LEXICAL TRANSFER AND RESOLUTION OF PRONOMINAL ANAPHORS

IN MACHINE TRANSLATION: THE ENGLISH-TO-KOREAN CASE

Ruslan Mitkov, Hiongun Kim, Kang-Hyuk Lee, Key-Sun Choi

Computer Science Department

Korea Advanced Institute of Science and Technology (KAIST)

373-1 Kusong-dong Yusong-gu

Taejon 305-701 KOREA

Email: ruslan@hanul.kaist.ac.kr, hgkim@csone.kaist.ac.kr,

khlee@hanul.kaist.ac.kr, kschoi@csking.kaist.ac.kr

TOPIC AREA: Machine Translation

ABSTRACT

There have been various results in English-to-Korean Machine Translation reported, but so far no attention has been paid to the problems of lexical transfer of proniminal anaphors and/or resolving pronominal reference and a default pronoun translation has been considered instead. In this paper we argue that pronouns cannot be handled trivially in English-to-Korean translation and one cannot bypass the task of resolving anaphoric reference if aiming at good and natural translation. In addition, we propose practical transfer rules for English-to-Korean anaphor translation and outline an anaphora resolution model for an English-to-Korean MT system in operation.

RESUMEN

En este artículo se consideran los problemos de la transferencia lexical de pronombres y la resolución anafórica en la traducción automática (TA), sobretodo desde el punto de vista de la traducción inglés-coreano. Después de introducir el problema general de la transferencia léxica, los autores argumentan la necessitad de la resolución anafórica en la traducción inglés-coreano, describen brevemente su sistema de TA y proponen reglas prácticas para la transferencia lexical de pronombres. Finalmente, se presenta un modelo de resolución anafórica que se realizirá en el sistema de traducción automática.

1. LEXICAL TRANSFER IN MACHINE TRANSLATION

In transfer-based Machine Translation (MT) systems, lexical transfer conisists of the replacement of a lexical item by a target lexical item. In case there is only one target language equivalent, lexical transfer does not seem to be a very critical problem. Equally unproblematic for lexical transfer are many-to-one translations. Unfortunately, in most of the language pairs there are a number of cases of one-to-many lexical item mappings available. The lexical transfer phase appears to be the most crucial step in MT because most of the difficult problems are caused by lexical differences beteween two languages.

According to [Hutchins & Somers 92] one-to-one lexical correspondences are expected to be common only in technical texts. We found out that even technical texts cannot be safe enough in lexical transfer. In reality, restricted sublanguages cannot always escape from the tough general language problems. In English-to-Korean translation, for instance, we face additional lexical transfer problems arising from the different possible alternatives when translating English pronouns[1]. And whereas in availability of choices lexical transfer usually requires inspection of surrounding text, in our case the correct choice of a pronoun depends on the information of its antecedent[2], i.e. the correct lexical transfer depends on successful anaphora[3] resolution.

2. TRANSLATION OF PRONOMINAL ANAPHORS

Assuming English to be a source language, English pronominal anaphors can be translated in different ways, depending on the target language.

Consider the translation of the pronoun "it" into different target languages. If the target language is French, Spanish or Italian, we have to choose between a feminine or masculine pronoun. If we translate into German, Greek and Slavic languages, then we have one more gender choice - neuter.

In some languages the pronoun 'it" has to be translated directly by its referent. In English to Malay translation for instance, there is a tendency of replacing 'it' with its referent. As an

---

[1] In this paper we only consider pronouns in 3rd person singular and plural. Pronouns in 1st and 2nd person in Korean are socially dependent, rare in technical sublanguages and are not subject to our current investigations.

[2] Also termed "referent" which is used as a synonym throughout the paper.

[3] Given the complexity of the problem, we have concentrated on pronominal type of anaphora and later in our paper each reference of ...

illustration, consider the following English text and its translation into Malay. In the Malay translation the anaphor *it* is translated by its referent *structure (struktur)*.

> For Malay texts, out of the 5 acceptable structures, we chose only one structure as the stable schema because *it* gave the highest percentage.

> Bagi teks bahasa Malaysia, daripada 5 struktur yang diterima, kami hanya memilih satu struktur sahaja sebagai skema stabil kerana *struktur* tersebut memberikan peratusan tertinggi

Very often pronominal anaphors are simply omitted in the target language. For example, though the English personal pronouns have their correspondences in Spanish, they are frequently not translated because of the typical Spanish elliptical zero-subject constructions. As an illustration, consider the following parallel English and Spanish texts [Korea 94].

> Songgyunkwan is located near the Taehangno Street, on the campus of Songgyunkwan university. *It* was originally the National Confucian Academy.

> Songgyunkwan está en el campus de la Universidad Songgyunkwan, cerca de la called Taehangno. Originalmente fue la Academia Nacional Conficiana.

Regarding English-to-Korean translation, the English pronouns can be omitted elliptically, translated by a definite noun phrase, by their referent, or by one or two possible Korean pronouns, depending on the syntactic information, the type and semantic class of the noun the anaphor refers. In section 4 we propose practical rules for lexical transfer of pronominal anaphors.

3. ANAPHORA RESOLUTION AND MACHINE TRANSLATION

The correct lexical transfer of anaphors in Machine Translation for target languages which mark the pronoun gender, requires anaphora resolution. Pronominal anaphoric reference has to be resolved before we decide which of the possible pronouns to use, because they take the gender of their antecedents.

Replacing a pronominal anaphor with its referent also implies that the translator (program) must be first able to recognize the referent. Even in cases of omitting the pronoun to be translated, usually the referent must be identified because of possible grammatical agreement(s) in the target language. In other words, we face inevitably the problem of

naphora resolution is a complicated problem in natural language processing. Considerable research has been done by computational linguists ([Carbonel & Brown 88], [Dahl & Ball 90], Frederking & Gehrke 87], [Hayes 81], [Hobbs 78], [Ingria & Stallard 89], [Rich & LuperFoy 8], [Robert 89]), but no complete theory has emerged which offers a resolution procedure with success guaranteed. Most approaches developed - even if we restrict our attention to pronominal anaphora, from purely syntactic ones to highly semantic and pragmatic ones, only provide a partial treatment of the problem.

Anaphora resolution within the domain of Machine Translation (MT) has its specific aspects. It also reflects two essential topics in Machine Translation: ambiguity in an MT context and translation of discourse instead of isolated sentences. Anaphora can be viewed as a sort of ambiguity, in that the antecedent of a given pronoun might be uncertain and referential relations are one of the means that constitute coherence of texts.

At least two English to Korean Machine Translation (MT) systems have been reported so far in operation ([Kim & Choi 93], [Lee & Kim 93]), but none of them has paid attention to the problems of lexical transfer of proniminal anaphors and/or resolving pronominal reference and default translation has been used for handling pronouns .

4. DO WE NEED ANAPHORA RESOLUTION IN ENGLISH-TO-KOREAN MACHINE TRANSLATION?

In Korean MT community, not much attention has been drawn to anaphora resolution problems. This is partly due to the complicated problem of anaphora resolution. But it is also due to the biased assumption that anaphoric expressions in the source language can be easily mapped to the corresponding anaphors in the target (Korean) language, or in many cases they can be simply ignored in the transfer phase.

Whereas in most European language pairs anaphora resolution is "compulsory" (or else we risk of rendering in certain cases quite unacceptable translations), there are certain cases in Korean where anaphora resolution may seem "optional".

Consider the sentences [Hutchins & Somers 92]:

    (1) The monkey ate the banana because it was hungry.
    (2) The monkey ate the banana because it was ripe.

In each case the pronoun "it" refers to something different: in (1) the monkey, in (2) the banana and in (3) - to the abstract notion of time. If we have to translate the above sentences in German, then anaphora resolution is inevitable, since the pronouns take the gender of their antecedents and since the German words "Affe" - (masculine, "monkey"), "Banana" (feminine, "banana") and "es" (neutral - "it" for time notion) are in different gender.

Consider the translation of the sentences (1)-(3) from English to Korean and their literal descriptions in English.

(1') 배고파서 원숭이는 바나나를 먹었다.

hungry-CAUSAL monkey-NOMINATIVE banana-ACCUSATIVE eat-PAST,DECLARATIVE.

(2') 익어서 원숭이는 바나나를 먹었다.

ripe-CAUSAL monkey-NOMINATIVE banana-ACCUSATIVE eat-PAST,DECLARATIVE.

(3') 티타임이어서 원숭이는 바나나를 먹었다.

tea time-CAUSAL monkey-NOMINATIVE banana-ACCUSATIVE eat-PAST,DECLARATIVE.

Note that in the above Korean translations there are no pronouns. These examples might seem encouraging that we could translate from English to Korean, bypassing the tough problem of anaphora resolution. However, such a conclusion would be too misleading.

The assumption that anaphoric expressions in the source language can be easily mapped to the corresponding anaphors in the target language, or in many cases they can be simply ignored in the transfer phase, is unfounded. It is not hard to find English sentences for which anaphora resolution is necessary in order to get their correct translation into Korean. Consider the sentences:

(4a) Although programmers usually write good programs, they may still make a mistake.

(4b) Although programs are usually written by good programmers, they may still contain mistakes.

In Korean, there are two types of pronominals corresponding to "they", one for human beings and the other for non-human beings. In order to assign the proper Korean pronominals to the English pronominal "they", the system should be able to resolve "they" between the two possible referents, "programmers" and "programs".

ihora resolution becomes a more serious business when we aim at achieving high-quality
ilation. The translation of (4a) and (4b) into Korean with the successful assignment of
ouns may still sound awkward to Koreans, because in Korean it is stylistically more
ral not to explicitly mention anaphors in subordinate clauses that are coreferential with
inal expressions in the main clause. It is somewhat similar to English participle
tructions whose subject is "understood." The best translation of (4b) in Korean could be
ribed in English literally as:

(5) Being usually written by good programmers, programs may still contain
mistakes. .

s, if we are able to get the translation of (4a) and (4b) without overt pronominals, we are
e likely to get better translation. This being so, anaphora resolution is very crucial in
lish-to-Korean MT because we must resolve the pronominal "they" to replace it by proper
iinal expressions.

·eover, "optional" anaphora resolution means preserving anaphoric ambiguity in case no
phora resolution is undertaken. It may seem that carrying ambiguities over translation is
n more "authentic" from the point of view of having a mirror translation of the source text.
resolving anaphoric ambiguity means that during the translation process text is not fully
lerstood. Generally speaking, however, analysis is aimed at producing an unambiguous
·rmediate representation [Isabelle & Bourbeau 85].

reover, a system strongly relying on the "ambiguity preservation" method, in addition to
·ring no computational advantage when ambiguity-preserving situations must be
ntified dynamically, is extremely vulnerable in situations where (i) the lexicon is growing
ile the system is in use or (ii) when additional languages must be introduced ([Nirenburg et
92]). Every new word sense added to the lexicon carries the potential of ruining the
isibility of retaining ambiguity in translation for all previous entries. All this means that
ra attention must be paid to the maintenance of the lexicons.

:NGLISH-TO-KOREAN MACHINE TRANSLATION AND MATES

: are currently investigating the anaphor lexical transfer problem and anaphora resolution
isibility with regards to possible extension of our English-to-Korean translation system
ATES/EK. MATES/EK has been developed through a co-research done by KAIST and SERI
·stems Engineering Research Institute) from 1988 to 1992, and is still under evolution in

result (parse tree) into an intermediate representation, and then transforms it into a Korean syntactic structure to construct a Korean sentence [Choi 94 et al]. MATES/EK has the following features:

- Morphological Analysis Using N-gram:

Category ambiguities are resolved by combining the N-gram and the rules.

- Augmented Context Free Grammars for English Syntactic Analysis:

An augmented context free grammar has been defined for general English syntactic analysis and the analyzer is implemented using Generalized parsing algorithm [Tomita 87].

- Lexical Semantic Structure (LSS) to represent the intermediate representation:

The result of the syntactic structure is transformed into an intermediate representation LSS, which is a dependency structure that is relatively independent to specific languages. In LSS, the constituents in a sentence are combined only in head-dependent relation based on the lexical categories, and there are no order relation between the constituents. Hence LSS is desirable for translation between English and Korean, two languages with fairly different syntactic structures.

- Grammar Writing Language and its Environment:

MATES/EK runs a series of tree transformations on the LSS structure from the English syntactic structure, in order to get a structure specific to Korean syntactic structure. To do this, we developed a grammar writing language, in which the rules describe the tree transformations, and its supporting system. During the tree transformation operations the system looks up in the English-Korean bilingual dictionary in order to get the Korean lexemes.

- Development Supporting Tools :

Taking into account the continuously growing property of an MT system, we developed a set of development supporting tools for grammar writing and editing, dictionary updating and translation testing.

[ATES/EK consists of a set of dictionaries, a set of grammar rules and processing modules. ranslation is carried out as sequential processing stages: English morphological analysis, nglish syntactic analysis, English semantic analysis, English-Korean lexical transfer, nglish-to-Korean structural transformation, Korean syntactic structure generation and .orean morphological generation. Figure 1 depicts the overall configuration of MATES/EK.
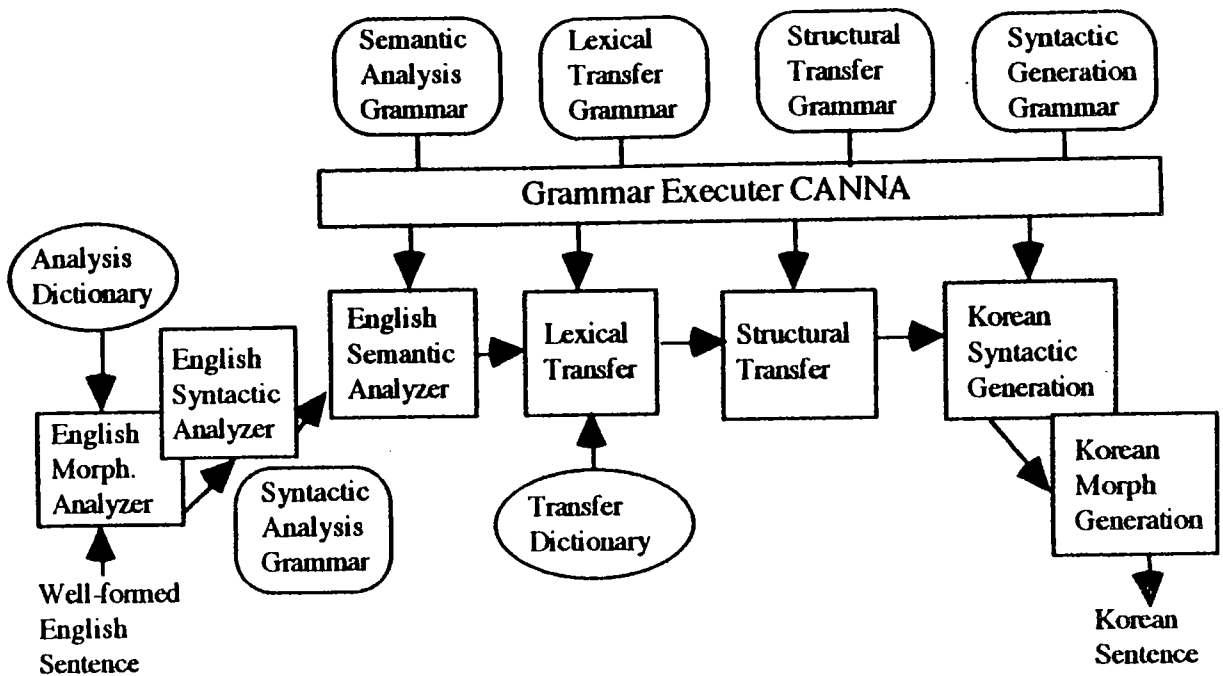

Figure 1. The system configuration of MATES/EK

This system is tested on 1,708 sentences, whose length is less than 26 words selected from 2500 sentences in the IEEE computer magazine September 1991. It shows about 95 percent of success rate for sentences shorter than 15 words, about 90 percent for 18 words, 80 percent for 21 words, and 75 percent for 26 words. This is a quite encouraging result since the magazine contains various kinds of texts of various styles.

6. LEXICAL TRANSFER RULES FOR ENGLISH-TO-KOREAN ANAPHOR TRANSLATION

Though MATES/EK has demonstrated very encouraging results, it is far from being a perfect automatic translator. We have not incorporated lexical transfer rules for pronominal anaphors and anaphora resolution module yet, and there are certain cases of complex and/or subsequent sentences in which the system will not produce the most natural translation.

We studied various texts from a Computer Science English corpus prepared for Machine Translation and on the basis of our observations we proposed practical anaphor translation rules for the needs of the lexical transfer in English to Korean Machine Translation. These

syntactic information, the type and semantic class of the noun the anaphor refers to, and can be very helful for the lexical choice during the transfer. We do not regard the described set of rules as complete: however, it provides a good starting point and gives an idea of the direction of our current research efforts. The following major cases are concerned:

5.1 Complex Sentence I (main clause + subordinate clause)

> When the subordinate clause in a complex sentence follows the main clause, the anaphor in the subordinate clause should not be translated.

(8) Cashes are one of the most important ideas in computer architecture because they can substantially improve performance by the use of memory.

메모리 사용에 의해 성능을 상당히 향상 시킬 수 있으므로 캐쉬는 컴퓨터 구조에서 가장 중요한 아이디어중 하나다.

memory use-INSTRUMENTAL performance-ACCUSATIVE substantially improve-CAUSAL cache-NOMINATIVE architecture-LOCATIVE most important ideas-among one-DECLARATIVE.

5.2 Complex Sentence II (subordinate clause + main clause)

> If the subordinate clause in a complex sentence is followed by the main clause, the anaphor(s) in the main clause is realized as the noun phrase(s) in the subordinate the anaphor refers to.

(9) As processors get faster, they will lose more and more of their performance to the memory system.

프로세서가 빨라질수록 프로세서의 메모리 시스템에 대한 성능이 상실될 것이다.

processor-NOMINATIVE fast-CONDITIONAL processor-POSSESSIVE memory system-DIRECTIVE performance-NOMINATIVE lose-FUTURE,DECLARATIVE.

5.3 Generalized Quantifiers

> If the anaphor refers to generalized quantifiers[4], it is translated into definite description patterns such as "그런 (kuren: 'such' or 'that') + nominal".

(10) Although <u>many approaches</u> may be technologically feasible, <u>they</u> must also be economically feasible to be applied in a market economy.

많은 접근방식들이 기술적으로 적당하더라도 그런 접근방식들은 시장경제에 적용될 때 경제적으로 적당해야 한다.

<u>many approaches</u>-NOMINATIVE technically feasible-CONCESSIVE <u>those approaches</u>-NOMINATIVE market economy-LOCATIVE apply-MODIFIER time-TEMPORAL economically feasible-DECLARATIVE.

5.4. Human "it" (child)

> If the antecedent of "it" is human, replace the pronoun with its antecedent (preceded by a definite article)

(11) The <u>child</u> is in the room, and <u>it</u> is playing with a doll.

아이는 방에 있는데, 그 아이는 인형을 가지고 놀고 있다.

<u>child</u>-NOMINATIVE room-LOCATIVE exist-CONJUNCTIVE, <u>the child</u>-NOMINATIVE doll-INSTRUMENT play-PRESENT,DECLARATIVE.

5.5 Non-human "she"

> If the antecedent of "she" is non-human, translate the pronoun by replacing it with its its antecedent (preceded by a definite article).

(12) There was a nice banquet on <u>a ship</u>. <u>Her</u> name was 'Pearl de Mer'.

배위에서 훌륭한 향연이 있었다. 그 배의 이름은 Pearl de Mer' 였다.

<u>ship</u>-LOCATIVE nice banquet-NOMINATIVE be-PAST,DECLARATIVE. <u>the ship</u>-POSSESSIVE name-NOMINATIVE 'Pearl de Mer' be-PAST,DECLARATIVE.

5.6 Inanimate "they"

> If the antecedent of "they" represents a set of non-human entities, translate the pronoun as "그것들 (kugotul)"

(13) When the system executes the erroneous <u>instructions</u> with certain data values, <u>they</u> cause a failure and the error becomes effective.

시스템이 어떤 데이타 값으로 오류가 있는 명령을 수행할 때, 그것들은 고장을 유발하고 오류가 효력을 발생한다.

System-NOMINATIVE certain data value-INCLUSIVE errors-NOMINATIVE exist-MODIFIER instructions-ACCUSATIVE execute-MODIFIER time-TEMPORAL, they-NOMINATIVE fault-ACCUSATIVE cause-CONJUNCTIVE error-NOMINATIVE effect-ACCUSATIVE occur-DECLARATIVE.

Every other case

In every other case, use the default translations : "he"-"그 (ku)", "she"-"그녀 (kunyo)", "it"-"그것 (kugot)" and "they"-"그들 (kugotul)"

(14) Mr. Han went out at 3:00 in the afternoon and he will come back late in the evening.

한씨는 오후 3시에 나갔는데 그는 저녁 늦게 돌아올 것이다.

Mr.Han-NOMINATIVE afternoon 3:00-TEMPORAL go_out-PAST,CONJUNCTIVE he-NOMINATIVE evening late return-FUTURE,DECLARATIVE.

he rules 5.1-5.7 are to be integrated into the lexical transfer module of MATES. These ractical rules are based on available information about the referent and before applying 1em, antecedents should have been already identified.

'he above rules are given in priority order and can be easily described in a more formal way.

. ANAPHORA RESOLUTION MODEL

Ne have seen that the correct lexical transfer of English pronouns in Korean requires dentifying their antecedents, i.e. anaphora resolution. In addition, since we are aiming at extending MATES into a system that can handle discourse translation (initially handling anaphoric references), we are studying different anaphora resolution strategies. At this stage, we have chosen a simplified version of our integrated anaphora resolution model proposed in [Mitkov 94]. Full implementation of this model, including center tracking inference engine, seems too costly for the immediate goals of our English to Korean translation system.

The main idea is that given the complexity of the problem, we think that to secure a comparatively successful handling of anaphora resolution one should adhere to the following
 domain (sublanguage) rather than focus on a particular natural

bonell & Brown 88], [Rich & LuperFoy 88]), ([Preuß et al 94]) do express the idea of
ti-level approach", or "distributed architecture", but their ideas a) do not seem to
ire enough discourse and heuristical knowledge and/or b) do not concentrate on and
tigate a concrete domain, and thus risk being too general. We have tried nevertheless to
porate some of their ideas into our proposal.

inaphora resolution model integrates modules containing different types of knowledge -
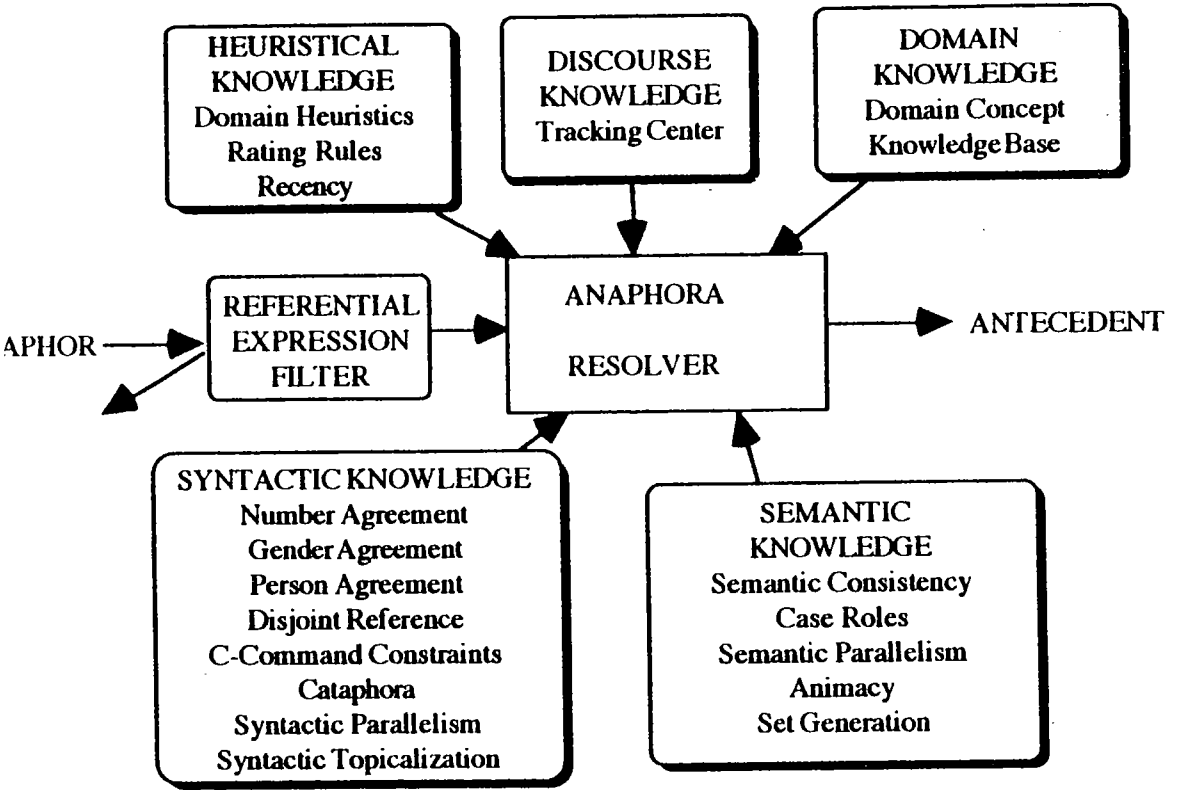ictic, semantic, domain, discourse and heuristical (Figure 2).

Figure 2: An integrated anaphora resolution architecture

syntactic module, for example, knows that the anaphor and antecedent must agree in
iber, gender and person. It checks if the c-command constraints hold and establishes
oint reference. In cases of syntactic parallelism, it prefers the noun phrase with the same
tactic role as the anaphor, as the most probable antecedent. It knows when cataphora is
sible and can indicate syntactically topicalized noun phrases, which are more likely to be
icedents than non-topicalized ones.

semantic module checks for semantic consistency between the anaphor and the possible

semantics or the animacy of the candidate. In cases of semantic parallelism, it prefers the noun phrase, having the same semantic role as the anaphor, as a most likely antecedent. Finally, it generates a set of possible antecedents whenever necessary.

The domain knowledge module is practically a knowledge base of the concepts of the considered domain and the discourse knowledge module knows how to track the center of the current discourse segment.

The heuristical knowledge module can sometimes be helpful in assigning the antecedent. It has a set of useful rules (e.g. the antecedent is to be located preferably in the current sentence or in the previous one) and can forestall certain impractical search procedures.

The use of common sense and world knowledge is in general commendable, but it requires a huge knowledge base and set of inference rules. At the present stage of our project, however, we do not envisage the development of this module.

The syntactic and semantic modules usually filter the possible candidates and do not propose an antecedent (with the exception of syntactic and semantic parallelism). Usually the proposal for an antecedent comes from the domain, heuristical, and discourse modules. The latter plays an important role in tracking the center and proposes it in many cases as the most probable candidate for an antecedent.

The referential expression filter is very useful in filtering out impersonal 'it'-expression (e.g. "it is important", "it is necessary", "it should be pointed out" etc.) , where 'it' is not anaphoric.

Initially, we envisage the implementation of the syntactic, semantic and heuristical modules, which together with the referential expression filter alone are helpful in solving practically most of the cases in our sublanguage.

A CONCLUDING NOTE

In our paper we have shown the importance of the correct lexical transfer and anaphora resolution in English-to-Korean Machine Translation. We would like to point out that the raised issues apply with equal importance to Korean text generation since the proposed lexical transfer rules, if correspondingly modified, can be regarded as lexical choice rules for generating Korean anaphoric expressions.

[4] Dik, S.C. (1968), *Coordination. Its implications for the Theory of General Linguistics*, Amsterdam, ed. J. Benjamin.

[5] Franchini, E. (1986), *Las condiciones gramaticales de la coordinación* copulativa en español, Berna, ed. Francke Verlag.

[6] Gili Gaya, S. (1987), *Curso superior de sintaxis española*, Barcelona, ed. Vox.

[7] Hernanz, M.LL. y Brucart, J.M., *La sintaxis. 1. Principios teóricos. La oración simple*, Barcelona, ed. Crítica.

[8] McCord,M.(1985) Using Slots and Modifiers in Logic Grammars for Natural Language. Artificial Intelligence, vol. 18

[9] McCord,M.(1990) Slot Grammar: A system for simpler construction of practical natural language grammars. Natural Language and Logic. International Scientific Symposium.

[10] McCord,M.;Shalom Lappin Anaphora Resolution in Slot Grammar. IBM. T.J. Watson Research Center

[11] Moreno Cabrera, J.C. (1991), *Curso universitario de lingüística general. Tomo I : Teoría de la gramática y sintaxis general*, Madrid, ed. Síntesis.

[12] Sola, J. (1972), *Estudis de sintaxi del català*. Barcelona, ed. 62, 2 vols.

ANEXO: EJEMPLOS

Resultado del análisis de una oración con extraposición.

La estructura *phrase* considerada es:

phrase(Posi,Posd,Posn,Nucleo,Features,Coms,Exts,Mizqs,Mders)

Posi: posición de la primera palabra anterior del sintagma.

Posd: posición de la última palabra del sintagma.

Posn: posición de la palabra núcleo del sintagma.

Núcleo: forma base y acepción de la palabra núcleo.

Features: características del sintagma.

Coms: slot frame.

Exts: huecos extrapuestos.

Mizq: modificadores por la izquierda.

Mder: modificadores por la derecha.

Cada modificador en la forma *slotllenado(sintagma modificador)*.

La oración analizada es:

. Os quiero ver de negro.

La estructura phrase obtenida para la frase completa es:

posizq: 0

posder: 5

posnucleo: 2

las siguientes condiciones.

I. P y N tienen características de concordancia incompatibles.

II. P está en el dominio del argumento de N.

III. P está en el dominio adjunto de N.

IV. P está en el argumento de una palabra núcleo H, N no es un pronombre, y N está conte

en H.

V. P está en el dominio de sintagma nominal de N.

VI. P es el determinante de un nombre Q, y N está contenido en Q.

Condición III: María llegó con ella.

Condición IV: Este es el hombre que dijo que él escribió.

El segundo algoritmo consiste en identificar un posible sintagma nominal que actúe

antecedente para un pronombre anafórico. Tomaremos como pronombres anafóricos los pronon

reflexivos. El concepto de hueco argumento (o complemento) mayor usado en la formulación del algo

se define en base a la jerarquía de huecos argumento siguiente:

subj > agent > obj > (iobj|pobj)

donde subj es el hueco sujeto, agent es el agente de una oración pasiva, obj es el hueco

directo, iobj es el hueco objeto indirecto y pobj es el objeto de un complemento preposicional de un

Asumimos las definiciones de dominio del argumento, dominio adjunto y dominio del sintagma nomi

Un sintagma nominal N es un posible antecedente para un pronombre anafórico A si N y

tienen incompatibilidades de concordancia y se cumple una de las siguientes cinco condiciones:

I. A está en el dominio del argumento de N, y N llena un slot argumento mayor que el que

A.

II. A está en el dominio adjunto de N.

III. A está en el dominio de sintagma nominal de N.

IV. N es el argumento de un verbo V, hay un sintagma nominal Q en el dominio del argu

o en el dominio adjunto de N de manera que Q no lleva determinante, y A es i) un argume

Q, o ii) A es un argumento de una preposición PREP y PREP es un adjunto de Q.

V. A es el determinante de un nombre Q, y i) Q está es el dominio del argumento de N y

un hueco argumento mayor que Q, o ii) Q está en el dominio adjunto de N.

BIBLIOGRAFIA

[1]   Gerardo Arrarte. SSG: Una gramática del español basada en el formalismo slot grammar. I
        de Tecnología Multimedia IBM. SEPLN VIII Congreso.

[2]   Barreras, J. ( ), *Resolución de elipsis y técnicas de parsing en una interficie de lenguaje*
        SEPLN VIII Congreso.

[3]   Cuenca, M.J. (1988), *L'oració composta (I) : la coordinació*, València, Servei de Publica
        la Universitat de València.

REFERENCES

[Barwise & Cooper 81] - Baswise J., Cooper R. - *Generalized quantifiers and natural languages*. Linguistics and Philosophy, 4, 1981

[Carbonell & Brown 88] J. Carbonell, R. Brown - *Anaphora resolution: a multi-strategy approach*. Proceedings of the 12. International Conference on Computational Linguistics COLING'88, Budapest, August, 1988

[Choi et al 94] Choi K.S. , Lee S.M. , Kim H.G. , Kim D.B, Kweon C.L., Kim G.C. - *An English-to-Korean Translator : MATES/EK*. Proceedings of the 15. International Conference on Computational Linguistics COLING'94, Kyoto, August, 1994 (to appear)

[Dahl & Ball 90] D. Dahl, C. Ball - *Reference resolution in PUNDIT*. Research Report CAIT-SLS-9004, March 1990. Center for Advanced Information Technology, Paoli, PA 9301

[Frederking & Gehrke 87] R. Frederking, M. Gehrke - *Resolving anaphoric references in a DRT-based dialogue system: Part 2: Focus and Taxonomic inference*. Siemens AG, WISBER, Bericht Nr.17, 1987

[Hayes 81] P.J. Hayes - *Anaphora for limited domain systems*. Proceedings of the 7th IJCAI, Vancouver, Canada, 1981

[Hobbs 78] J. Hobbs - *Resolving pronoun references*. Lingua, Vol. 44, 1978

[Hutchins & Somers 92] - *An Introduction to Machine Translation*, Academic Press, 1992

[Ingria & Stallard 89] R. Ingria, D. Stallard - *A computational mechanism for pronominal reference*. Proceedings of the 27th Annual Meeting of the ACL, Vancouver, British Columbia, 26-29 June 1989

[Isabelle & Bourbeau 85] Isabelle P. and L. Bourbeau - *TAUM-AVIATION: Its technical features and some experimental results*, Computational Linguistics, 11, 1985

[Kim & Choi 93] Kim C.G., Choi K.S. - *Machine translation environment system for English to Korean*, CAIR-TM-93-19, 1993

[Korea 94] - *Traveler's Korea/Viajeros a Corea*. Korea National Tourism Corporation, 1994

[Lee & Kim 93] - Lee H.G, Kim Y.T. - *An idiom-based approach to Machine Translation*. Proceedings of TMI'93, Kyoto, Japan, 14-16 July, 1993

[Mitkov 94] Mitkov R. - *An integrated model for anaphora resolution*. Proceeedings of the 15th International Conference on Computational Linguisticss COLING'94, Kyoto, Japan, 5-9 August 1994 (to appear)

[Nirenburg et al. 92] Nirenburg S., Carbonell J., Tomita M., Goodman K. - *Machine Translation: a knowledge-based approach*, Morgan Kaufmann Publishers, 1992

[Preuß 94 et al] Preuß S., Schmitz B., Hauenschild C., Umbach C. - *Anaphora Resolution in Machine Translation*. In W. Ramm, P. Schmidt, J. Schütz (eds.) Studies in Machine Translation and Natural Language Processing, Volume on "Discourse in Machine Translation" (to annaar)

[Rich & LuperFoy 88] Rich E., LuperFoy S. - *An architecture for anaphora* Proceedings of the Second Conference on Applied Natural Language Processing, A 9-12 February 1988

[Robert 89] Robert M. - *Résolution de formes pronominales dans l'interface d' d'une base de données.* Thèse de doctorat. Faculté des sciences de Luminy, 1989

[Tomita 87] Tomita M. - *An efficient augmented context-free parsing* Computational Linguistics, 13, 1987