Using Computational Semantics for Chinese Translations

Evelyne Viegas, Wanying Jin, Stephen Beale
New Mexico State University
Computing Research Laboratory
Las Cruces, NM 88003
USA
viegas,wanying,sb@crl.nmsu.edu

Abstract

In this paper, we show how a computational semantic approach is best fitted to address the translation of highly isolating languages. We use Chinese as an example and present the overall process of translation from Chinese to English. within the framework of Knowledge-Based Machine Translation (KBMT), using computational semantics while de-emphasizing syntax. We focus here on Word Sense Disambiguation (WSD) after addressing acquisition issues for Chinese and English lexicons.

1 Introduction

In this paper, we present results of a theoretical and an applied investigation on the translation of Chinese texts into English. Because Chinese is among the most isolating languages, it presents a challenging situation for Machine Translation (MT): grammatical functions can be irrelevant; morpho-syntactic markers are sparse, and are very often elided (Li & Thompson, 1981). We argue that for isolating languages syntactic analysis should, at the very least, be de-emphasized to let semantics overtly take over in the process. By de-emphasizing syntactic analysis, we mean that there is no need to produce the N-best syntactic parses. In this paper, we show how computational semantics along with lexico-syntactic dependencies coded in the lexicon, can account for syntactically ambiguous parses. In section 2, we briefly present the type of information encoded in the lexicon. In section 3, we address acquisition issues in building the Chinese and English lexicons. In section 4. we present results on the task of Word Sense Disambiguation (WSD).

The present work is part of Mikrokosmos,[1] a knowledge-based machine translation system under development at the Computing Research Laboratory, New Mexico State University jointly with the Department of Defense. The system uses an interlingual approach Nirenburg *et al.* (1992) to represent the deep meaning of the text using concepts from the ontology (Nirenburg, 1995; Mahesh, 1996). Source languages are related to target languages through text meaning representations (TMR) (Nirenburg and Defrise, 1991). As a component of the MikroKosmos system, the Chinese lexicon provides syntactic and semantic knowledge. Ontological concepts are used to represent the lexicon meaning. Various methodologies (Onyshkevych and Nirenburg, 1994; Viegas and Raskin, 1998; Viegas, 1998) have been investigated over the years carrying out the structure of a computational lexicon entry within a knowledge-based framework. In addition, a set of machine translation aided tools has also been developed to support the knowledge acquisition through the access of various on-line resources (Ogden *et al.*, 1995). A web lexicon acquisition interface is specially designed for constructing Chinese and English semantic lexicons (Viegas *et al.*, 1999) (see Figure 1 and Figure 2 below).

2 An Overview on Our Computational Lexicons

The lexicons used by our systems use information which is distributed among various levels of lexical information, relevant to phonology, orthography, morphology, syntax (SYN), semantics (SEM)[2], syntax-semantic linking (SYNSEM), stylistics, paradigmatic and

[1] See MikroKosmos web site at http://crl.nmsu.edu/Research/Projects/mikro/.

[2] The semantics is expressed as a frame as in, for instance, Fillmore (1985).

syntagmatic information, and also database type management information.

The Mikrokosmos ontology includes 1) **concepts** about the world or domain; 2) **property** of the concept; and, 3) **relationships** among them. In an interlingual MT system, the ontology provides the building blocks to represent the text meaning that enables lexicons in different languages to share the world knowledge. It also allows the analyzer and generator to share the knowledge regardless of the specific language distinctions. In MikroKosmos the ontology also stores selectional-like restrictions and other pieces of world knowledge; fills the gaps in the text meaning by making inferences based on the content of conceptual knowledge of the ontology. During processing, it resolves semantic ambiguities by using the topology of the ontology to measure the semantic affinity between meanings (Onyshkevych, 1997). In a multilingual environment, the main practical advantage of connecting the lexicon to an ontology is cost-effectiveness, as only the "language-dependent" properties have to be acquired when adding new natural languages to the system. Currently, in the MikroKosmos system, the ontology contains about 5,000 concepts covering a wide range of categories in the world. The ontology is organized in a tangled hierarchy with ample interconnections across the branches. Each concept, on average, has 14 relation links such as IS-A; SUBCLASS; AGENT; THEME-OF; HEADED-BY; HAS-MEMBER. The top level of the ontology differentiates between OBJECT with subclasses PHYSICAL-OBJECT, MENTAL-OBJECT and SOCIAL-OBJECT; EVENT with subclasses PHYSICAL-EVENT, MENTAL-EVENT and SOCIAL-EVENT; and PROPERTY with subclasses ATTRIBUTE and RELATION. For a detailed description, see (Mahesh, 1996).

We illustrate below relevant aspects, for this paper, of a lexicon entry via the description of one sense of the Chinese word 削减 (reduce) in Table 1, which is an EVENT, and the noun 配额 (quota) in Table 2, which is mapped to an ATTRIBUTE which DOMAIN is OBJECT.

削减-V1

syn:				
	root:	0		
	subj:	1	cat:	NP
	obj:	2	cat:	NP
sem:	DECREASE			
	agent:	11	Human	
	theme:	21	Object	
synsem:				
	subj:	1	sem:	11
	obj:	2	sem:	21

Table 1: Sense Entry for the Chinese lexical item 削减.

配额-N1

syn:				
	root:	0		
	mods:	1	cat:	NP
sem:	QUOTA			
	domain:	11	Object	
	range :	21	Any Number	
synsem:				
	domain:	1	sem:	11
	range :	2	sem:	21

Table 2: Sense Entry for the Chinese lexical item 配额.

The SYN Zone essentially amounts to an underspecified piece of a syntactic parse of a sentence using the lexeme.

The SEM Zone maps to an ontological concept in the Mikrokosmos ontology.[3]

We give below two samples of concepts for the EVENT DECREASE and the ATTRIBUTE QUOTA.

```
Concept Name:          DECREASE
DEFINITION:            "a change in quantity where
                       the final value is lesser
                       than the initial value"
IS-A:                  CHANGE-IN-QUANTITY
SUBCLASSES:            SUBTRACT-FROM
DIRECTION-OF-CHANGE:   NEGATIVE
AGENT:                 HUMAN
THEME:                 OBJECT
INITIAL-VALUE:         ALL
FINAL-VALUE:           ALL
TYPE-OF-CHANGE:        MENTAL PHYSICAL SOCIAL
```

[3]See http://crl.nmsu.edu/Research/Projects/mikro/index.html.

```
Concept Name:  QUOTA
DEFINITION:    "a specified amount of
                something"
IS-A:          AMOUNT
DOMAIN:        OBJECT
RANGE:         ANY-NUMBER
```

The SYNSEM Zone provides the syntax-semantic linking. For instance, in Table 1., we know that the SUBJ:1 is linked to the semantics AGENT:11 of type HUMAN. It is better, for acquisition purposes, to leave this information transparent to the user.

The information in the SEM zone is later used to generate a TMR which will be used to generate the English text.

3 Lexicon acquisition

On-line news articles 新华日报 (XinHua Daily News) are used as a resource for building the Chinese lexicon. The NMSU segmenter (Jin, 1994; 1995) identifies individual words in the texts. An on-line English-Chinese dictionary provides the phoneticism (拼音) and English translations. The lexicon acquisition procedure is semi-automatic. It includes two phases: pre-acquisition and acquisition. In the pre-acquisition phase, all the senses of each Chinese word are extracted from a dictionary automatically; next, mapping word senses into ontological concepts is done manually by referring to the CRL ontology where we aim at parsimony. In the acquisition phase, the WWW lexicon interface is used to derive syntactic patterns for each lexical entry and to make links between syntactic arguments and semantic elements.

The lexicon acquisition process requires a lot of effort. Besides the acquisition itself, it involves the intensive training for the acquirers and the development of tools to construct the lexicon structures and to access various on-line resources. The Chinese lexicon acquisition is done semi-automatically. To reach the highest cost-effect, the procedure is divided into two phases: Pre-acquisition and Acquisition as proposed in (Viegas and Raskin 1998).

The types of mapping **Mapping-tags** is defined as follows:

- **direct mapping** uses a concept without constraint (e.g., *table*);
- **constrained mapping** uses a concept with constraint (e.g., *eat, drink*);
- **attitude** expresses a speaker attitude with respect to the lexeme (e.g., *good*);
- **modality** expresses a speaker modality with respect to the lexeme (e.g., *must*);
- **deverbal** represents nouns and adjectives which meaning is a composition of a filler and an event (e.g.*bombing, readable*);
- **aspectual** for truly aspectuals (e.g.*begin*) and also with actions expressing aspectuality (e.g. *stare*, duration: prolonged)
- **unknown** is used when none of the above applies to the lexeme under study.

3.1 Pre-acquisition

The goal of the Chinese pre-acquisition has been to collect information about Semantics: Mapping-Tag; Lexeme; PinYin; POS; Frequency and Translations. To do so, for each entry, one must find the phonetics (PinYin) and translations from an Chinese-English dictionary; assign a POS; capture all meanings and find the ontological head concept with/without constraints. The procedure involves:

- Segmenting the Chinese texts (Jin 1995);
- Providing the word frequency count;
- Searching the Chinese-English dictionary for the phonetics and translations;
- Searching on-line corpora for finding all the frequent usages of a word;
- Searching a concept in the ontology to which the word can be mapped;
- Creating the pre-acquisition files;

The definition of a concept along with all its ontological slots and the hierarchy must be checked. If the concept chosen fits the meaning of the word, a direct mapping is found. If some slots need further restrictions, a constrained mapping is found. All constraints also must be ontological concepts. The pre-acquisition is complete when all information is acquired and an output file containing Semantics; Mapping-Tag; Lexeme; PinYin; POS; Frequency and Translations is made. There is still time at acquisition time to revise the pre-acquisition files if needed.

3.2 Acquisition

The task of acquisition consists in:

- verifying the semantics acquired in the pre-acquisition phase;
- providing a syntactic structure;

- creating the link between syntax and semantics;
- generating and formatting the lexicon file.

Since the pre-acquisition is done semi-automatically, the manual effort sometimes introduces errors resulting in the selection of inappropriate concepts, inappropriate case roles or incorrect constraints. The acquisition phase automatically verifies all the case roles (e.g. AGENT) and its values (e.g. HUMAN) for the selected head concepts (e.g DECLARE). Then, all possible syntactic patterns for each entry of the source languages are provided in order to establish the syntactic parse for a sentence. Once the syntactic information is provided, it is important to link all the subcategorization variables such as SUBJ, OBJ with the semantic roles such as AGENT, THEME, etc.

The procedure of acquisition includes the following steps:

- check the preliminary information provided by the pre-acquisition phase, i.e. concept; Chinese characters; PinYin; POS and Mapping-Tag;
- verify all the values of all case roles chosen in the pre-acquisition phase. Each case role and its value is verified automatically referring to the ontology;
- acquire the syntax. The syntax must correspond to the case roles chosen in the semantics, i.e. if two case roles are selected, no more than two syntactic arguments can be chosen. For instance the second Syntactic argument can be OBJ, COMP or PP-ADJUNCT.

The results from the acquisition process is a file containing accurate information about the Concept(with/without constraint). Once the acquisition is complete, all entries are reorganized in a hierarchy. All entries with the same lexeme are organized under one super entry.

To summarize, our lexicons include selectional restriction type information along with semantico-syntactic dependencies per word.

4 Computational Semantics for Word Sense Disambiguation

The Mikrokosmos analyzer has produced high quality semantic analyses of article-length Spanish texts in the domain of company mergers and acquisitions. The average text was 17 sentences long; average sentence length was over 21. For evaluation purposes, correct senses for all the open class words were tagged by a native speaker. Mikrokosmos selects the right sense of open-class words about 97% of the time. Syntactic analysis contributed to about 38% of word sense disambiguation (that is, these were the cases when different senses of a lexeme had different subcategorizations) (see Mahesh *et al.*, 1997).

In the following, we address the task of WSD on Chinese sentences, using a semantic analyzer which can perform WSD at a very high percentage of correctness, with hardly any need for syntactic information. We first go through a simple example to exemplify the process of WSD.

Consider the Chinese sentence:

美 单方 削减 中国 纺织品 出口 配额

美 (propn, mei3, US)
单方 (adv, dan1fang1, unilaterally)
削减 (v, xue1jian3, reduce)
中国 (propn, zhong1guo3, China)
纺织品 (n, fang3zhi1pin3, textile product)
出口 (n, chu1kou3, export)
配额 (n, pei4e2, quota)

The US unilaterally reduced the quota of Chinese textile exports.

This example is simple from a WSD viewpoint because we only have one word 出口 which is ambiguous in the lexicon with two senses (OPENING ISA OBJECT and EXPORT ISA EVENT). The word 出口 was correctly disambiguated (see below for the WSD process). The f-structure below (Bresnan 1982) shows the syntactic parse for the sentence above. Note that the compound is left ambiguous: all modifiers MODS are at the same level. It would be expensive and vain to produce the exhaustive list of combinations inside the compound, whereas only some of them would be semantically valid. This is why we let the semantic processor decide on the relations expressed between the words.

```
(((SUBJ ((ROOT 美) (CAT N) (GLOSS "US")))
  (MODS ((ROOT 单方) (CAT ADV)
         (GLOSS "unilaterally")))
  (ROOT 削减) (CAT V) (GLOSS "reduce")
  (OBJ  ((MODS ((ROOT 中国) (CAT N)
                (GLOSS "China")))
         (MODS ((ROOT 纺织品) (CAT N)
```

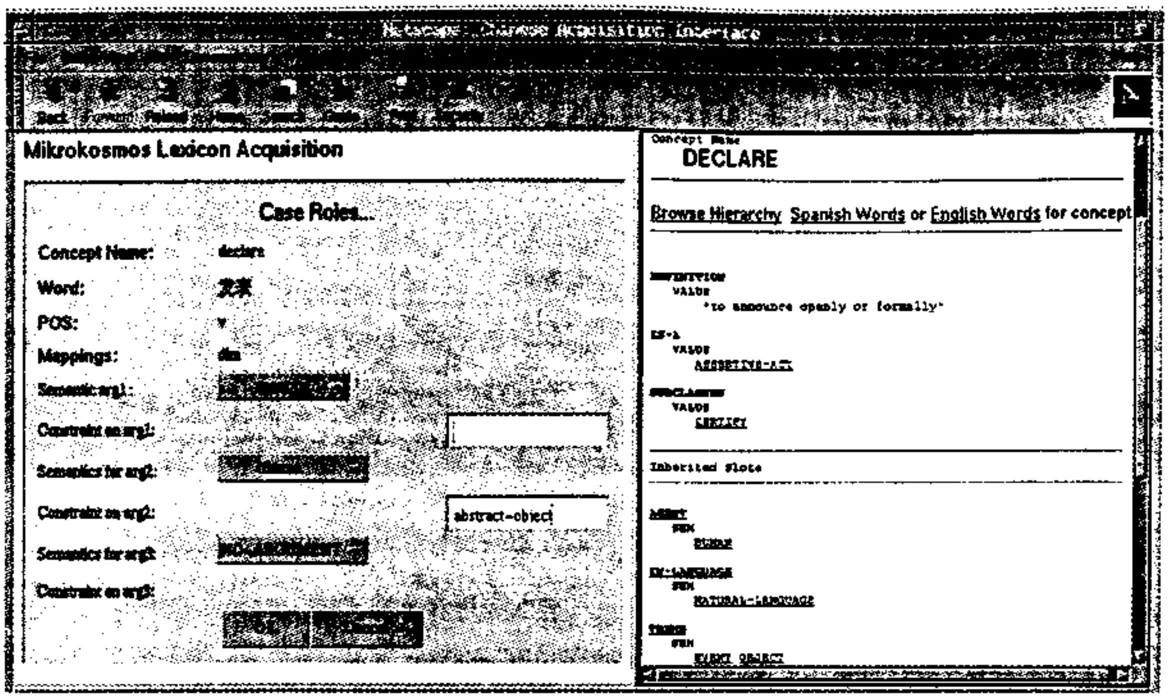

Figure 1: Interface for the selection of semantic case role.

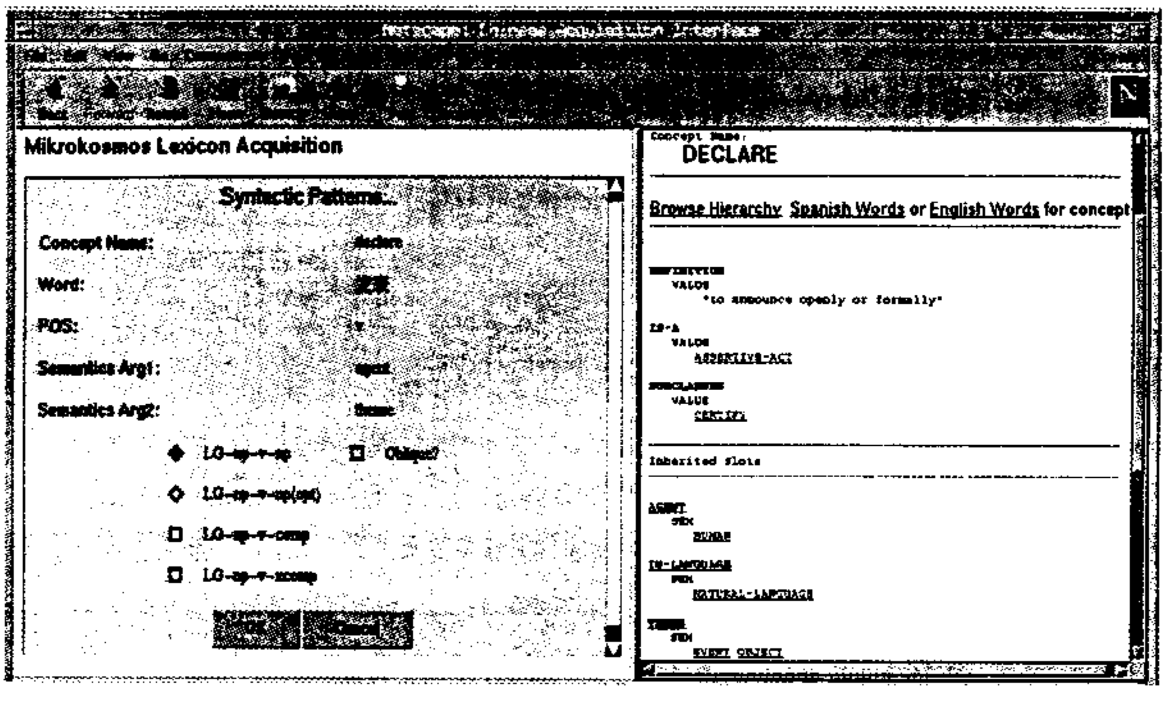

Figure 2: Interface fot selection of syntactic patterns.

```
          (GLOSS "textile")))
    (MODS ((ROOT 出口) (CAT N)
          (GLOSS "export")))
    (ROOT 配额) (CAT N)
          (GLOSS "quota") ))))
%(((SUBJ ((ROOT mei3)  (CAT N)
%           (GLOSS "US")))
%  (MODS ((ROOT dan1fang1)  (CAT ADV)
%           (GLOSS "unilaterally")))
%  (ROOT xui1jian3)  (CAT V)
%  (GLOSS "reduce")
%  (OBJ  ((MODS ((ROOT zhong1guo2)  (CAT N)
```

```
%            (GLOSS "China")))
%       (MODS ((ROOT fang3zhi1pin3)  (CAT N)
%            (GLOSS "textile")))
%       (MODS ((ROOT chu1kou3)  (CAT N)
%            (GLOSS "export")))
%       (ROOT pei4e2)  (CAT N)
%       (GLOSS "quota") ))))
```

削减 (reduce) subcategorizes for a SUBJ and an OBJ in its lexicon entry and is mapped to the concept DECREASE (see Table 1.). SYNSEM links the SUBJ to AGENT: HUMAN and OBJ to THEME: OBJECT. Moreover the grammar of Chinese tells us that the last N in a compound can be identified as a head, as discussed in Huang, (1997).[4] Therefore DECREASE will try to match the semantics of the last N (quota) as the OBJ (grammatical object) of the V which expects a semantics of type OBJECT. Note that 配额 (quota) an ATTRIBUTE with OBJECT in its DOMAIN, and will still be selected thanks to inferences, as shown below in INFERENCE63.

In this sentence we had 7 open class words, out of which 5 had only one sense. The number of senses left after syntactic binding was 8, out of which 6 had one sense. In other words, the semantic processor was left with 2 senses for one word, which it disambiguated correctly. 100% of the words were correctly disambiguated. Obviously this was a simple example and only one word was 2-ways ambiguous; see below results on up to 5-ways ambiguous.

We illustrate in Figure 3. the analysis process and the static knowledge sources.

The syntactic trees are provided to CRL by the University of Maryland in LFG-like syntactic structure (Bresnan, 1982). Lexical entries from the Mikrokosmos lexicon are instantiated for each of the root forms in the syntactic structure. Syntactic variables in the instantiations of the various lexicon entries are bound to one another using the syntactic patterns in the lexical entries to establish syntactic dependencies and map them to semantic dependencies. In addition, ontological concepts referred to the semantic zones of the lexical entries are instantiated.

In the next step, selectional constraints are retrieved from the ontology and added to those encoded in the lexicon. Individual selectional constraints are checked by an ontological search program called Onto-Search

[4] "Heads are final with respect to modifiers." (Huang, 1997).

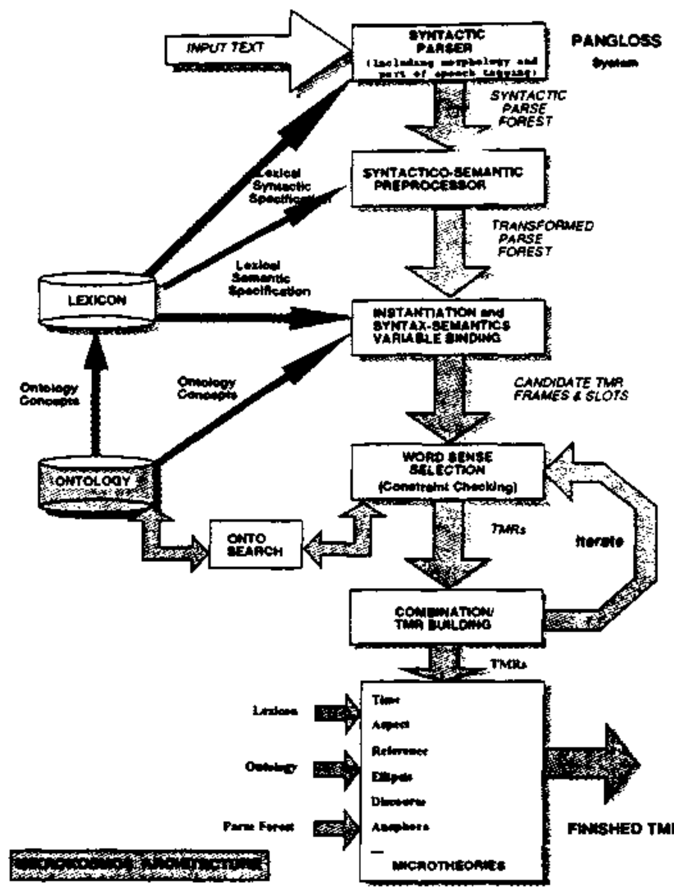

Figure 3: The Mikrokosmos Analysis Architecture.

(Onyshkevych, 1997). The resulting preference values for each constraint are combined in an efficient control and search algorithm called Hunter-Gatherer that combines constraint satisfaction, branch and bound, and solution synthesis techniques to pick the best combination of word senses of the entire sentence in near linear time, as described in (Beale, 1997). Chosen word senses are assembled into TMR frames using lexical semantic representations from the lexicon. Finally, a variety of microtheories are applied to further analyze elements of text meaning such as time, aspect, propositions, sets, coreference, and so on, to produce the final TMR.[5]

[5] A few microtheories are already in place and several

The semantic analysis of sentence "98000" above is presented below in a frame-based representation, followed by its natural language output produced by the generator:

```
prop-59
time: time-58
head: decrease-51

prop-61
time: time-60
head: export-57

decrease-51
agent: united-states-of-america-54
theme: quota-52

quota-52
theme-of: decrease-51 inference63
RELATION: export-57 fabric-55 china-56

unilateral-53
domain: decrease-51

united-states-of-america-54
agent-of: decrease-51 inference65

fabric-55
RELATION-OF: quota-52

china-56
RELATION-OF: quota-52

export-57
RELATION-OF: quota-52

INFERENCE63
type: METONYMY
      object64
      DOMAIN-OF: quota-52

INFERENCE65
type: METONYMY
      human66
      MEMBER-OF: united-states-of-america-54
```

The USA unilaterally reduced China's textile export.

Note that the semantic processor does more than matching selectional constraints or finding the distance along ISA links. The search inside the ontology also involves looking for metonymic type links, as identified in INFERENCE65 where *US* is in a metonymic relation with HUMAN, and INFERENCE63 where *quota* is an ATTRIBUTE with a DOMAIN: OBJECT, which enables it to "match," via relaxation, to the grammatical object of DECREASE which has a semantic type OBJECT.

Also note that the semantic analyzer signals when there are RELATIONs between the Ns inside the compounds, although, at this point, it does not further go down the hierarchy of concepts to recover the most specific one which could apply between the Ns. The generator takes care of the recovery on a needed basis (e.g. when an English compound cannot be used).

The generator produced the following output from the semantic representation above:

```
(((SUBJ
  ((ROOT united-states-of-america)
   (PERSON third) (NUMBER singular)
   (LEX-ENTRY united-states-of-america-C1)
   (INPUT united-states-of-america-58)
   (CAT N)))
  (ROOT decrease)
  (OBJ
   ((MODS
     ((MODS
       ((MODS
         ((ROOT china) (PERSON third)
          (NUMBER singular) (CAT N)
          (LEX-ENTRY china-C1)
          (INPUT china-57)))
        (ROOT fabric) (PERSON third)
          (NUMBER singular) (CAT N)
          (LEX-ENTRY fabric-C1)
          (INPUT fabric-56)))
        (ROOT export) (PERSON third)
          (NUMBER singular) (CAT N)
          (LEX-ENTRY export-C1)
          (INPUT export-55)))
     (ROOT quota) (PERSON third)
       (NUMBER singular)
       (LEX-ENTRY quota-C1)
       (INPUT quota-54) (CAT N)))
  (CAT V) (LEX-ENTRY decrease-C1)
  (INPUT decrease-53) (NUMBER singular)
  (PERSON third)))
```

The United States of America decreased China fabric export quota.

In the absence of a "microtheory" of compounding one cannot generate anything like "the export quota of China fabric." See Viegas *et al.* (1999) for the treatment of Chinese compounds.

Finding out which relation holds between words with a good confidence in retrieving the

are currently being developed by the Mikrokosmos team.

right one from the ontology requires to first get a high confidence rate on the WSD in general and on the words entering in the relations in particular. We present below results on the task of WSD on a very complex sentence "98000" which has 28 words.

在 第 21 届 东新澳 中央银行组织 行长 研讨会 上 中国人民银行 副 行长 殷介炎 就 资本 大量 流入 情况 下 宏观 经济 政策 的 协调 问题 发表 意见

98000: "At the 21st Southeast-Asia-New-Zealand-Australia Central Bank Organization Presidents' Symposium, the vice president of the People's Bank of China, Yin Jie Yan, expressed his opinion on the issue of the macro-economic policy coordination under the condition that capital inflow in large amounts."

In 98000, 7 words (out of 28) were 2-ways ambiguous and 1 was 4-ways ambiguous among open-class words. The table below is for open-class words only.

Sentence Number: 98000

Number of senses in lexicon:	1	2	3	4	5	6
correct:	12	5	0	0	0	0
incorrect:	0	1	0	1	0	0
total:	12	6	0	1	0	0

We ran larger experiments on texts of about 350 words, among which about 100 open class words were up to 5-ways ambiguous. Our first run (i.e. with no modification of f-structures and no lexical tuning) gave us about 90% correctness in the WSD task. The 10% left came from wrongly parsed syntactic structures, in which case we "de-emphasized" syntax, elevating, for instance, all MODS under the root; or, wrong mapping in the lexicon, which was then corrected. It should be noticed that in our approach, lexical tuning is part of the lexicon acquisition task. Once we have acquired a lexicon, we use it with analyzers and generators and "correct" it if necessary. In some cases, the acquirer left the constraints too broad which did not enable the analyzer to pick up the right sense. Our last results on the task of "revising" lexicon entries and syntactic parses yielded the following results:

total senses in lexicon: 333
total number (num.) of open class words: 187
num. of words with only one sense: 95
num. of senses left after syntax binding: 195
num. of words with 1 sense after syntax: 157
num. ambiguous (amb.) senses after syntax: 72
num. amb. words after syntax: 30
num. amb. words correctly disambiguated: 24
num. amb. words incorrectly disambiguated: 6

number of ambiguous senses:	2	3	4	5	6	7	8	9
correct:	19	1	3	1	0	0	0	0
incorrect:	4	2	0	0	0	0	0	0
total:	23	3	3	1	0	0	0	0

The semantic analyzer disambiguated correctly in 96.8% of the cases, with an error rate of about 3.2%.

5 Conclusions - Perspectives

In this paper, we showed the advantage of adopting computational semantics for MT when the language does not offer morpho-syntactic clues.

The work reported here is being conducted on a small scale (about 100 sentences) using a Chinese lexicon of about 2500 word sense entries and an English lexicon of about 20,000 word sense entries. Results on WSD are promising (above 90% of open-class words correctly disambiguated in complex sentences). The Chinese effort has involved the tagging of all the words in our training corpus of 10 texts to identify the "right" sense with respect to our static knowledge sources. Note that there is yet no absolute measures to identify the "real" performance of a system in terms of sense assignment, as discussed recently in WSD meetings (e.g. SENSEVAL). We read the numbers for WSD as qualitative measures showing our progress (programs debugging and development, modification/tuning of lexical resources.)

Acknowledgments:

This work has been supported in part by DoD under contract number MDA-904-92-C-5189. We are grateful to the Mikrokosmos team, and in particular Sergei Nirenburg, Arnim Ruelas, Julia Lonergan, Jeff Longwell, Ron Dolan, Boyan Onyshkevych and Victor Raskin.

References

Beale, S. 1997. Hunter-gatherer: Applying constraint satisfaction, branch-and-bound and solution synthesis to computational semantics. *Ph.D. Diss., Carnegie Mellon University.*

Beale, S., S. Nirenburg and K. Mahesh. 1995. Semantic Analysis in the Mikrokosmos Machine Translation Project. In Proceedings of SNLP-95, Bangkok, Thailand.

Beale, S., E. Viegas and S. Nirenburg. 1997. Breaking Down Barriers: The Mikrokosmos Generator. In *Proc. of the Natural Language Processing Pacific Rim Symposium*, Phuket, Thailand.

Beale. S., S. Nirenburg, E. Viegas and L. Wanner. 1998. De-Constraining Text Generation. In Proc. of the *9th INLG*, Niagara on the Lake, Ontario, Canada.

Bresnan, J. (ed.) 1982. *The Mental Representation of Grammatical Relations.* Cambridge, MA: MIT Press.

Fillmore, C.J. (1985). Frames and the Semantics of Understanding. *Quaderni de Semantica*, VI.2.

Jin, W. 1994. Chinese Segmentation Disambiguation. In *Proc. of the International Computational Linguistics.* Japan.

Jin. W. 1995. Identify unknown words in Chinese Corpus. In Proc. *Natural Language Processing Pacific Rim Symposium '95*, Seoul, Korea.

Li, C.N. and S.A. Thompson. 1981. *Mandarin Chinese: a Functional Reference Grammar.* University of California Press.

Mahesh, K. 1996. *Ontology Development: Ideology and Methodology.* TR MCCS-96-292: NMSU, CRL.

Mahesh, K., S. Nirenburg, S. Beale, E. Viegas. V. Raskin and B. Onyshkevych. 1997. Word sense disambiguation: why statistics when we have these numbers In Proceedings of the *7th International Conference on TMI*, Santa Fe, NM.

Meyer, I, Onyshkevych, B. and L. Carlson. 1990. *Lexicographic Principles and Design for Knowledge-based Machine Translation.* TR CMU-CMT-90-118, Carnegie Mellon University.

Nirenburg. S. and C. Defrise. 1991. Practical Computational Linguistics. In Johnson and Rösner (eds.) *Computational Linguistics and Formal Semantics.* CUP.

Nirenburg, S., J. Carbonell. M. Tomita and K. Goodman. 1992. *Machine Translation: A Knowledge-Based Approach*, San Matro, CA, Morgen Kaufmann.

Nirenburg, S., V. Raskin and B. Onyshkevych. 1995. Apologiae Ontologiae. Proc. of *The Conference on TMI*, Leuven, Belgium.

Ogden. W, et. al. 1995. Cibola: A Translator's Support System, MCCS-95-285. NMSU, CRL.

Onyshkevych. B and S. Nirenburg. 1994. The Lexicon in the Scheme of KBMT Things. TR MCCS-94-277, NMSU, CRL.

Onyshkevych, B. (1997). *An Ontological-Semantic Framework for Text Analysis.* Ph.D. Diss., School of Computer Science, Carnegie Mellon University.

Viegas, E. and V. Raskin. 1998. *Computational Semantic Lexicon Acquisition: Methodology and Guidelines.* TR MCCS-98-315. NMSU, CRL.

Viegas, E. 1998. Multilingual Computational Semantic Lexicons in Action: The WYSINNWYG Approach to NLP. Proc. *36th ACL and the 17th COLING*, Montréal, Canada.

Viegas, E., A. Ruelas, J. Lonergan, S., Beale and S. Nirenburg. 1999. Developing a Large-scale Semantic LKB to Suit an Intelligent Planner. Proc. of *the 7th European Workshop on NLG.* Toulouse, France.

Viegas, E. 1999. An Overt Semantics with a Machine-guided Approach for Robust LKBs. In the Proc. *SIGLEX99 Standardizing Lexical Resources*, ACL99, University of Maryland.