Tenets for an Interlingual Representation of Definite NPs

Montserrat Meya
CDS Barcelona
Siemens SA, Spain
e-mail: Montserrat_Meyallopart@eurokom.ie

1. Introduction

The main goal of this paper (as in Keenan and Stavi 1986) is to characterize the possible determiner denotations in order to develop a computational approach that makes explicit use of this information. To cope with the constraints that languages impose when generating determiners, a computational model has to follow the laws that map definiteness to structures and strings and viceversa.

In the following proposal I distantiate from K. Bühlers Deixis Theory and Weinrichs (76) proposal where indefinites suggest subsequent information, while definite point out facts from the previous information. This very general position is insufficient if we want to formalize NP-definiteness.

The semantics of NP definiteness must be captured adequately in computational frameworks for such tasks as answering quantified NL- questions, or in a MT system to convert NPs from one language into another.

In the first part of this paper I draw a typology of definiteness; later I reflect on the definiteness of NPs in an IL-representation. The major result is given by the determiner generators. Definiteness should be evaluated in a Q-A system and in MT. The extensive functionality of definiteness is first elaborated in the parsing and results in an IL-representation; finally the determiner generators create correct morphological determiners and right determiner structures.

2. Parsing definite NPs

Weinrich (1976) distinguishes three types of anaphoric and cataphoric information: context, situation and code. Code is the paradigmatic meaning that speakers find in the lexicon (standard definitions).

The NPs parsing strategy- that copes with these three information types- comes down to these main questions:

- the question of building adequate semantic interpretation for the functional features fixed during analysis
- The question of reference resolution for definite NPs

To achieve that goal the Np grammar must work with a sufficient:

- determiner subcategorisation
- strategy about constraints depending upon the semantic features of the nouns, and on the context (PPs, Rel-S, compounding, etc.)

This strategy, of course, can not rely on the presupposition of existence and uniqueness, i.e. the "accommodation" model of Lewis

1

(1979) or the Russellian analysis of "the" so closely related to the analysis of "only", that would fail for sentences like: (MacCawley 81. p.177):

1. "The restaurant on the Clark street is excellent"

because Russell's analysis did not allow truth value gaps or interactions between logical form and context (MacCawley. p.177), if there is more than one restaurant on the Clark Street, sentence 1 expresses a false proposition.

Determiner Subcategorisation

The list of determiners is surprisingly long (Keenan 86) and there is even the possibility to build complex determiners by adding modifiers to determiners. According to keenan and Stavi there are 3 types of determiners: cardinal, logical and non logical.

Dets are taken into consideration according to the fact of being terminal or pre-terminal nodes. This subcategorisation entails the following additional feature-value pairs.

Terminal determiners are subcategorized in the lexicon according to the features:

- definite (logical) vs. indefinite (non-logical)
- demonstrative vs. possessive
- type of quantification {collective, distributional, numeral, amount, partitive, negation}
- comparatives

Pre-terminal determiner features

When NP structures are built during parsing, the intermediate nodes must come up with values for the features:

- definiteness {proper name, bounded NP}
- negation

- topicalization
- generics
- head-specifier relation
  - apposition {Titel,...}
  - genitives {possessive, part-of,...}
  - predication {equality vs. copula}
  - type of structure {PP, Rel-S,...}
- semantics of the NP {mass vs. count noun, temporal, local, argument,...}

Each feature configuration corresponds with a NP structure with full decoration (in a MT system), or a compositional formula representation (in understanding systems); in both cases it is built up rule-to-rule.

The outgoing object of the parser is a IL-representation. Later on the generators will define, according to the idiosyncratic laws of each language, the morpheme assignment for this interlingua (IL).

Determiner functionality

The semantic value of the tags {generics, definiteness, topic} for NPs is obtained during analysis and it is the result of taking into consideration:

- the whole NP structure {attribution, apposition, genitives, head-specifier relation in compounds is relevant for definiteness.
- the NP typology together with the sentence aspect and time is relevant for generics.
- the semantic primitives tie the lexical item to an ontological class and is relevant for the possible word readings
- word order and negation is relevant for topic

Generics

Generics do not underly universal or nearly universal quantifiers. The reference of generics mirrors that of habitual propositions. They are treated as predication of individuals though

2

allowing the clauses in question to take different truth values at different times.

Generics are expressed in the present tense with the aspect value habitual (usually, occasionally, always). The 3 different generic types reflect the different quantification of the determiners in propositions like:

1. "Men do not cry"
   (Quant.=distrib)(=usually every man does not cry)

2. "Germans are bad husbands"
   (Quant.= distr)

3. "a man does not cry"
   (Quant.=mass)(whoever has the property of being)

4. "the lion is a mammal"
   (Quant=coll) (The species...)

5. "I like lamb"

Contrary to MacCawley I think that indefinite plural generics as in (1) are basically references to sets of objects. Indefinite singular generic NPs are references to properties of individuals (3), and definite singular NPs with count nouns (4) or indefinite (without article) mass nouns (5) are basically references to kinds.

The final value "generics" on the NP is calculated from the verbal tense value (PR) and from the aspect value (habitual) together with the determiner of the NP, the semantic type of the noun, and the whole NP syntactic structure.

6. "Wine is a drink"
   vs. "el vino es una bebida"
   [$SUBJ mass]

7. "I drink wine"
   vs. "bebo (0) vino"
   [$OBJ mass]

The same generic NP in examples 6/7 behave differently in spanish according to the sentence argument.

Both the generics and the syntactic role in the sentence have an effect on the article generation in different languages.

Head-Specifier Structure

The addition of features providing information about certain syntactic structures is very useful in reflecting the nominal definiteness behavior. The following structures assist in distinguishing (between general and specific reference) definite NPs:

* acronyms and proper names:
  "The Union"
* appositions:
  "The Siemens company"
* complex proper names (dates, time:
  2 o'clock, 2.3.1990,...)
* attribution (with adjectives, relative clauses, genitive constructions, prepositional phrases, compounding,...)

Here some examples with the value definiteness =True:

"The second paragraph", "the cleanest water", "The man that I remember", Peter's hatred", "the truth about his past", "the picture on the wall", "the king's daughter",..

Genitives in whatever morphological form are always definite constructs:

* The king's daughter
  Die Koenigstochter
  la hija del rey
* House door
  Haustür
  la puerta de la casa

Germanic languages only have one determiner in compounds, however the genitive relation must be captured to reflect the definiteness of the specifier in another language.

During NP parsing, constituents are marked according to their head- specifier function together with their semantic type.In that way,

PPs, adjectival phrases and compounds are represented in a uniform way. Additional feature constraints on head and specifier regulate definiteness, depending on their semantic primitives:

- Druckfestigkeit
  resistence a la compresion
- Druckknopf
  bouton a presion
- Einkommensteuer
  impot sur la revenue
- Einfuhrsteue
  impot d'importation

Directional and local compounds, or the corresponding PPs in other languages contain the feature definiteness because they are usually tied to some deictic situation (See Loebner 1985, semantic vs. pragmatic definiteness).

- im Palast
  en el palacio
- zum Palast
  al palacio

Topicalization

The series "definite article + Lexeme + indefinite article + Lexeme" reflects a methodological progress in the information situation. If this direction changes, the direction of the predication also changes. That means that the noun (or NP) will be focused and therefore has a definite reference.

8. the FBI employed American citizens
   (la FBI emplea 0 a ciudadanos)

9. American citizens are employed by the FBI

10. Los ciudadanos americanos son empleados por la FBI

11. a: Dem Kind gab ich den Ball
    b: Es al niño a quien di la pelota

12. The man I love

Topicalized NPs must be marked as such. Voice in example 9 topicalize the NP and demands in spanish definiteness for the NP. The bare topicalization in german in example (11a) corresponds a marked extraposition in spanish (11b). Languages make a discriminating use of this feature when generating determiner requirements.

Predicative constructions

In predicative constructions the relevant criteria for the use of determiners are the type of copula (NP expresses a property (14-15), or a reference (16)), while with other verbs the relevant criteria is the aspect (process vs. accomplishment or achievement)

A sentence involving a noun X with definite article in predicate position does not assert that the subject is X. We have to distinguish between a "be" of identity and that of copula.

13. Felipe es (el) presidente de España
    a: $(\iota x: presidente(x, España))$
    $=(Felipe, x)$

14. Alex is an artist vs. Alex es artista

15. Felipe es el presidente que prefiero
    a: $(\exists x)\ (PE\ x, (\forall y) \supset$
    $(\sim=yx, \sim PEy), Felipe, x)$

English demands article for copula cases like (14), existing the opposite referential value for the article "the" (the teacher). Referential examples are treated the same in english and spanish for identity predicate.

At the other hand arguments of predicates in accomplishments or achievements (not as a process) are (16) individualized and as such demand definiteness. The sole exception are lexicalized expressions and sentences with function verbs (17)

16. "voy a la playa" vs. "I go to the beach" (achievement)
    "bese un tigre"(al tigre) vs. I kissed

4

the tiger (accomplish.)
"me gusta besar tigres" vs."I like to kiss tigers" (process)

17. "pongo la mesa" vs. "Ich decke den Tisch"
"tengo razon" vs.Ich habe Recht"

3. Interlingual representation of definiteness

Definiteness is a means to satisfy coherence requirements. Definite NPs have a scope just like quantifiers and they may be involved in scope ambiguities as Russell suggested (MacCawley p.178). The most reliable test to decide about the definiteness of an NP is to use Peterson's (1979) test of there-insertion. There insertion requires that the existential quantifier be immediately above the clause into which there is to be inserted. If an existential context is possible then there is no definiteness.

18. there are two men in the garden
*there are the two men in the garden
(no possibility of there)

Definiteness in an extensional model it is a function from properties into sets of properties.

NPs in the IL representation used in MT are flat tree structures with decorations containing sets of attribute-value pairs. These attributes were sketched above. Feature configurations are equivalent to well formed formulas after lambda conversion, so that this approach can be used in a NL understanding system as well.

Below a rough outline of a PSR for nominals:

1: NP <-- 2: DET + 3:NO
conditions: ......
structure:
1: nu = 2:nu
1:Quant = !VP.aspect & Qtype of

determiner
1:Generics = !VP.tense & VB.aspect (i)
1:semtype= = !NO.semtype
1:definiteness = !{DET,NO.Compl} (ii)
1:comparative = !NO.compl !PREP (iii)

• (i) if tense = present and aspect = habitual and NO is plural then the quantification is distributional; if the tense is present but NO is singular then the quantification of the det is collective

• (ii) if the determiner type is possessive or demonstrative, or the NP contains genitives, relative clauses, etc..the NP is definite

• (iii) this feature keeps track for predicative constructions like:

19. he works as a teacher
vs. trabaja como 0 maestro

The final value for definiteness and quantification is the result of the above mentioned calculus that is percolated to the NP nodes.

4. Generators for definiteness

It is straightforward to show that for any definiteness configuration there is a determiner assignment in a specific language. The generation grammar is the inverse process of the analysis. Below some rules:

Rule 1 :
if
definiteness = T
NU = Sg
semtype = ~ABS
focus = T
then
Ger. = definite DET + NO
span = demonstrative DET + NO
fran = demonstrative DET + NP
"esta bolsa vs. die Tüte"

Rule 2:
if
generics = T

NU = PL
Quant = distrib.
then
    Germ. = no DET + NP pl
    eng. = demonstrative DET + NO pl
    spn. = Definite DET + NO pl
"los hombres no lloran"

Rule 3:
if
    generics = T
    NU = sg
then
    Germ. = Indefinite DET + NO sg
    Spn. = Indefinite DET + NO sg
"un hombre no llora"

The difference between rules 2 and 3 is that 3 is a statement about the fact that something having the property of being a man does not cry (is not allowed to cry).

5. Summary

A computational system must analyze and generate definiteness even for those cases where there is no unique referent to look for. Here I assume that definiteness is not the only function of definite determiners, and that it would be inadequate to use a brute force approach in looking for either uniqueness of referents nor uniqueness of roles as proposed by Bosch & Geurts 89. Definiteness as a semantic value of definite determiners is one of its possible values (the logical value of the quantifier "only"); more common values are distributional, collective quantification, or a surface phenomena within generic constructions, or in connection with the predicative value. Definiteness is unmarked when placed before proper names, and becomes an existential quantifier when accompanied by an abstract noun. NPs in directionals and locals usually have a pragmatic value (for more detail (Loebner 1985)).

Romance languages have a higher use of "definiteness" than germanic languages. While the latter use no article the former use them; while the latter use definite determiners the former use demonstratives.

The internal value, however, remains the same. Compare:

20. Freedom is essential
    vs. la libertad es esencial

21. Love does not exist
    vs. el amor no existe

Or is "love" and "freedom" more "definite" in Romance languages?

References

Alonso,A . 1961 : Estilistica y gramatica del articulo en espa?ol,en Estudios LInguisticos. Madrid

Barwise,J & R.Cooper: 1981. Generalised Quantifiers and Natural Language. Linguistics and Philosphy, 4. p.159-219

Benthem van ,J &A.ter Meulen (eds) Generalised Quantifiers. Dordrecht

Bosch,P. & B.Geurts 1989: Proccessing definite NPs. IBM. Lilog Project.

Gardenfors, P (ed) 1987: Generalised Quantifiers. Reidel.Dordrecht

Geurts, Bart 1985 : Generics, in .:Journal of Semantics. vol.4. p.247

Grimm, H.J 1987: Lexikon zum Artikelgebrauch. Leipzig

Keenan,E. & J.Stasvi 1986: A semantic characterisation of Natural Language Determiners in: Linguistics and Philosophy, 4 p.159-219

Loebner Sebastian 1985 : Definites. In.Journal of Semantics. vol.4.p.279

6

MacCawley,J.D 1981 : Everything that linguistis have always wanted to know about logic. Univ. Chicago Press

Peterson, Ph. 1979 : On the logic of few, many and most. Notre Dame Journal of Formal Logic 20. p. 155-179

Weinrich, H. 1976 : Sprache in Texten. Klett. Stuttgart

7