Lund University, Dept of linguistics Working Papers 36 (1990), 139-150

Reflexivization in Referent Grammar

Bengt Sigurd and Barbara Gawrońska-Werngren

Abstract

The interpretation of reflexive pronouns has been the topic of extensive linguistic research. This paper illustrates how different types of reflexive pronouns can be interpreted and generated within the framework of Referent Grammar, a grammar which can be run directly on the computer. The rules illustrated are to be used in the machine translation project SWETRA which translates between English, Russian, Swedish and German (Sigurd & Gawrońska-Werngren 1988). Some of the translation problems are also discussed.

INTRODUCTION

All languages include special markers which signal reference to the same referent as the subject refers to. Such pronominal markers are called reflexive pronouns. In English the word *himself* in the sentence *The detective hit himself* indicates that it is the same person as the subject who is hit, i.e. the detective. In the sentence *The detective hit him*, a different person must be referred to, a person who possibly was mentioned in a preceding sentence such as *A thief entered the room*. The Swedish equivalent to the first sentence is *Detektiven slog sig (själv)*. If the pronoun is to refer to somebody else a Swede would say *Detektiven slog honom*.

Possessive pronouns may also be used for reference to the subject (reflexively), as in *The detective hit his dog*, but English possessive pronouns are ambiguous between reflexive and non-reflexive uses. In our example it could have been the detective's dog or the thief's dog. In Swedish, however, one would have known from the shape of the pronoun if the dog belonged to the detective (*sin hund*) or to the thief (*hans hund*).

Reflexive pronouns are used more loosely as a kind of verb particle rather than an object np proper in some cases, as illustrated by French *se promener* (stroll; literally push oneself forward, which no French person thinks about). French has a number of reflexive verbs which from an English and a Swedish point of view could be called "pseudo-reflexive", e.g. *se passer* (happen, Swedish *hända*), *se baigner* (bathe, Swedish *bada*). The Swedish equivalent of English *move*, i.e. *röra sig* is a pseudo-reflexive from an English point of view. Similarly, there are often "frozen" or lexicalized expressions including reflexive pronouns which do not follow the main rules and do not refer to the subject. The following are some Swedish examples: pa sin tid (literally: in his time; in the past), i sinom tid (in due time). Similarly, Russian reflexive pronouns are used in "frozen" expressions like v svoe vremja (when the right moment comes), svojego roda (of its kind).

Reflexive pronouns are furthermore known to be the origin of the passive markers in some languages, e.g. the Nordic languages. A remnant of the reflexive pronoun (s) is seen in Swedish verbs, e.g. *målas* (be painted). In Old Icelandic the passive ended in *-sk* which is even more clearly reminiscent of the reflexive pronoun (*sik*). All this is very well known and reflexivization has been described in separate languages as well as from general linguistic and typological points of view (for a recent article on reflexivization in English with a comprehensive bibliography, see Zribi-Hertz 1989; for many insightful comments on the acceptability of different Nordic constructions, see Diderichsen 1939 and Wellander 1946; for typological aspects, see Shopen 1985:117-118).

This paper will not survey all the literature and the problems of reflexivization. The purpose of the paper is to indicate how reflexivization can be handled within the framework of Referent Grammar (RG), a functional generalized phrase structure grammar (Sigurd 1987). We will show that the existence of abstract objects called "referents" in the RG representations of noun phrases makes it easy to analyze and generate reflexive pronouns adequately. The meaning of a reflexive pronoun is simply the referent of the subject np, which is what grammarians of all times have realized. We will use English, Swedish and Russian for illustration.

THE NATURE OF REFLEXIVIZATION

It is clear from the facts of reflexivization that some constituents must "know" the subject of the sentence in order to chose the right pronoun – the pronoun which "reflects" the subject. This goes for the object np in *Detektiven målade sig* (The detective painted himself) as well as the possessive determiner of the object of the infinitive in Swedish: *Detektiven ville sparka sin hund* (The detective wanted to kick his (own) dog) and for the np (object) inside a prepositional phrase as in: *Detektiven undersökte mordet på sin fru* (The detective investigated the murder of his (own) wife).

The situation is, however, more complicated – that is why linguists have devoted so much time and space to reflexivization. In e.g. Detektiven undersökte miljonärens mord på sin fru (The detective investigated the millionaire's murder of his wife) the reflexive pronoun sin/his is ambiguous, but preferably refers to the millionaire. A Swede could try to disambiguate the sentence by saying Detektiven undersökte miljonärens mord på hans fru, but this does not work perfectly: hans (his) will now be ambiguous as in English, i.e. between the detective, the millionaire and some other person, e.g. the thief. This paper will not focus on all such acceptability questions, however.

MORPHOLOGY

As has been illustrated already it is necessary to distinguish between reflexive personal pronouns and reflexive possessive pronouns. The English 3rd person reflexive personal pronoun varies with the sex and number of the referent (*himself-herself-itself: themselves*). The equivalent Swedish reflexive personal pronoun does not vary with sex and number, but it is inflected for grammatical person, as in English: 1st person sg *mig*, 2nd sg *dig*, 3rd sg and pl *sig*.

English does not distinguish between reflexive and non-reflexive possessive pronouns, but the sex and number of the owner is shown: *his, her, its:their*. Swedish makes a distinction between reflexive and non-reflexive possessive pronouns, and the non-reflexive personal pronoun varies with sex, gender and number: *hans* (his), *hennes* (her), *dess* (its): *deras* (their). The Swedish possessive reflexive pronoun varies with the gender and number of the possessed item: *sin hund* (*hund* 'dog' is non-neuter), *sitt bord* (*bord* 'table' is neuter), *sina hundar* (plural).

Since English distinguishes morphologically between masculine, feminine and non-human personal reflexive pronouns it is possible to make some identifications which cannot be made in Swedish. In the somewhat strange sentence *The detective promised the girl to get her dog* it is natural to believe that the dog belongs to the girl – unless the detective is known to be a woman. But in the corresponding – equally strange – Swedish sentence *Detektiven lovade flickan att få sin hund*, it is not quite clear whose dog it is, but semantically it is reasonable to assume that the girl has asked for her dog and therefore that it is the girl's dog. The Swedish sentence *Detektiven lovade flickan att få hennes hund* seems to indicate that the detective was female. In the sentence *Detektiven lovade*

flickan att rasta sin hund the detective is naturally the subject of the infinitive and in that case sin must refer to the detective. The English equivalent *The detective promised the girl to walk her dog* is ambiguous, while the sentence *The detective promised the girl to walk his dog* is not. Thus, the different morphological features of the personal and reflexive pronouns allow different interpretations in Swedish and English.

In Russian there is also a distinction between possessive and non-possessive reflexive pronouns. The possessive reflexive pronoun *svoj* can be used for all grammatical persons – in 1st and 2nd person alternatively with possessive pronouns *moj* (my), *tvoj* (your), *naš* (our), *vaš* (your+pl). It is therefore possible to say both *ja vzjal moju knigu* (I took my+fem+ sg+ack book+ack) and *ja vzjal svoju knigu* (I took reflposs+fem+sg+ack book+ack). The possessive pronoun agrees with its head noun in number, gender and case – as shown above.

The non-possessive reflexive pronoun has two variants: a weak (clitic) and a strong (phonologically independent) one. Neither is inflected for person. In its weak, unstressed variant the reflexivity marker cliticizes onto the verb: it has the form *-sja* after consonants and *-s'* after vowels, e.g. *brit'sja* (shave+refl – 'to shave (oneself)', like Swedish *raka sig*), 1st person sg *ja brejus'*, 2nd person sg *ty breeš'sja*. In many verbs, the marker *-sja* may be interpreted as the object of the action expressed by the verb. In such cases the verb usually has a transitive, non-reflexive variant – like *brit'* 'to shave (somebody)'. But *-sja* verbs are often "pseudo-reflexives", where the particle has lost its function of a logical object – e.g. *smejat'sja* (to laugh), *soglasit'sja* (to agree). The reflexive particle *-sja* can also make a verb non-personal, i.e. unable of taking a nominative subject. Compare *rabotaet* (work+3sg) and *rabotaetsja* in the sentences: *on ne rabotaet* (he not work+2sg – 'he does not work') and *segodnja ne rabotaetsja* (today not work+refl – 'the work does not go on well today').

The strong variant of the non-possessive reflexive pronoun is se-bja/sebe/soboj – case inflected according to the same pattern as personal pronouns (cf. ja – I, genitive/accusative menja, dative/prepositional mne, instrumentalis mnoj), but having no nominative form, as it cannot function as sentential subject. The connections between the verb and the reflexive pronoun may be more or less close in Russian as in English and Swedish. Some verbs seem to be lexically subcategorized for the reflexive pronoun – they assume a certain meaning when combined with sebja and their valency frame changes. Compare *čuvstvovat'* – 'to feel (something)'

and čuvstvovať sebja – 'to feel (in a certain way)'. In ja čuvstvuju sebja ploxo (I feel+1sg reflpro bad – 'I feel bad') sebja cannot be replaced by an np having reference of its own: *ja čuvstvuju Ivana ploxo (*I feel Ivan badly) and cannot be omitted *ja čuvstvuju ploxo. The other word for 'to feel' – čuvstvovať has a different valency frame – it takes objective nps: ja čuvstvuju bol' (I feel pain).

In other cases, the pronoun *sebja* may be more or less freely replaced by a non-reflexive np, without any changes of the meaning or the valency of the verb – as in *rasskazyvat' o sebelo Marine* (to tell (something) about oneself/about Marina).

In the RG approach, weak non-possessive reflexive markers are treated as parts of the predicate. Their morphological status as well as the fact that they often occur in pseudo-reflexives makes it natural to interpret them as particles rather than referential expressions. Thus *brit'sja* and *smejat'sja* are both treated as intransitive verbs. In cases like *čuvstvovat'* the verb is lexically subcategorized for the reflexive pronoun (*sebja*) and an adverb as obligatory arguments when used in a specific meaning (here: 'feel in a certain way'). If the pronoun *sebja* is not a specific part of the lexical subcategorization of the verb, it receives a referential interpretation, i.e. the same referent number as the syntactically closest subject. We will eleborate the RG-treatment in more detail below.

REFLEXIVIZATION IN REFERENT GRAMMAR

Some typical syntactic positions of reflexive pronouns have been illustrated above. We will now examine the different cases and describe how they can be treated in Referent Grammar.

Personal (non-possessive, objective) reflexives 1. Object np. Ex: *The detective struck himself,* Swedish: *Detektiven slog sig (själv).*

As mentioned above an np must be sensitive to the subject of the sentence. Within the generative rules of Referent Grammar, directly written in the Definite Clause Grammar (Prolog) formalism, this can be achieved by inserting the representation of the subject np (Nps) into other constituents. The following (simplified) rule illustrates this:

sent(s(subj(np(Nps)),pred(P),obj(np(Npo)),adv(A))) -->
nps(Nps),vt(P),npo(Nps,Npo),adv(Nps,A).

This rule says that a sentence may consist of a subject noun phrase (nps), followed by a transitive verb (vt), another np (npo) and an adverbial phrase (adv). The variable Nps guarantees that the object and the adverbial phrase "know" the subject. What is important to know about the subject varies with the language as illustrated above – English requires knowledge about sex and number which Swedish does not. But most important is the identification of the referent of the subject which is done by assigning referent numbers in Referent Grammar. When used, the variable Nps has to be resolved in order to extract the following information: m(R,f(Sex,Gender,Number)), where R is the referent number of the subject referent and f(Sex,Gender,Number) is the bundle of features of the noun which is the head of the subject np. We disregard differences in person here. R may be: 1, 2, 3, etc. As can be seen, this knowledge about the subject np is percolated down into the object np and the adverb.

The following are simplified general rules for reflexive pronouns in English (enpo, erefl) and Swedish (snpo, srefl).

enpo(m(R,f(Sex,Gend,Numb)),R) --> erefl(f(Sex,Gender,Numb)).
erefl(f(ma,_,sg)) --> [himself].
erefl(f(fe,_,sg)) --> [herself].
erefl(f(thing,_,sg)) --> [itself].

snpo(m(R,f(Sex,Gend,Numb)),R) --> srefl(f(Sex,Gend,Numb)).
srefl(f(_,_,_) --> [sig].

These rules are just a formalization of what has been said above. The sex, gender and number of the head noun is of no interest in Swedish, which we have made explicit by writing the rule in this way. By using the referent variable (R) as the representation of the functional representation (meaning) of reflexive pronouns we indicate that they do not mean any-thing. They just reflect the identity of the subject referent.

The sentence *The detective hit himself* would receive the following simplified representation in RG: s(subj(1,detective),pred(hit),obj(1)), where 1 is the referent number.

It is now necessary to say something about the non-reflexive pronouns in Referent Grammar. Their characteristic feature is that they refer to something, but only exceptionally to the subject referent. In order to capture this, we simply represent their meanings with the English pronouns. Their reference is not governed by syntactic rules, but special rules for the identification of discourse referents have to apply. The rules needed are only the following:

 $enpo(m(R,him)) \rightarrow [him].$ $enpo(m(R,her)) \rightarrow [her].$ $snpo(m(R,him)) \rightarrow [honom].$ $snpo(m(R,her)) \rightarrow [henne].$

The simplified functional representation of: *The detective hit him* will then be: s(subj(1,detective),pred(hit),obj(2,him)). The system will assign a new referent number to the noun phrase *him*, but discourse rules are needed to identify this referent number with a previous referent number. It is clear that it cannot be the same as the referent number of the detective this time. But there are cases where it might be possible to identify the referent of a pronoun with the referent of the subject, although the reflexive pronouns should have been used in that case according to the main rule.

Languages which may have the object before the verb are known to have certain restrictions. Thus Swedish allows the "fronting" of the object as illustrated in *Pojken/Honom gav flickan boken* (Literally: The boy/Him the girl gave the book). But not even Swedish allows a reflexive objective pronoun in this position: **Sig gav flickan boken* (Literally: Herself the girl gave the book). It is however possible to use a reflexive possessive pronoun in a fronted noun phrase, as in *Sin hund gav flickan allt*. (Literally: Her dog the girl gave everything). This fact can simply be taken care of in RG by stating that a preposed npo cannot be just a reflexive pronoun.

2. Object in a prepositional phrase. Ex: The detective looked upon himself (in the mirror), Swedish: Detektiven såg på sig (i spegeln).

In these cases the reflexive pronoun is within a prepositional phrase. The prepositional phrases also carry a variable (Nps) containing the referent

number of the subject and the grammatical features of the head noun. It is therefore possible to write the following rule, in which the information is percolated down to the object np (npo).

 $pp(Nps,pp(P,Np)) \rightarrow p(P),npo(Nps,Np).$

If the npo is a reflexive pronoun, the meaning representation of the npo (Np) will simply be the subject referent, as shown above.

There are many constructions of this type in which there is an adverbial or preposition-like particle in Swedish. Such constructions often take on special meanings and stress the preposition (adverbial particle), as illustrated by: *Flickan ordnade till sig* (Literally: The girl ordered to herself – The girl put herself in order), *Detektiven lockade till sig hunden* (Literally: The detective called to himself the dog).

3. Attributive prepositional phrases. Ex: The detective investigated the murder of his wife, Swedish: Detektiven undersökte mordet på sin fru.

The thing to do here is to percolate the subject referent down into the postnominal prepositional phrase as indicated by the following additional np-rule using the rules given above.

npo(Nps,np(Np,Pp)) --> np(Nps,Np),pp(Nps,Pp).

Possessive reflexives

Ex. The detective investigated the millionaire's murder of his wife, Swedish: Detektiven undersökte miljonärens mord på sin fru.

As observed above the problem here is that the wife might belong to the millionaire or the detective in this case where the head noun is a derived verbal noun (vnoun). We may illustrate the different solutions by the following simplified rule.

npo(Nps,nsent(subj(G),pred(N),obj(Pp)))-->
gen(G),n(N),pp(X,Pp),{vnoun(N),(G/= [],X=G; G=[],X=Nps)}.

This means that if the noun (N) is a verbal noun, the construction containing this noun, an (optional) genitive modifier (gen(G)) and a prepositional

phrase (Pp) is treated as a "small clause" (nominal sentence, nsent). The verbal noun N (in our example the word *murder*) is the predicate of the sentence. If there is a non-empty genitive modifier (G/=[]), then the referent of this modifier (G) might be interpreted as the subject of the nsent. In such a case, G is the first available subject referent to be per-colated into the prepositional phrase (pp) and, subsequently, into the npo inside the pp. The subcondition G/=[],X=G means: if the genitive modifier is not empty, the possessive pronoun inside the prepositional phrase may co-refer with G. This is the preferred solution in *miljonärens mord på sin fru: sin* and *miljonärens* receive the same referent number. When no genitive modifier is present (G=[]), the first available subject referent is Nps, as in our example corresponding to the referent of the np *detektiven*.

In infinitive clauses

Infinitives with auxiliary or aspectual verbs. The detective started to hit his dog, Swedish: Detektiven började slå sin hund

What is needed here is a way of carrying the subject referent down into the infinitive clause, and this may be done by having an additional argument in the infinitive which is set at the value of the subject, as illustrated by the following simplified rules.

sent(s(subj(Nps),pred(V),obj(F))) --> nps(Nps),aspv(V),isunt(Nps,F). isunt(Nps,s(subj(Nps),pred(V),obj(Npo))) --> trinf(V),npo(Nps,Npo).

The rules mentioned above for npo will carry the identity of the subject down properly.

As is well known languages may have long series of verbs if the modal infinitives are concatenated as is illustrated by the following case: *The detective wanted to start to try to hit his dog/Detektiven ville börja försöka slå sin hund.* Such examples are taken care of by rules percolating the subject referent down to the lowest verb.

Infinitives with control verbs require that the subject of the embedded infinitive is inserted so that the other constituents of the infinitives can "know" who is the subject. If sentences are ambiguous, several solutions must be offered. But we will not show how this can be done.

147

Participial and infinitival clause abbreviations

There are many types of clause abbreviations, but we will only illustrate a few in order to show how the subject must percolate down into such constructions in order to handle reflexive pronouns.

Present participles with or without a conjunction (subjunction), such as *while* or *without*, are very common in English, but also occur in Swedish. But in most cases such constructions in English have to be changed into a full clause in translatation (by transfer rules) which we will not go into in this paper. They serve just as adverbial clauses and are treated as a kind of adverbial in Referent Grammar. Examples: (*While*) washing his hands, he... (Swedish: medan han tvättade sina händer...).

The following rule illustrates how the subject is percolated down into an abbreviated clause consisting of a conjunction, a transitive present participle and its object.

adv(Nps,C,s(subj(Nps),pred(V),obj(Npo))) -->
conj(C),vtprpart(V),npo(Nps,Npo).

Note the variable Nps which recurs in the object noun phrase (npo) and guarantees that the grammar will generate the proper reflexive pronoun using the reflexive rules illustrated above. The value of the conjunction (C) is placed before the f-representation of the subordinated clause. Since the value of the present participle (V) will be placed as the predicate we will get representations with non-finite values as predicate and some languages will prefer to make such clauses into finite clauses. This means that a clause such as *when repairing his car*, would be rendered as *when he repaired his car*. This change means inserting the proper tense, which might not be all that easy.

Phrases like satisfied with himself/(with) all his money in the bank (he left the town) are also seen as abbreviated clauses which function as adverbials in the functional representation of the main clause. Again the value of the subject must be percolated down into the components of the abbreviated clause in order to ascertain the right reflexive pronoun.

Infinitival clause abbreviation may be illustrated by the example: *The detective called in order to inform his wife*. The phrase: *in order to inform his wife* is classified as an adverbial phrase in Referent Grammar and it is then possible to percolate the knowledge of the subject down into this infinitival clause by rules similar to those illustrated.

REFLEXIVE PRONOUNS AS A PROBLEM IN MACHINE TRANSLATION

The differences between Swedish and English reflexive pronouns discussed above illustrate the problems which might appear when translating between languages. When translating from Swedish into English the problems are minor. The general Swedish pronoun *sig* has to be rendered by *himself*, *herself*, *itself*, *themselves* depending on the features of the head of the subject. These features are available in the representations and the rules illustrated above can be used. Similarly the Swedish possessive reflexive pronouns *sin*, *sitt*, *sina* which vary with the gender and number of the head noun of the np should be rendered by *his*, *her*, *its*, *their* in English. This is little problem since the gender features of the head noun are known in the rules. Referent Grammar has no problems in handling the common differences in agreement due to different morphological categorization.

The situation is worse when translating from English into Swedish because Swedish requires disambiguation of the English possessive pronouns: his, her, its, their. In sentences such as The detective liked his dog, there is a choice between Swedish Detektiven tyckte om sin hund and Detektiven tyckte om hans hund. How does one know? Only the context can provide an answer. If it is known that the detective was a woman, it is clear that his is not reflexive. If it is known that the detective has a dog and there is no other dogowner mentioned the case is clear. It is also clear that the situation must contain clues for the reader if he is to understand the pronoun correctly. We will not outline a solution to the whole problem here, but only indicate that a representation of the contents of the text (discourse) is necessary, above all the contents of the immediate context. We refer the reader to Gawrońska-Werngren (1990), where similar problems are treated in relation to the problem of inserting proper definite and indefinite articles when translating from Russian into Swedish.

Russian possessive personal pronouns differ distributionally from their Swedish counterparts, which sometimes causes problems in translation. In Swedish, the possessive reflexive pronoun must be used when talking about things/persons which the subject has a very close relationship to, e.g. han träffade sin bror/sin granne (he met his brother/his neighbour). There is no such need in Russian: on vstretil brata/soseda (lit. he met brother/ neighbour). Such problems are solved in SWETRA by providing

149

the nouns denoting "inalienable property" (parts of the body, relatives etc.) a special feature (iprop) in the lexicon and inserting the appropriate reflexive pronoun on the transfer stage.

Another translation problem is caused by the fact that Russian possessive reflexive pronouns often have to be translated as of (one's) own or Swedish egen/eget (own) when stressed, as in u menja svoja mašina (at me own car – I have a car of my own). Difficulties like these can to a certain extent be handled on the transfer stage, e.g. by rules like: if a Russian possessive reflexive pronoun is found in a possessive construction u X Y, like u menja svoja mašina, it should be treated as emphasized and translated into Swedish/English as egen/eget respectively phrases with own. As can be seen, reflexivization poses many interesting problems to linguistic theory and to machine translation.

REFERENCES

- Diderichsen, P. 1939. 'Om Pronominerne sig og sin'. Acta Philologica Scandinavica 13.
- Gawrońska-Werngren, B. 1990. 'Translation great problem on the problem of inserting articles when translating from Russian into Swedish'. Proc Coling 1990. Helsinki (to appear)
- Shopen, T. (ed). 1985. Language typology and syntactic description, vol.1: Clause structure. Cambridge University Press.
- Sigurd, B. 1987. Referent Grammar. A generalized phrase structure grammar with built-in referents. *Studia Linguistica* 41, 115-35.
- Sigurd, B. & B. Gawrońska-Werngren, 1988. 'The Potential of SWETRA - A multilanguage MT system'. *Computers and Translation* 3, 237-50.
- Wellander, E. 1946 (3rd ed). Riktig svenska. Stockholm: Svenska Bokförlaget.
- Zribi-Herz, A. 1989. 'Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse'. *Language* 65:4, 695-727.

Lund University, Dept. of Linguistics Working Papers 36 (1990), 151-162

Understanding Coordination by Means of Prolog

Bengt Sigurd and Per Warter

Abstract

Coordinated structures are very frequent in texts, but generally grammatical theories have little to say on the subject. This paper describes the different types of coordination and shows how they can be analyzed - and understood - using Prolog. The work has been carried out within the automatic translation project SWETRA (Sigurd & Gawrońska-Werngren 1988).

INTRODUCTION

Almost all sentence constituents can be coordinated, and coordination is very common in texts, a fact which is not reflected duly in the treatment of coordination in traditional or modern grammar. The following sentence illustrates coordination of phrases and words: *The rich, nice and beautiful* (adjectives) *boys and girls* (nouns) *who live in Florida and in New York* (prep phrases) *can and do* (auxiliaries) *give or throw* (main verbs) *parties when and where they want* (subjunctions)

over and over. (lexicalized coordinated phrase)

Coordination of clauses may be illustrated by the following sentence: The boy, whom the parents loved and (whom) the neighbours hated, took the car and drove to New York in order to work and (to) have fun. This sentence illustrates coordinated relative clauses, coordinated main clauses, and coordinated infinitives, the latter with optional to. The sequence: took the car and drove to New York can alternatively be considered as a case of coordinated verb phrases.

There are at least some words which can hardly be coordinated, however, e.g. articles: ?A or the boy may come. Similarly, some pronouns can hardly be coordinated with anything, e.g. who: ?The boy, who or that