

The treatment of Scope and Negation in Rosetta *

Elly van Munster

Philips Research Laboratories
P.O. Box 80.000, 5600 JA Eindhoven, The Netherlands

Abstract

This paper deals with the treatment of Scope and, in particular, Negation in Rosetta, a Machine Translation system which translates between Dutch, English and Spanish (Spanish only as a target language). It will be argued that the SOV- versus SVO-character of a language has important consequences for its possibilities of reflecting scope through word order. A description will be given of the problems that arise translating from one type of language (the SOV-language Dutch) to the other (the SVO-languages English and Spanish). The extent to which these problems can be solved will be outlined.

The paper has been divided into two main sections. In section one the phenomena are described linguistically, in section two a general idea is given of how these phenomena are dealt with in Rosetta.

1 Linguistic phenomena

1.1 Expression of Scope

Scope bearing elements can be divided into two classes:
1) NEG, containing both the adverbs *niet*(Dutch)/*not*(English)/*no*(Spanish) and quantifiers with morphologically incorporated negation, such as *niets*(D.)/*nothing*(E.)/*nada*(S.).

2) NPs and adverbials containing a quantifier (from now on Q-elements)¹, like *veel kinderen*(D.)/*many children*(E.)/*muchos niños*(S.), *een vis*(D.)/*a fish*(E.)/*un pez*(S.), *vaak*(D.)/*often*(E.)/*muchas veces*(S.), *in sommige gevallen*(D.)/*in some cases*(E.)/*en algunos casos*(S.), etc.

The question I am concerned with is how the scope order of Q-elements and NEG can be determined. In a Montague Grammar of the PTQ type, (1) would have two interpretations, one with *Every man* having wide scope and one with *two women* having wide scope /Montague 1973/, /Dowty 1981/:

(1) Every man loves two women.

*This paper condenses part of the content of the author's thesis, /Van Munster 1985/, extended with the results of further research. This research was partially sponsored by Nehem (Nederlandse Herstructureringsmaatschappij).

¹In this paper I limit myself to Q-NPs, although a parallel can be drawn between Q-NPs and Q-Adverbials (including Q-PPs). Furthermore only extensional contexts are taken into consideration; intensional contexts are ignored.

However, the interpretation with *Every man* having wide scope is far more natural. Therefore, in accordance with Jackendoff's principle, /Jackendoff 1972/, I make the simplifying assumption that *surface order* in principle represents the most *plausible scope order* of Q-elements and NEG in the following sense:

"A Q/NEG element has scope over the Q/NEG elements on its right and is itself inside the scope of the Q/NEG elements on its left".

Starting from this principle implies that other important factors are not taken into consideration, such as:

- *Intonation*, which is not visible in a written text. Therefore sentences are considered under neutral stress and intonation.
- *Context*. Currently only isolated sentences are taken into consideration.

Essential for translation is that, even if we assume that (1) is ambiguous between two scope readings, the sentence in the target language will have this same ambiguity as long as the Q-elements have the same surface order as in the source language. Furthermore, both sentences will have the same 'most plausible reading'.

1.2 SVO versus SOV

Rosetta translates between two types of languages, namely the SOV-type (Dutch) and the SVO-type (Spanish and English). This SOV- versus SVO-character has important consequences for the expression of scope. I claim that in both types of languages the position of NEG is as close to the left-hand side of the verb as possible: it only precedes possible Q-elements that are within the scope of NEG /Van Munster 1985/.

Consider the following scheme:

Dutch:	S	O-(O)	NEG-V
English/Spanish:	S	NEG-V	O-(O)

In an SOV-language the verb (in basic position) is in sentence-final position, while in an SVO-language the verb is in second position. Consequently, in an SVO-language only two elements (one in subject position and one in 'shift-position', i.e. the position to the left of the subject)² can precede NEG; the objects are to the right of NEG in basic position. In an SOV-language like Dutch, however, the objects are to the left of NEG in basic position. In principle there is no restriction to the number of elements that can appear to the left of NEG.

In general it can be said that, especially if the sentence

contains a NEG, an SOV-language is more 'suited' to express scope through word order than is an SVO-language.

This basic difference between Dutch and Spanish/English can cause problems when translating from one type of language into the other. Consider e.g. (2), where (2)a cannot be translated into (2)b since the relative order of NEG and the Q-NP is not the same:

- (2) a De kinderen aten veel snoepjes niet op.
 'The-children-ate-many-sweets-not'
 b The children didn't eat many sweets.

In the English sentence the object has to be topicalized in order to get the correct scope relations:

- (2) c Many sweets the children didn't eat.

1.3 Subdivision of Quantifiers

I argue that the following subdivision of NPs can be made:

1A) Q-NPs sensitive to scope, i.e. the surface order of Q-NPs and NEG is crucial for the interpretation. (Dutch: *iemand* ('everybody'), *een N* ('a N'), *veel N* ('many N'), *alle N* ('all N'), *twee N* ('two N'); English: *many N*, *three*; Spanish: *muchos* ('many'), *dos N* ('two N'), etc.).

E.g. (3)a does not mean the same as (3)b:

- (3) a Niet veel mensen houden van vis.
 'Not-many-people-like-fish'
 b Veel mensen houden niet van vis.
 'Many-people-like-not-fish'

1B) Q-NPs not sensitive to scope, i.e. the surface order of Q-NPs and NEG is not crucial for the interpretation; these NPs always have wide scope, irrespective of their position. (e.g. *sommige N* ('some'), *most N*, *alguien* ('someone'), *something*, etc.) Since these NPs do contain a quantifier, however, there is a strong preference for a surface order which reflects the scope. Therefore, (4)b is a much more natural word order than (4)a, although both sentences have in fact the same meaning. (NB. For some speakers (4)a is even out).

- (4) a Niemand gelooft sommige opmerkingen.
 'Nobody-believes-some-remarks'
 b Sommige opmerkingen gelooft niemand.
 'Some-remarks-believes-nobody'

2) definite NPs (e.g. *Jan* ('John'), *het boek* ('the book'), *the many linguists*, etc.). Surface order is irrelevant for scope-interpretation. If in (4)a and (4)b

²Another term would be 'topicalization-position'. However, this term can cause confusion since in Rosetta a distinction is made between 'scope-shift' (treated in this paper; the sentence still has a neutral intonation), and 'topicalization' (the sentence has a non-neutral intonation; not the surface order but the original position of the topicalized Q-element reflects the scope.) Both types of shift go to *shift-position*. Scope-shift is a transformation, Topicalization a rule. (For these terms see section 2.1).

sommige is replaced by the definite *deze*, (a) and (b) not only have the same meaning but there is no difference in naturalness either. Definite NPs have the feature [-Q].

1A and 1B together are the NPs traditionally called quantifiers. I will indicate the two types with resp. the features [+S] and [-S].

Note that the subdivision of Q-NPs is *language specific*, e.g. *iemand* in Dutch is scope-sensitive, while the Spanish and English equivalents (resp. *alguien* and *someone*) are not.

Spanish and English seem to have much more [-S] elements than Dutch does, which, as I claim, relates to the fact that they are less suited to express scope through word order (see section 1.2).

1.4 Two approaches to scope

In TG-oriented theories a distinction is often made between different types of negation: S-negation and VP-negation (e.g./Jackendoff 1972/, /Lasnik 1972/, a.o.), with the special addition of TVP- and V-negation for Dutch (e.g./Hoekstra 1985/) ³. To illustrate briefly what is meant by the different types of negation, let me give an example of each:

- (5) a *Jan heeft geen boek gelezen.*
 'John-has-no-book-read'
 (S-negation)
 b *Niet veel mensen hebben een boek gelezen.*
 'Not-many-people-have-a-book-read'
 (S-negation)
 c *Iemand heeft niet gereisd.*
 'Somebody-has-not-travelled'
 (VP-negation)
 d *Veel mensen hebben geen boek gelezen.*
 'Many-people-have-no-book-read'
 (TVP-negation)
 e *Jan heeft veel mensen iets niet verteld.* ⁴
 'John-has-many-people-something-not-told'
 (V-negation)

It is assumed that the constituents to the right of NEG, including the verb, are within the scope of NEG. In (a) NEG follows the subject, but since *Jan* is definite, S-negation is equivalent to VP-negation: as a general rule the position of NEG is *after* a definite (unless the sentence is contrastive). Both (a) and (b), however, can be paraphrased by 'it is not the case that ...' which is a proof of S-negation /Jackendoff 1972/. In (c) NEG has scope over the VP containing an intransitive verb, in (d) over the VP containing a transitive verb plus direct object and in (e) merely over the transitive verb. In logical terms, however, these sentences merely differ in the relative scope order of NEG and Q-elements (i.e. of scope operators). In a semantic, Montague-like theory a verb (unless it is a modal) is not a scope

³For a treatment of scope-ambiguity in TG-framework see /May 1977/.

⁴In the sample sentences the perfect tense is used since this renders a word order with the main verb in basic, i.e. sentence final, position.

operator and as such not relevant for the scope interpretation of the sentence. The fact that Dutch, but not Spanish/English, has TVP- and V-negation is a logical consequence of the fact that NEG can appear in more surface order positions in Dutch (being an SOV-language).

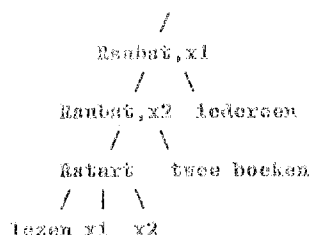
In Rosetta scope in a sentence is expressed in so-called *syntactic derivation trees* (see section 2.1).

2 Scope in Rosetta

2.1 The Rosetta Framework

In Rosetta *Compositional Grammars* of the Montague type are used. This means that sentences are built up starting from basic expressions by applying syntactic rules which prescribe how bigger expressions can be constructed from smaller ones. The grammars have to obey the *Compositionality Principle*: every rule and every basic expression has a well-defined meaning. This derivation process can be shown in a so-called *syntactic derivation tree*. Consider e.g. the strongly simplified syntactic derivation tree of (6)a which contains two Q-NPs.

(6) iedereen leest twee boeken.



The moment of substitution of a Q-NP indicates its scope-domain. In (6) *iedereen* is substituted later than *twee boeken*, so it has wider scope. In other words, for the most plausible reading of a sentence we want the Q-elements to be substituted from right to left in generation (*right-left generation*). Since the Substitution rules apply freely in principle, it is essential that the conditions on the applicability of Substitution- and Negation rules force this order, basically in the following way:

(a) An argument-substitution rule ($R_{subst,x}$) only applies if there are no free variables (from now on VARs), to the right of the variable to be substituted.

(b) A negation rule (R_{neg}) only applies if there are no free VARs to the right of the position where NEG is inserted.

For (6) this means that the output of the rules is as follows (details omitted):

Rstart:	x1	x2	lezen
Rsubst,x2:	x1	twee boeken	leest ^b
Rsubst,x1:	iedereen	twee boeken	leest

(A 'Verb-2-transformation' puts the verb in second position: - iedereen leest twee boeken)

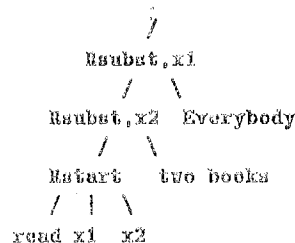
^bBetween Rstart and Rsubst,x the Time-rules apply.

Note that $R_{subst,x1}$ cannot apply before $R_{subst,x2}$ under requirement (a) stated above, since x2, a free VAR to the right of x1, has not been substituted for yet. In other words, $R_{subst,x1}$ is blocked.

An advantage of the derivation-tree method is that these trees can represent the scope order of Q-elements and NEG in a sentence in a natural way (as they can be assigned an interpretation in terms of logical expressions).

Translation is done by means of *isomorphic grammars*, which means that the grammars of the languages concerned are attuned to each other in the following way: for each basic expression in one language there must be at least one corresponding expression in the other language with the same meaning. For each syntactic rule in one language there must be at least one corresponding syntactic rule in the other language with the same meaning operation. Two sentences are a translation of each other if they are derived from corresponding basic expressions by application of corresponding rules. Scope can be maintained in translation if in SL and TL the Substitution- and Negation rules are applied in the same order. Consider now the English derivation tree of (7), corresponding to the Dutch one:

(7) Everybody reads two books.



Rstart:	x1	read	x2
Rsubst,x2:	x1	reads	two books
Rsubst,x1:	Everybody	reads	two books

In this example there is no problem making an isomorphic derivation for English (or Spanish), because in the English translation the order of Q-NPs is the same as in Dutch.

In Rosetta a distinction is made between *rules*, which are meaningful and relevant for translation, and *transformations*, which are language-specific, meaningless and not relevant for translation. Since in the derivation tree only the *rules* are represented, the corresponding trees have exactly the same geometry.

If the sentence contains a negation, this negation is treated at the *projection path*⁶ on sentence level, i.e. not constituent-internal, wherever possible.⁷

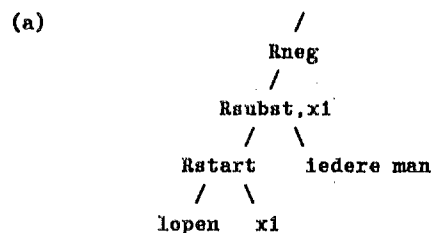
e.g. in (8) the position of NEG is syntactically inside the NP but will be put in this position (generatively) by means of a transformation. The derivation tree is:

⁶For this notion see /Appelo et al./ et al.

⁷To give an example of a sentence where the NEG is constituent (namely ADV-) internal, in (1) *not* has only scope over *without a reason*, since *someone*, (which is [-S]), and *not anyone* follows:

(1) Not without a reason I punished *someone*.

- (8) a Niet iedere man loopt.
'Not-every-man-walks'



Rneg applies at sentence-level.⁸

(For more theoretical details about the Rosetta framework see /Appelo 1987 et al./).

2.2 Translating Scope

Now, there may be various reasons why the right-left substitution order causes problems, both within one language and in translating from one language to another. In the subsections 1 and 2 the problems will be sketched, in 3 a general strategy for a solution in Rosetta will be given.

2.2.1 Switch of arguments

Problems within one language arise if the arguments have been switched with respect to the order of the verbpattern (i.e. the argument structure of the verb), in order to express the correct scope relations in the sentence. Consider e.g. (9):

- (9) Veel boeken leest iedereen.
'Many-books-reads-everybody'

Recall that in analysis the Q-arguments are substituted from left to right (cf. section 2.2), i.e. *veel boeken* (= x2) before *iedereen* (= x1). Now, the output of the generative rules is as follows:

Rstart: x1 x2 lezen
Rsubst,x1: (blocked)

Rsubst,x1 has to apply first but is blocked since there is a free VAR (x2) to the right.

This type of switch also occurs in translating from one language into the other, namely if the verb in the TL has a different order of arguments than the verb in the SL. Consider e.g. the following verbpatterns:

Spanish: x1 dar x2 x3
Dutch: x1 x3 x2 geven

Again assuming that surface-order reflects scope order, (10)a and (10)b are not a correct translation of each other:

- (10) a Jan geeft iedereen een boek.
'John-gives-everybody-a(='some')-book'
b Juan da un libro a todo el mundo.
'John-gives-a('certain')-book-to-everybody'

The order of Q-NPs in the Spanish sentence has to be switched somehow.

⁸NEG is introduced syncategorematically, although it could have been a basic expression as well.

2.2.2 SOV/SVO problems

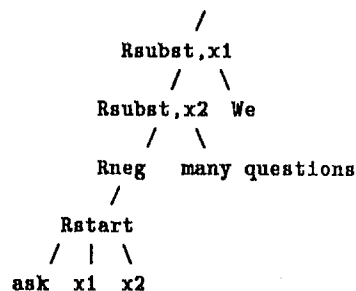
If a sentence containing a NEG-element has to be translated from an SOV-language (like Dutch) into an SVO-language (like English/Spanish) problems may arise. Recall that the position of NEG is closely related to the position of the verb (cf. scheme in section 2). In principle no problems arise if NEG does not follow a Q-object in Dutch. (11)a and (12)a can simply be translated into (11)b and (12)b respectively:

- (11) a Niet iedereen komt.
b Not everybody comes.
(12) a Veel mensen krijgen geen kado.
b Many people don't get a present.

However, as I explained in section 1.2, in Dutch (an SOV-language), NEG may occur to the right of a non-topicalized Q-object, as in (13):

- (13) Wij stelden veel vragen niet.
'We-asked-many-questions-not'

The corresponding syntactic derivation tree is as follows (since the Dutch and English trees are isomorphic, I only give the English or target one):



Schematically, the output of the rules is:
Dutch:

Rstart:	x1	x2		stellen
Rneg:	x1	x2	niet	stellen
Rsubst,x2:	x1	veel vragen	niet	stellen
Rsubst,x1:	Wij	veel vragen	niet	stellen

English:

Rstart:	x1	ask	x2
Rneg:	(blocked)		

Since x2 is to the right of the verb (and thus of the NEG-position), *Rneg* is blocked. Note that this blocking is justified: without blocking the result of applying the English rules would be (14), which is not a correct translation of (13):

- (14) We didn't ask many questions.

In other words, the wrong output is blocked but how can a correct translation be obtained?

