Translation Relations in Semantic Syntax

Pieter A.M. Seuren

Institute of General Linguistics
PO Box 9103
6500 HD Nijmegen - Holland
email: pseuren@vms.uci.kun.nl

subject keywords: Generative Grammar
Sentence Generation
Semantic Analysis
Machine Translation

Summary

SEMANTIC SYNTAX (SeSyn) is a direct continuation of work done in the '60s and '70s under the name of GENERATIVE SEMANTICS. The main component of SeSyn is a rule system (the syntax) which transforms the Semantic Analysis (SA) of any given sentence into a Surface Structure (SS) of that sentence. SAs are formulated in a higher order Predicate Calculus, according to a small set of context-free Formation Rules and a Lexicon. SA-trees have simple branchings and deep S-embeddings. They are input to the Transformation Rules, which deliver a corresponding SS. The Transformation Rules fall into two classes, those of the Cycle, and those of the Postcycle. The former apply cyclically, starting with the most deeply embedded S and ending with the top-S. They are mostly lexicon-driven: predicates are lexically marked for the cyclic rules they induce. The largely structure-driven postcyclic rules apply in linear order as defined by the grammar. Morphology is largely, and phonology totally, left out of account. It is claimed that the fully implementable theory of SeSyn (significant parts have been implemented by Henk Schotel) makes for a subtler and more precise coverage of the facts of the languages treated than any other grammar system on the market. It moreover appears that language-specific differences amount largely to different postcyclic rule orderings, to different parameter settings in otherwise identical rules, to different lexical inductions of cyclic rules, or to different positions in the formation rules. When more grammars of different languages are available a parametrization of the general theory should be feasible.

SA-trees are not entirely universal: they are to some extent still language-specific. They are, moreover, to some extent semantically redundant. They are language-specific in so far as different structures express identical meanings (eg. *he could have eaten* vs Dutch *hij had kunnen eten*). The structural differences must be manifest at SA-level if the syntax is to remain non-arbitrary. They are redundant in so far as scope sometimes has to be assigned where it is semantically irrelevant. For the *grammar* to work, adverbials and the tenses, for example, require precise scope assignments with respect to each other. For the *semantics*, however, it is sometimes indifferent which comes first and which last. The paper illustrates how, in principle, structurally different translation equivalents between English and Dutch can be handled in a principled way. It must be seen as a small part of a large project extended over many years.

1. Introductory

Semantic Syntax develops the components **SA**, **P**, **G** and **SS** in the following configuration:

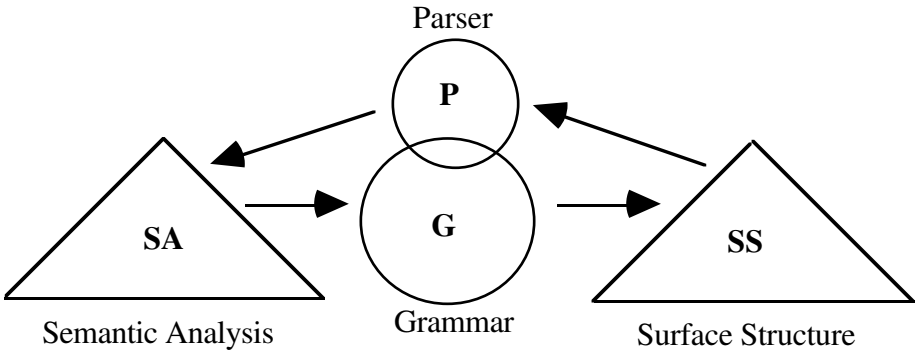


Figure 1

SA is subject to configurational restrictions described in the (context-free) Formation Rules. **G** is generative and transforms SA-trees to SS-trees in two stages, the **Cycle** and the **Postcycle**. **P** reduces SS-trees to SA-trees. **G** and **P** are related but not identical: **P** is strongly lexically driven and takes shortcuts through **G**, together with any **G**-independent rules and principles found useful.

Lexical meanings are not analysed cognitively, but cognitively oriented work in lexical meaning analysis and semantic selection and planning processes should link up naturally with the SA-trees. SA-structures, moreover, specify tenses, modalities, mood, logical operators (e.g. negation, quantifiers), and in general all material that takes scope over the lexical nuclear structure of a sentence. What Semantic Syntax does is make sure that all this semantic information gets properly expressed in well-formed surface structures of the language concerned. This purely grammatical aspect of sentence generation is often treated somewhat dismissively in the literature.[1] Yet without a proper grammatical machinery the very idea of sentence or text generation makes no sense. And such machinery has so far not been made available. An adequate machinery for the generation of well-formed surface structures from a semantic input simply means a few pieces put into place in the total jigsaw puzzle of natural language processing.

From a formal semantic point of view, SAs behave in a much more orderly fashion than SSs: they are formulated in an (n^{th} order) Predicate Calculus in VSO-format. But they are not good enough yet for semantic computation, i.e. the mapping onto a pre-existing **Discourse Domain** (**DD**) via an **Incremental Calculus**. Certain aspects of SA-trees are irrelevant for the incremental calculus: SA-trees are *semantically redundant*, and hence also to some extent *language-specific*, due to the requirements of the syntax of each language. A level of universal semantic analysis, or **USA**, at which the various meaning equivalents are bartered, is therefore useful in any principled account of translation relations.

The weeding out of language-specific elements from SAs to produce USAs involves at least three steps: (a) logical scope must be eliminated wherever it is irrelevant, (b) bits of SA-structure expressing a single meaning may have to be taken together or a single SA-element may have to be expanded, and (c) the argument structure of lexical predicates (the lexical 'nucleus')

must to some extent be neutralized. USAs will thus be heavily underdetermined from a syntactic point of view. The point is that, given some USA, the grammar and the lexicon of each language can be plugged in and supply the missing syntactic information on the basis of both the formation rules and the syntactic information supplied with the lexical items of the language.

This position gives rise to the more integrated configuration of fig.2.

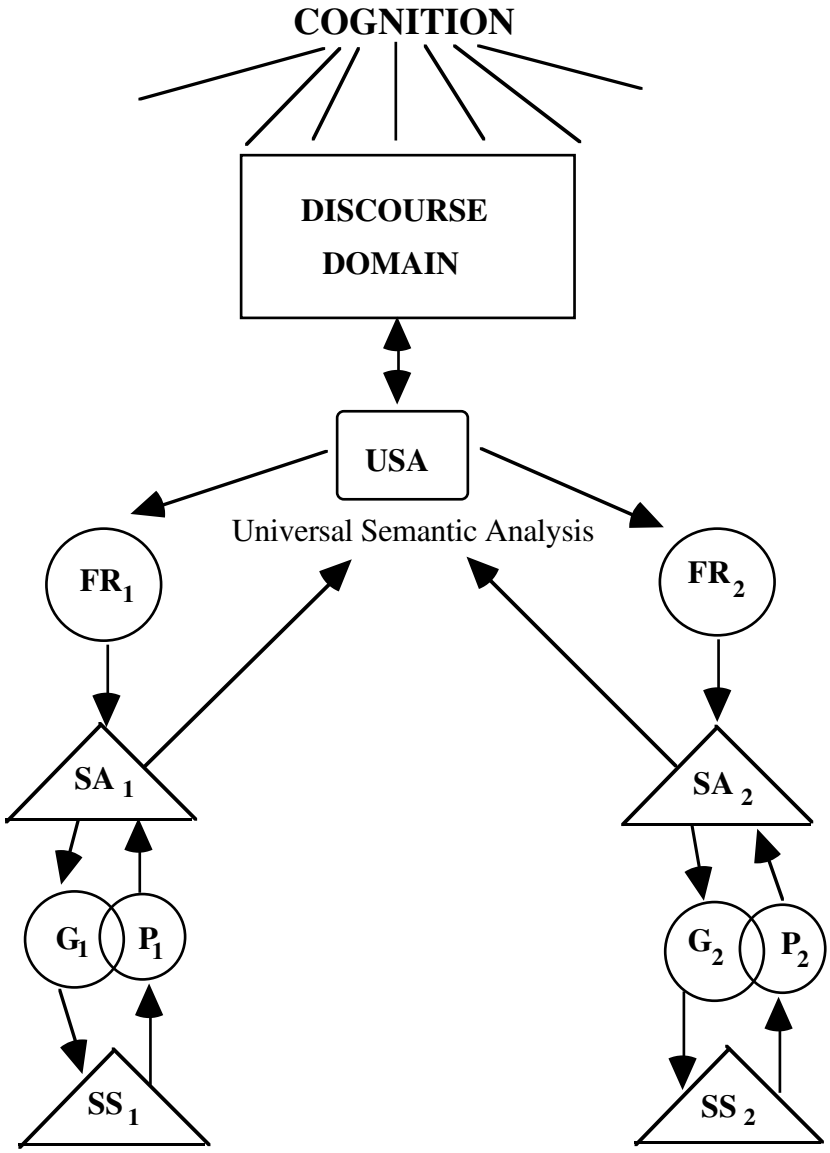


Figure 2

2. The SAs of English and Dutch

Here follows a simplified fragment of the Formation Rules for English and Dutch. Some amount of lexicon is also added. Note that the Formation Rules are context-free rewrite rules. No NP-syntax is provided. The cyclic and postcyclic transformation rules are not gone into here, as they are irrelevant for translation relations. For an introduction to the generation process (the $\mathbf{G_i}$s in Figure 2) see the paper by Schotel in this volume[2].

Some Formation Rules for English

(1)a	S''^{ADV_2}	$\rightarrow$	$V_{Adv_2} + S''^{(ADV_2)}$ **or:** $V_{Prep_2} + S''^{(ADV_2)} + NP$
b.	S''	$\rightarrow$	$V_{t_1} + S'^{M}/S^{oTAU}/S'$
(2)a.	S'^{M}	$\rightarrow$	$V_{M} + S'^{(ADV_2)}$
b.	S'^{ADV_2}	$\rightarrow$	$V_{Adv_2} + S'^{(ADV_2)}$ **or:** $V_{Prep_2} + S'^{(ADV_2)} + NP$
c.	S'	$\rightarrow$	$V_{t_2} + S^{oADV_3}/S^{oPROGR}/S^{oPASS}/S^{o}$
(3)a.	S^{oTAU}	$\rightarrow$	$V_{Tau} + S^{oADV_3}/S^{oPASS}/S^{o}$
b.	S^{oADV_3}	$\rightarrow$	$V_{Adv_3} + S^{oADV_3}/S^{oPROGR}/S^{oPASS}/S^{o}$
			or: $V_{Prep_3} + S^{oADV_3}/S^{oPROGR}/S^{oPASS}/S^{o} + NP$
c.	S^{oPROGR}	$\rightarrow$	$V_{Progr} + S^{oPASS}/S^{o}$
d.	S^{oPASS}	$\rightarrow$	$V_{Pass} + S^{o}_{[PASS]}$
e.	S^{o}	$\rightarrow$	V_{LEX}+ **<**lex.arg.frame**>**

Some lexicon

SA-category	**Fillers**	**Surface category**	**Cyclic rules**
V_{Adv_2}	occasional, normal, often, usual, clever, stupid,...	Adverb	$L_{S/V}$
	not, always, just, never, recent, soon,...	Adverb	L_{V}
	today, yesterday, tomorrow, recent, soon,...	Adverb	$L_{S/right}$
V_{Adv_3}	fast, careful, good, clever, stupid,...	Adverb	L_{right}
V_{Prep_2}	in, on, at, with, about, to, during, near, before,...	Preposition	OI,$L_{S/right}$
	by, in, after, before,...	Preposition	OI,SD,$L_{S/right}$
V_{Prep_3}	in, at, with, about, for, over, ...	Preposition	OI,L_{right}
V_{t_1}	PRES, PAST	Affix	SR, L_{V}
V_{M}	will, may, must, shall, can,...	Verb	L_{V}
V_{t_2}	Ø	Verb	L_{V}
	have	Verb	PaP, L_{V}
V_{Tau}	be, [$_{V}$[be] $_{Prt}$[going]]	Verb	SR^{3}
V_{Progr}	be	Verb	PrP, L_{V}
V_{Pass}	be	Verb	PaP, L_{V}

V_{LEX}:

SA-cat.	**Fillers**	**Argument frame**	**Surface cat.**	**Cyclic rules**
V_{Adj}	small, red, old, round,...	+ NP/S^{NOM}	Adjective	— —
V_{Adj}	probable, possible, true,...	+ $NP/_{NP}[S'']$	Adjective	— —
V_{Adj}	(un)likely	+ $NP/S'/_{NP}[S'']$	Adjective	SR
V_{Adj}	rumoured	+ $NP/S'/_{(NP)}[S'']$	Adjective	SR
V_{Adj}	due	+ NP/S'	Adjective	SR
V_{Verb}	laugh, sleep, swim,...	+ NP	Verb	— —
V_{Verb}	read, eat, drink, read, write,...	+ NP (+ NP)	Verb	— —
V_{Verb}	give, sell,...	+ NP (+ NP) + NP	Verb	— —
V_{Verb}	seem, appear,...	+ S'/S''	Verb	SR
V_{Verb}	tend,...	+ S'	Verb	SR
V_{Verb}	happen,...	+ NP/S'	Verb	SR
V_{Verb}	$continue_1$, $start_1$,...	+ $NP/S'_{ø}$	Verb	SR
V_{Verb}	$continue_2$, $start_2$,...	+ $NP_x + S'_{ø}$	Verb	SD

V_{Verb} need,...	+ S^o	Verb	SR
V_{Verb} follow, make sense,...	+ $NP/_{NP}[S"]$	Verb	— —
V_{Verb} believe,...	+ NP + $NP/S'/_{(NP)}[S"]$	Verb	SR
V_{Verb} think,...	+ NP + S"	Verb	SR
V_{Verb} expect,...	+ NP_x + NP/S'/S"	Verb	[SD / SR]
V_{Verb} want, like,...	+ NP_x + $NP/S'_{ø}$	Verb	[SD / SR]
V_{Verb} try,...	+ NP_x + $NP/S'_{ø}$	Verb	SD
V_{Verb} $help_1$	+ NP + NP_x + S^o	Verb	SD,([-to])[4]
V_{Verb} $help_2$	+ NP + S^o	Verb	SR,[-to]
V_{Verb} tell,...	+ NP + NP_x + $NP/S'_{ø}/S"$	Verb	SD
V_{Verb} know,...	+ NP + $NP/S'/_{NP}[S"]$	Verb	SR
V_{Verb} cause,...	+ NP + $NP/S'_{ø}$	Verb	SR
V_{Verb} make,...	+ NP + $NP/S'_{ø}$	Verb	SR, [-to]
V_{Verb} assert, deny,...	+ NP + $NP/_{NP}[S"]$	Verb	— —
V_{Verb} see, hear,...	+ NP + $NP/_{(NP)}[S"]/S^o$	Verb	SR, [-to]

The symbol '**<**lex.arg.frame**>**' in Formation Rule (3e) is an instruction to follow the Argument Frame restrictions formulated in the lexicon for the V_{LEX}, i.e. lexical predicate (V_{Adj} or V_{Verb}), chosen. Ss occurring as argument in the argument frame of a lexical verb are complement-Ss. They occur in three varieties: S", S' and S^o, each of which may be headed by an NP. An S" (i.e. S with two tenses) is developed into a full clause. S' and S^o-structures (i.e. with one tense or no tense, respectively) become surface infinitivals. The special type $S'_{ø}$ is an S'-complement with the restriction that the tense to be selected (i.e. V_{t_2}) must be Ø (i.e. 'simultaneous'), and cannot be *have* (i.e. 'preceding').

Note that the position of the English modal verbs in the Formation Rules, between V_{t_1} and V_{t_2}, automatically accounts for their defective paradigm: they can receive only the PRES or PAST tense, can never be embedded as an infinitive or a participle (as S'^{M} is not a possible complement-S type), but can be followed by simple or perfective infinitivals. The defective paradigm of the English modals is thus accounted for at no extra cost.

The Cyclic Rules that figure in the rightmost column of the lexical entries are mentioned even though they are not relevant in the context of this paper. The rules in question, mainly L (LOWERING, where the landing site of the lowered element is specified by a subscript), OI (OBJECT INCORPORATION), SD (SUBJECT DELETION), PaP (PAST PARTICIPLE), PrP (PRESENT PARTICIPLE), SR (SUBJECT RAISING), PR (PREDICATE RAISING), are the rules induced ('triggered') by the predicates in question. The rule PR does not occur in the English complementation system. In the auxiliary system it occurs only once, for the copula verb *be*, not incorporated here. English is an almost exclusively SR-language, as against Dutch, which is almost exclusively a PR-language (see note 5 below).

Some Formation Rules for Dutch

(1)a. $S''^{ADV_2} \rightarrow V_{Adv_2} + S''^{(ADV_2)}$ **or:** $V_{Prep_2} + S''^{(ADV_2)} + NP$

b. $S'' \rightarrow V_{t_1} + S'$

(2) $S' \rightarrow V_{t_2} + S^{oADV_3}/S^{oPASS}/S^{o}$

(3)a. $S^{oADV_3} \rightarrow V_{Adv_3} + S^{oADV_3}/S^{oPASS}/S^{o}$

b. $S^{oPASS} \rightarrow V_{Pass} + S^{o}_{[PASS]}$

c. $S^{o} \rightarrow V_{LEX}$+ **<**lex.arg.frame**>**

Some lexicon

SA-category	Fillers	Surface category	Cyclic rules
V_{Adv_2}	toevallig, vaak, niet, al, vandaag, graag, morgen,...	Adverb	L_{right}
V_{Adv_3}	vlug, goed, slecht, rechtop, dom,...	Adverb	L_{right}
V_{Prep_2}	in, op, aan, met, over, naar, voor, door, ...	Preposition	OI,L_S
V_{Prep_3}	in, op, aan, met, over, naar, voor, door, ...	Preposition	OI,L_{right}
V_{t_1}	PRES, PAST	Affix	SR, L_V
V_{t_2}	Ø	Verb	L_V
	hebben/zijn	Verb	PaP, L_V
V_{Pass}	worden	Verb	PaP, L_V

V_{LEX}:

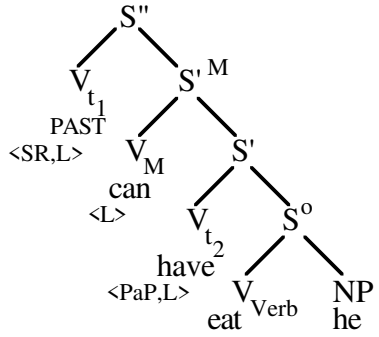
SA-cat.	Fillers	Argument frame	Surface cat.	Cyclic rules
V_{Adj}	klein, rood, oud, rond,...	+ NP/S^{NOM}	Adjective	— —
V_{Adj}	waarschijnlijk, mogelijk, waar,...	+ $NP/_{NP}[S'']$	Adjective	— —
V_{Verb}	lachen, slapen, zwemmen,...	+ NP	Verb	— —
V_{Verb}	lezen, eten, drinken, schrijven,...	+ NP (+ NP)	Verb	— —
V_{Verb}	geven, verkopen,...	+ NP (+ NP) + NP	Verb	— —
V_{Verb}	schijnen, lijken,...	+ S'/S"	Verb	PR[5]
V_{Verb}	moeten, kunnen, zullen,...	+ S'	Verb	PR[-te]
V_{Verb}	$beginnen_1$, $beloven_1$,...	+ $NP/S'_{ø}$	Verb	PR
V_{Verb}	$beginnen_2$, $beloven_2$,...	+ $NP_x + S'_{ø}$	Verb	SD
V_{Verb}	volgen, erop lijken,...	+ $NP/_{NP}[S'']$	Verb	— —
V_{Verb}	geloven,...	+ NP + $NP/_{(NP)}[S'']$	Verb	— —
V_{Verb}	denken,...	+ NP + NP/S'/S"	Verb	SD,(PR)
V_{Verb}	verwachten,...	+ NP + NP/S'/S"	Verb	SD
V_{Verb}	willen,...	+ NP_x + $NP/S'_{ø}$/S"	Verb	SD,PR[-te]
V_{Verb}	verlangen,...	+ NP_x + $NP/S'_{ø}$/S"	Verb	SD
V_{Verb}	proberen,...	+ NP_x + $NP/S'_{ø}$	Verb	SD,(PR)
V_{Verb}	$helpen_1$	+ NP + S^o	Verb	PR[-te]
V_{Verb}	$helpen_2$	+ NP + NP_x + S^o	Verb	SD,(PR[-te])
V_{Verb}	doen,...	+ NP + S^o	Verb	SR, [-te]
V_{Verb}	zeggen, beweren, weten,...	+ NP + $NP/_{NP}[S'']$	Verb	— —
V_{Verb}	zien, horen,...	+ NP + $NP/_{(NP)}[S'']/S^o$	Verb	PR[-te]

3. Some troublesome translation equivalents

Let us now consider the English sentence *He could have eaten* and its Dutch equivalent *Hij had kunnen eten.* Both their surface structures and their SAs differ in interesting ways, as is shown in (1) and (2):[6]

(1)a. He could have eaten

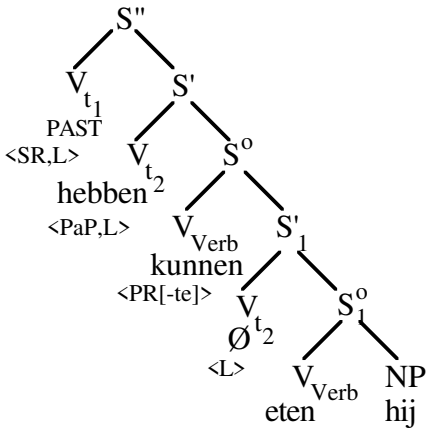
b.



(FR 1b)
(FR 2a)
(FR 2c)
(FR 3e)

(2)a. Hij had kunnen eten

b.



(FR 1b)
(FR 2)
(FR 3c)
(FR 2)
(FR 3c)

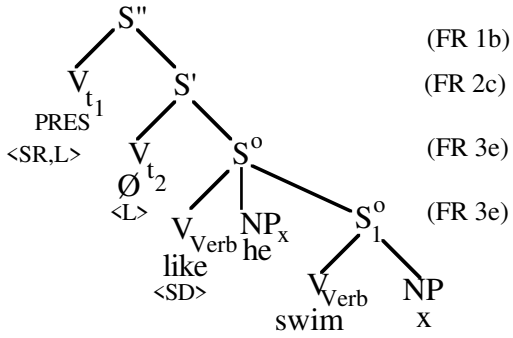
Note that the Dutch sentence contains a complement-S' under the main lexical verb *kunnen*, as specified in the Dutch lexicon. The complement-S' is characterized by the subscript 1. The English sentence, on the other hand, uses the modal verb *can* and does everything with the means made available within the auxiliary system.

It is time for some semantics now. In (1b) the terminal element *PAST* is related to a meaning description 'COUNTERFACTUAL'.[7] *Can* stands for 'POSSIBLE', *have* is contextually defined as 'AT THAT PAST TIME',[8] *eat* is semantically rendered as 'EAT', a shorthand notation for the lexical satisfaction conditions of this predicate, and *he* is rendered as 3sgM' (i.e. 'third person singular masculine'). In the Dutch SA-structure (2b) *PAST* is again related to a meaning description 'COUNTERFACTUAL', *hebben*, as the English *have*, to 'AT THAT PAST TIME', *kunnen* to 'POSSIBLE', *eten* to 'EAT', and *hij* again as '3sgM'. The t_2 Ø in S_1' has the normal meaning of 'SIMULTANEOUS'. This element is semantically vacuous here but required by the syntax.

A second example, well-known in machine translation circles, is the pair of English (3a) and Dutch (4a), with their corresponding SA-trees:

(3)a. He likes to swim

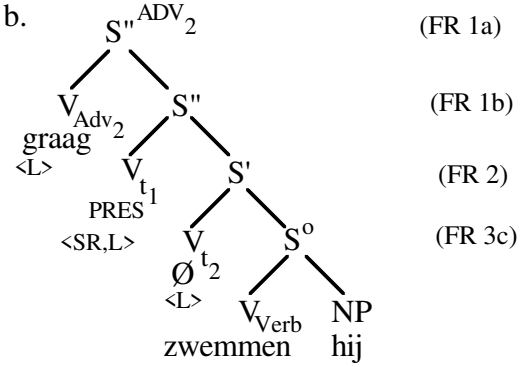
b.



(FR 1b)
(FR 2c)
(FR 3e)
(FR 3e)

(4)a. Hij zwemt graag

b.



(FR 1a)
(FR 1b)
(FR 2)
(FR 3c)

The Dutch manner adverb *graag* can be semantically described as a function that takes a proposition A and delivers A restricted by 'WITH PLEASURE'. The English lexical complement taking verb *like* has, in one of its uses, precisely the same meaning description. The grammatical difference between English and Dutch is that English expresses the propositional function 'WITH PLEASURE' as a lexical verb (*like*) that takes a complement-S, while Dutch expresses it as a sentence adverb (*graag*), i.e. within the auxiliary system. In this sense we have thus the oppo–site here of what we found with the pair (1) and (2). Meaning descriptions like the ones given here can be used at a level of representation at which the various translation equivalents can be bartered.

Note that the pair (5) and (6):

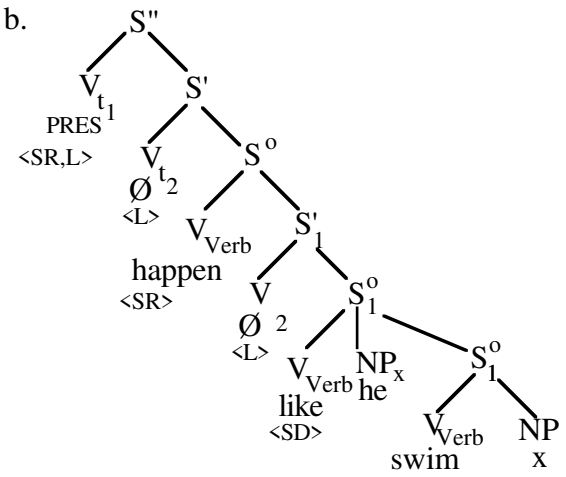
(5) He happens to swim
(6) Hij zwemt toevallig

is entirely analogous to (3) and (4), respectively. All that needs to be done is replace *like* by *happen*, and *graag* bij *toevallig*, with, of course, the concomitant argument structures.

Finally, we shall consider the pair (7) and (8), which combines the two previous pairs:

(7)a. He happens to like to swim

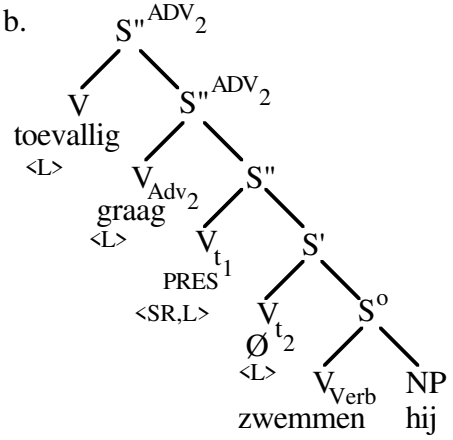
b.



(FR 1b)
(FR 2c)
(FR 3e)
(FR 2c)
(FR 3e)
(FR 3e)

(8)a. Hij zwemt toevallig graag

b.



(FR 1a)
(FR 1a)
(FR 1b)
(FR 2)
(FR 3c)

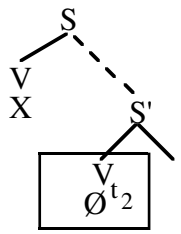
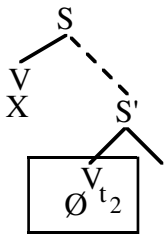
4. Translation equivalents via USA

A formal procedure is envisaged converting translation equivalent SAs in two different languages into each other. In view of the considerable syntactic differences among languages even at SA-level, this is done via the intermediary station of a universal semantic analysis or **USA**. In accordance with fig.2 above, the conversion process goes from SA_1 of the source language L_1 to USA, and from USA via the Formation Rules of the target language L_2 to SA_2. The transition from SA_1 to USA is mediated by the **semantic lexicon** of L_1, which provides, ideally, semantic descriptions of SA-elements in some universal cognitive descriptive language or notation. Analogously, the output from USA to SA_2 passes through the semantic lexicon of L_2, warranting optimal semantic equivalence. Given these items, the Formation Rules are consulted for their proper introduction into the equivalent SA-structure of L_2.

It is, of course, well-known that any fully adequate formal translation system will have to go deeper than this, since situational and world knowledge are sometimes required for a proper interpretation of a sentence. Linguistic interpretation is to some extent underdetermined by the linguistic material that carries the message. No attempt is made here to achieve a

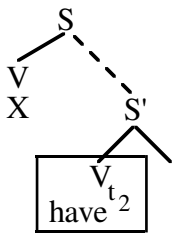
formalization of the cognitive structures and processes that are required for a fully adequate formal translation procedure. The procedure sketched here should help for cases where no 'deep' cognitive appeal is required.

In the following simplified sample of a semantic lexicon between English and Dutch the double arrow indicates the description of SA-elements in the semantic lexicon, in terms of a universal semantic, i.e. cognitive, metalanguage. The description (necessarily incomplete here) is given between single quotes, sometimes with a convenient abbreviatory symbol after the colon. Sometimes a partial tree structure is presented as an SA-element of a given language. The actual item or items described are then presented in a box; the environment is outside the box. Lexical translations are schematized as words in small caps, as before.

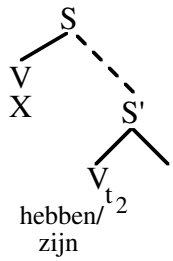
	English		**Semantic metalanguage**		**Dutch**
(1)	PRES	⇔	'time of speaking': TS	⇔	PRES
(2)	PAST	⇔	'at discourse-defined past time': PA	⇔	PAST
(3)		⇔	'same time as [[X]]': SIM	⇔	

NB: $_{V}$[X] is the nearest time predicate up: V_{t_1} (PRES or PAST) or (for English) $_{V_M}$[will]
[[X]]: semantic value of X

(4)

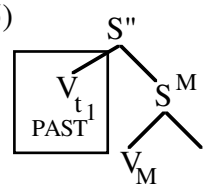


⇔ 'before [[X]]': PREC ⇔



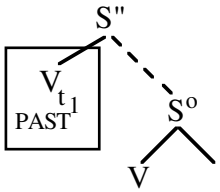
NB: $_{V}$[X] is the nearest time predicate up: V_{t_1} (PRES or PAST) or (for English) $_{V_M}$[will]
[[X]]: semantic value of X

(5)



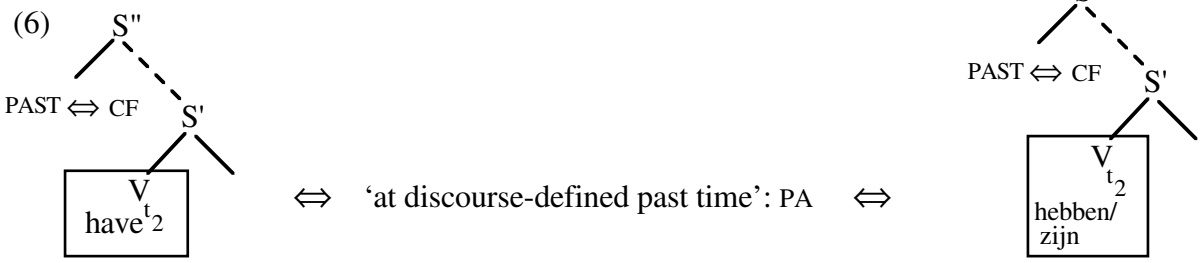
will/can/shall

⇔ 'COUNTERFACTUAL': CF ⇔



zullen/kunnen/moeten

NB: in this environment the modal verbs are scope-insensitive with regard to the tenses, but not with regard to PAST = CF

(6) 

(7) can ⇔ 'POSSIBLE': POSS ⇔ kunnen

(8) eat ⇔ 'EAT' ⇔ eten

(9) swim ⇔ 'SWIM' ⇔ zwemmen

(10) like ⇔ 'WITH PLEASURE': PL ⇔ graag
(scope-insensitive with regard to the tenses)

(11) happen ⇔ 'BY COINCIDENCE': COIN ⇔ toevallig
(scope-insensitive with regard to the tenses)

We can now show how the translation equivalences of the sentence pairs discussed above are in principle formally negotiable with the help of the semantic translation lexicon and an intermediate universal semantic structure (**USA**). First we take the pair **<**(1),(2)**>**.The direction of the translation is irrelevant. Let us take it to be from English to Dutch.

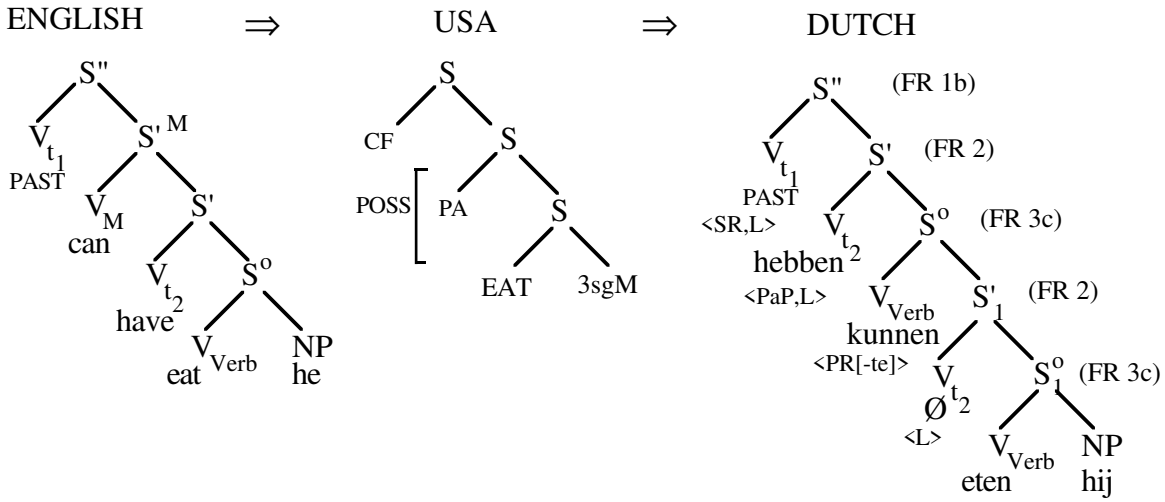


Figure 3: the translation from (1b) into (2b)

In the English SA-structure (which is taken to be the result of the parser **P**, not elaborated here) the element PAST translates into USA 'CF', according to entry (5). This again translates as PAST in Dutch, with the proviso that somewhere down the tree a modal verb occurs. This condition is fulfilled two Ss down in the Dutch tree. English *can* translates as 'POSS' in USA, and from there as *kunnen* in Dutch, according to entry (7). Likewise, the V_{t2} *have* goes to 'PA' (i.e. contextually defined past time), and from there to Dutch *hebben/zijn* as per entry (6). The translations of *eat* and *he* are trivial. The USA-tree contains 'POSS' as a modal operator whose

scope relation vis-à-vis the tenses is semantically irrelevant, under CF as specified in entry (5). This is indicated in the USA-tree as shown.

The point is now that the Dutch lexical entries can be looked up in the lexicon that goes with the Dutch grammar. There it is specified that PAST is a V_{t1} and is therefore produced by Formation Rule (1b). This gives us the top of the Dutch SA-tree. Then, *hebben* is a V_{t2} and thus produced by FR (2), while *kunnen* is a lexical verb, produced by FR (3e). The argument frame of *kunnen* requires an S' and hence a V_{t2} which, by default, is Ø. $_S$[Eten,hij] follows trivially. From here the grammar of Dutch takes over and transforms the Dutch SA into the well-formed surface structure *Hij had kunnen eten.* Refer to note 2.

Now consider the pair **<**(3),(4)**>**, which we will now translate from Dutch into English.

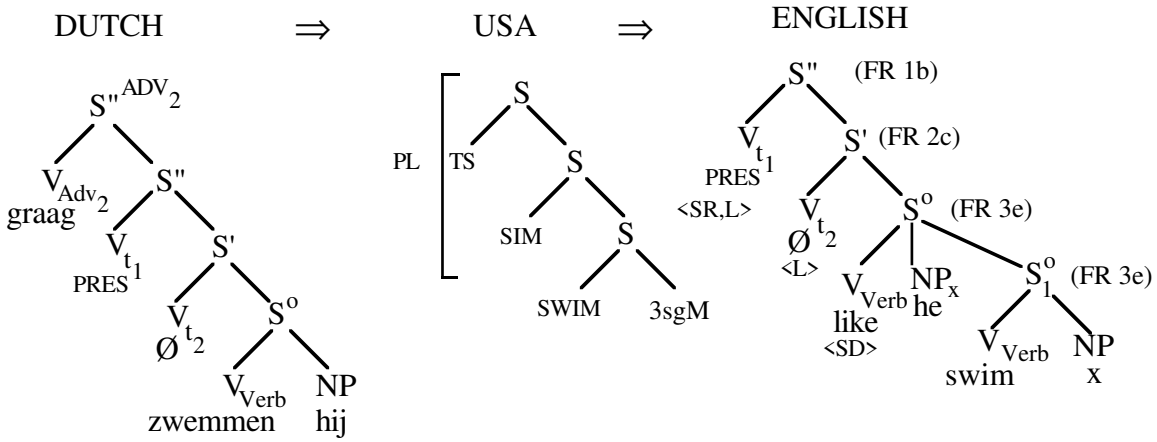


Figure 4: the translation from (4b) into (3b)

The semantic lexicon translates *graag* as 'WITH PLEASURE' or 'PL', and again into English as the lexical verb *like* (entry (10)). The entry says that this entry is scope-insensitive with respect to the tenses, as indicated in the USA-tree. This gives it the freedom to occur below the two tenses in the English SA. The tenses translate trivially. The S-structure [SWIM,3sgM] is treated as follows. As specified in the English lexicon, the verb *like* takes an NP-subject and induces the cyclic rule of SD (SUBJECT DELETION). The S-structure [SWIM,3sgM] has to be attached to *like* as object complement-S. The procedure is such that the subject of SWIM, i.e. 3sgM, is transferred to *like* and will then control the application of SD on the *like*-cycle. The result is the English SA of fig.4., which is input to the generator for English.

Again, the lexical entries put out by the semantic lexicon of the target language L_2 find their own way in the grammatical lexicon of L_2, so that the output tree builds itself up automatically.

Finally, we consider **<**(7),(8)**>**, this time from English into Dutch.

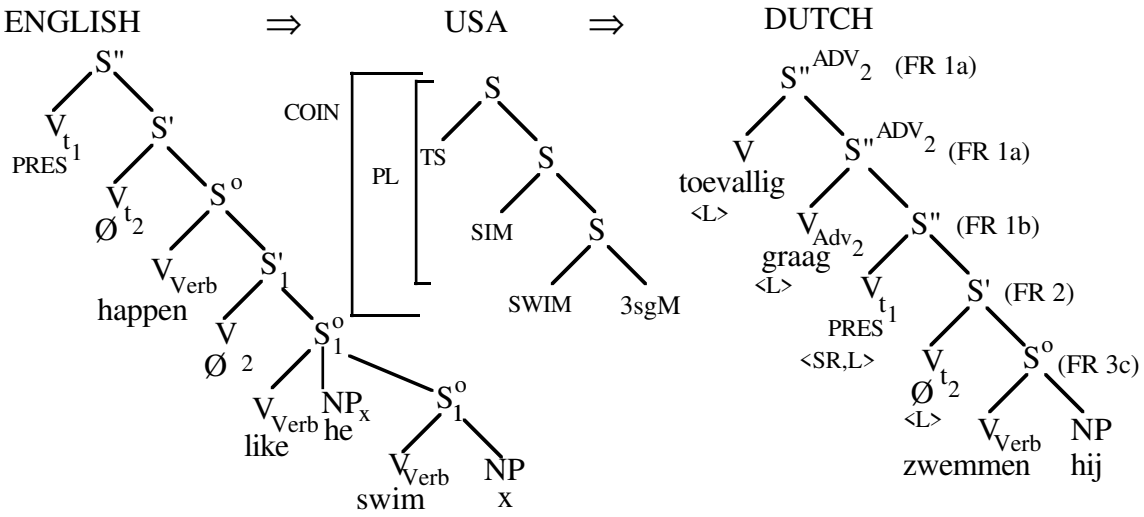


Figure 5: the translation from (7b) into (8b)

The translations of the individual items are as specified in the semantic lexicon. But note that although COIN and PL are scope-insensitive with regard to the tenses they are not scope-insensitive with regard to each other. The Dutch equivalent of COIN, *toevallig*, must therefore be processed before (i.e. higher in the tree) *graag*, the equivalent of PL. Since FR (1a) of the Dutch grammar is recursive, this procedure determines the top two S-layers of the Dutch SA-tree in fig.5. The rest is as before. Essentially the same procedure makes it possible to translate (9) and (10) into each other:

(9) He likes to happen to swim
(10) Hij zwemt graag toevallig.

Again, the grammars of the languages concerned convert the SAs into well-formed corresponding surface structures.

References

de Rooij, J. 1991: Ik had al moeten eindigen vs. ik moest al geëindigd hebben. *Gramma*, 15.3:235-246

Hutchins, W.J. 1986: *Machine Translation: Past, Present, Future.* Ellis Horwood, Chichester-Halsted Press (John Wiley & Sons), New York-Chichester-Brisbane-Toronto.

Schotel, H. 1994: *SeSynPro. Towards a Workbench for Semantic Syntax*. This volume.

Seuren, P.A.M. 1986: Helpen en helpen is twee. *Glot* 9.1/2:110-117.

Seuren, P.A.M. (to appear): *Semantic Syntax.* Blackwell, Oxford.

Zock, Michael, & Gérard Sabah (eds.) 1988: *Advances in Natural Language Generation. An Interdisciplinary Perspective.* Printer Publishers, London.

[1] For example, Hutchins (1986:267-8): 'there are relatively few problems attached to particular constructions or linguistic phenomena'. Or McDonald in Zock & Sabah (1988:X): 'There are, of course, technicalities: the generator takes care of making the number of subject and verb agree and gets the right case on the pronouns. The more important contribution of a generator, however, is that it provides a framework that makes the process of assembling and couching the message a text conveys exceptionally flexible.' McDonald's mere 'technicalities' of grammatical treatment have, till the present day, never been worked out in a principled way. Moreover, they involve a great deal more than just getting case and agreement right (and even that is far from simple).

[2] For a much fuller version of the syntax (formation and transformation rules) for English and Dutch (and a few more languages), together with a full discussion of the formal and empirical details, see Seuren (to appear).

[3] The TAU-type of S^o corresponds to sentences with the idiosyncratic verbs *be to* or *be going to*, as in *He is (going) to leave soon.* These verbs occur only in the simple present or past, not in the perfect tenses and not under the modals. They do occur in the subjunctive mood, however: *If he were to leave soon,...* As far as the tenses are concerned, the treatment presented here automatically accounts for the restrictions noted. The subjunctive requires a separate treatment not gone into here.

[4] See Seuren (1986) on the two verbs *help*, in English as well as in Dutch.

[5] Note that the cyclic rule of SUBJECT RAISING (SR) does not occur in the complementation system of Dutch, only in the auxiliary system, and there only once (for V_{t_1}).Where English uses SR in its complementation system, Dutch systematically makes use of PREDICATE RAISING (PR).

[6] Some Dutch dialects also have the English variant *Hij kon hebben gegeten.* In Flemish Dutch this variant is even dominant (De Rooij 1991).

[7] In English,Dutch and many other languages, *PAST* may signify counterfactuality. When in construction with some predicate of futuricity or possibility this meaning tends to be the only one possible.

[8] The meaning description 'AT THAT PAST TIME' is normally assigned to the first tense PAST. However, when PAST functions as a counterfactuality marker the second tense takes over the function(s) of the first. Thus, in such cases, Ø means 'AT THE TIME OF SPEAKING' (=PRES), and *have* means 'AT THAT PAST TIME' (=PAST). When PRES and PAST have their normal meaning of 'AT THE TIME OF SPEAKING' and 'AT THAT PAST TIME', respectively, Ø means 'SIMULTANEOUS WITH THE TIME INDICATED BY THE FIRST TENSE', and *have* stands for 'PRIOR TO THE TIME INDICATED BY THE FIRST TENSE'.