf. Section 4). Subsequently, we report on and evaluate several experiments which we have performed by feeding our translation system with a Turkish child story text (cf. Section 5). The discussion ends with some remarks and with a summary of the paper (cf. Section 6).

2 Models for Rule-Based Translation

2.1 Direct Translation

The most straightforward MT strategy is the so-called direct translation. Basically, the strategy is to translate each word into its target language counterpart while proceeding word-by-word through the source language text or speech. If the only difference between two languages were due to their lexical choices, this approach could be a very easy way of producing high quality translation. However, languages differ from each other not only lexically but also structurally.

In fact, the direct translation strategy works very well only for very simple cases like the following:

(1) Turkish: Direct Translation to English:

Köpek-ler havlar-lar. $\implies$ Dogs bark.

dog-pl   bark-3pl

In this example, the direct translation strategy provides us with a perfect translation of the Turkish sentence (interpreted as a kind-level statement about dogs). But, consider now the following example:

(2) Turkish: Direct Translation to English:

Kadın   onu   tanı-yor.

woman $\left\{ \begin{array}{c} him \\ her \\ it \end{array} \right\}$ *know-cont-3* $\implies$ Woman $\left\{ \begin{array}{c} him \\ her \\ it \end{array} \right\}$ $\left\{ \begin{array}{c} is \\ are \end{array} \right\}$ knowing.

Supposing that the referent of the pronoun is a male person, the expected translation for the given Turkish sentence would be the following:

(3) Correct Translation:

The woman knows him.

The direct translation approach fails in this example in the following respects: First, the translation results in a subject-object-verb (SOV) ordering, which does not comply with the canonical SVO ordering in English. SOV is the basic word order in Turkish. Second, the subject does not have the required definite article in the translation. The reason for this is another typological difference between the two languages: Turkish lacks a definite article. Third, the word-by-word translation leaves the English auxiliary verb ambiguous with respect to number, as the Turkish verb does not carry the number information. Fourth, the verb *know* is encoded in the progressive aspect in the translation, which is unacceptable as it denotes a mental state. This anomaly is the result of directly translating the Turkish continuous suffix *–yor* to the English suffix *–ing*. Fifth, the pronoun is left ambiguous with respect to gender in the translation, as Turkish pronouns do not bear this information.

2.2 Transfer Approach

2.2.1 Syntactic Transfer

As Jurafsky and Martin [6] point out, examples like those above suggest that the direct approach to MT is too focused on individual words and that we need to add phrasal and structural knowledge into our MT models to achieve better results. It is through the transfer approach that a rule-based strategy incorporates the structural knowledge into the MT model. In this approach, MT involves three phases: *analysis*, *transfer*, and *generation*. In the analysis phase, the source language text is parsed into a syntactic and/or semantic structure. In the transfer phase, the structure of the source language is transformed to a structure of the target language. The generation phase takes this latter structure as input and turns it to an actual text of the target language.

Let us first see how the transfer technique can make use of syntactic knowledge to improve the translation result of the example discussed above. Assuming a simple syntactic paradigm, the input sentence can be parsed into the following structure:

(4)
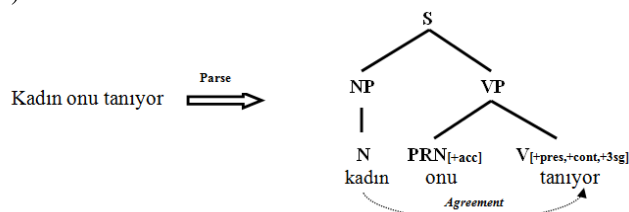
Once the sentence has been parsed, the resulting tree will undergo a *syntactic transfer* operation to resemble the target parse tree and this will be followed by a lexical transfer operation to generate the target text:

(5)
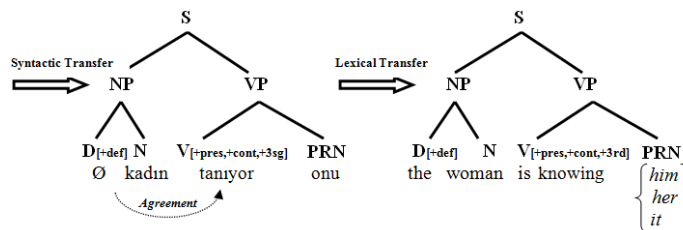
The syntactic transfer exploits the following facts about English: a singular count noun must have a determiner and the subject agrees in number and person with the verb. Collecting the leaves of the target parse tree, we get the following output:

(6)   <u>Translation via Syntactic Transfer:</u>

The woman is knowing $\left\{ \begin{matrix} him \\ her \\ it \end{matrix} \right\}$.

This output is free from the first three defects noted with the direct translation. However, the problem of encoding the mental state verb in progressive aspect and the

gender ambiguity of the pronoun still await to be resolved. These require *meaning-related knowledge* to be incorporated into the MT model.

2.2.2 Semantic Transfer

The *context-independent* aspect of meaning is called *semantic meaning*. A crucial component of the semantic meaning of a natural language sentence is its *lexical aspect*, which determines whether the situation that the sentence describes is a (*punctual*) *event*, a *process* or a *state*. This information is argued to be inherently encoded in the verb. Obviously, knowing is a mental state and, hence, cannot be realized in the progressive aspect.

We can apply a shallow semantic analysis to our previously obtained syntactic structure, which will give us a tree structure enriched with aspectual information, and thereby achieve a more satisfactory transfer:

(7)
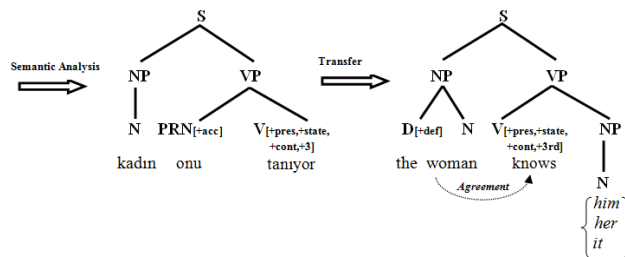
The resulting translation is the following:

(8)   Translation via Semantic Transfer:

2.3 Interlingua Approach

There are two problems with the transfer model: it requires contrastive knowledge about languages and it requires such knowledge for every pair of languages. If the meaning of the input can be extracted and encoded in a language-independent form and the output can, in turn, be generated out of this form, there will be no need for any kind of contrastive knowledge. A language-independent meaning representation language to be used in such a scheme is usually referred to as an *interlingua*.

A common way to visualize the three approaches to rule-based MT is with *Vauquois triangle* shown below (adopted from [6]):
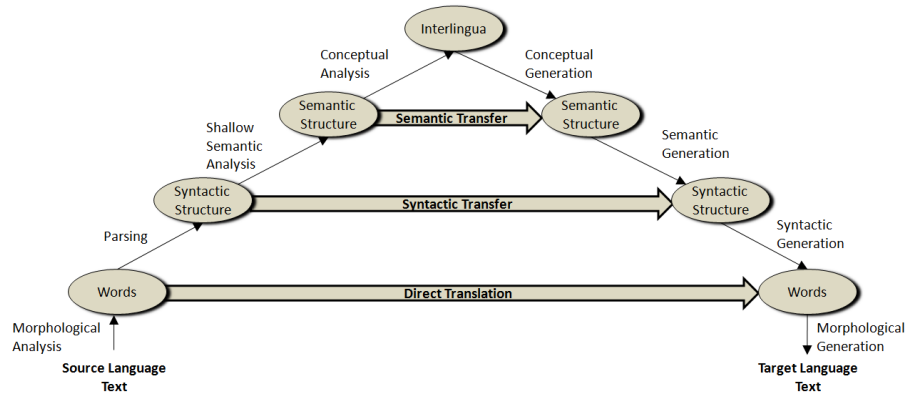
Fig. 1. The Vauquois triangle.

As Jurafsky and Martin point out:

> [t]he triangle shows the increasing depth of analysis required (on both the analysis and generation end) as we move from the direct approach through transfer approaches, to interlingual approaches. In addition, it shows the decreasing amount of transfer knowledge needed as we move up the triangle, from huge amounts of transfer at the direct level (almost all knowledge is transfer knowledge for each word) through transfer (transfer rules only for parse trees or thematic roles) through interlingua (no specific transfer knowledge). (p. 867)

3 Lattice-Based Interlingua Strategy

A question left open above is that of what kind of representation scheme can be used as an interlingua. There are many possible alternatives such as predicate calculus, Minimal Recursion Semantics or an event-based representation. Another interesting possibility is to use lattices built using Formal Concept Analysis (FCA) as meaning representations to this effect.

FCA, developed by Ganter & Wille [5], assumes that data from an application are given by a formal context, a triple (G, M, I) consisting of two sets G and M and a so called incidence relation I between these sets. The elements of G are called the objects and the elements of M are called the attributes. The relation I holds between g and m, $(g, m) \in I$ if and only if the object g has the attribute m. A formal context induces two operators, both of which usually denoted by $'$. One of these operators maps each set of objects A to the set of attributes A' which these objects have in common. The other operator maps each set of attributes B to the set of objects B' which satisfies these attributes. FCA is in fact an attempt to give a formal definition of the notion of a 'concept'. A formal concept of the context (G, M, I) is a pair (A, B) such that $G \supseteq A = A'$ and $M \supseteq B = B'$. A is called the *extent* and B the *intent* of the concept (A, B). The set of all concepts of the context (G, M, I) is denoted by $C(G, M, I)$. This set is ordered by a subconcept – superconcept relation, which is a partial order relation denoted by $\leq$. If (A_1, B_1) and (A_2, B_2) are concepts in $C(G, M, I)$, the former is said to

be a subconcept of the latter (or, the latter a superconcept of the former), i.e., $(A_1, B_1) \le (A_2, B_2)$, if and only if $A_1 \subseteq A_2$ (which is equivalent to $B_1 \supseteq B_2$). The ordered set $C(G, M, I; \le)$ is called the concept lattice or (Galois lattice) of the context (G, M, I). A concept lattice can be drawn as a (Hasse) diagram in which concepts are represented by nodes interconnected by lines going down from superconcept nodes to subconcept ones.

Priss [15], rewording an idea first mentioned by Kipke & Wille [8], suggests that once linguistic databases are formalized as concept lattices, the lattices can serve as an interlingua. She explains how a concept lattice can serve as a bridge between two languages with the aid of the figure below (taken from [13]):
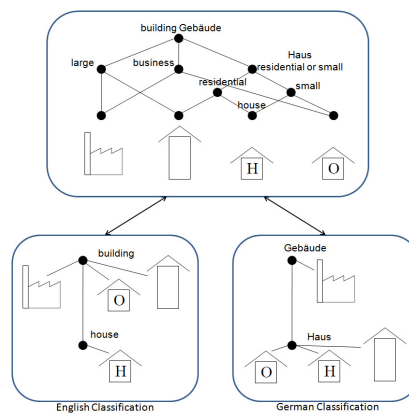

Fig. 2. – A concept lattice as an interlingua.

[This figure] shows separate concept lattices for English and German words for "building". The main difference between English and German is that in English "house" only applies to small residential buildings (denoted by letter "H"), whereas in German even small office buildings (denoted by letter "O") and larger residential buildings can be called "Haus". Only factories would not normally be called "Haus" in German. The lattice in the top of the figure constitutes an information channel in the sense of Barwise & Seligman [2] between the German and the English concept lattice. ([15] p. 158)

We consider Priss's approach a promising avenue for interlingua-based translation strategies. We suggest that this approach can work not only for isolated words but also even for text fragments. In what follows, we will sketch out a strategy with interlingual concept lattices serving as filters for refining translation results. The strategy proceeds as follows: 1) Compile a concept lattice from a data source like WordNet. 2) Link the nodes of the lattice to their possibly corresponding expressions in the source and target language. 3) Translate the input text into the target language with no consideration of the pragmatic aspects of its meaning. 4) Integrate the concepts derived from the input text into the concept lattice. The main motivation behind this strategy is to refine the translation results to a certain extent by means of pragmatic knowledge structured as formal contexts.

4 A Translation System with Interlingual Concept Lattices

4.1 A Concept Lattice Generator

Concept lattices to be used as machine translation filters should contain concept nodes associated with both functional and substantive words. All languages have a finite number of functional words. Therefore, a manual construction of the lattice fragments that would contain them would be reasonable. However, manually constructing a concept lattice for lexical words would have considerable drawbacks such as the following:

- It is labor intensive.
- It is prone to yielding errors which are difficult to detect automatically.
- It generates incomplete lists that are costly to extend to cover missing information.
- It is not easy to adapt to changes and domain-specific needs.

Taking these potential problems into consideration, we have developed a tool for generating concept lattices for lexical words automatically. As this is an FCA application, it is crucial to decide on which formal context to use before delving its implementation details.

   Priss & Old [16] propose to construct concept neighborhoods in WordNet with a formal context where the formal objects are the words of the synsets belonging to all senses of a word, the formal attributes are the words of the hypernymic synsets and the incidence relation is the semantic relation between the synsets and their hypernymic synsets. The neighborhood lattice of a word in WordNet consists of all words that share some senses with that word.[1] Below is the neighborhood lattice their method yields for the word *volume*:
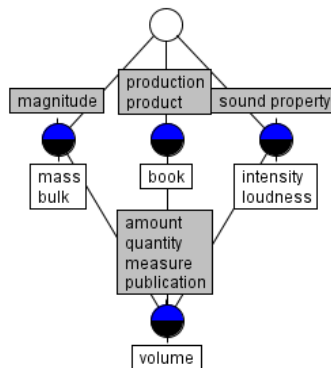

Fig. 3. – Priss and Old's neighborhood lattice for the word *volume*.

---

[1] As lattices often grow very rapidly to a size too large to be visualized, Wille [18] describes a method for constructing smaller, so-called "neighborhood" lattices.

Consider the bottom node. The concept represented by this node is not a naturally occurring one. Obviously, the adopted formal context causes two distinct natural concepts to collapse into one single formal concept here. The reason is simply that WordNet employs one single word, i.e., *volume*, for two distinct senses, i.e., *publication* and *amount*. This could leave a translation attempt with the task of disambiguating this word. In fact, WordNet marks each sense with a single so-called synset number.

When constructing concept lattices in WordNet, we suggest two amendments to the formal context adopted by Priss and Old. First, the formal objects are to be the synset numbers. Second, the formal attributes are to include also some information compiled from the glosses of the words. The first change allows us to distinguish between the two senses of the word *volume*, as shown in Fig. 4a. But, we are still far from resolving all ambiguities concerning this word, as indicated by the presence of two objects in the leftmost node. The problem is that the hypernymic attributes are not sufficiently informative to differentiate the *3-D space* sense of the word *volume* from its *relative amount* sense. This extra information resides in the glosses of the word and once encoded as attributes it evokes the required effect, as shown in Fig. 4b.
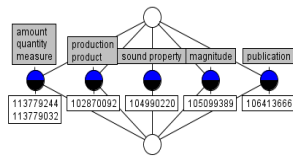

Fig. 4a. – A neighborhood lattice with the objects being synset numbers.
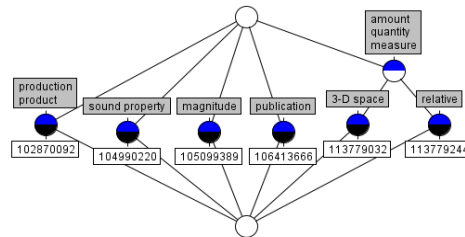
Fig. 4b. – A more fine-grained neighborhood lattice with the objects being synset numbers.

Each gloss, which is most likely a noun phrase, is parsed by means of a shift-reduce parser to extract a set of attributes. Having collected the objects (i.e. the synset numbers) and the associated attributes, the FCA algorithm that comes with the FCALGS library [9] is used for deriving a lattice-based ontology from that collection. FCALGS employs a parallel and recursive algorithm. Apart from its being parallel, it is very similar to Kuznetsov's [10] Close-by-One algorithm.

However, even the lattice in Fig4.b is still defective in at least one respect. The names of the objects denoted are lost. To remedy this problem, we suggest to encode the objects as tuples of synset numbers and sets of names, as illustrated below.
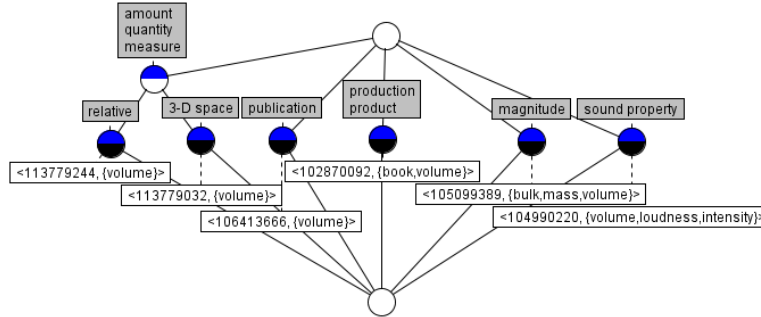
Fig. 5. – A neighborhood lattice including the names of the objects.

Another point to note is that the name of a synset serves as the attribute of a subconcept. For example, 'entity' is the name of the topmost synset. But, as everything is an entity, any subconcept must treat it as an element of its set of attributes.

4.2 A Sense Translator

Each WordNet node is associated with a set of synonymous English words, which is referred to as its synset. Each synset, in effect, denotes a sense in English. Thus, one task to accomplish is to translate synsets into Turkish to the furthest possible extent. We should, of course, keep in mind that some synsets (i.e. some senses encoded in English) may not have a counterpart in the target language. To find the Turkish translation of a particular synset, the Sense Translator first downloads a set of relevant articles via the links given in the disambiguation pages Wikipedia provides for the words in this set. It searches for the hypernyms of the synset in these articles. It assigns each article a score in accordance with the sum of the weighted points of the hypernyms found in this article. More specifically, if a synset has N hypernyms, the Kth hypernym starting from the top is assigned $Weight_K = K/N$. Let $Frequency_K$ be the number of occurrences of an item in a given article, then the score of the article is calculated as follows:

$$\textbf{Article Score} = \textbf{Weight}_1 * \textbf{Frequency}_1 + ... + \textbf{Weight}_N * \textbf{Frequency}_N. \qquad (1)$$

If the article with the highest score has a link to a Turkish article, the title of the article will be the translation of the English word under examination. Otherwise, the word will be left unpaired with a Turkish counterpart. Figure 6 visualizes how the word *cat* in WordNet is translated into its Turkish counterpart, kedi, via Wikipedia.
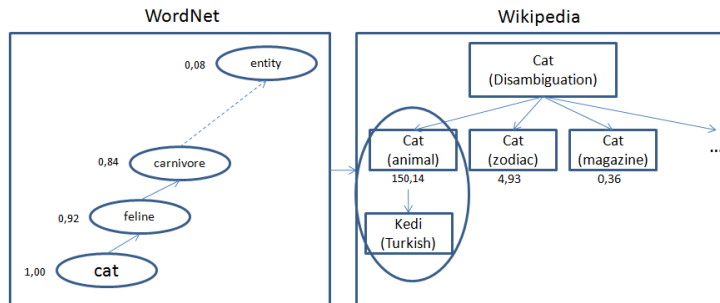
Fig. 6. - Translating the word *cat* into Turkish via Wikipedia.

The Turkish counterparts will be added next to the English names, as shown below:

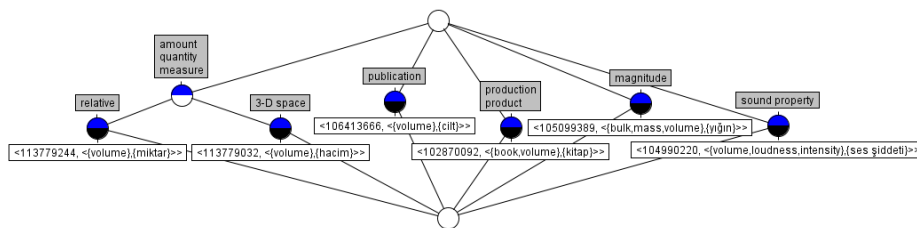**Fig. 7.** - A neighborhood lattice including the Turkish counterparts of the English names.

4.3 A Rule-Based Machine Translator

We have designed a transfer-based architecture for Turkish-English translation and implemented the translator in SWI-Prolog which is an open-source implementation of the Prolog programming language. Below is a figure representing the main modules of the translator:
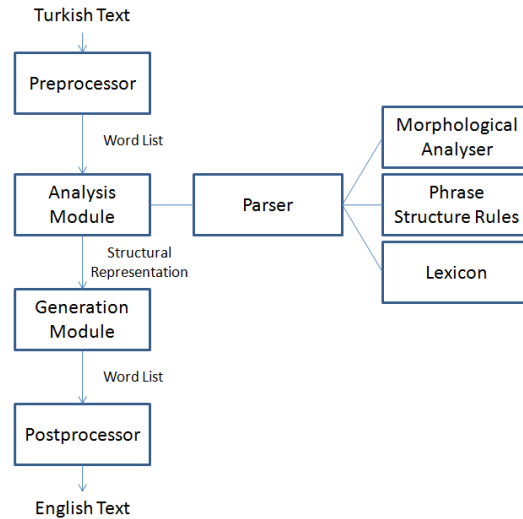
Fig. 8. - The main modules of the rule-based machine translator.

The word list extracted by the Preprocessor is used as an input to the Analysis Module. We have devised a shift-reduce parser in the analysis phase for building up the grammatical structure of expressions. Briefly, a shift-reduce parser uses a bottom-up strategy with an ultimate goal of building trees rooted with a start symbol [1]. The Generation Module first rearranges the constituents using transformation rules. Afterwards, all the structures are lexically transferred into English using a bilingual dictionary.

4.4 Filtering Translation Results with the Concept Lattice

Let us turn to our exemplary sentence introduced in (2) (i.e. *Kadın onu tanıyor*). Failing to take the context of the sentence into account, the rule-based translator generates the result in (8) (i.e. *The woman knows him/her/it*), where the pronoun is left ambiguous with respect to gender.

Our claim is that we can resolve such ambiguities using FCA and thereby refine our translations. To this effect, we propose to generate transient formal concepts for noun phrases. We make the following assumptions. Basically, personal pronouns, determiners and proper names introduce formal objects whereas adjectives and nouns encode formal attributes.

Suppose that our sentence is preceded by (the Turkish paraphrase of) a sentence like '*A man has arrived*'. The indefinite determiner evokes a new formal object, say *obj1*. As the source text is in Turkish, all attributes will be Turkish words. The Turkish counterpart of the word *man* is *adam*. Thus, the transient concept for the subject of this sentence will be ({*obj1*}, {adam}). The task is now to embed this transient concept into the big permanent concept lattice. To do this, a node where the Turkish counterpart of the synset name is 'adam' is searched for. Immediately below this node is placed a new node with its set of objects being {*obj1*} and with no additional attributes. As this is a physical object, the subconcept of this new node has to be the

lowest one. As for the second sentence, the NP *kadın* (the woman) will be associated with the transient concept ({X},{kadın}) and the pronoun *onu* (him/her/it) with the transient concept ({Y},{entity}). X and Y are parameters to be anchored to particular formal objects. In other words, they are anaphoric. It seems plausible to assert that the attributes of an anaphoric object must constitute a (generally proper) subset or hypernym set of the attributes of the object serving as the antecedent. Assume that X is somehow anaphorically linked to an object *obj2*. Now, there are two candidate antecedents for Y. The woman, or the object *obj2*, is barred from being antecedent of the pronoun by a locality principle like one stated in Chomsky's [3] Binding Theory: roughly stated, a pronoun and its antecedent cannot occur in the same clause. There remains one single candidate antecedent, *obj1*. As its attribute set is a hyponym set of {entity}, it can be selected as a legitimate antecedent. The concept node created for the man will also be the one denoted by the pronoun with Y being instantiated with *obj1*. In the concept lattice constructed in WordNet, the concept named as 'man' includes 'male person' in its set of attributes. Hence, the ambiguity is resolved and the pronoun translates into English as 'him'.

It is worth noting that in case there is more than one candidate antecedent, an anaphora resolution technique, especially a statistical one, can be employed to pick out the candidate most likely to be the antecedent. The interested reader is referred to Mitkov [12] for a survey of anaphora resolution approaches in general and to Kılıçaslan et al [7] for anaphora resolution in Turkish.

The gender disambiguation process can also be carried out for common nouns. Consider the following fragment taken from a child story:

(9)

Kız kardeş-ler-i   Sinderella-ya   gül-müş-ler.

girl $\left\{\begin{array}{l}\text{brother}\\\text{sister}\end{array}\right\}$ -pl-acc Cinderella-dat   laugh-pst-3pl

'His/her sisters laughed at Cinderella.'

Kardeş-ler-i   ayakkabı-nın ona

$\left\{\begin{array}{l}\text{brother}\\\text{sister}\end{array}\right\}$ -pl-poss3 shoe-gen $\left\{\begin{array}{l}\text{him}\\\text{her}\\\text{it}\end{array}\right\}$

uy-ma-yacağ-ı-na   inan-ıyor-lar-mış.
fit-neg-fut-poss3-dat   believe-cont-agr3pl-pst

'His/her sisters/brothers believed that the shoe would   not fit him/her.'

Turkish, leaves not only pronouns but also many other words ambiguous with respect to the gender feature. The word 'kardeş' in this example is ambiguous between the translations *sister* and *brother*. This ambiguity will be resolved in favor of the former interpretation in way similar to the disambiguation process sketched out for pronouns above.

In fact, the problem of sense disambiguation is a kind of specification problem. Therefore, it cannot be confined to gender disambiguation.   For example, given that we have somehow managed to compile the attributes listed in the column on the left-hand side, our FCA-based system generates the translations listed on the right-hand side:

zehirli, diş 'poisonous, tooth'	*fang*
zehirli, mantar 'poisonous, mushroom'	*toadstool*
sivri, diş 'sharp, tooth'	*fang*
arka, koltuk 'rear, seat'	*rumble*
acemi, asker 'inexperienced, soldier'	*recruit*

It will, of course, be interesting to try to solve other kinds of translation problems with FCA-based techniques. We leave this task to accomplish in the light of further research in the future.

5   Results and Evaluation

In the early years of MT, the quality of an MT system was determined by human judgment. Though specially trained for the purpose, human judges are prone to suffer at least from subjectivity. Besides, this exercise is almost always more costly and time consuming. Some automated evaluation metrics have been developed in order to overcome such problems. Among these are BLEU, NIST, WER and PER.

BLEU [14] and NIST [4] are rather specialized metrics. They are employed by considering the fraction of output n-grams that also appear in a set of human translations (n-gram precision). This allows the acknowledgment of a greater diversity of acceptable MT results.

As for WER (Word Error Rate) and PER (Position-independent Word Error Rate), they are more general purpose measures and they rely on direct correspondence between the machine translation and a single human-produced reference. WER is based on the Levenshtein distance [11] which is the edit distance between a reference translation and its automatic translation, normalized by the length of the reference translation. This metric is formulated as:

$$WER \;\; = \;\; \frac{S+D+I}{N} \qquad (2)$$

where N is the total number of words in the reference translation, S is the number of substituted words in the automatic translation, D is the number of words deleted from the automatic translation and I is the number of words inserted in the reference not appearing in the automatic translation.

Although WER requires exactly the same order of the words in automatic translation and reference, PER neglects word order completely [17]. It measures the difference in the count of the words occurring in automatic and reference translations. The resulting number is divided by the number of words in the reference. It is worth noting that PER is technically not a distance measure as it uses a position-independent Levenshtein distance where the distance between a sentence and one of its permutations is always taken to be zero.

We used WER to evaluate the performance of our MT system. This is probably the metric most commonly used for similar purposes. As we employed a single human-produced reference, this metric suits well to our evaluation setup. We fed our system

with a Turkish child story involving 91 sentences (970 words).[2] We post-edited the resulting translation in order to generate a reference. When necessary calculations were done in accordance with formula (1), the WER turned out to be 38%.

The next step was to see the extent to which the performance or our MT system could be improved using concept lattices as filters for the raw results. To this effect, we devised several concept lattices like that in figure 3 and filtered the lexical constituents of each automatic translation with them.

A considerable regression in error rate is observed in our system supplemented with concept lattices: the WER score is reduced down to a value around 30%.

One question that comes to mind at this point is that of whether the improvement achieved is statistically significant or not. To get an answer we had recourse to the Wilcoxon Signed-Rank test. This test is used to analyze matched-pair numeric data, looking at the difference between the two values in each matched pair. When applied to the WER scores of the non-filtered and filtered translation results, the test shows that the difference is statistically significant ($p < 0.005$).

Another question is that of whether the results are practically satisfactory. To get some insight to this question, we should employ a baseline system for a comparison on usability. Google Translate, a statistical MT system that stands out with its wide range of users, can serve for this purpose. The WER score obtained employing Google Translate on our data is 34%. Recalling that the WER score of our system supplemented with concept lattices is 30%, we seem to be entitled to argue for the viability of rule-based MT systems. Of course, we need to make this claim tentatively since the size of the data on which the comparisons are made is relatively small. However, it should also be noted that we have employed a limited number of concept lattices of considerably small sizes. It is of no doubt that increasing the number and size of filtering lattices would improve the performance of our MT system.

More importantly, we do not primarily have an NLP concern in this work. Rather, we would like the results to be evaluated from a computational linguistics perspective. Everything aside, the results show that even a toy lattice based ontology can yield statistically significant improvement for an MT system.

6 Conclusion

In this paper, we have illustrated some translation problems caused by some typological divergences between Turkish and English using a particular example. We have gone through the direct translation, syntactic transfer and semantic transfer phases of the rule-based translation model to see what problem is dealt with in what phase. We have seen that a context-dependent pragmatic process is necessary to get to a satisfactory result. Concept lattices appear to be very efficient tools for accomplishing this pragmatic disambiguation task. Supplementing a rule-based MT system with concept lattices not only yields statistically significant improvement on the results of the system but also enables it to compete with a statistical MT system like Google Translate.

---

[2] This is the story where the example in (9) comes from.

References

1. Aho, A.V., Ullman, J.D.: *The Theory of Parsing, Translation, and Compiling*, Vol. 1., Prentice Hall (1972)
2. Barwise J., Seligman, J.: *Information Flow. The Logic of Distributed Systems*. Cambridge University Press (1997)
3. Chomsky, N.: *Lectures on Government and Binding*, Foris, Dordrecht (1981).
4. Doddington, G.: "Automatic Evaluation of Machine Translation Quality Using N-gram Co-occurrence Statistics". In *Proceedings of HLT 2002 (2nd Conference on Human Language Technology)*. San Diego, California, 128-132 (2002)
5. Ganter, B., Wille, R.: *Formale Begriffsanalyse: Mathematische Grundlagen*. Berlin: Springer (1996)
6. Jurafsky, D., Martin, J. H.: *Speech and Language Processing*, 2nd Edition, Prentice Hall (2009)
7. Kılıçaslan, Y., Güner, E. S., Yıldırım, S.: Learning-based pronoun resolution for Turkish with a comparative evaluation, Computer Speech & Language, Volume 23, Issue 3, p. 311-331 (2009)
8. Kipke, U., Wille, R.: *Formale Begriffsanalyse erläutert an einem Wortfeld*. LDV–Forum, 5 (1987)
9. Krajca, P., Outrata, J., Vychodil, V.: *Parallel Recursive Algorithm for FCA*. In: Belohlavek R., Kuznetsov S. O. (Eds.): Proc. CLA 2008, CEUR WS, 433, 71–82 (2008)
10. Kuznetsov, S.: *Learning of Simple Conceptual Graphs from Positive and Negative Examples*. PKDD 1999, pp. 384–391 (1999)
11. Levenshtein, V. I.: "Binary codes capable of correcting deletions, insertions, and reversals," Tech. Rep. 8. (1966)
12. Mitkov, R.: Anaphora Resolution: The State of the Art. Technical Report, University of Wolverhampton (1999)
13. Old, L. J., Priss, U.: *Metaphor and Information Flow*. In Proceedings of the 12th Midwest Artificial Intelligence and Cognitive Science Conference, pp. 99-104 (2001)
14. Papineni, K., Roukos, S., Ward, T., Zhu, W. J.: "BLEU: a method for automatic evaluation of machine translation" in *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics* pp. 311–318 (2002)
15. Priss, U.: Linguistic Applications of Formal Concept Analysis, Ganter; Stumme; Wille (eds.), *Formal Concept Analysis, Foundations and Applications*, Springer Verlag, LNAI 3626, pp. 149-160 (2005)
16. Priss, U., Old, L. J.: "Concept Neighbourhoods in Lexical Databases.", In Proceedings of the 8th International Conference on Formal Concept Analysis, ICFCA'10, Springer Verlag, LNCS 5986, p. 283-295 (2010)
17. Tillmann C., Vogel, S., Ney, H., Zubiaga A., Sawaf, H.: Accelerated DP based search for statistical translation. In *European Conf. on Speech Communication and Technology*, pages 2667–2670, Rhodes, Greece, September (1997)
18. Wille, R.: *The Formalization of Roget's International Thesaurus*. Unpublished manuscript (1993)