LINKING PROPOSITIONS

D. S. Brée & R. A. Smit

Rotterdam School of Management, Erasmus University, P.O.B. 1738, NL-3000 DR Rotterdam, The Netherlands.
This research was partially funded by the Netherlands Organization for the Advancement of Pure Research.

KEYWORDS: Semantics, Subordinating conjunctions, Language processing, Language universals.

ABSTRACT

The function words of a language provide explicit information about how propositions are to be related. We have examined a subset of these function words, namely the subordinating conjunctions which link propositions within a sentence, using sentences taken from corpora stored on magnetic tape. On the basis of this analysis, a computer program for Dutch language generation and comprehension has been extended to deal with the subordinating conjunctions. We present an overview of the underlying dimensions that were used in describing the semantics and pragmatics of the Dutch subordinating conjunctions. We propose a Universal set of Linking Dimensions, sufficient to specify the subordinating conjunctions in any language. This ULD is a first proposal for the representation required for a computer program to understand or translate the subordinating conjunctions of any natural language.

1. Introduction

Languages provide speakers with the means to express propositions and to link these propositions. Propositions are expressed in language by means of clauses. These can form sentences in themselves, or they may be linked together within one sentence, either at the same level or embedded one within the other. It is this last category of linking, by means of subordinate conjunctions, in which we are interested. The questions that we ask are:

- Do languages provide a systematic way of expressing the possible subordinating links between propositions?
- If they do, which dimensions can be used to capture the means that are provided in all languages for these links?
- What is this Universal set of Linking Dimensions (ULD)?

We have attempted to provide a systematic description of the subordinating conjunctions (SCs) in Dutch. It is this description that we will use to propose a ULD. At the top level we have divided the SCs into just four types: inferential, temporal, causal and manner SCs. This is fewer than the dozen or so types found in the traditional and modern grammars, which give a 'flat' taxonomic tree, making all the obvious distinctions at the first level. At each branch in the taxonomic tree, we have tried to make as few divisions as possible, in order to make the motivation for each split clear.

These four categories were chosen because they enable quite different kinds of relationship to be set up between a main and a subordinate proposition. They each indicate a different function that the subordinate event/state has in relation to the main predication. The most abstract relationship is that of **inference**, in which the speaker uses the sub proposition to give the grounds for his belief concerning the truth status of the main proposition. The other three types indicate more than simply a relationship between beliefs (propositions); they convey the speaker's beliefs about relationships that exist 'in reality'. Two of these are quite specific: **time** and **cause**. The fourth category of SC, **manner**, also serves to indicate that there is a relationship 'in reality' between the sub and main events/states; however, this relationship is one that depends highly on the schema that the observer uses to codify what he sees.

We will now describe the sub trees for each of these four types of linkage. The meanings of the Dutch SCs, taken from Van Wijk and Kempen (1980), have been determined using:

- sentences taken from a Dutch corpus (Uit den Boogaard, 1975; shown as e.g. 1.2345);
- the authoritative Algemene Nederlandse Spraakkunst - ANS (Geerts et al., 1984).

2. Inferential linkage

An inference relationship exists between two propositions when the truth value of one can be deduced from the other. The grounds for the deduction are left open. They may be based on some causal model of reality:

If metal is heated it expands.

But they can also be purely definitional:

If two angles of a triangle are equal, their opposite sides are equal.

The types of inferential linkage depend in the first instance on the truth status of the main proposition. This may be either true, probably true but with the possibility of an escape, hypothetically true or counterfactual. A false main proposition is not indicated by an SC but by the use of the past tense and/or a modal auxiliary verb.

True. If the main proposition is true, then the inference relation from the sub proposition may be used, denied or deemed irrelevant.

The **use** of the inference is indicated in Dutch by the SC aangezien (since):

De rector had besloten de school te sluiten aangezien het verbod was overtreden. (ANS, p.655)

(The principal decided to close the school since the ban had been contravened.)

It has been, and still is, customary to classify aangezien as a causal SC (ANS, p.655). This is incorrect. Causal SCs can be topicalized, inferential SCs cannot. The reason for this distinction is that causal SCs say something about reality, whereas inferential SCs are used to make an inference. This making cannot be topicalized. Aangezien, however, like non-temporal since, cannot be topicalized, so it is not causal:

*Het is aangezien het verbod was overtreden, dat de rector besloot de school te sluiten.

An inference relationship may be **denied** using the SC hoewel (although). Then the normal inference is from the sub proposition to the falsity of the main proposition:

Hoewel het verbod was overtreden, besloot de rector de school niet te sluiten.

(Although the ban had been contravened, the principal decided not to close the school.)

The **irrelevance** of any inference relationship is indicated by ongeacht (whether ... or not):

Ongeacht of het verbod was overtreden, zou de rector hebben besloten de school te sluiten.

(Whether the ban had been contravened or not, the principal would have decided to close the school.)

Escape linkages. If the speaker wishes to indicate that the main proposition is not certainly, but only probably, true, then an SC can be used to indicate the circumstances under which the main proposition may indeed by false. Such SCs provide an ESCAPE from the speech act being made in the main clause. This speech act may be an assertion, in which case the escape is from the truth of the main proposition; but any other type of speech act, such as a promise, may also be escaped from. The escape may be either when the sub proposition is true, with tenzij (unless), or when it is false, using mits (provided that):

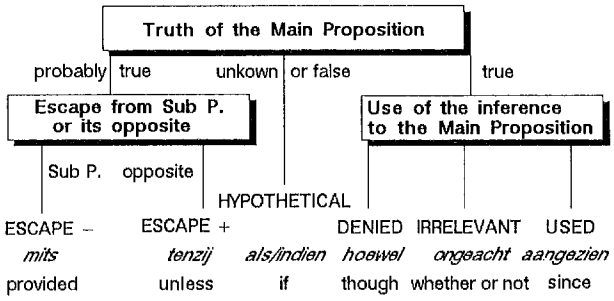
De oecumene zal niet slagen tenzij het gesprek met Israel ... wordt gevoerd. (2.3040)
(Ecumenism will not succeed unless there is a dialogue with Israel ...)

Jongeren kennen een normale behoefte aan gezag en normen, mits zij er de zin van weten te ontdekken. (5.3341)
(Youngsters have a normal need for authority and norms, provided they can discover their sense.)

Hypothetical linkage. If the truths of the main and sub propositions are unknown, an inference relationship from the sub to the main proposition can be shown by using either als or indien (if). Of these two SCs, als is the more common, but it is ambiguous between several uses (see below); indien is more formal and emphatic:

Als/Indien het verbod is overtreden, zal de rector de school sluiten.
(If the ban has been contravened, the principal will close the school.)

Figure 1. INFERENTIAL SCs



Counterfactuals. Just as with the true inferential linkages, the counterfactual inference may be simply used or denied. The inference, in this false world, may be **used** to infer a main proposition which is true in this false world but false in the actual world:

If Eve hadn't given Adam the apple, he wouldn't have eaten it.

The inference may also be **denied** to give a main proposition which is true in both worlds:

Even if Eve hadn't given Adam the apple, he still would have eaten it.

This is a semifactual. The same SC, als (if), is used both for hypotheticals and counter/semi-factuals. This is not the case in all languages, e.g. Polish, Japanese.

In order to indicate that the sub clause does not correspond with the truth status of the sub proposition, the tense of sub clause verb is placed one step further into the past than would normally be the case. That is to say: if the tense would normally be past, past-perfect or present-perfect then it is set to past-perfect; otherwise it is set to past. With the true counterfactuals (as opposed to the semifactuals) the fact that the main clause also does not correspond with the truth status of the main proposition is indicated by using the past-future, i.e. using the past form of the verb zullen as the auxiliary finite verb form:

Als ik geld had, zou ik op reis gaan.
(If I money had, would I travelling go.)

Moreover the simple (or perfect) past can also be used to indicate counterfactuality:

Als ik geld had, ging ik op reis.
(If I had money, I went travelling.)
Als ik geld gehad had, had ik meer gereisd.
(If I had had money, I had more travelled.)

The ANS (p.468) provides no rule for choosing between using the simple (or perfect) past and using the modal auxiliary zullen. There are, however, systematic differences (Nieuwint, 1984). More problematic is that the simple past tense may indicate one of two things: the Time Of Reference (TOR) is in the past or we are dealing with a counterfactual. This ambiguity can be resolved by the context: if the TOR is already in the present, then the past tense indicates counterfactuality.

Summarizing, we find four types of inferential linkage, depending on the truth value of the main proposition:

- true: the inference from a true sub proposition may be used, denied or deemed irrelevant;
- probably true: this truth can be escaped, either from the sub proposition or from its negation;
- hypothetical inference from a sub proposition whose truth is uncertain;
- counterfactuals: an inference from a sub proposition that is known to be false, to either a false or a true main proposition.

3. Temporal linkage

The temporal SCs specify the time of the main predication in relation to the time of the event indicated in the sub clause. The system we have used to represent the different possible temporal linkages is based on two dimensions: the relative temporal order of the main and sub events, and the place of the main event within this restricted time range.

Relative temporal order. A sub clause introduced by a temporal SC is used to restrict the time during which the main proposition is true: the time of the main event may be at a time that is either earlier than, or later than, the time of the sub event, or it may be coincidental with the time of the the sub event.

The position within the range. The place of the main event within this restricted time range is the second dimension. It may be either:

- at some time within the proximity of the sub event:
  - either immediately adjoining the sub event;
  - or in the vicinity of the sub event time.
- at some time within an interval. The way the bounds of this time interval are indicated depends on the relationship between the TOR and the interval itself. If the TOR is to fall within the interval, then the duration of the sub event determines the interval. Otherwise the TOR marks one bound of the interval, the other bound being set by the time of the sub event. Again this category has two alternatives:
  - either the main event occurs at some moment within the interval;
  - or it occurs for the whole of the interval, in which case the event must be able to have a duration or be repeatable.

These four different places within the time range, together with the three ways of specifying the range,

give twelve different possibilities for indicating the time relationship between the main and sub events.

It is not to be expected that any language will have SCs to distinguish between all these twelve possible temporal linkages. In Dutch three of the relationships cannot be expressed using an SC. Moreover, the sub-distinctions made in the second dimension are not always made.

Figure 2. TEMPORAL SUBORDINATING CONJUNCTIONS

Relation of main to sub event.	Main event in proximity of sub: Next to	Nearby	Main event in interval of TOR & sub event: Sometime	Durative
Earlier than:			voor (before)	totdat (until)
Coincidental:		toen/als wanneer/nu (when)	terwijl (while)	zolang (as long as)
Later than:	zodra (as soon as)	nadat (after)	sinds (since)	sinds (since)

The difference between the four SCs als, toen, wanneer and nu requires an explanation:

- toen is used to set the TOR to some point before the time of utterance, and so only occurs when the TOR is in the past (cf. the use of when as discussed by Kamp, 1981). The TOR is set to the time of the sub event;
- nu is used when the TOR has already been fixed, and an event, the sub event, which happens to be coincidental with the TOR, is a cause or reason for the main event;
- als and wanneer are used:
  - for a temporal coincidence after the TOR, without bringing the TOR forward;
  - to indicate a repeated or repeatable temporal coincidence.

Wanneer (when), which hardly ever occurs in spoken Dutch except as an interrogative, is temporal. Als (if) is not confined to a temporal role, being used also for manner and inferential linkages. So its use puts the burden of interpretation onto the addressee.

We have seen that the temporal linkage is specified on two principle dimensions:

- the time range to which the main event is restricted may be before, after or coincidental with the time of the sub event;
- the time of the main event may be related either directly to the time of the sub event, or it may fall within an interval. In the former case the proximity may be indicated. In the latter case the main event may be considered to occur once in the interval or during the whole of the interval. The interval itself is bound between the sub event and the TOR, unless the TOR falls within the time period of the sub event, in which case the interval is equivalent to the duration of the sub event itself.

4. Causal linkage

A temporal linkage is not the only relationship that can be indicated in reality between the sub and main propositions. A causal linkage can also be made from the sub event or state, to the main event or state. There are two main types of causal linkage: teleological and ateleological.

An ateleological cause is a purely physical link, i.e. mechanistic, in the sense that no will is posited. The mechanism operates inevitably, e.g. gravitation that controls the motion of the planets. E.g.

> De pechdag voor de NS werd gisteren nog gecompleteerd, doordat op het centraal station in Utrecht twee machinisten van aansluitende posttreinen bij het wisselen van trein allebei precies in de verkeerde stapten. (1.5847)
>
> (The day of troubles on the railways yesterday was even more complete, because at the central station in Utrecht two drivers of connecting post trains, when changing trains, each stepped into the incorrect train.)

An ateleological link may also be proportional: the more there is of some sub property the more there will be of the main property, as in:

> De dagen lengen naarmate de nachten korten.
>
> (The days lengthen as the nights shorten.)

A teleological link, on the other hand requires that a will be present. They are volitional. The being that exerts this will has two components of interest: a perception of his own state and an awareness of his own goal. There are, correspondingly, two types of teleological cause: reason and motive. Reason is primarily state controlled, e.g.

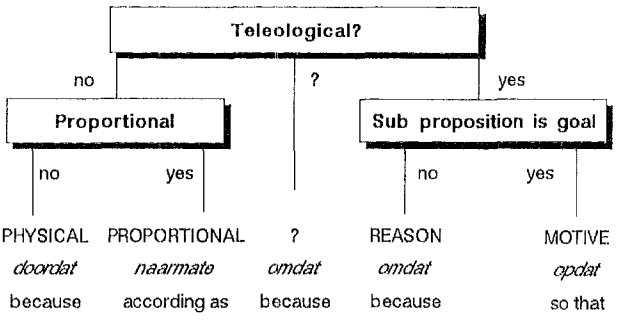
> Het kwam hem voor dat hij, juist omdat hij zo gewoon mogelijk wilde doen, zich zo ongewoon voelde. (4.1610)
>
> (He realized that he, just because he wanted to behave as normally as possible, felt himself to be so abnormal.)

Motive is primarily goal controlled, e.g.

> De regering heeft het bedrag van de steun verhoogd opdat de armsten geen honger zullen lijden. (Donaldson, 1984, p.195)
>
> (The government has increased the amount of the support so that the poorest people will not suffer from hunger.)

Note that the description of the goal state is not true, as it has yet to be achieved. The standard way of indicating this is to use the conditional auxiliary (zullen - will) in the subordinate clause. As an illustration of the contrast: 'feeling ill' is a reason for going to bed, 'to get better' is a motive for going to bed.

Figure 3. CAUSAL SC s



Omdat may be used either for a teleological reason or for an ateleological mechanistic link. As a result doordat is used to emphasize an ateleological cause. In, e.g.

> Because there were several new dancers in the troupe, the form of the ballet (was) changed.

because would be translated by doordat to indicate that the new dancers themselves caused the change; whereas omdat would indicate that the choreographer made the change to accommodate the ballet to the new dancers.

We have found that there is a distinction between mechanistic and volitional causal linkages. Mechanistic links may also be proportional. Volitional, or teleological, links may be based either on reason or on motive.

5. Manner linkage

The fourth and last type of linkage is the least specific. The sub proposition indicates something about the manner of the main proposition. A manner SC is used to add a descriptor, which includes the sub event or state, to the main proposition. The principle distinction to be made is whether this sub event/state actually exists or is (perhaps) imaginary, i.e. whether the sub proposition is true or of unknown truth value.

True sub proposition. If the sub proposition is true, then either the manner of the main proposition is specified as being restricted to the same as the manner in the sub:
Hij speelt viool zoals hij piano speelt.
(He plays violin in the same way as he plays piano.)
or the sub proposition is an addition to the main one:
Hij speelt viool evenals zijn vader dat heeft gedaan.
(He plays the violin, just as his father did.)

Non true sub proposition. If the sub proposition is false or of unknown truth value then alsof is used:
Hij speelt viool alsof hij piano speelt.
(He plays the violin as if he plays the piano.)
Whether the sub proposition is false or merely of unknown truth value must be determined using the context. The speaker can indicate a false value by using the past tense, just as with conditional counterfactuals:
Hij speelt viool alsof hij piano speelde.
(He plays the violin as though he played the piano.)

Figure 4. MANNER SUBORDINATING CONJUNCTIONS

Sub proposition:	True	Non-true
Restriction	zoals (same way as)	
Addition	evenals (just as)	alsof (as though)

6. Conclusion.

In this analysis of the semantics of Dutch SCs, we have concentrated on the main aspects in order to distinguish the wood from the trees. We have ignored SCs that are archaic, dialectal or formal. We have also ignored secondary uses of certain SCs, e.g. that the temporal SC terwijl (while) can be used to highlight a contrast. We believe that this is not prejudicial to our case. Secondary uses are just that; they are not different meanings, as we have argued elsewhere for the non-standard uses of if (Brée & Smit, 1985).

The definition of the meanings of the Dutch SCs is specific enough to be implemented in a sentence generation program. We have demonstrated this using Kempen's Incremental Procedural Grammar (Hoenkamp, 1983). Our extension (see Brée, Smit & Schotel, 1984) allows a user to enter two or more propositions and the type of relationship between them (inferential, temporal, causal, manner). Then it asks questions corresponding to the semantic tree for the corresponding type of SC, in order to select the appropriate kind of SC (e.g. hypothetical, denial, etc.). The program then uses the selected kind to find the correct SC in Dutch. So the selection of the kind of SC is independent of the language; the actual SC is selected from a table of SCs built up from the semantic definition of each SC within one language. Our program can also take a sentence as input and break it down into main and sub propositions, replacing the linking SC by its language independent semantic definition. The important point is that the definition of the kinds of SC is language independent.

We set out with the aim of establishing a Universal set of Linking Dimensions (ULD) that speakers use in linking propositions. It is the semantic trees that provide us with the ULDs. There are two levels at which we could hypothesize universality. The first, the strong hypothesis, is that the trees are the same in all languages. Then the kinds of SCs should be the same in all languages. For Dutch and English this is more or less the case. The differences in the SCs in the two languages come about in the different ambiguities that arise because the same SC is used for more than one kind of relationship (e.g. als ==> if/when, since ==> sinds/aangezien). As these ambiguous uses are not the same in the two languages, difficulties arise for translation programs. However, this does not negate the strong hypothesis.

If it does turn out that there are languages with SCs that cannot be defined using these trees, then a second, weaker, thesis may hold, namely that the building blocks from which the trees are made, are the same in all languages. That Polish and Japanese have SCs specifically for counterfactual inferences, leads us to suspect that it is this second thesis that will be found to hold. It will provide the basis for constructing a means for representing the functions performed by SCs in all languages in linking propositions.

In either case, why is it that people relate propositions using the ULD? Is it because their brains are so constructed, or because their minds reflect the nature of the environment in which they find themselves? Is the ULD a natural or artificial phenomenon (Simon, 1981)?

REFERENCES

Brée, D.S. & R.A. Smit 1985. Non-standard uses of if. In The proceedings of the 2nd Conference of the European Chapter of the Association for Computational Linguistics. Geneva: 218-225.

Brée, D.S., R.A. Smit & H.P. Schotel 1984. Generation and comprehension of Dutch subordinating conjunctions by computer. In O'Shea, T., ed., Proceedings of the 6th European Conference on Artificial Intelligence. Elsevier, Amsterdam: 205-208.

Dik, S.C. 1981. Functional grammar. Foris, Dordrecht.

Donaldson, B.C. 1984. Dutch reference grammar. Nijhoff, Leiden.

Geerts, G., W. Haeseeryn, J. de Rooij, & M.C. van den Toorn, eds, 1984. Algemene nederlandse spraakkunst. Wolters-Noordhoff, Groningen, Holland.

Hoenkamp, E.C.M. 1983. Een computermodel van de spreker: psychologische en linguistische aspecten. Ph.D. thesis. Katholieke Universiteit van Nijmegen.

Kamp, H. 1981. A Theory of truth and semantic representation. In Groenendijk, J.A.G., T.M.V. Janssen, & M.B.J. Stokhof, eds, Formal methods in the study of language, vol. 1. Mathematisch Centrum, Amsterdam: 277-322.

Nieuwint, P.J.G.M. 1984 Werkwoordstijden in nederlandse "counterfactuals". De Nieuwe Taalgids, 77(6): 542-555.

Simon, H.A. 1981 Sciences of the artificial. 2nd ed. MIT press, Cambridge.

Uit den Boogart, P.C. 1975. Woordfrequenties van geschreven en gesproken Nederlands. Oosthoek, Scheltema en Holkema, Utrecht.

Wijk, C. van & G. Kempen 1980. Funktiewoorden: een inventarisatie voor het Nederlands. Review of Applied Linguistics, 47, 53-68.