Connecting conditionals: a corpus-based approach to conditional constructions in Dutch

Reuneker, A.

Appendices

APPENDIX A

Annotation guidelines (features)

A.1 Introduction

This appendix includes the annotation guidelines that were written for interns in the project. The guidelines aim to provide clarity about both the technical procedure and the (linguistic) criteria for feature annotation.

In section A.2 general instructions are provided. In sections A.3 to A.11, the guidelines for the annotation of clause order, syntactic integration, tense, modality, aspect, person and number, sentence type, negation, and focus particles are provided respectively.

A.2 General instructions

Before presenting the annotation instructions, some remarks are in order. Make sure to read these instructions carefully before beginning your annotation work. In case of questions, send an email. It is better to ask a question than to provide incorrect or imprecise annotations. For contact details, see section A.2.5 below.

A.2.1 Natural language data are messy

The instructions below are readily applicable to non-problematic cases, but when dealing with natural-language data, utterances do not always adhere to clearly defined patterns. Do not panic, as each feature is presented with known problem cases. If these instructions do not clarify the issue at hand, you can

always use the general label 'NA', which stands for 'not available' or 'non-applicable'. In such cases, write down the reason you used this label in the available comment column.

A.2.2 Interpretational features

Not all features are created equal. Some features are explicit and grammatical, such as clause order, and differences in coding will most likely be the result of temporary loss of attention. Other features, such as type of modality, are more interpretative. Still, try to be as consistent as possible in assigning labels for those features, because reliability will be calculated and problem cases will be discussed.

A.2.3 Practical advice

Some practical advice is to annotate one feature at a time. It is not efficient, nor beneficial for consistency to code one item at a time for all features. Make sure you schedule your annotation work in blocks of a fixed time span (20 or 30 minutes). Take a small break in between and repeat. For a specific implementation of such a time-management technique, see for instance the *Pomodoro* technique (Cirillo, 2009). Mostly, coding for more than half a day is not only extremely repetitious, but also bad for the quality of your annotations.

A.2.4 File format

The data are presented in a spreadsheet in so-called 'wide format', meaning that each row in the spreadsheet represents one observation (here: one conditional sentence) and multiple properties of that sentence, such as the metadata (source, mode, genre, register et cetera), and the features to be annotated (see Gries, 2013, pp. 20–26). You can add your annotations using any spreadsheet software compatible with CSV files (Comma Separated Values), such as *LibreOffice Calc* (free, see https://www.libreoffice.org/discover/calc) or *Microsoft Excel* (paid, available on University computers). Each sentence is presented with its preceding and following sentence. Only the sentence itself is to be annotated. The co-text is provided to be able to interpret the sentence in context. The data are prepared for you and the order is randomised. Please do not re-order the data, as this will make combining annotations more tedious.

A.2.5 Contact

In case of questions or comments, please contact Alex Reuneker at a.reuneker@hum.leidenuniv.nl.

A.3 Clause order

A.3.1 Introduction

The feature *clause order* represents the order of the antecedent and consequent of conditional sentences. In canonical conditionals, the antecedent is introduced by *als* (*if*). The coding uses the position of the antecedent only. Four values for this feature are possible: sentence-initial antecedent, sentence-final antecedent, sentence-medial antecedent and insubordinate antecedent. These four options are exemplified below.

(1) **Sentence-initial antecedent**
Als je op de knop drukt(,) gaat het licht aan.
If you press the button, the light will switch on.

(2) **Sentence-final antecedent**
Het licht gaat aan(,) als je op de knop drukt.
The light will witch on(,) if you press the button.

(3) **Sentence-medial antecedent**
Dat is, als ik het zo mag zeggen, nogal een flauwe opmerking.
That is, if you'll excuse me, a rather dull comment.

(4) **Insubordinate antecedent**
Als jij nou even koffie zet (...)
If you make some coffee (...)

A.3.2 Instructions

For each item, determine the position of the antecedent with respect to the consequent. Annotate using the appropriate label. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the numerical labels after the semicolon. They will be converted to their full counterparts after you are done annotating.

Sentence-initial (si; 1)
The antecedent precedes the consequent.

(a) Als er genoeg water bij Lobith binnenkomt, staat de stuw open.
If enough water enters Lobith, the weir is open.

Sentence-final (fi; 2)
The antecedent follows the consequent.

(a) Gemeenten kunnen de witte scholen niet uitbreiden met extra lokalen, als die schooluitbreiding een gevolg is van witte vlucht.
Municipalities cannot expand the white schools with extra classrooms, if the school expansion is a result of 'white flight'.

Sentence-medial (sm; 3)
The antecedent is inserted into the consequent.

(a) Vervolgens neemt de verzekeraar dan, als de aankoopnota in orde lijkt, de koopprijs als waarde in de polis over.
Subsequently, the insurer will then, if the purchase invoice appears to be in order, adopt the purchase price as value in the policy.

Insubordinate (in; 4)
The antecedent is used without expressing a consequent.

(a) Als u uzelf even kort introduceert en uw vraag stelt...
If you introduce yourself briefly and ask your question...

A.3.3 Problem cases

Please take note of the following known problem cases and annotate accordingly.

Incomplete utterance ('NA')
If the utterance is incomplete, use the 'NA'label. For instance, an incomplete consequent such as below is not an instance of insubordination. Do check the available co-text for possible parts of the conditional.

(a) Als Nicolaas en Jacobien uh als ik die uitnodig.
If Nicolaas and Jacobien uh if I invite them.

Running astray ('NA')
Especially in spoken data, utterances can run astray. Use the 'NA'label in such cases.

(a) Voorzitter zou ik de heer De Wit mogen vragen stel nou dat of jee als de uitkomsten van de evaluatie of die nou tweeduizend plaatsvindt of eerder wat uw verzoek is stel nou dat daaruit komt dat een onderdelen wellicht na een redelijk goed uitvoerende goed werkende uh wet is?
Chairman, could I ask Mr De Wit now that whether or if the results of the evaluation whether it takes place in 2000 or earlier what your request is, suppose that a part may be after a reasonably well executed, well-functioning uh law?

Embedded conditionals
Conditionals are sometimes embedded in matrix clauses. In such cases, treat the embedded conditional as an autonomous sentence and code accordingly. In the example below, the annotation should thus read *si* for a sentence-initial antecedent.

(a) Het CDA vindt dat als hij eenmaal koning koning is rol moet blijven spelen bij kabinetsformaties.
The CDA believes that if [when] he finally is king, he should continue to play a role in cabinet formations.

Non-declarative sentences
If the conditional has a non-declarative consequent, such as an interrogative consequent, or is embedded in a question or command, please annotate according to the regular instructions above. In the example below, the annotation should thus read *si* for a sentence-initial antecedent.

(a) En als ze het doen hoe doen ze het?
And if they do it how do they do it?

Co-construction
If parts of the conditional are produced by different authors, please use the regular instructions.

Crossing sentence-borders
All conditionals are presented with the preceding and following co-text. If one of the parts of the conditional is in the co-text, treat it as if it were in the regular item-slot.

A.4 Syntactic integration

A.4.1 Introduction

The feature *syntactic integration* represents the type of syntactic dependency between the consequent and the antecedent. Syntactic integration is reflected in the word order of the consequent and the occurrence of the resumptive particle *dan* 'then'. It is important to keep in mind that syntactic integration is not independent of clause order. This feature should therefore only be annotated in case of sentence-initial antecedent in combination with a declarative consequent. Furthermore, embedded and insubordinate conditionals are excluded from the annotation of this feature. The possible values of this feature are exemplified below.

(5) **Integrative**
Als je op de knop drukt, gaat het licht aan.
If you press the button, the light will switch on.

(6) **Resumptive**
Als je op de knop drukt, dan gaat het licht aan.
If you press the button, then the light will switch on.

(7) **Non-integrative**
Als je op de knop drukt, het licht gaat aan.
If you press the button, the light will switch on.

A.4.2 Instructions

For each item, determine the type of syntactic integration. Annotate using the appropriate label. Below, the coding instructions are presented, together with examples from the corpus. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the numerical labels after the semicolon. They will be converted to their full counterparts after you are done annotating.

Integrative (int; 1)
The consequent follows the antecedent and features subject-verb inversion.

(a) Als de regering-Schroder daartoe inderdaad besluit, komt de regering-Balkenende met haar bezuinigingsbeleid in Europa nog meer alleen te staan.
If the Schroder government does indeed decide to do so,the Balkenende government stands alone even more with its economic policy in Europe.

Resumptive (res; 2)
The consequent follows the antecedent, is introduced by the resumptive particle *dan* 'then' and features subject-verb inversion.

(a) Als iemand werkelijk gelukkig is dan moet deze persoon in het bezit zijn van het goede.
If someone is really happy then this person must be in possession of the good.

Non-integrative (non; 3)
The consequent follows the antecedent and does not feature subject-verb inversion or a resumptive particle.

(a) Als je kijkt wat er de laatste zes, zeven jaar over ons is geschreven: ik ben niet anders gewend.
If you look at what has been written about us in the last six or seven years: I am not used to anything else.

A.4.3 Problem cases

Please take note of the following known problem cases and annotate accordingly.

Incomplete utterance ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

A.5 Tense

A.5.1 Introduction

The feature *tense* represents the grammatical tense of the verb phrase in a clause. For this feature, Broekhuis, Corver and Vos's (2015a, p. 157) adaptation of te Winkel's (1866) and Verkuyl's (2008) *Binary Tense Theory* is used (see chapter 5), in which two binary features determine tense: ±past (*present*, *past*) and ±perfect (*perfect*, *imperfect*), which results in four basic tenses: simple present (present, imperfect), present perfect (present, perfect), simple past (past, imperfect), and past perfect (past, perfect). Please keep in mind that in this perspective on tense, *zullen* '*will*' is a modal auxiliary, not a future auxiliary. It will be treated as a modality marker in the annotation guidelines for modality. This means that for a sentence such as in example (8), the tense is *simple present*, not, as is common in in traditional grammar, *present future*.

(8) Ik *zal* wandelen
I will *walk.*

The four tenses are exemplified by examples below. Observe that both the antecedent and consequent can have the same tense, but do not have to.

(9) **Simple present**
Als er genoeg water bij Lobith *binnenkomt*, *staat* de stuw open.
If enough water enters *at Lobith, the weir* is *open.* (simple present, simple present)

(10) **Present perfect**
Als Li dit inderdaad *heeft gezegd*, wat *bedoelde* hij dan?
If Li indeed has said *this, what* did *he* mean*?* (present perfect, simple past)

(11) **Simple past**
De leraren *maakten* bezwerende gebaren als de uitbundigheid binnen of buiten te groot *werd*.
The teachers made *bewildering gestures if there* was *too much exuberance inside or outside.* (simple past, simple past)

(12) **Past perfect**
De Amerikanen zelf *zouden* nooit akkoord *gaan* als Europa een dergelijk voorstel *had gedaan*.
The Americans themselves would *never agree if Europe* had made *such a proposal.* (simple past, past perfect)

(13) **Infinitival phrase**
Rationeel zou zijn om te geloven dat Socrates gestorven is (of om je *te onthouden van* een oordeel, als het je niets *kan interesseren*).
It would be rational to believe that Socrates died (or to refrain from *judgement if you* are *not interested).* (simple present, infinitival)

An important point of attention with respect to verb tense is the combination of the verb *zijn* with a participle, in which case it can either be a copular verb or an auxiliary verb. In the first case, the predicate describes what the subject *is*, in the latter, what the subject *does*. The difference is not always easy to tell and is interpretative. The difference is important for the classification of verb tense, because in case of a copula, the verb tense is present or simple past, in case of an auxiliary, it is present or past perfect.

The first test to use is to reformulate, if necessary, the predicate into a subordinate clause and testing for the acceptability of the so-called 'red and green word order' (cf. Pauwels, 1953; Haeseryn et al., 1997; de Sutter, 2005) (examples adapted from Haeseryn et al., 1997).

(14) Ik heb je toch gezegd dat mijn moeder al jaren {dood is/*is dood}.
I have told you that my mother has been dead for years.

(15) Ik heb je toch gezegd dat mijn moeder al jaren {overleden is/*is overleden}.
I have told you that my mother died years ago.

(16) Ik heb je toch gezegd dat mijn moeder in 1981 {overleden is/is overleden}.
I have told you that my mother died in 1981.

What can be seen in the subordinate clause in (14) is that the adjective has to precede the finite verb – the reverse order is not available. The finite verb in (15) is used as a copula and the participle acts as an adjective, which is reflected in the availability of one word order only – the participle has to precede the finite verb. In (16), however, both the order finite verb-participle – the 'red order' – and participle-finite verb – the 'green order' – is possible, indicating the status of the finite verb as an auxiliary. Consequently, the verb tense in

(15) is *simple present*, while in (16), the verb tense is *present perfect*. Other test that help, although not determine, are the presence or possibility to add duration to the clause, which indicates that the 'being' interpretation is more prominent, consequently viewing the verb *zijn* 'be' as copula and classifying the clause as 'simple present'. Conversely, when the clause has a prepositional phrase indicating an actor or the possibility to add such a phrase, such as '... by her nephew', the most prominent interpretation is that of 'doing' instead of 'being'. Remember that these are only aids in determining the right label, they are by no means perfect tests.

Another complexity is formed by embedded clauses, as in the corpus examples below.

(17) Mohammed is van plan om zijn opleiding op te pakken als hij weer beter is en zich goed voelt.
Mohammed is planning to resume his education if [when] he is well and feels good again.

(18) De lowbudget-maatschappij Ryanair dreigt het populaire vliegveld Charleroi te verlaten als de Europese Commissie haar een boete geeft.
The low-budget airline Ryanair threatens to leave the popular Charleroi airport if the European Commission hands out a fine.

Both in (17) and (18) the question is what the consequent of the conditional is. Is it the full complex clause, or only the embedded clause? In (17), it seems to be the case that Mohammed is planning to get back to school if or when he feels better. It does not seem plausible that he starts planning at the moment he feels better. The same goes for (18). Does the airline company threaten {to leave if the European Committee fines the company}, or does the airline company threaten to leave if the European Committee fines the company? Here too, the former seems more plausible, as it is the fining and leaving between which the conditional connection holds, not between fining and threatening. This results in the extra label 'infinitival' for the tense in the consequent, as the subordinate clause is an infinitival clause.

A.5.2 Instructions

For each item, determine the verb tense of the verb phrase. Annotate the according label. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the numerical labels after the semicolon. They will be converted to their full counterparts when you are done annotating.

Simple present (spr; 1)
The situation takes place during the present-tense interval including speech time. This tense is realised by the finite verb in present tense

and an optional participle. Default time is speech time, but adverbial modification can cancel this. This imperfect tense includes the (simple) future, as can be seen below.

(a) Ik *wandel*.
I {walk/am walking}.

(b) Ik *zal wandelen*.
I will *walk.*

(c) Ik *ben* aan het *wandelen*.
I am walking.

Simple past (spa; 2)

The situation takes place during the past-tense interval. This tense is realised by the finite verb in past tense and an optional participle. Default time is speech time in the past. This imperfect tense includes future in the past, in which the situation takes place in the non-actualised part of the paste-tense interval.

(a) Ik *wandelde*.
I walked.

(b) Ik *zou* vandaag/morgen *wandelen*.
I would *walk today/yesterday.*

(c) Ik *was* aan het *wandelen*.
I was *walking.*

Present perfect (prp; 3)

The situation takes place in the actualised part of the present-tense interval. This tense is realised by one of the auxiliaries *hebben* 'have' or *zijn* 'be' in present tense and a past participle. The situation is completed before speech time, but this default interpretation can be cancelled by adverbial modification, as in (c). This tense includes the future perfect. The focus of this tense is on the completion of a situation.

(a) Ik *heb gewandeld*.
I have walked.

(b) Ik *zal hebben gewandeld*.
I will have walked.

(c) Ik *heb* je paper morgen zeker *gelezen*.
I will *certainly* have read *your paper tomorrow.*

(d) Ik *heb* dat niet *durven doen*.
I have *not* dared doing *that.*

Past perfect (pap; 4)

The situation takes place before the speech-time in the past and is completed within this time span. This tense is realised by the one of the

auxiliaries *hebben* 'have' or *zijn* 'be' in past tense and a past participle. This tense includes the future perfect in the past, in which the situation takes place and is completed in the non-actualised part of the past-tense interval. The focus of this tense is on the completion of a situation.

(a) Ik *had gewandeld*.
I had walked.

(b) Ik *zou hebben gewandeld*.
I would have walked.

(c) Ik *had* dat niet *durven doen*.
I had *not* dared doing *that.*

Infinitival phrase (inf; 5)

In case of an embedded clause with an infinitival phrase (*te* 'to' + infinitive), determine whether the full complex clause or only the embedded clause forms the consequent. In (a) below, the consequent is 'zijn opleiding op te pakken' ('to resume his education'), not 'Mohammed is van plan om zijn opleiding op te pakken' ('Mohammed is planning to resume his education').

(a) Mohammed is van plan om zijn opleiding *op te pakken* als hij weer beteris en zich goed voelt.
Mohammed is planning to resume *his education if [when] he is well and feels good again.*

A.5.3 Problem cases

Please take note of the following known problem cases and code accordingly.

Copula or auxiliary, simple or perfect

In case the finite verb *zijn* 'be' is combined with a participle, it can either be a copula or an auxiliary. See the introduction of this section for instructions.

Incomplete utterance ('NA')

See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')

See the instructions for dealing with incomplete utterances in section A.3.3 above.

A.6 Modality

A.6.1 Introduction

The feature *modality* represent the type(s) of modality expressed in the antecedent and consequent. Modality, is hard to define (see section 5.5), but the following working definition will suffice as a starting point. Modality is the view a speaker presents on the situation expressed, either in relation to reality, or in relation to her attitude.

In this study, we distinguish between four categories of modality, which are somewhat easier to define. The categories are *epistemic*, *evidential*, *deontic* and *dynamic modality*. It is important to remember that modality can be expressed by modal auxiliaries, as well as modal adverbs/adverbial phrases. The four types of modality are exemplified below.

(19) **Epistemic modality**
Als ik erachter *zou* komen dat een school regels stelt om groepen leerlingen te weren, *zou* ik onmiddellijk ingrijpen.
If I were *to find out that a school sets rules to exclude groups of students, I* would *intervene immediately.*

(20) **Evidential modality**
Als *blijkt* dat het geen Sars is, dan wordt er toch veelal een rekening ingediend en dat kunnen boeren onmogelijk betalen.
If it turns out *not to be Sars, then a bill is usually submitted and farmers cannot afford that.*

(21) **Deontic modality**
Als dat zo is *moeten* de depots worden gesloopt.
If that is the case the depots must *be demolished.*

(22) **Dynamic modality**
Als mijn man over mijn buik aait, *wil* ik braken.
If my husband rubs my belly, I want *to vomit.*

A.6.2 Instructions

For each item, identify the antecedent and consequent. For both the antecedent and the consequent, first determine whether or not there modal marking occurs. In case of modal marking, annotate the dominant interpretation of the modal markers in the clause, choosing from the types listed below. The labels are presented between parentheses. The parentheses are not to be included in your annotation.

When you are not sure of the correct label for a certain clause, for instance when the modal marking is ambiguous, please include a short description of the trouble you run into into the comment column. Remember that, in most cases, this is part of the annotation process and the under-specifying nature of language.

Epistemic (epi; 1)
The clause is marked for expressing the speaker's judgement of the factual status of the proposition.

(a) Jan *kan/moet* in zijn kantoor zijn.
John may/must *be in his office.*

(b) Jan is *waarschijnlijk* in zijn kamer.
John is probably *in his office.*

Evidential (evi; 2)
The clause is marked for expressing the speaker's direct or indirect evidence for the factual status of the proposition.

(a) Ik *kan* zien dat Jan in zijn kamer is.
I can *see John is in his office.*

(b) *Volgens Willem* is Jan in zijn kamer.
According to William*, John is in his office.*

Deontic (deo; 3)
The clause is marked for directive meaning, trying 'to get others to do things'.

(a) Jan, je *kan/moet* naar je kantoor gaan.
John, you can/must *go to your office.*

Dynamic (dyn; 4)
The clause is marked for expressing ability or willingness.

(a) Jan *kan* naar zijn kantoor gaan.
John can *go to his office.*

(b) *Hopelijk* gaat Jan naar zijn kantoor.
Hopefully*, John will go to his office.*

A.6.3 Problem cases

Please take note of the following known problem cases and annotate accordingly.

Incomplete utterance ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')

See the instructions for dealing with incomplete utterances in section A.3.3 above.

Ambiguity of *mogen* 'may'

The modal verb *mogen* 'may' can be used to express deontic modality, epistemic modality (especially in past tense, i.e. *mochten* 'should', and sometimes evidential modality), as in the examples below respectively. Please take the whole clause into account when deciding on the most appropriate type of modality.

(a) Als ik dezelfde achternaam had, met al een site op die naam, dan *mocht* je echt iets anders gaan verzinnen.
If I had the same last name, with a site already under that name, then you should *you really come up with something else.*

(b) Alleen als de Arnhemmers onverhoopt *mochten* degraderen, gaat de reddingsplan op de helling.
Only if the Arnhemmers were *to be relegated unexpectedly, the rescue plan will be overruled.*

(c) Als ik Kelly *mag* geloven ga ik het feest van het jaar missen, maar ik heb 't er maar mee te doen.
If I may *believe Kelly I am going to miss the party of the year, but there's nothing I can do about it.*

Double modal marking

Sometimes, one clause contains more than one marker of modality, as in the consequent of the conditional in the example below, in which *moeten* 'must' marks deontic modality (i.e. obligation), but combined with *waarschijnlijk* 'likely' clearly concerns knowledge.

(a) Als ik bijvoorbeeld evenveel reden heb om te denken dat mijn kat in Doos 1 kroop dan dat het Doos 2 was, dan *moet* ik het even *waarschijnlijk* achten dat de kat in Doos 1 zit, als dat ze in Doos 2 zit.
If I for example have as much reason to think that my cat crawled into box 1 as into box 2, then I must *consider it as* likely *that the cat is in box 1 as it is in box 2.*

The most appropriate annotation here is thus to use the label for epistemic modality.

Isolating clauses

A problem that may arise in annotation of modality is that it 'spills over' from one clause to another. Yet, the task at hand is to look at modal marking in isolated clauses. So, in the example below, one could argue that the consequent is marked for dynamic modality, as 'making the trading' is dependent on the ability referred to in the antecedent.

(a) Als we *konden* ruilen, *zou* ik het doen.
If we could *trade, I* would *do it.*

However, when viewed in isolation, the consequent is marked for epistemic modality by means of the auxiliary *zou* 'would'. Therefore, the most appropriate annotation for the consequent is here to use to label for epistemic modality.

A.7 Aspect

A.7.1 Introduction

The feature *aspect* concerns the internal-temporal characteristics of the situations presented in the antecedent and in the consequent.[1] The type of aspect to be annotated is also known as *actionality*, *lexical aspect*, *situational aspect*, and *Aktionsart*. It is not marked formally, but part of the lexical semantics of the verb (phrase). The grammatical part of aspect consists of perfective and imperfective aspect and is not considered here.

In this study, we distinguish between four main categories ('Vendler classes') of aspect: *states*, *activities*, *achievements* and *accomplishments*. The category is based on the combination of three binary properties. First, a situation is stative (no change over time) or dynamic (change over time). Second, a situation has a duration, i.e. it is durative (extends in time) or it is punctual (one point in time). Third, a situation has a telicity value, i.e. it is bounded (natural endpoint) or unbounded (no natural endpoint). The features of all four types are summarised in the table below.

Table A.1:
Characteristics of situation types (aspect)

	Change	Duration	Telicity
State	Stative	Durative	Atelic
Activity	Dynamic	Durative	Atelic
Accomplishment	Dynamic	Durative	Telic
Achievement	Dynamic	Punctual	Telic

The four types of aspect are exemplified below. In (23) both the antecedent and the consequent express stative situations referring to characteristics (liking jazz, being a fan of Sonny Clark). In (24), both clauses express express activities, as both running and being distracted are durative, change over time, but

[1] This guideline is based on collaborative work with M.P.M. Bogaards, who, as part of a research internship, annotated a number of features and developed an extensive guideline for annotating aspect. This section presents a practical, somewhat shortened version of this extense guideline.

have no inherent endpoint. In (25), both clauses express accomplishments, as running is an activity with an inherent endpoint. In (26), finally, both clauses express achievements, as winning and losing involve a change of states and refer to an inherent endpoint, but are not durative.

(23) Als je van jazz houdt, dan ben je vast fan van Sonny Clark.
If you like jazz, then you will appreciate Sonny Clark.

(24) Als ik aan het hardlopen ben, wil ik niet worden afgeleid.
If I am running, I do not want to be distracted.

(25) Als zij een kilometer rent, rent hij er twee.
If she runs for a kilometre, he runs for two.

(26) Als zij de wedstrijd wint, heeft hij verloren.
If she wins the game, he will have lost.

A.7.2 Instructions

For each item, identify the antecedent and consequent. For both the antecedent and the consequent, indicate the type of event expressed by the *main verb*, its direct object (if the verb is transitive), and the grammatical subject. In case of non-main verbs, make sure not to identify the correct main verb. In the consequent of the conditional in (27) below, the main verb is *halen* 'get', not *kan* 'can', a modal auxiliary.

(27) [...] We hebben brood in huis maar je kan nog wat brood bij halen als je wilt.
We have bread at home, but you can get some more bread if you want.

Choose from *state*, *activity*, *achievement* or *accomplishment*. In (27), the antecedent expresses a state (you want bread), and the consequent express an accomplishment (you get some more bread). Below the coding instructions are presented, together with examples from the corpus. The labels are presented between parentheses. The parentheses are not to be included in your annotation.

When you are not sure of the correct label for a certain clause, for instance when the type of event is ambiguous or there are multiple states in one clause, please include a short description of the trouble you run into. Remember that, in most cases, this is part of the annotation process and the under-specifying nature of language.

State (sta; 1)
The event does not change over time (stative), extends over time (durative) and has no natural endpoint (unbounded). The event cannot be expressed with a present progressive ('I am V-ing') as an answer to the question 'What are you doing?', e.g., 'What are you doing? I am loving

Mary'. The event can be used in the question 'For how long has Subject already Vpart?', as in 'For how long has John already loved Mary?' The present tense cannot be interpreted habitually, e.g., 'John is ill' cannot mean that John is ill often or recurrent. Adverbially modifying the verb by 'in an hour' leads to incoherence, e.g., 'John was ill in an hour'. Finally, the event cannot be a complement of 'to finish', e.g., 'John finished being ill'.

Examples of verbs that typically denote states are *begrijpen* 'to understand', *bezitten* 'to own', *haten* 'to hate', *hebben* 'to have', *horen* 'to hear', *geloven* 'to believe', *houden van* 'to love', *kennen* 'to know', *leven* 'to live', *verlangen* 'to desire' and *weten* 'to know' (cf. Broekhuis, Corver & Vos, 2015a, p. 37).

(a) Als je nou een heel kaal huis hebt dan is dat wel leuk.
If you have a very empty house then that is nice.

Activity (act; 2)

The event changes over time (dynamic), extends over time (durative) and has no natural endpoint (unbounded/non-telic). Adverbially modifying the verb by 'almost' or 'in an hour' leads to incoherence, e.g., 'John almost ran', 'John ran in an hour'. The event cannot be a complement of 'to finish', e.g., 'John finished looking for a restaurant'. The event cannot be used with 'within an hour', as in 'John ran within an hour', but it can be used with 'during an hour', as in 'John ran during an hour'.

Examples of verbs typically denoting activities are *bibberen* 'to shiver', *denken (over)* 'to think (about)', *dragen* 'to carry', *duwen* 'to push', *hopen* 'to hope', *eten* (intransitive) 'to eat', *lachen* 'to laugh', *lezen* (intransitive) 'to read', *luisteren* 'to listen', *praten* 'to talk', *rennen* 'to run', *schrijven* (intransitive) 'to write', *sterven* 'to die', *wachten (op)* 'to wait (for)', *wandelen* 'to walk', *zitten* 'to zit' (cf. Broekhuis, Corver & Vos, 2015a, p. 37).

(a) Als jij een spelletje doet, ga ik maar tegelijk even pokeren.
If you play a game, I will play poker at the same time.

Accomplishment (acc; 3)

The event changes over time (dynamic), extends over time (durative) and has a natural endpoint (bounded/telic). The event cannot be used in the question 'For how long did ... V?', as in 'For how long did John run a kilometre?' The event cannot be used with 'during an hour', as in 'John ran a kilometer during an hour', but it can be used with 'within an hour', as in 'John ran a kilometre within an hour'.

Examples of verbs typically denoting accomplishments are *bouwen* 'to build', *eten* (transitive) 'to eat', *koken* (transitive) 'to cook', *lezen* (transitive) 'to read', *opeten* 'to eat up', *schrijven* (transitive) 'to write', *oversteken* 'to cross', *verbergen* 'to hide', *verorberen* 'to consume' and *zingen* (transitive) 'to sing' (cf. Broekhuis, Corver & Vos, 2015a, p. 37).

(a) En als we dan toch de computer gaan halen bij Gertie en Jeroen dan kunnen we misschien gelijk ik weet niet of je dat vandaag lukt om de computer van ouders af te maken.
And if we do go and get the computer from Gertie and Jeroen then we might be able to, I don't know if you can do that today, finish the parents' computer.

Achievement (ach; 4)

The event changes over time (dynamic), does not extend over time (punctual) and has a natural endpoint (bounded/telic). Adverbially modifying the verb by 'for an hour' leads to incoherence, e.g., 'John ran a mile for an hour'. Adverbially modifying the verb by 'almost' leads to ambiguity, e.g., 'John almost ran a mile' may mean either 'John almost started running a mile' or 'John ran a distance close to a mile'. The event cannot be a complement of 'to finish', e.g., 'John finished reaching the top'. Adding the verb 'stop' leads to incoherence (e.g., 'John stopped reaching the top'). The event cannot be used in the question 'For how long has Subject already Vpart?', as in 'For how long has John already reached the top?'

Examples of verbs typically denoting achievements are *aankomen* 'to arrive', *beginnen* 'to start', *bereiken* 'to reach', *botsen* 'to collide', *herkennen* 'to recognize', *ontploffen* 'to explode', *ontvangen* 'to receive', *overlijden* 'to die', *zich realiseren* 'to realize', *stoppen* 'to stop', *opgroeien* 'to grow up', *vinden* 'to find', *winnen* 'to win' and *zeggen* 'to say' (cf. Broekhuis, Corver & Vos, 2015a, p. 37).

(a) Als mijn wifi dan een keer niet overschakelt, wil ik zelf kunnen overrulen.
If my WiFi does not switch for once, I want to be able to overrule it myself.

A.7.3 Problem cases

Please take note of the following known problem cases and code accordingly.

Incomplete utterance ('NA')

See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')

See the instructions for dealing with incomplete utterances in section A.3.3 above.

Non-literal use of verbs

Sometimes, verbs are used non-literally to refer to a situation, as in the figurative use of *slepen* 'to drag' below.

(a) 'Iedere bezoekende buitenlandse delegatie dreigt Beshir en Taha dat ze voor een internationaal gerechtshof *worden gesleept* als ze de Janjaweed in Darfur niet ontwapenen', zegt hij met genoegen. (WR-P-P-G-0000096092)
'Every visiting foreign delegation threatens Beshir and Taha to be dragged *before an international court if they do not disarm the Janjaweed in Darfur', he says with delight.*

Here, 'to drag' does not refer to a physical activity, but to a decision and thus an achievement. Another eample is the durative verb *to say* being used to express a (punctual) decision, as in the example below.

(b) En als dan *gezegd wordt* ja maar dan hoeven we de lasten niet nog meer te verhogen voorzitter dan denk ik dat een gemeente welke dat ook is want dan denk ik dat een gemeente ook moet kijken wat hun ambitieniveau kost. (fn000151)
And if it is said *yes but then we do not have to increase the burden even more, chairman, then I think that a municipality whichever that is, because then I think that a municipality should also look at what their level of ambition costs.*

As the intended interpretation is one of decision, annotate such examples as achievements.

(Semi-)aspectual non-main verbs

Non-main verbs may express aspectual information. Such semi-aspectual non-main verbs are the following posture auxiliaries: *zitten* 'sitting', *staan* 'standing', *liggen* 'lying', *lopen* 'walking' and *hangen* 'hanging' plus *te* 'to' and an infinitive, as in the example below.

(a) Kijk als ik hier avonds zo effe *zit te lezen* of *TV zit te kijken* joh dan hoor ik ze lachen of weet ik veel wat maar da niet erg.
Look if I sit here and read or watch TV *then come on I hear them laughing or I don't know but that's OK.*

In such cases, include 'semi-aspectual non-main verb' in the comment column and annotate the main predicate, here *zit te lezen of TV zit te kijken* 'sit and read or watch tv', i.e. an activity. The same goes for the

aspectual non-main verbs *gaan* 'going', *komen* 'come', *blijven* 'stay' + infinitive, *bezig zijn te*+infinitive, and *aan het* + infinitive + *zijn*, as in the following example.

(b) Je moet goed uitkijken als je *aan het schommelen bent*.
You should be very cautious if [when] you're playing on the swings.

In this example, the main verb *schommelen* 'playing on the swings' refers to an activity.

A.8 Person and number

A.8.1 Introduction

In this study, the feature *person and number* is defined by the combination of the person and number of the grammatical subject, i.e. the noun phrase congruent with the finite verb. The feature *person* represents the grammatical distinction between speaker (first person), addressee (second person) and other entities talked about (third person). The feature *number* represents the grammatical category that refers to quantity in a binary fashion, either *singular* or *plural*. The six combinations of person (first, second, third) and number (singular, plural) are exemplified by corpus examples below.

(28) **First-person singular**
Als *ik* een pistool of mes had gehad, had *ik* dat gebruikt. (1ps, 1ps)

(29) **First-person plural**
Als *we* dat weten, kunnen *we* besluiten het wel of niet te doen. (1pp, 1pp)

(30) **Second-person singular**
Als *je* te weinig rendement toont, word *je* eruit gegooid.
Als *je* niet doet wat ik zeg, rot dan maar op.(2ps, 2ps)

(31) **Second-person plural**
Als *jullie* dan ook nog op of andere frommelmatras liggen dan volgens mij doet dan *niemand* een oog dicht. (2pp, 3ps)

(32) **Third-person singular**
Als *mijn broertje* een paar blauwe plekken had, dacht *ze* meteen aan leukemie. (3ps, 3ps)

(33) **Third-person plural**
Als *het* drukker wordt, gaan *vader en moeder* gewoon wat harder werken. (3ps, 3pp)

As can be seen in (32), both the grammatical subject *ze* 'she' and the finite verb *dacht* 'thought' are used to distinguish between singular and plural subjects.

A.8.2 Instructions

For both the antecedent and consequent in each item, find the grammatical subject and finite verb and determine the person and number of the grammatical subject. Annotate the according label. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the numbers after the semicolon. They will be converted to their full counterparts when you are done annotating.

First-person singular (1ps; 1)
The subject is a noun phrase with the pronoun *ik* 'I' as its head and refers to the speaker in singular form.

(a) *Ik* zal heel blij zijn als *ik* dit altijd uit mijn mouw schud.
I *will be very happy if* I *always do this easily.*

First-person plural (1pp; 2)
The subject is a noun phrase with either the pronoun *we* 'we' or *wij* 'we' as its head and refers to the speakers or speaker and associated entities in plural form.

(a) We zouden wel een inconsistentie krijgen als we ook het volgende zouden aannemen.
We *would have an inconsistency if* we *also assumed the following.*

Second-person singular (2ps; 3)
The subject is a noun phrase with either the pronoun *je* 'you', *jij* 'you' or *u* 'you' (polite) as its head and refers to the addressee or to an unspecified entity ('generic or impersonal *you*', only with *je* and *u*) in singular form.

(a) Als *je* de nieuwste features niet belangrijk vindt, verdien *je* hier geld mee en kun je een klasse groter krijgen.
If you *do not deem the latest features important,* you *will earn money with this and you can increase the class*

Second-person plural (2pp; 4)
The subject is a noun phrase with the pronoun *jullie* 'you' as its head and refers to the addressees or addressee and associated entities in plural form.

(a) Laten *jullie* het even weten als *jullie* klaar zijn met praten over voetbal?
Will you *let us know if [when]* you *have finished talking about football?*

Third-person singular (3ps; 5)
The subject is a noun phrase with a noun or pronoun as its head and refers to an entity that is not speaker or addressee in singular form. The subject can also be an infinitival construction or a subject clause.

(a) Als *een oplichtend natriumatoom* als een soort trillend elektrisch deeltje kan worden opgevat, zal *een magneet* die beweging opsplitsen in twee iets verschillende trillingen.
If a glowing sodium atom *can be seen as some sort of vibrating electrical particle,* a magnet *will split that movement into two slightly different vibrations.*

Third-person plural (3pp; 6)
The subject is a noun phrase with a noun or pronoun as its head and refers to an entity that is not speaker or addressee in plural form. The subject can also be an infinitival construction or a subject clause.

(a) En als *varkens* eenmaal bloed hebben geproefd, willen *ze* meer.
And if [once] *pigs* have tasted blood, *they* want more.

A.8.3 Problem cases

Please take note of the following known problem cases and code accordingly.

Incomplete utterance ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Embedded infinitival clauses
If the conditional is embedded in a clause introduced by implicit or explicit *om*, frequently introduced by communicative or mental verbs, as in the example below, and the consequent is an infinitival clause, use the 'NA'label.

(a) Wij adviseren u dan ook [om] cd 1 pas terug te sturen als *u* het hele boek uit heeft. (2ps, NA)
We therefore advise you [to] only return CD 1 if [when] you *have finished the entire book.*

Covert grammatical subject
In case a sentence does not include an overt subject, try to use the finite verb to determine person and number. If the finite verb allows for multiple interpretations, use 'NA', as would be the case in the corpus example in (a) below, in which the finite verb *opzijzet* 'sets aside' can be congruent with a first-person, second-person or third-person singular subject.

(a) Dus als dat *opzijzet* dan is dat goed.
so if that sets aside *then that's good.*

Imperative consequent

In case of imperative clauses, no subject is overtly present. In line with the literature on imperatives, the implicit subjects of imperatives are annotated here as second-person singular.

(a) Als *een verslaafd kind* je alles van je weg heeft geroofd en een psychish wrak van je heeft gemaakt, okee, bekijk dan nog eens opnieuw wat de opties zijn. (3ps, 2ps)
If an addicted child *has stolen everything from you and turned you into a psychic wreck, okay, check out the options again.*

Embedded clauses

If the antecedent and/or consequent have embedded clauses, only the subject and finite verb of the matrix clause are to be considered.

(a) Als *ik* denk dat zij dat al weten, dan heb *ik* er geen zin meer in. (1ps, 1ps)
If I *think that they already know,* I *don't feel like it anymore.*

A.9 Sentence type

A.9.1 Introduction

The feature *sentence type* represents the illocution of a sentence, which is reflected mainly in the word order of the consequent. In this feature, four types are distinguished, exemplified below: *declarative*, *imperative*, *interrogative* and *exclamatory* sentences.

(34) **Declarative sentence**
Als het mooi weer is, *(dan) gaan we wandelen.*
If the weather is nice, we will go for a walk.

(35) **Imperative sentence**
Als het mooi weer is, *pak (dan) je wandelschoenen.*
If the weather is nice, get your hiking boots.

(36) **Interrogative sentence**
Als het mooi weer is, *gaan we (dan) wandelen?*
If the weather is nice, will we go for a walk?

(37) **Exclamatory sentence**
Als het mooi weer is, *hoe mooi kan het leven (dan) zijn!*
If the weather is nice, how wonderful can life be!

A.9.2 Instructions

Determine the most appropriate sentence type. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the numbers after the semicolon. They will be converted to their full counterparts after you are done annotating.

Declarative sentence (dec; 1)
A declarative consequent makes an assertion of a proposition. As we are dealing with conditionals, the assertion is dependent on the antecedent. The word order of the consequent can be that of a regular declarative sentence (i.e. a main clause with subject, verb, object (svo) order, as in the second example below), but it can also have an integrative word order with subject-verb inversion (vso), as in first example below, and a resumptive element featuring *dan* 'then' and integrative word order.

(a) Als je de baby neerzet, *gaat ze schreeuwen.*
If you put the baby down, she'll scream.

(b) Als je tijd hebt, *Marie staat te wachten.*
If you have time, Mary is waiting.

Imperative sentence (imp; 2)
An imperative consequent gives a command or makes a request. There is no overt subject, as in (a), or there is a second-person subject. Next to the standard form of the imperative, there are three other possibilities: infinitival, participial, and adverbial imperatives, as in the examples below respectively.

(a) Als u de aangifte nog niet heeft verstuurd, *doe dat dan zo snel mogelijk.*
If you haven't already sent in the tax return, do so as soon as possible.

(b) Als de bakplaat heet is, afblijven!
If the baking tray is hot, hands off!

(c) Als je geen virusscanner hebt, *opgepast!*
If you haven't got an antivirus programme, be warned!

(d) Als jullie nu nog niet weg zijn, *naar buiten!*
If you still haven't left, outside!

Interrogative sentence (int; 3)
An interrogative consequent presents a question. The question begins with a *wh*-word (*wat* 'what', *wie* 'who', *wanneer* 'when', *waarom* 'why' et cetera), as in the first example below, or a finite verb, as in the second example.

(a) Als je op de knop drukt, *wat gebeurt er dan?*
If you press the button, what will happen?

(b) Als je op de knop drukt, *gaat dan het alarm af?*
If you press the button, will the alarm go off?

Exclamatory sentence (exc; 4)
An exclamatory consequent expresses an emotion. It can begin with a *wh*-word, a subordinate conjunction or a qualification of the addressee. Note that the term *exclamation* here refers to a functional category. The term is not as strict as with 'pure exclamatives' see, Broekhuis and Corver (2016, pp. 1481–1486).

(a) Als ik in je kamer kijk, *wat een stof ligt daar!*
If I look into your room, what an amount of dust!

(b) Als ik in je kamer kijk, *dat je daar kunt leven!*
If I look into your room, that you live there!

(c) Als ik in je kamer kijk, *sloddervos die je bent!*
If I look into your room, you slob!

A.9.3 Problem cases

Please take note of the following known problem cases and annotate accordingly.

Incomplete utterance ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

One-word consequent (wrd; 5)
In some cases, the consequent consists of only one word or word group, like the adverb *jammer* 'pity' in (a). The first step in these cases is to whether or not we are dealing with an imperative, as in (b), or an exclamation, as in (a), or an interrogative consequent, as in (c). If any of these options is available, annotate accordingly. If this is not the case, as in the first example below, we use label the sentence type 'one-word consequent', although the consequent can also consist of more than one word or constituent, as in the last example below.

(a) Als je dan niet kunt, *jammer!*
If you can't make it then, pity!

(b) Als de politie komt, *stop!*
If the police comes, stop!

(c) Als je een van die relschoppers bent, *waarom?*
If you are one of those hooligans, why?

(d) *Ja*, als je tenminste bedoelt dat ik dan mee mag.
Yes*, if at least you mean that I can come too.*

(e) Als je denkt dat ik een processor kan overklokken, *nee man.*
If you think that I can overclock a processor, no man.

A.10 Negation

A.10.1 Introduction

The feature *negation* represents the *polarity* of the antecedent and the consequent. In this feature, three types of negation are distinguished: *morphological*, *syntactic*, and *implicit* or *lexical* negation, as exemplified below.

(38) **Morphological negation**
Als de deur dicht is, is het *on* mogelijk om binnen te komen.
If the door is closed, it is im *possible to enter.*

(39) **Syntactic negation**
Als de deur dicht is, is het *niet* mogelijk om binnen te komen.
If the door is closed, it is not *possible to enter.*

(40) **Lexical negation**
Als de deur dicht is, ben je *buitengesloten*.
If the door is closed, you are locked out*.*

A.10.2 Instructions

Determine whether or not a clause contains negation and if so, which type is the most appropriate. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the letters after the semicolon. They will be converted to their full counterparts after you are done annotating.

Morphological negation (mor; 1)
The clause has an element with one of the following prefixes: *anti-*, *de-*, *on-*, *dis-*, *mis-*, *non-*, *niet-*, *in-*, *a-*, *ir-*, or *wan* or one of the following suffixes: *-loos*, *-arm*, or *-vrij*.

(a) Als het piept, moet je het *de* monteren.
If it squeals, you must disassemble it.

(b) Als je dat denkt, zeg ik je dat het *on* waar is.
If you think so, I'll tell you it's untrue.

(c) Als je zo werkt, noemen we dat *dis* functioneren.
If you work like that, we call that dysfunctional.

(d) Als je het zo aanpakt, zal het *mis* lukken.
If you do it like that, it will fail.

(e) Als je de aantallen bekijkt, zie je dat een van de soorten *non*-existent is.
If you look at the numbers, you'll see that one of the species is non-existent.

(f) Als je de foto's ziet, word je vanzelf een *niet*-roker.
If you see the pictures, you'll automatically become a non-smoker.

(g) Als je net begint, ben je waarschijnlijk *in* capabel.
If you are just starting out, you're probably incapable.

(h) Als je dat op straat gooit, vind ik dat *a* sociaal.
If you throw that on the street, I find that unsocial.

(i) Als zelfs jij dat denkt, is het misschien *on* waar.
If even you think so, it might be untrue.

(j) Als het niet rationeel is, is het *ir* rationeel.
If it is not rational, it is irrational.

Syntactic negation (syn; 2)

The clause includes one of the following explicit negations: *niet*, *geen*, *niets*, *nooit*, *niemand*, or *nergens*. In cases of double negation, take the negation with widest scope.

(a) Als u de aangifte nog *niet* heeft verstuurd, doe dat dan zo snel mogelijk.
If you have not yet sent the declaration, do so as soon as possible.

(b) Als u *geen* oven heeft, gebruik dan de magnetron.
If you do not have an oven, use the microwave.

(c) Als je *niemand* wilt zien, verstop je dan.
If you do not want to see anyone, then hide.

(d) Als je alles opeet, is er *niets* meer over.
If you eat everything, there will be nothing left.

(e) Als zij *nooit* naar buiten gaat, is ze misschien wel ziek.
If she never goes out, she might be sick.

(f) Als je *geen* landkaart hebt, kom je *nergens*.
If you do not have a map, you will get nowhere.

Lexical negation (lex; 3)

The clause includes a lexical item expressing a negative meaning. The following list provides a guideline of what counts as lexical negation, but

no exhaustive set of negative items: *achterwege laten*, *aflopen*, *afstaan*, *alleen (maar)*, *allerminst*, *amper*, *anders*, *behalve*, *behoeden (voor)*, *buiten*, *enkel*, *evenmin*, *gebrek (aan)*, *het minst*, *hoogstens*, *maar*, *minder*, *missen*, *moeilijk*, *nauwelijks*, *negatief*, *niemendal*, *niettemin*, *niks*, *nimmer*, *noch*, *onthouden (van)*, *ophouden*, *opraken*, *opschorten*, *pas*, *slecht*, *slechts*, *stoppen*, *stopzetten*, *tegen*, *tegenvallen*, *ternauwernood*, *twijfelen*, *uitsluiten*, *verbieden*, *verbreken*, *verdwijnen*, *vergeten*, *verliezen*, *vervallen*, *verwerpen*, *voorkomen*, *weghalen*, *wegnemen*, *wegvallen*, *weinig*, *zelden*, *zinloos*, *zomin*, and *zonder*. Please use the space in the comments section to mark unclear cases.

(a) Als Petra zoveel werkt, heeft ze *amper* tijd voor iets anders.
If Petra works so much, she hardly has time for anything else.

(b) Je verdient *allerminst* een bonus als je zo tekeergaat.
You least of all deserve a bonus if you go berserk like that.

(c) Als er geen uitnodigingen worden verstuurd, komen er maar *weinig* bezoekers.
If no invitations are sent, there will be few visitors.

(d) Hij kon het maar *moeilijk* verwerken als zij hem weer eens bedroog.
He could hardly handle it if she deceived him once again.

(e) De computer hoeft *slechts* het pad te berekenen als op de knop wordt gedrukt.
The computer only needs to calculate the path if the button is pressed.

A.10.3 Problem cases

Please take note of the following known problem cases and annotate accordingly.

Incomplete utterance ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Delimiting lexical negation
Once one starts annotating lexical negation, the boundary between what is and what is not negation may start to shift, as can be seen in the example below.

(a) Er gaat pas een significante stroom lopen als de kring *gesloten* is.
A significant current will only start if the circuit is closed.

In (a), one can interpret 'a circuit being closed' as the negation of 'a circuit being open'. The question then arises where such 'negation' ends. In this case, discussion led to removing the label for negation, but there is no principled boundary. In cases like (a) therefore, document your decision in the comments column.

A.11 Focus particles

A.11.1 Introduction

The feature *focus particle* represents whether or not the conditional is in scope of a focus particles like *alleen* 'only', *zelfs* 'even' and *altijd* 'always', as exemplified below.

(41) *{Zelfs/ook/behalve}* als de deur dicht is, tocht het.
{Even/also/except *if the door is closed, there is a draft.*

(42) *{Alleen/altijd/zeker/slechts/vooral/enkel/pas/met name}* als de deur open is, tocht het.
{Only/always/certainly/only/especially/only/.../especially} *if the door is open, there is a draft.*

(43) Hij benut elke kans, *bijvoorbeeld* als hij iets met winst kan verkopen.
He uses every opportunity, for example *if can sell something with a profit.*

(44) Daar heeft hij helemaal gelijk in, als je *tenminste* naar de consumptie-voetafdruk kijkt.
He is absolutely right about that, if at least *you look at the consumption footprint.*

Although there are three types of focus particles (restrictive, additive and recurrent), it is sufficient to annotate the particle itself.

A.11.2 Instructions

This is a lexical feature and it has been automatically indexed. The following particles were included: *al* 'already', *alleen* 'only', *altijd* 'always', *behalve* 'except', *bijvoorbeeld* 'for example', *elke keer* 'every time', *enkel* 'only', *helemaal* 'completely', *iedere keer* 'every time', *juist* 'especially', *meestal* 'mostly', *met name* 'especially', *ook* 'also', *pas* 'only', *precies* 'precisely', *slechts* 'only', *telkens* 'every time', *tenminste* 'at least', *vooral* 'especially', *zeker* 'certainly' and *zelfs* 'even'. However, the script used may result in a number of false positives, especially because the scope of the particle is not assessed. As it is often, but not always the case that the focus particle directly precedes the conjunction *als* 'if', the main instruction is to remove annotations for sentences in which the lexical item indexed does not function as a focus particle.

A.11.3 Problem cases

Please take note of the following known problem cases and annotate accordingly.

Incomplete utterance ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Running astray ('NA')
See the instructions for dealing with incomplete utterances in section A.3.3 above.

Scope
In certain cases, the focus particle does not (directly) precede the conditional conjunction. See (44) for an example. Determine whether or not the conjunction is in scope of the particle by placing the particle in front of the conjunction. In case of (44), this is the case, as can be seen below.

(a) Daar heeft hij helemaal gelijk in, *tenminste als* je naar de consumptie-voetafdruk kijkt.
He is absolutely right about that, at least if *you look at the consumption footprint.*

APPENDIX B

Feature distributions

B.1 Introduction

In this appendix, the distributions of features (clause order, syntactic integration, verb tense, modality, aspect, person and number, sentence type, negation and focus particles) are presented for detailed 'table look-up' (see section 4.6). See chapter 5 for discussion of the features.

B.2 Feature distributions by mode and register

Table B.1:
Distribution of clause orders by mode and register

		Clause order								
Mode	Register	Initial	%	Medial	%	Final	%	Insub.	%	Total
Spoken	Formal	710	*63.68*	63	*5.65*	329	*29.51*	13	*1.17*	1115
	Informal	660	*59.57*	28	*2.53*	312	*28.16*	108	*9.75*	1108
	Total	1370	*61.63*	91	*4.09*	641	*28.83*	121	*5.44*	2223
Written	Formal	655	*53.89*	23	*1.86*	553	*44.81*	3	*0.24*	1234
	Informal	676	*55.87*	23	*1.90*	481	*39.75*	30	*2.48*	1210
	Total	1331	*54.46*	46	*1.88*	1034	*42.31*	33	*1.35*	2444
Total		2701	*57.87*	137	*2.94*	1675	*35.89*	154	*3.30*	4667

Note. Percentages are row-based.

Table B.2:
Distribution of degrees of syntactic integration by mode and register

		Degree of integration						
Mode	Register	Integration	%	Resumption	%	Non-int.	%	Total
Spoken	Formal	230	*38.02*	348	*57.52*	27	*4.46*	605
	Informal	155	*25.62*	431	*71.24*	19	*3.14*	605
	Total	385	*31.82*	779	*64.38*	46	*3.80*	1210
Written	Formal	463	*75.65*	144	*23.53*	5	*0.82*	612
	Informal	449	*77.55*	1191	*20.55*	11	*1.90*	579
	Total	912	*76.57*	263	*22.08*	16	*1.34*	1191
Total		1297	*54.02*	1043	*43.40*	62	*2.58*	2401

Note. Percentages are row-based.

Table B.3:
Distribution of verb tenses by mode and register

Mode	Register	Tense (a)	Tense (c)										
			Simple present	%	Present perfect	%	Simple past	%	Past perfect	%	Infinitive	%	Total
Spoken	Formal	Simple present	839	*91.10*	11	*1.19*	52	*5.65*	3	*0.33*	16	*1.74*	921
		Present perfect	36	*85.71*	4	*9.52*	1	*2.38*	0	*0.00*	1	*2.38*	42
		Simple past	42	*39.62*	0	*0.00*	58	*54.72*	3	*2.83*	3	*2.83*	106
		Past perfect	0	*0.00*	0	*0.00*	4	*21.05*	14	*73.68*	1	*5.26*	19
	Informal	Simple present	787	*95.74*	9	*1.09*	22	*2.68*	1	*0.12*	3	*0.36*	822
		Present perfect	27	*87.10*	2	*6.45*	2	*6.45*	0	*0.00*	0	*0.00*	31
		Simple past	20	*25.64*	0	*0.00*	53	*67.95*	3	*3.85*	2	*2.56*	78
		Past perfect	0	*0.00*	0	*0.00*	0	*0.00*	13	*100*	0	*0.00*	13
		Total	1751	*86.17*	26	*1.28*	192	*9.45*	37	*1.82*	26	*1.28*	2032

Note. Percentages are row-based. Due to size, this page presents part 1 of 2 parts of the full table.

Mode	Register	Tense (a)	Tense (c)										
			Simple present	%	Present perfect	%	Simple past	%	Past perfect	%	Infinitive	%	Total
Written	Formal	Simple present	884	*93.35*	6	*0.63*	38	*4.01*	0	*0.00*	19	*2.01*	947
		Present perfect	73	*96.05*	1	*1.32*	1	*1.32*	0	*0.00*	1	*1.32*	76
		Simple past	41	*28.08*	2	*1.37*	98	*67.12*	3	*2.05*	2	*1.37*	146
		Past perfect	3	*7.89*	0	*0.00*	9	*23.68*	26	*68.42*	0	*0.00*	38
	Informal	Simple present	890	*91.47*	7	*0.72*	56	*5.76*	0	*0.00*	20	*2.06*	973
		Present perfect	44	*97.78*	0	*0.00*	1	*2.22*	0	*0.00*	0	*0.00*	45
		Simple past	13	*17.81*	0	*0.00*	54	*73.97*	4	*5.48*	2	*2.74*	73
		Past perfect	3	*21.43*	0	*0.00*	2	*14.29*	9	*64.29*	0	*0.00*	14
		Total	1951	*84.39*	16	*0.69*	261	*11.29*	42	*1.82*	44	*1.90*	2312
Total			3702	*85.22*	42	*0.97*	451	*10.38*	79	*1.82*	70	*1.61*	4344

Note. Percentages are row-based. Due to size, this page presents part 2 of 2 parts of the full table.

Table B.4:
Distribution of modality by mode and register

			Modality (c)										
Mode	Register	Modality (a)	Epistemic	%	Evidential	%	Deontic	%	Dynamic	%	No	%	Total
Spoken	Formal	Epistemic	37	*40.22*	1	*1.09*	10	*10.87*	7	*7.61*	37	*40.22*	92
		Evidential	4	*17.39*	1	*4.35*	1	*4.35*	2	*8.70*	15	*65.22*	23
		Deontic	5	*21.74*	0	*0.00*	0	*0.00*	0	*0.00*	18	*78.26*	23
		Dynamic	9	*11.54*	1	*1.28*	13	*16.67*	10	*12.82*	45	*57.69*	78
		No	151	*17.34*	16	*1.84*	92	*10.56*	85	*9.76*	527	*60.51*	871
	Informal	Epistemic	15	*34.88*	0	*0.00*	2	*4.65*	5	*11.63*	21	*48.84*	43
		Evidential	0	*0.00*	0	*0.00*	1	*20.00*	2	*40.00*	2	*40.00*	5
		Deontic	1	*4.76*	1	*4.76*	2	*9.52*	6	*28.57*	11	*52.38*	21
		Dynamic	9	*9.18*	0	*0.00*	24	*24.49*	17	*17.35*	48	*48.98*	98
		No	91	*11.79*	14	*1.81*	67	*8.68*	79	*10.23*	521	*67.49*	772
		Total	322	*15.89*	34	*1.68*	212	*10.46*	213	*10.51*	1245	*61.45*	2026

Note. Percentages are row-based. Due to size, this page presents part 1 of 2 parts of the full table.

			Modality (c)										
Mode	Register	Modality (a)	Epistemic	%	Evidential	%	Deontic	%	Dynamic	%	No	%	Total
Written	Formal	Epistemic	38	*42.70*	3	*3.37*	8	*8.99*	5	*5.62*	35	*39.33*	89
		Evidential	7	*26.92*	0	*0.00*	2	*7.69*	3	*11.54*	14	*53.85*	26
		Deontic	9	*21.43*	0	*0.00*	6	*14.29*	2	*4.76*	25	*59.52*	42
		Dynamic	17	*20.99*	2	*2.47*	11	*13.58*	8	*9.88*	43	*53.09*	81
		No	166	*17.53*	23	*2.43*	66	*6.97*	97	*10.24*	595	*62.83*	947
	Informal	Epistemic	17	*45.95*	1	*2.70*	2	*5.41*	3	*8.11*	14	*37.84*	37
		Evidential	0	*0.0*	0	*0.00*	1	*12.50*	0	*0.00*	7	*87.50*	8
		Deontic	7	*18.92*	2	*5.41*	1	*2.70*	4	*10.81*	23	*62.16*	37
		Dynamic	26	*20.47*	1	*0.79*	16	*12.60*	16	*12.60*	68	*53.54*	127
		No	146	*16.35*	12	*1.34*	56	*6.27*	97	*10.86*	582	*65.17*	893
		Total	433	*18.93*	44	*1.92*	169	*7.39*	235	*10.28*	1406	*61.48*	2287
Total			755	*17.51*	78	*1.81*	381	*8.83*	448	*10.39*	2651	*61.47*	4313

Note. Percentages are row-based. Due to size, this page presents part 2 of 2 parts of the full table.

Table B.5:
Distribution of aspect by mode and register

							Aspect (c)				
Mode	Register	Aspect (a)	State	%	Activity	%	Accomplisment	%	Achievement	%	Total
Spoken	Formal	State	175	*47.17*	59	*15.90*	27	*7.28*	110	*29.65*	371
		Activity	93	*40.97*	51	*22.47*	12	*5.29*	71	*31.28*	227
		Accomplishment	45	*48.91*	8	*8.70*	10	*10.87*	29	*31.52*	92
		Achievement	175	*47.81*	40	*10.93*	21	*5.74*	130	*35.92*	366
	Informal	State	170	*49.13*	54	*15.61*	37	*10.69*	85	*24.57*	346
		Activity	84	*44.68*	41	*21.81*	15	*7.98*	48	*25.53*	188
		Accomplishment	30	*38.96*	11	*14.29*	17	*22.08*	19	*24.68*	77
		Achievement	15	*50.16*	42	*13.42*	22	*7.03*	92	*29.39*	313
		Total	929	*46.92*	306	*15.45*	161	*8.13*	584	*29.49*	1980
Written	Formal	State	198	*45.21*	76	*17.35*	32	*7.31*	132	*30.14*	438
		Activity	92	*41.26*	52	*23.32*	8	*3.59*	71	*31.84*	223
		Accomplishment	34	*36.56*	13	*13.98*	17	*18.28*	29	*31.18*	93
		Achievement	179	*41.72*	68	*15.85*	27	*6.29*	155	*36.13*	429
	Informal	State	170	*35.27*	101	*20.95*	51	*10.58*	160	*33.20*	482
		Activity	100	*44.64*	48	*21.43*	15	*6.70*	61	*27.23*	224
		Accomplishment	28	*33.33*	11	*13.10*	21	*25.00*	24	*28.57*	84
		Achievement	123	*42.27*	37	*12.71*	20	*6.87*	111	*38.14*	291
		Total	924	*40.81*	406	*17.93*	191	*8.44*	743	*32.82*	2264
Total			1853	*43.67*	712	*16.78*	352	*8.29*	1327	*31.27*	4244

Note. Percentages are row-based.

Table B.6:
Distribution of person and number by mode and register

			Person & number (c)											
Mode	Register	Person & number (a)	1ps	%	1pp	%	2ps	%	2pp	%	3ps	%	3pp	%	Total
Spoken	Formal	1ps	42	*41.58*	3	*2.97*	7	*6.93*	0	*0.00*	44	*43.56*	5	*4.95*	101
		1pp	9	*13.04*	28	*40.58*	2	*2.90*	0	*0.00*	24	*34.78*	6	*8.70*	69
		2ps	16	*6.23*	9	*3.50*	112	*43.58*	0	*0.00*	102	*39.69*	18	*7.00*	257
		2pp	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	0
		3ps	46	*9.91*	32	*6.90*	41	*8.84*	1	*0.22*	275	*59.27*	69	*14.87*	464
		3pp	6	*3.43*	8	*4.57*	9	*5.14*	0	*0.00*	69	*39.43*	83	*47.43*	175
	Informal	1ps	89	*59.73*	2	*1.34*	12	*8.05*	0	*0.00*	42	*28.19*	4	*2.68*	149
		1pp	5	*8.93*	20	*35.71*	8	*14.29*	0	*0.00*	15	*26.79*	8	*14.29*	56
		2ps	37	*8.98*	11	*2.67*	210	*50.97*	0	*0.00*	128	*31.07*	26	*6.31*	412
		2pp	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	2	*100.00*	0	*0.00*	2
		3ps	53	*20.31*	13	*4.98*	42	*16.09*	1	*0.38*	129	*49.43*	23	*8.81*	261
		3pp	14	*19.44*	2	*2.78*	9	*12.50*	2	*2.78*	22	*30.56*	23	*31.94*	72
		Total	317	*15.71*	128	*6.34*	452	*22.40*	4	*0.20*	852	*42.22*	265	*13.13*	2018

Note. Percentages are row-based. Due to size, this page presents part 1 of 2 parts of the full table.

Mode	Register	Person & number (a)	Person & number (c) 1ps	%	1pp	%	2ps	%	2pp	%	3ps	%	3pp	%	Total
Written	Formal	1ps	46	*62.16*	0	*0.00*	0	*0.00*	0	*0.00*	23	*31.08*	5	*6.76*	74
		1pp	1	*1.59*	24	*38.10*	2	*3.17*	0	*0.00*	32	*50.79*	4	*6.35*	63
		2ps	5	*3.05*	3	*1.83*	89	*54.27*	0	*0.00*	61	*37.20*	6	*3.66*	164
		2pp	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	0	*0.00*	0
		3ps	18	*2.86*	22	*3.49*	19	*3.02*	0	*0.00*	486	*77.14*	85	*13.49*	630
		3pp	4	*1.53*	7	*2.67*	3	*1.15*	0	*0.00*	135	*51.53*	113	*43.13*	262
	Informal	1ps	136	*55.97*	2	*0.82*	19	*7.82*	2	*0.82*	79	*32.51*	5	*2.06*	243
		1pp	10	*21.28*	18	*38.30*	2	*4.26*	0	*0.00*	15	*31.91*	2	*4.26*	47
		2ps	42	*10.69*	10	*2.54*	187	*47.58*	2	*0.51*	136	*34.61*	16	*4.07*	373
		2pp	0	*0.00*	0	*0.00*	4	*44.44*	1	*11.11*	4	*44.44*	0	*0.00*	9
		3ps	85	*25.45*	16	*4.79*	68	*20.36*	1	*0.30*	152	*45.51*	12	*3.59*	334
		3pp	16	*24.24*	2	*3.03*	10	*15.15*	0	*0.00*	23	*34.85*	15	*22.73*	66
		Total	363	*15.89*	104	*4.55*	403	*17.64*	6	*0.26*	1146	*50.15*	263	*11.51*	2285
Total			680	*15.80*	232	*5.39*	855	*19.87*	10	*0.23*	1198	*46.43*	528	*12.27*	4303

Note. Percentages are row-based. Due to size, this page presents part 2 of 2 parts of the full table.

Table B.7:
Distribution of sentence types by mode and register

		Sentence type (c)								
Mode	Register	Declarative	%	Imperative	%	Interrogative	%	Other	%	Total
Spoken	Formal	1051	*95.03*	2	*0.18*	44	*3.98*	9	*0.81*	1106
	Informal	926	*92.23*	4	*0.40*	43	*4.28*	31	*3.09*	1004
	Total	1977	*93.70*	6	*0.28*	87	*4.12*	40	*1.90*	2110
Written	Formal	1171	*95.13*	3	*0.24*	49	*3.98*	8	*0.65*	1231
	Informal	991	*83.91*	43	*3.64*	93	*7.87*	54	*4.57*	1181
	Total	2162	*89.64*	46	*1.91*	142	*5.89*	62	*2.57*	2412
Total		4139	*91.53*	52	*1.15*	229	*5.06*	102	*2.26*	4522

Note. Percentages are row-based.

Table B.8:
Distribution of types of negation by mode and register

			Negation (c)						
Mode	Register	Negation (a)	Syntactic	*%*	Morphological	*%*	No	*%*	Total
Spoken	Formal	Syntactic	35	*22.73*	6	*3.90*	113	*73.38*	154
		Morphological	3	*17.65*	3	*17.65*	11	*64.71*	17
		No	128	*14.00*	11	*1.20*	775	*84.79*	914
	Informal	Syntactic	18	*19.57*	2	*2.17*	72	*78.26*	92
		Morphological	0	*0.00*	0	*0.00*	0	*0.00*	0
		No	104	*12.11*	7	*0.81*	748	*87.08*	859
		Total	288	*14.15*	29	*1.42*	1719	*84.43*	2036
Written	Formal	Syntactic	49	*25.00*	14	*7.14*	133	*67.86*	196
		Morphological	12	*27.27*	3	*6.82*	29	*65.91*	44
		No	143	*14.71*	37	*3.81*	792	*81.48*	972
	Informal	Syntactic	41	*25.95*	2	*1.27*	115	*72.78*	158
		Morphological	3	*37.50*	0	*0.00*	5	*62.50*	8
		No	141	*14.78*	7	*0.73*	806	*84.49*	954
		Total	389	*16.68*	63	*2.70*	1880	*80.62*	2332
Total			677	*15.50*	92	*2.11*	3599	*82.39*	4368

Note. Percentages are row-based.

Table B.9:
Distribution of types of focus particles by mode and register

		Focus particle								
Mode	Register	Additive	*%*	Restrictive	*%*	Iterative	*%*	No	*%*	Total
Spoken	Formal	43	*3.63*	35	*2.95*	4	*0.34*	1104	*93.09*	1186
	Informal	26	*2.13*	18	*1.48*	33	*2.70*	1143	*93.69*	1220
	Total	69	*2.87*	53	*2.20*	37	*1.54*	2247	*93.39*	2406
Written	Formal	89	*7.18*	74	*5.97*	8	*0.65*	1069	*86.21*	1240
	Informal	21	*1.72*	42	*3.44*	13	*1.06*	1146	*93.78*	1222
	Total	110	*4.47*	116	*4.71*	21	*0.85*	2215	*89.97*	2462
Total		179	*3.68*	169	*3.47*	58	*1.19*	4462	*91.66*	4868

Note. Percentages are row-based.

APPENDIX C

Data preparation

C.1 Introduction

In this appendix, the data preparation for clustering is elaborated. In section C.2, features are tested on feature independence and skewedness of value distributions. The (re)coded variables, their distributions, and their deviations from the mode (DM) are presented in section C.3. Then, in sections C.4 to C.6, the technical details of the initial variable selection, distance calculation, and final variable selection are discussed, and the results of these preparatory steps are presented. For the main discussion of data preparation for clustering, see chapter 6.

C.2 Feature independence and skewedness

Before processing the data, feature independence and possible skewedness of value distributions were checked. Both steps are discussed in this section. For a general discussion of these pre-processing steps, see section 6.3.2.

First, it was inspected whether all features are, in theory at least, independent of each other. For most of the features discussed in the previous chapter, this poses no problems. Negation, for instance, is not dependent on other features, as any type of negation can, in theory, be used in combination with any tense, clause order or other feature. There was, however, one set of three features that showed internal dependency, namely clause order, syntactic integration and sentence type. A feature like syntactic integration can, as was discussed in section 5.3, only be annotated for sentence-initial antecedents. A resumptive

pattern, for example, is not available for sentence-final antecedents, meaning that the features clause order and syntactic integration are dependent. To solve this problem, clause order, syntactic integration and sentence type were merged into the new variable 'syntactic pattern', which was comprised of the levels 'integration', 'resumption', 'non-integration', 'sentence-medial', and 'sentence-final'. The low-frequent group of embedded conditionals were grouped together with sentence-medial conditionals, and conditionals with non-declarative consequents were grouped together with non-integrative conditionals, because their word order does not allow for the other patterns in this variable.

Second, the distribution of some features was skewed, either showing signs of trait prevalence, or a large number of values and low frequencies per value. Focus particles are an example of this problem. Conditionals with a focus particle were initially annotated with the different particles as values, which led to a large number of values with low frequencies i.e., the different particles in the corresponding variable (see section 5.10). To retain the information while bringing down the number of levels, the values were classified into the categories discussed, namely *additive*, *restrictive* and *iterative particles*. To inspect the balance between the number of levels in a variable, and the distribution of these values, dispersion was measured for each variable. The measurement of dispersion is common for numerical data, for which all sorts of well-known measures of statistical dispersion are available, such as *range* and *standard deviation*. It is less common for qualitative data (nominal, ordinal), which is the type of data frequently encountered in corpus linguistic studies, such as this study. Wilcox reflects as follows on this problem.

> All standard statistics texts discuss the measurement of variation in a univariate distribution when the variable under consideration satisfies the requirements of an ordinal, interval, or ratio scale. [...] However, a discussion of the measurement of variation with nominal-scale data is usually conspicuous by its absence. (Wilcox, 1973, p. 325)

Wilcox (1973) therefore proposes a number of measurements of 'qualitative variation', among which 'Deviation from the Mode' or *DM*, presented in (1).

(1) $DM = 1 - \frac{\sum_{i=1}^{k} (f_m - f_i)}{N(K-1)}$

The basic principle is that (1) stands for an index of deviation from the modal frequency, 'analogous to the variance as a measure of deviation from the mean' (Wilcox, 1973, p. 325). This measure was therefore used to assess the dispersion of each variable over all conditionals in the corpus. For a detailed overview of the (re)coded variables, their distribution and DM-values are included in table C.1 in Appendix B on page 488. Please note that the deviation from the mode was calculated twice for features which suffered from the 'missing data-problem' discussed in section 4.5: once with without 'NA'-values, and once

with those values as 'no'-values.[1] What we see in the results is that a number of tense in the antecedent and consequent, modality in antecedents, negation in antecedents and consequents and particularly focus particles have low DM-values. For modality, negation and focus particles this is especially the case when absence of those features is considered a level ('no'). To deal with this, we will implement variable selection in the next section.

C.3 Coded variables and deviation from the mode

In table C.1, the (re)coded variables, their distributions, and their deviations from the mode (DM) are presented.

[1] This was not done for the features *negation (a)* and *negation (c)*, because no measure of dispersion can represent a variable with only one level.

Table C.1:
Coded variables, levels and Deviation from the Mode (DM)

Feature(s)	Section(s)	Variable	Levels	*%*	*% NA*	*DM*	*DM NA*
Clause order, syntactic integration, Sentence type	5.2, 5.3, 5.8	syntactic_pattern	integration	*13.1*		*0.85*	
			resumption	*5.5*			
			non-integration	*27.9*			
			sentence-medial	*0.2*			
			sentence-final	*39.8*			
Verb tense	5.4	tense_a	simple present	*84.2*		*0.21*	
			simple past	*9.3*			
			present perfect	*4.5*			
			past perfect	*2.0*			
		tense_c	simple present	*87.5*		*0.18*	
			simple past	*10.6*			
			present perfect	*1.0*			
			past perfect	*1.9*			
Modality	5.5	modality_a	epistemic	*6.13*	*32.1*	*0.24*	*0.72*
			evidential	*1.46*	*7.6*		
			deontic	*2.73*	*14.3*		
			dynamic	*8.79*	*46.0*		
			no	*80.90*			

Note. Due to size, this page presents part 1 of 3 parts of the full table.

Feature(s)	Section(s)	Variable	Levels	*%*	*% NA*	*DM*	*DM NA*
Modality	5.5	modality_c	epistemic	*17.64*	*45.3*	*0.49*	*0.73*
			evidential	*1.70*	*4.4*		
			deontic	*9.03*	*23.2*		
			dynamic	*10.54*	*27.1*		
			no	*61.09*			
Aspect	5.6	aspect_a	state	*38.7*		*0.82*	
			activity	*20.3*			
			accomplishment	*8.0*			
			achievement	*32.9*			
		aspect_c	state	*43.5*		*0.82*	
			activity	*16.8*			
			accomplishment	*8.3*			
			achievement	*31.3*			
Person & number	5.7	subject_a	1ps	*13.1*		*0.73*	
			1pp	*5.5*			
			2ps	*27.9*			
			2pp	*0.2*			
			3ps	*39.8*			
			3pp	*13.5*			
		subject_c	1ps	*16.0*		*0.65*	
			1pp	*5.5*			
			2ps	*19.9*			
			2pp	*0.2*			
			3ps	*46.0*			
			3pp	*12.4*			

Note. Due to size, this page presents part 2 of 3 parts of the full table.

Feature(s)	Section(s)	Variable	Levels	*%*	*% NA*	*DM*	*DM NA*
Negation	5.9	negation_a	syntactic	*13.77*	*89.27*	*0.23*	*0.21*
			morphological	*1.65*	*10.73*		
			no	*84.57*			
		negation_c	syntactic	*15.14*	*87.73*	*0.25*	*0.26*
			morphological	*2.12*	*12.27*		
			no	*82.75*			
Focus particles	5.10	focus_type	additive	*3.7*	*43.2*	*0.11*	*0.85*
			iterative	*1.2*	*14.5*		
			restrictive	*3.6*	*42.3*		
			no	*91.4*			

Note. Due to size, this page presents part 3 of 3 parts of the full table.

C.4 Initial variable selection

In this section, the details of the initial variable selection are presented. For a general discussion of this step, see section 6.3.2.

Assessing variable importance in clustering is not an easy task. Talavera (2005, p. 440) argues that feature selection for clustering is not addressed often, mostly because there is no consensus on how to evaluate the results of a clustering algorithm. One of the reasons for this is that clustering is an unsupervised machine-learning technique, which means that there is no objective class assignment for each observations against which the results of the clustering can be tested, in contrast to supervised techniques generally called 'classification' in the machine-learning literature (see e.g., Berry, Mohamed & Yap, 2019, for an introduction and up-to-date overview). Silvestre, Cardoso and Figueiredo (2013) explain the difference clearly:

> In supervised learning, namely in classification, feature selection is a clearly defined problem, where the search is guided by the available class labels. In contrast, for unsupervised learning, namely in clustering, the lack of class information makes feature selection a less clear problem and a much harder task. (Silvestre, Cardoso & Figueiredo, 2013, pp. 331–332)

The major difference between variable selection for classification on the one hand and variable selection for clustering on the other, is thus that in supervised machine learning, the labels or types are known for (at least) a part of the dataset. Algorithms can be trained by estimating those labels or types based on the features in the dataset. The estimated labels are then compared to the existing labels, and the accuracy of the predictions can then be measured. When accuracy is sufficient, the algorithm can be used to label the part of the dataset that has not been assigned labels manually.

For unsupervised techniques, no labels are present *a priori*, making it less clear how to determine the accuracy of the results of the learning algorithm chosen. As a result, no standard approach is available for feature selection in clustering (cf. Questier et al., 2002, p. 311; see also Li et al., 2017). The fact that there is no training set available for evaluation directly affects variable selection methods. For supervised techniques, various models can be generated by starting with only one variable as predictor and incrementally adding features (a 'forward' approach) or by starting with a full model and incrementally removing features (a 'backward' approach). Features that sufficiently increase the predictive power of the model are kept, while those that do not are left out of the final model. As discussed above, in clustering techniques, the labels are not known, making it impossible to directly assess the contribution of each variable. Before moving on to strategies to deal with this problem, it is deemed necessary here to mention an added complication in this study, namely that the literature available on feature selection for clustering is mainly targeted at clustering numerical variables, not categorical variables (for an overview, see e.g.,

Li et al., 2017, p. 36). Furthermore, existing unsupervised variable selection methods make use of conventional distance metrics (cf. Liu & Zhang, 2016), such as Euclidian distance for numerical data or Gower's distance for categorical data. Such metrics do not take into account the distributional properties of the dataset, as discussed at length in section 6.3.

The above means that insights from different methods have to be combined for variable selection in the current study.[2] Results will therefore be interpreted with caution. To do so, the following approach was chosen: first, the internal distributions of the variables are evaluated and the results are combined with an informal ranking of theoretical importance of variables. This will constitute the initial variable selection. Second, after the distance calculation in the next section, insights from the initial variable selection are used to measure the impact of variables on the stress of dimension-reducing models of the dataset.

To inspect the informativeness of variables statistically, two main approaches are available, so-called 'filter methods' and 'wrapper methods' (see e.g., Dy & Brodley, 2004; Xiaofei, Deng & Partha, 2005; see Alelyani, Tang & Liu, 2013, for a recent overview). Filter methods assess the qualities of variables by evaluating their internal variation and distribution. Whereas in filter methods the internal distributional characteristics of the variables are used to assess their possible contribution to subsequent clustering steps, wrapper methods work fundamentally different (see Talavera, 2005, p. 441). Wrapper methods are based on feature selection in supervised classification (see Kohavi, John et al., 1997; Dy & Brodley, 2004, pp. 847–848), and take subsets of the feature space to generate clustering solutions, which are then evaluated according to an internal quality criterion, such as an information-theoretic criterion like the Akaike Information Criterion (AIC, see Akaike, 1969; Akaike, 1974; Bozdogan, 1987) or the Bayesian Information Criterion (BIC, see Schwarz, 1978). The contribution of each variable can be assessed by looking at its contribution to the quality of the model. A wrapper method thus uses a form of clustering itself to form groups with which the influence of each of the variables in measured, evaluated and ranked. The type of clustering used is (a form of) model-based clustering, which assumes that 'the data is generated by a mixture of underlying probability distributions' (Vermunt & Magidson, 2002, pp. 89–90) and a likelihood function is used to maximise the likelihood of the expression data, i.e., the probability of a group of observations coming from one distribution, while another group comes from another probability distribution. These algorithms have been tested largely on numerical datasets. Model-based clustering assumes that the data originate 'from a finite mixture of underlying probability distributions' (Blattberg, Kim & Neslin, 2008, p. 414; cf. Fraley & Raftery, 1998). Because each cluster comes from a different (Gaussian) distribution, the contribution of a variable in identifying these clusters can be evaluated. This is, however, no trivial assumption for the data in this study, and forms a serious drawback,

[2]While a number of new methods have been proposed, most have not been implemented and tested thoroughly yet. See for instance Fop and Murphy (2018).

as wrapper methods may evaluate variable importance based on clusters that were formed on different grounds than used in the (non model-based) clustering approaches in the following steps. Although wrapper methods usually provide a more informative picture of variable importance, its reliance on model-based clustering is an argument in favour of a simpler, but less model-dependent filter approach to perform the initial assessment of the information value of each variable.

Within filter approaches to variable selection, two distinct types of measures can be distinguished: univariate and multivariate selection methods. As the names suggest, univariate methods assess variables individually using an evaluation criterion based on the internal distribution of the variable, for example in terms of entropy, divergence or mutual information. Possible dependencies or interactions between variables thus are not taken into consideration. In multivariate methods, such dependencies are evaluated. Although a number of these methods are available (see Tabakhi, Moradi & Akhlaghian, 2014; Solorio-Fernández, Carrasco-Ochoa & Martínez-Trinidad, 2020, for overviews), most are suited for supervised tasks as they depend on class labels. Furthermore, most methods are limited to or tested on numerical variables only. The initial variable selection in this study was therefore performed by combining a simple univariate method, namely the calculation of the frequency ratio (FR) of each variable, with a ranking based on the theoretical importance of the variables. Although in the methodological literature, expert-selection of variables does not appear often, it can be found in studies applying machine-learning methods (see e.g., van den Berge et al., 2017). In this initial feature selection, no variables will actually be removed from the dataset, but the insights will be used as indications of potentially problematic features. In section 6.3.6, these insights are used together with a multivariate method for the final feature selection.

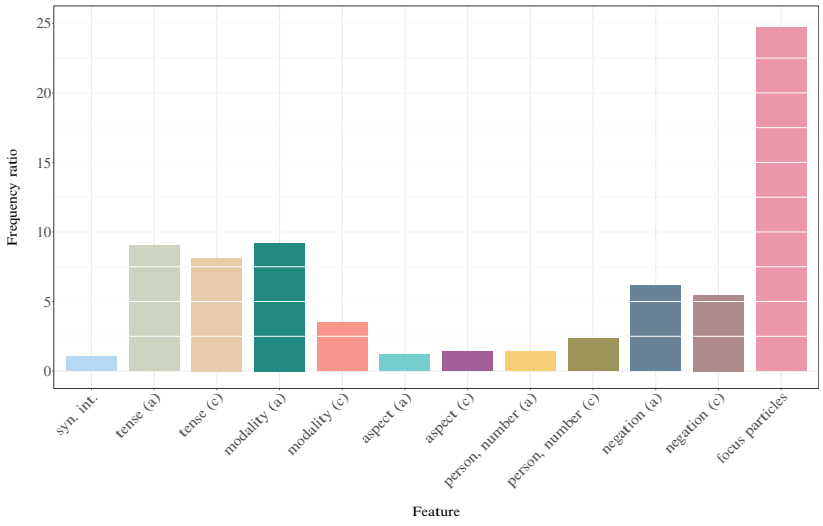
As mentioned above, the goal at this point is to indicate which variables have non-informative distributions or have less theoretical relevance. For the first step, the frequency ratio of each variable was assessed using the formula in (2) below. As can be seen, this simple calculation only divides the frequency of the most frequent value (f_i) by the frequency of the second-most frequent value (f_j). The reasoning here is that if the ratio between the frequency of the most frequent value and the second-most frequent value is large, it may be better to remove the variable from the model (see Kuhn & Johnson, 2013, p. 45).[3]

$$(2)\quad \text{FR} = \frac{f_i}{f_j}$$

[3]There is another criterion that must be met before considering a variable uninformative, namely that the percentage of unique values is less than 20% of the number of observations (see Kuhn, 2008, p. 4; Kuhn & Johnson, 2013, p. 45). This requirement is met by all variables. See section C.2.

In Figure 6.1 presented in section 6.3, and repeated below in Figure C.1, the frequency ratio of each variable is presented, As this ratio is the frequency of the most common value of a feature divided by the second most common value, the higher the ratio is, the bigger the prevalence of the most common value is.

Figure C.1:
Frequency ratio per feature



In this figure, it can be seen that some ratios are higher than others. While it is up to the analyst to set thresholds for the frequency ratio, no hard boundaries are needed to see there is one clear outlier, namely focus particles. This can be explained by the fact that most conditionals do not feature a focus particle, so the 'no'-value has a much higher frequency than any of the other values (i.e., types of focus particles, see section 5.10). For this particular feature, absence accounted for 3757 of the values, while the second most frequent value, additive focus particles, had a frequency of only 152. Using the formula in (2), this results in $3757/152 = 24.72$. This is a problem for any subsequent step in the analysis, as this variable introduces complexity into the model, while explaining very little variation. It can also be seen that modality in the antecedent, and tense in both clauses have somewhat higher frequency ratio's than the other features. For modality, this has the same cause as for focus particles, namely that around 80% of all antecedents is not modalised. This is interesting when contrasted with modalisation of consequents, which has a much lower frequency ratio, mainly due to a lower number of non-modalised clauses and a secondary prevalence of epistemic modality.

C.5 Distance calculation

Before the insights from the previous section can be used for the final variable selection, distance calculations and evaluations thereof are needed. Various distance measures are elaborated in this section. For a general discussion of this step, see sections 6.3.3 to 6.3.5.

C.5.1 Basic distance calculation

In (3) below, the formula for *Gower's Distance* is presented.

(3) $S_{ij} = \frac{\Sigma_{k=1}^{N} W_{ijk} S_{ijk}}{\Sigma_{k=1}^{N} W_{ijk}}$

Here, W_{ijk} is the weight for variable k between conditionals i and j, and S_{ijk} is the distance between conditionals i and j with respect to that variable k. Weight W is 1 by default and is a constant value per variable. The distances or *dissimilarities* between conditionals are calculated by subtracting Gower's similarity score from 1. Using the measure in (3), a dissimilarity matrix can calculated, consisting of the dissimilarities between all individual conditionals on all features.

The reason for the somewhat explicit elaboration on distance calculation here is twofold. First, the calculation of the distance matrix can have profound effect on any subsequent analysis, and although *distance* might seem to be an objective measurement, the researcher has several choices to make, such as the choice for a metric used to calculate distance. Typically, one can choose from *Euclidean*, *Manhattan* or *City block*, and *Gower's distance* (see e.g., Anderberg, 1973, Chapter 5). The first two metrics are only applicable to numerical datasets (or data transformed to numerical values), and as in corpus linguistics the data are often collected on the categorical level, this leaves Gower's Distance. A second choice that, to my knowledge, is not mentioned in earlier corpus linguistics studies, even in those which explicitly mention the step of distance calculation, is how missing values are dealt with – probably because most implementations of Gower's metric allow for such values to be included.[4] This, however, is non-trivial and can have, as will be shown in what follows, severe impact on the distance matrix and subsequent analyses. Second, the distance matrix itself is a source of information for the researcher, and can be used to answer questions concerning the level of homogeneity of the dataset, and, for instance, to identify the most and least representative examples (see section 6.3.7).

Distance calculation will be explained examples (397) to (400) from section 6.3, repeated below for convenience.

(397) If you flick the switch, the light will go on.

[4] As the reader will notice, the subject of missing values is a recurrent issue throughout this thesis. See chapter 5.

(398) If he attacks the enemies, they strike back.

(399) The water is not cold, if it is boiling.

(400) Even if we work hard, we may not leave early today.

In Table 6.1 from section 6.3, repeated below in Table C.2, it is reflected that that two of the conditionals have modal marking in the consequent.

Table C.2:
Data structure for examples in (397) to (400)

Example	Clause order	Person & Number (a)	Modality (c)
(397)	initial	2ps	epistemic
(398)	initial	3ps	no
(399)	final	3ps	no
(400)	initial	1pp	deontic

As (398) and (399) have *no* as a value for this feature, Gower's metric considers them to have a distance of 0 for this feature, i.e., they are identical on this feature. This seems right, but a possible bias arises when the total distance between these two conditionals is calculated. Their distance is 0.33, because they share two of the three features, namely person and number, and modality. The other distances are presented in Table 6.2 in section 6.3, repeated below in Table C.3.

Table C.3:
Distance matrix for examples in (397) to (400)

	Ex. (397)	Ex. (398)	Ex. (399)	Ex. (400)
Ex. (397)	0.00			
Ex. (398)	0.67	0.00		
Ex. (399)	1.00	0.33	0.00	
Ex. (400)	0.67	0.67	1.00	0.00

Looking at the examples, however, one could also argue that conditionals in (398) and (399) have only one feature in common (i.e., person and number), as the absence of a feature (modality) is hardly grounds for similarity. This problem is discussed in general terms by Anderberg as follows.

> [...] There is the question of what to do with 0-0 matches. [...] For example, suppose the data units are animals and the variables are "has feathers," and "has webbed feet." Dogs and cats and many other animals would fall into cell d [not possessing either attribute,

> AR] because there is no way they could have such attributes. It would be misleading to allow these 0-0 matches to contribute to the measure of association between cats and dogs. (Anderberg, 1973, p. 88)

As most conditionals are not marked for modality (see section 5.5), should the absence of this feature contribute to the similarity index? As discussed, conditionals without a focus particle could be annotated using 'NA'. The result is that in most implementations of Gower's metric (see the formula in (3) above), the feature is ignored completely in the comparison of two conditionals of which at least one has 'NA' for this feature, while it still adds to the (dis)similarity of conditionals that do have this relatively infrequent feature. This would then result in a distance of not 0.33, but 0.50 between the conditionals in (398) and (399), because they would only share one of two features present. This seems more appropriate, as can be seen in the distance matrix in C.4.

Table C.4:
Distance matrix with 'NA' for 'no' in Table C.2

	Ex. (397)	Ex. (398)	Ex. (399)	Ex. (400)
Ex. (397)	0.00			
Ex. (398)	0.50	0.00		
Ex. (399)	1.00	0.50	0.00	
Ex. (400)	0.67	0.50	1.00	0.00

Several differences can be seen between the distance matrix in Table C.3 and C.4 above. Although the distance between (398) and (399) is corrected for agreement on an absent feature, there is another change, namely that the distance between the conditionals in (397) and (398) has become 0.5, because (398) has 'NA' for the focus particle feature, removing it from the distance calculation. For this small example corpus, in which half of the observations actually have this feature, one could argue for both including these no-values or excluding them, but remembering the low frequency of a feature such as modal marking in especially the antecedent (see section 5.5), this would mean inflating the similarity between conditionals by including highly prevalent 'no'-values. On the other hand, treating 'no'-values as 'NA' introduces the problem that conditionals with 'NA' for certain features may be considered more identical than other conditionals, while such a result is debatable.

C.5.2 Evaluation of distance matrices by multimodality

In this section, histograms of distances per measurement are presented and evaluated in terms of multimodality, because multiple modes in the distribution of distances indicates multiple clusters in the data (see Ackerman, Adolfsson & Brownstein, 2016; Adolfsson, Ackerman & Brownstein, 2019).

Before discussing the distributions, please note that the distances were normalised after calculation. The reason for doing so, was that distance measures produce results on different scales which do no necessarily fall between 0 and 1. The distances were normalised using the simple so-called 'min-max normalisation' in (4) below, which was applied to each distance distribution, resulting in a comparable scale from 0 to 1 for each distance distribution, while keeping the internal distribution the same.

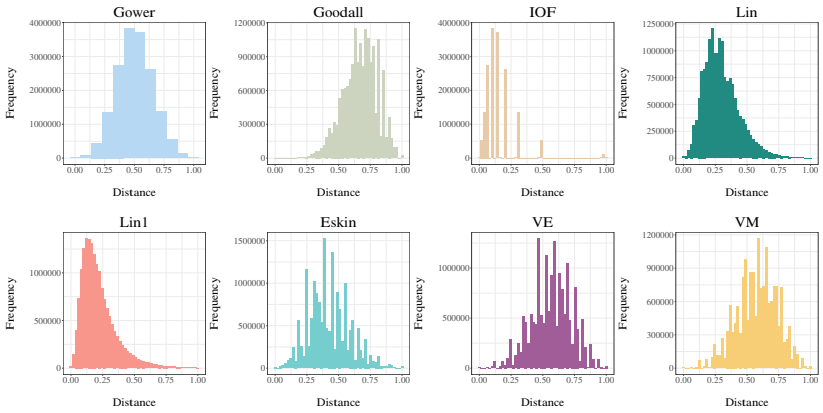
(4) $z_i = \frac{x_i - min(x)}{max(x) - min(x)}$

A further step to maximise the clusterability of the data was to identify and remove outliers from the distance matrix. As Almeida et al. (2007, p. 209) argue, data with outliers 'are difficult to tackle with most clustering algorithms', because the data structure becomes 'less defined' and may have a negative impact on clustering results (especially in case of single and average linkage, see also section 6.4.4). For the current purposes, a simple strategy was used, in which all distances outside a threshold value (here, 5 times the standard deviation) were standardised.[5]

The histograms of the distance matrices are presented in Figure 6.2 in section 6.3, and repeated in Figure C.2 below.

[5]More elaborate tests for outlier-detection are described by, among others, Grubbs (1969), Dean and Dixon (1951). See also Tietjen and Moore (1972) for discussion. These tests were not used here, because they rely more heavily than the current approach on a normal distribution of the data, whereas tests for multimodality test for the opposite.

Figure C.2:
Distribution of distances per measure



Looking at these histograms, it can be seen that the different distance calculations unfortunately do not produce clear multimodal distributions.[6] For Gower's measure, the general shape of a normal distribution can be seen, but we also clearly see the result of the simple calculation, as there are only twelve distinct values in the complete distance matrix due to a limited number of possible distances based on the Gower (or 'Simple Matching') measure. The distribution of Goodall distances shows a negative skew, meaning that the left tail is longer and mass of the distribution is on the right of the figure. The IOF measure produces the tendency of a right-tailed (i.e., positive skewness) distribution, from which it is not clear whether it is suffering from the same problem associated with the discrete nature of the data as Gower's measure, or whether the tail indeed shows separate modes. Both the Lin and Lin1 measures, show a unimodal, but right-skewed distribution. The Eskin measure features a number of smaller modes, but as these figures are based on discrete data, these figures should be interpreted with caution. Like the *Lin* and *Lin1* measures, the VE and VM measures produce similar, but not identical distance distributions, with the modes of the former being slightly more dispersed than the latter.

As a more formal check, each distribution was subjected to a multimodality test. The general idea of applying such a test on a distance matrix is that it assumes that the data comes from a unimodal distribution, which functions as the null hypothesis. Given the test chosen (see below), a large p-value ($>= 0.05$) indicates no significant diversion from the (nearest) unimodal distribution, i.e., only a single mode is present in the data. A small p-value (< 0.05) on the other

[6] For comparison, see the examples by Ackerman, Adolfsson and Brownstein (2016, p. 5), which show a number of truly multimodal distance distributions.

hand, questions the assumption that the data are unimodally distributed, and indicates that there is evidence for multiple modes in the data, which could reflect multiple clusters (Adolfsson, Ackerman & Brownstein, 2019, pp. 6–7). Of course, the histograms were already visually inspected for unimodality, but a statistical test may determine in a more formal way whether or not further clustering steps are warranted. One such test is the *dip* test, resulting in a dip statistic which reflects the maximum distance between the distribution in question and the closest unimodal distribution (see Hartigan & Hartigan, 1985, p. 70; Hartigan, 1985; Maechler, 2016, for implementation), and provides a corresponding p-value, indicating whether or not the null hypothesis that the distances come from a unimodal distribution may be rejected. The test takes each set of distances, and compares these distributions to the closest normal distribution, indicating whether the data contain one or more peaks or *modes* (cf. Chamalis & Likas, 2018).

The dip tests performed provided an unexpected result, indicating that all distributions differ significantly from a unimodal distribution, i.e., all p-values are less than 0.05.[7] The reason for reporting this finding is first that this problem was found yet in the literature on clusterability, and second that the actual histograms may provide a clue to the cause of what at least looks like false-positive results. As can be seen in the histograms in Figure 6.2 (see page 6.3 in section 6.3), especially that for the Gower distances, the discrete nature of the data is reflected in the number of distinct distances. In fact, as mentioned with respect to the distance matrix for Gower's measure, only 12 distinct distances are present, which can be explained by the fact that this simple distance measure, in which correspondence and non-correspondence simply amount to a distance of 0 or 1, has a limited set of output distances. The *dip* statistic, however, is tested on numeric (i.e., non-discrete) data, which does not suffer from this problem.[8] A possible explanation is that the 'gaps' between the discrete distances are picked up by the statistic as deviations from the closest normal distribution, resulting in significant deviations from the null hypothesis. This, of course, is problematic and leads to misleading results, because we can see the distribution of distances in most cases actually does closely resemble the bell curve of a normal distribution. Because of this, the visual assessment the distributions will be used with caution, and a second approach to evaluating the distance matrices is presented in the next section.

[7] In fact, all rounded p-values are 0.00.

[8] Unfortunately, another frequently used multimodality test, the Silverman test (Silverman, 1981) suffers from the same problem with categorical data, i.e., it indicates that all distributions in Figure 6.2 on page 353 in section 6.3 deviate significantly from a unimodal distribution.

C.5.3 Evaluation of distance matrices by dimension-reduction

The goodness-of-fit value used for evaluating the dimension reduction results was calculated using the following formula, commonly called *Kruskal's Stress.*

(5) $$\text{Stress} = \sqrt{\frac{\Sigma_{ij}(d_{ij} - \hat{d}_{ij})^2}{\Sigma_{ij} d_{ij}^2}}$$

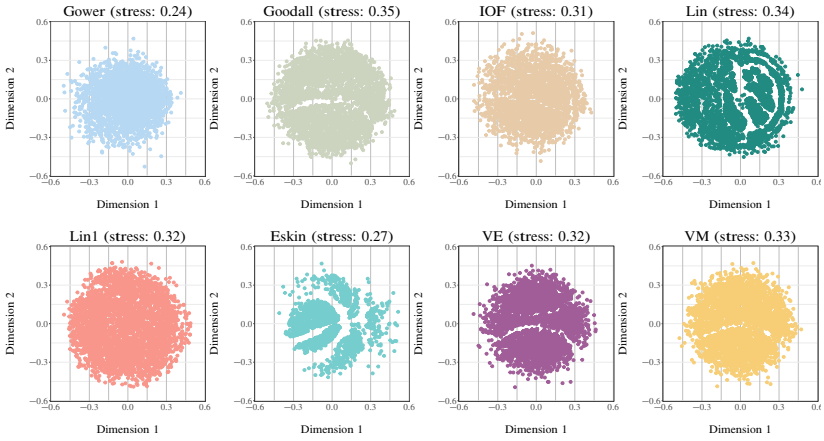
Here, d_{ij} is the distance between observations i and j, and $\hat{d}_{ij}$ is the distance between those observations in the model. The lowest stress value is 0, which indicates 'complete accordance between all rank order distances in the input data and the final ordination' (Dexter, Rollwagen-Bollens & Bollens, 2018, p. 435). The greater the value, the worst the fit of the model is to the actual distance matrix. A common guideline is that stress values higher than 0.2 are considered 'poor and potentially uninterpretable' (Tyler & Kowalewski, 2014, p. 5). Clarke (1993, p. 126) provides the following 'rules of thumb' for interpretation of stress levels:

(6) < 0.05 = excellent ordination, no prospect of misinterpretation
< 0.1 = good ordination, no real risk of misinterpretation
< 0.2 = usable ordination, risk of misinterpretation
> 0.2 = dangerous to interpret
> 0.35 = effectively randomly placed

As Dexter, Rollwagen-Bollens and Bollens (2018) argue, however, such guidelines 'do not account for the mathematical relationship which links ordination stress to sample size' and they show how large data sets may suffer from increased stress levels.

Below, in figure C.3, the NDMS-configurations for each of the distance measurements is presented, together with the stress index. Please note that for each configuration, two dimensions were used. While increasing the number of dimensions generally decreases stress, the configuration becomes harder to visualise and interpret (Dexter, Rollwagen-Bollens & Bollens, 2018, p. 434). Keeping the number of dimensions at two both conforms to the standards in the literature, and allows for comparison and easier interpretation of results.

Figure C.3:
NMDS configurations and stress levels for distance matrices (full feature set)



Note. All configurations are based on two-dimensional ordination.

What can be seen here, is that all stress levels are above 0.20, and only two are between 0.20 and 0.30.[9] Again, one should be careful in interpreting these figures in isolation, but together with the inspection of the distributions of the distances, it seems that no measure indicates strong clusterability. The Gower and Eskin measures indicate a relatively low stress level. Especially the Eskin measure seems able to produce a configuration with well-separated groups of observations. The above does not mean that no further steps can be taken in the exploration of the feature space of Dutch conditionals.

A first possible cause for the results reported above became clear in the previous chapter, namely that a number of features showed significant and sometimes strong associations to mode and register. It therefore could be argued that the overall analysis of the data may be troubled by these factors. Therefore, distance matrices were also calculated per mode-register combination (spoken-formal, spoken-informal, written-formal, written-informal). The corresponding distributions of distances per measure and the NMDS configurations are presented below.

[9] Although Dexter, Rollwagen-Bollens and Bollens (2018, pp. 437–438) show a clear 'asymptotically increasing relationship between ordination stress and sample size', using both field-derived and simulated data, as we will see in what follows, sample-size does not seem to be the main cause of the high stress levels here.

Figure C.4:
Distribution of distances per measure (spoken-formal sub-corpus)

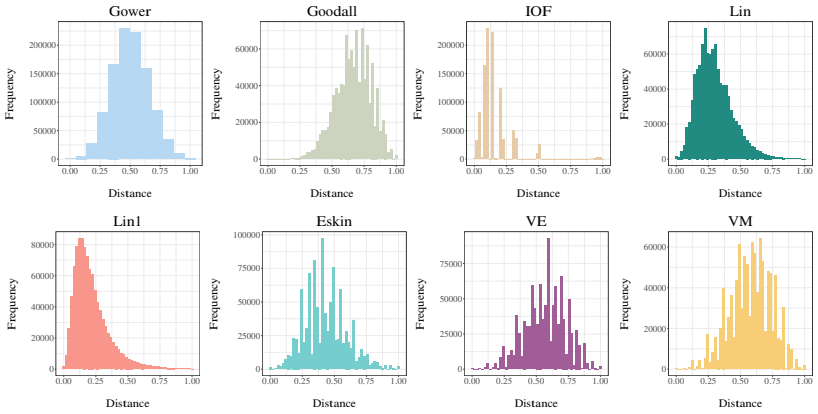
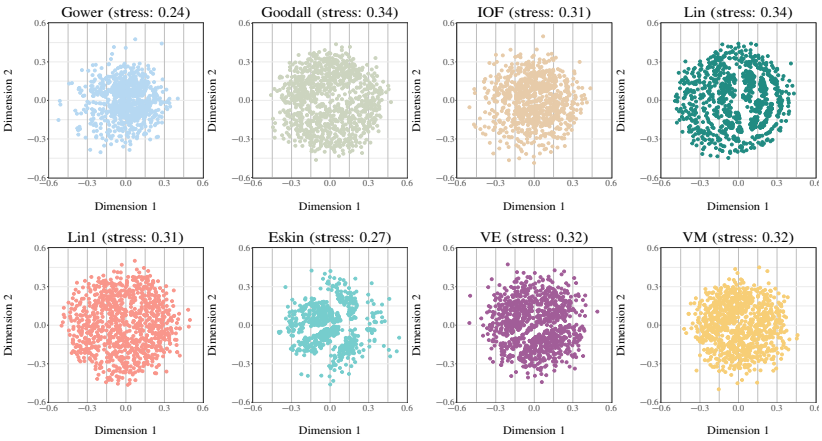


Figure C.5:
NMDS configurations and stress levels for distance matrices (spoken-formal sub-corpus)



Note. All configurations are based on the full feature set and on two-dimensional ordination.

Figure C.6:
Distribution of distances per measure (spoken-informal sub-corpus)

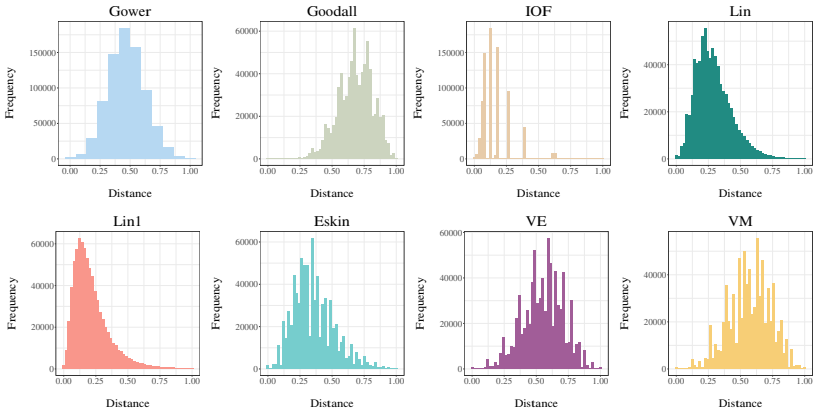
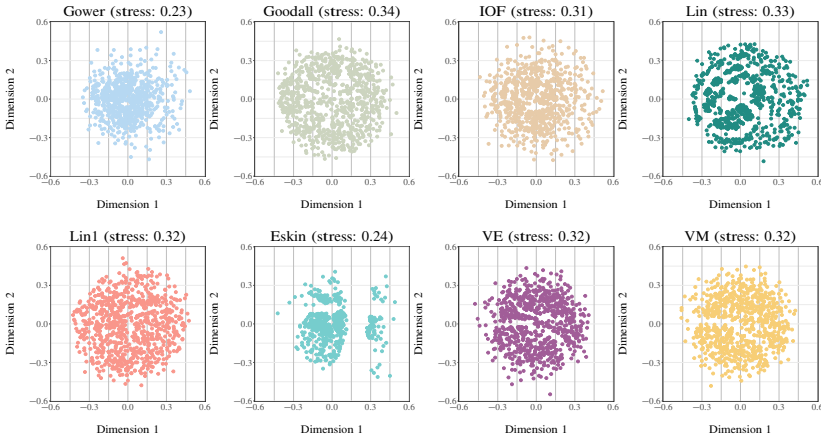


Figure C.7:
NMDS configurations and stress levels for distance matrices (spoken-informal sub-corpus)



Note. All configurations are based on the full feature set and on two-dimensional ordination.

Figure C.8:
Distribution of distances per measure (written-formal sub-corpus)

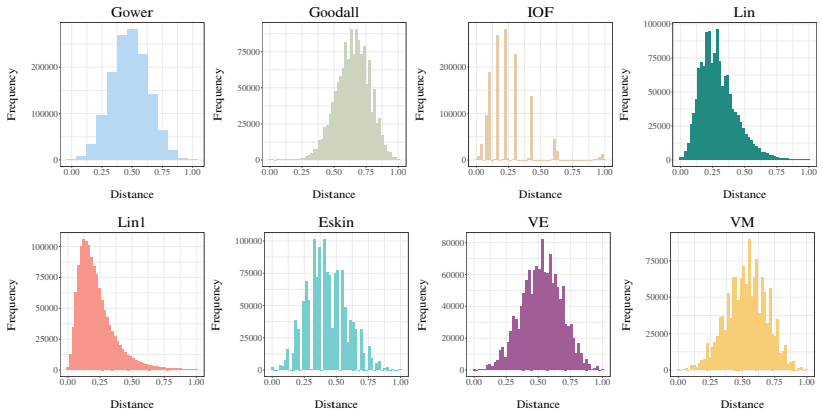
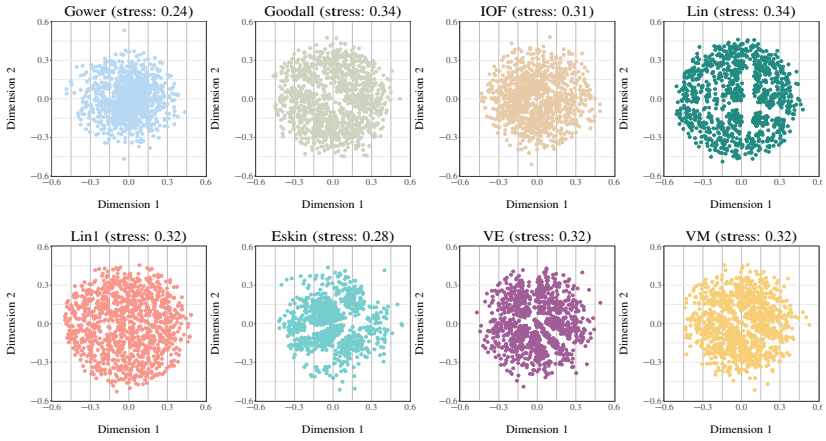


Figure C.9:
NMDS configurations and stress levels for distance matrices (written-formal sub-corpus)



Note. All configurations are based on the full feature set and on two-dimensional ordination.

Figure C.10:
Distribution of distances per measure (written-informal sub-corpus)

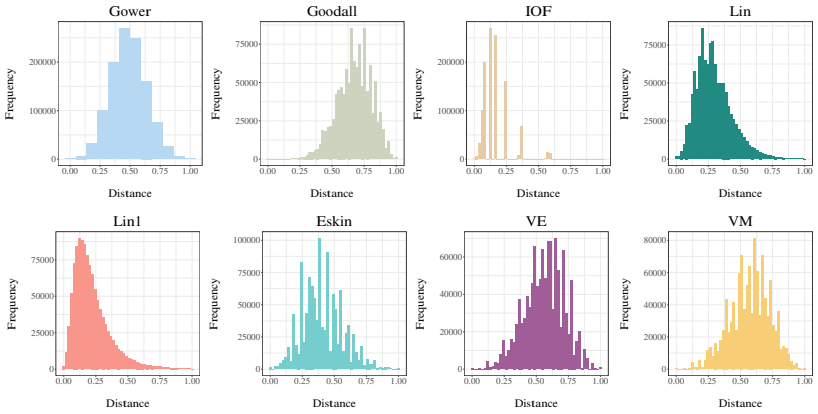
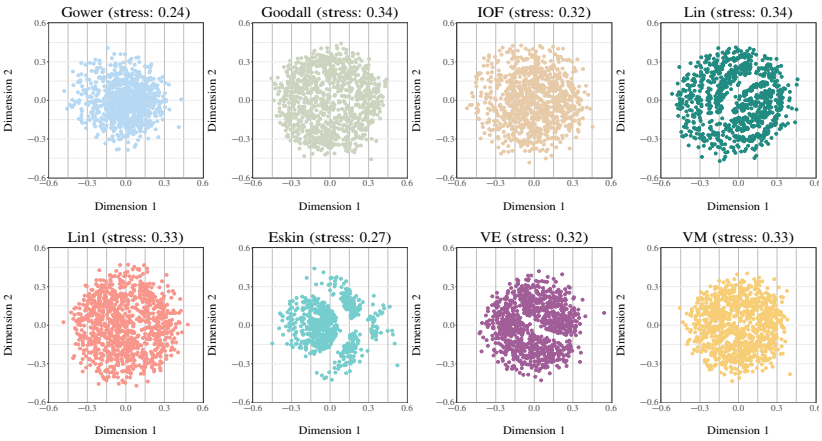


Figure C.11:
NMDS configurations and stress levels for distance matrices (written-informal sub-corpus)



Note. All configurations are based on the full feature set and on two-dimensional ordination.

As these figures show, there is no indication that mode and register have a systematic impact on the distributions of distances and stress levels in dimension reduction. In other words, calculating distances per sub-corpus does not seem to increase the clusterability of the data set.

A second possible cause was already discussed in section 6.3, namely that clustering algorithms may suffer from datasets including variables that are not relevant to the set of variables that indeed do show signs of underlying structure. Therefore, the next section returns to variable selection in order to investigate whether clusterability can be improved by removing variables from the feature space of conditionals.

C.6 Final variable selection

To evaluate the contribution of each variable in the dataset, a number of models was generated, including a full model with all variables, an informed model using only those variables suggested by the initial variable selection discussed above, supplemented, for testing purposes, with a number of random models. For a general discussion of this step, see section 6.3.6.

For reasons of computation time, the models were based on a random sample of 500 conditionals from the corpus. First, all distance measures discussed in section 6.3.4 were calculated for this sample. The histograms were plotted to check comparability to distances for the full dataset, and, crucially, the same NMDS-procedure was followed. As may be expected by sample size, the results were almost identical to those presented in figure C.3.[10] This indicates that the sample is representative of the complete dataset. The resulting NMDS configurations and goodness-of-fit values may therefore be used as a baseline for further variable selection. The next step was to follow the same procedure as above, but for an 'informed model' i.e., the set of features indicated by the initial variable selection. This model involves seven instead of twelve variables, namely syntactic integration, negation in the antecedent and in the consequent, modality in the antecedent and in the consequent, and tense in the antecedent and in the consequent. Directly comparing the ordination results from this set to that of the full set of variables may be criticised however, as a lower number of variables provides less variation to be explained by a model. Therefore, five random sets of seven variables were selected and put through the same procedure. The results are presented below.[11]

[10] The rounded stress values were 0.23 (Gower), 0.35 (Goodall), 0.31 (IOF), 0.33 (Lin), 0.31 (Lin1), 0.27 (Eskin), 0.33 (VE) and 0.33 (VM).

[11] The random variable sets were the following.
Random model 1: aspect (a), aspect (c), focus particle, modality (c), negation (c), subject (c), tense (a)
Random model 2: aspect (c), modality (a), modality (c), negation (a), subject (a), subject (c), syntactic integration
Random model 3: focus particle, modality (a), negation (a), negation (c), tense (c), subject (a), syntactic integration

Table C.5:
Goodness-of-fit values for NMDS configurations

Model	Gower	Goodall	IOF	Lin	Lin1	Eskin	VE	VM
Full	0.23	0.35	0.31	0.33	0.31	0.27	0.33	0.33
Informed	0.15	0.28	0.24	0.27	0.24	0.17	0.24	0.24
Random 1	0.20	0.31	0.30	0.31	0.27	0.23	0.28	0.28
Random 2	0.21	0.32	0.31	0.32	0.30	0.23	0.30	0.30
Random 3	0.17	0.29	0.27	0.30	0.24	0.18	0.25	0.25
Random 4	0.19	0.31	0.29	0.32	0.25	0.21	0.28	0.28
Random 5	0.17	0.30	0.28	0.30	0.26	0.21	0.26	0.26

Note. Goodness-of-fit values are reported in terms of stress values.

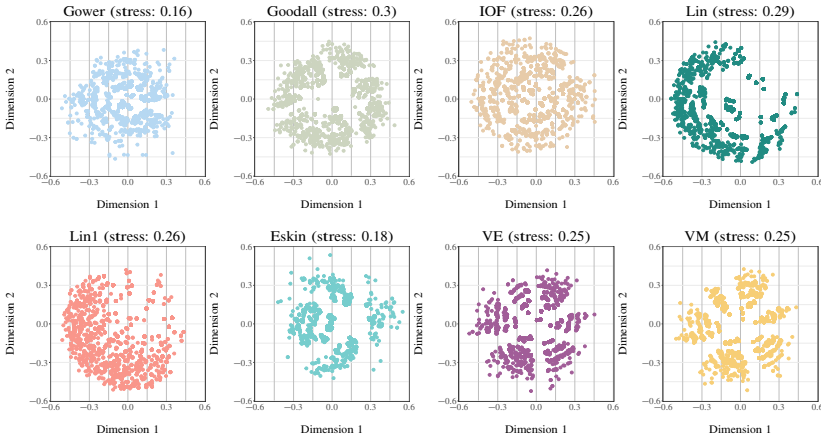
As we can see in this table, there indeed seems to be an effect of number of variables on ordination stress. However, it can also be observed that none of the models involving a random selection of seven variables performs as well as the informed model. We can see, however, that 'random model 3' comes close to the informed model, which can be explained by the fact that only two variables of this model do not appear in the informed model (focus particles, and subject of the antecedent). As an intermediate conclusion, it can be seen that removing features indicated as problematic by either a high frequency ratio (focus particles) or theoretical relevance (aspect, person and number) indeed improves the model. Further variables could be added or deleted in a stepwise-fashion until the lowest stress-values have been acquired, but as discussed in the sections above, there is a risk involved in not having an agreed upon measure of quality of a clustering solution. Furthermore, as less variables introduce less variation to be explained, a smaller model is not preferred per se. Although removing another feature from the informed model does slightly improve the fit of the NMDS-configuration, most notably when removing syntactic integration (resulting in stress values between 0.11 and 0.24), subsequent analyses would not take into account this feature, while it has been linked to conditional connections convincingly in the literature, placing it high in the theoretical ranking presented in section 6.3. By removing this variable, any variation concerning it introduces would not used for clustering, and would not be explained. While this may not be a perfect way to go about feature selection, as was already mentioned based on the literature in section 6.3, feature-selection for unsupervised machine-learning using categorical variables is problematic. As no agreed upon and robust methods have been found to evaluate unsupervised machine-

Random model 4: aspect (a), focus particle, negation (a), negation (c), subject (a), tense (a), syntactic integration
Random model 5: aspect (a), aspect (c), modality (a), negation (c), subject (c), tense (a), tense (c)

learning outcomes based on categorical data, the variable selection from the informed model in table C.5 was selected for the subsequent steps in the analyses.

The results of dimension reduction on this feature set on the complete dataset are presented in the NMDS configurations in Figure C.12 below.

Figure C.12:
NMDS configurations and stress levels for distance matrices (reduced feature set)



Note. All configurations are based on two-dimensional ordination.

As can be seen here, stress levels are lower for all metrics, and lowest for Gower and Eskin, which both indicate 'usable ordination', albeit with a risk of misinterpretation (see the guidelines listed in (6) on page 501).

APPENDIX D

Cluster evaluations

D.1 Introduction

In this appendix, the measures for cluster evaluations are discussed in technical detail (section D.2). In sections D.3 and D.4, the evaluations of hierarchical and partitional cluster solutions are discussed. In section D.5, finally, the dendrogram of the hierarchical clustering solution is presented. For the main discussion of clustering results and their evaluations, see chapter 6.

D.2 Measures of cluster evaluation

First of all, the homogeneity within clusters (i.e., within-cluster variance or *purity*) was measured using the *Within-Cluster Entropy* coefficient (WCE). The within-cluster variability in the k-cluster solution is calculated using the formula by Šulc and Řezanková (2019, p. 65) below, in which n_g is the number of observations in cluster g, and n_{gcu} is the number of observations in cluster g having value u for variable c.

$$(1)\ WCE(k) = \sum_{g=1}^{k} \frac{n_g}{n \cdot m} \sum_{c=1}^{m} \left(- \sum_{u=1}^{K_c} \left(\frac{n_{gcu}}{n_g} ln \frac{n_{gcu}}{n_g}\right)\right)$$

A low WCE-value reflects low within-cluster variability, i.e., high within-cluster homogeneity. As this value is influenced by the number of clusters, a higher number of clusters will result in a lower WCE-value (cf. Ladds et al., 2018, p. 10), which is one of the reasons hinted at above to use a number of indices, rather than just one measure of clustering quality.

Second, *separation* of clusters, i.e., between-cluster variance (see Sevcik, Rezankova & Husek, 2011), was measured using *Pseudo F Coefficient based on Entropy* (PsFE), as discussed by Šulc (2016, p. 33) and applied by Ladds et al. (2018), presented in the formula below.

(2) $PsFE(k) = \frac{(n-k)[nWCE(1)-nWCE(k)]}{(k-1)nWCE(k)}$

The PsFE-value depends on the number of observations n, the number of clusters k, variability in the complete dataset *nWCE(1)* and within-cluster variability *nWCE(k)*, where k is the number of clusters in the solution. The higher the PsFE-value, the better the grouping distinguishes between groups.

Third, *consistency* was measured by means of average silhouette width (the *Silhouette Coefficient*). The silhouette width of an observation, calculated using the formula from Kaufman and Rousseeuw (1990, p. 85) below, in which i is an observation, *a(i)* is the 'average dissimilarity of i to all other objects of A' and *b(i)* is the closest neighbouring cluster, i.e., the cluster that has the lowest average dissimilarity to cluster A.

(3) $\mathrm{s(i)}=\frac{b(i)-a(i)}{max\{a(i),b(i)\}}$

Here, the Silhouette Coefficient (or 'average silhouette width'; see Kaufman & Rousseeuw, 1990, p. 87) was used to assess the consistency of a clustering solution given k clusters. The solution with the highest Silhouette Coefficient is indicated to be the most appropriate solution for the given dataset. As silhouette widths are normalised values between -1 and 1, and as this coefficient is not restricted to particular algorithms, it can be used to assess the clustering solutions in more absolute terms. Kaufman and Rousseeuw (1990, p. 88) provide the following interpretation guidelines, which, as with other such guidelines, should be used with caution and attention to the underlying data. Silhouette Coefficient between 0.71 and 1.00 suggest a 'strong structure' has been found, while values between 0.51 and 0.70 suggest a 'reasonable structure'. Values between 0.26 and 0.50 indicate a 'weak structure' and it is advised to try applying additional methods on the dataset. Values lower than 0.26 indicate 'no substantial structure'. Negative values indicate observations are grouped in the wrong cluster.

Fourth, a measure used during data preparation (see section C.2 in appendix C), namely *deviation from the mode* (DM), was used as a counterweight to silhouette widths, as in testing the algorithms and evaluations, some silhouette coefficients indicated high consistency for solutions with extremely big or small clusters. This was especially the case for solutions with a low number of clusters. As it is, of course, unwanted to select a solution with such a skewed membership distribution, *DM*-values were used as an index of dispersion over clusters formed. Please note that a high DM-value is not preferable per se, as a cluster solution does not require comparable frequencies per cluster, but a very low DM-value is an indication of extreme size differences between clusters.

Fifth and finally, the stability of clustering solutions was evaluated using a bootstrapping procedure. In this case, 100 samples of the dataset were taken. Each of these samples consisted of randomly selected points from the dataset. Each point could be selected more than once. Clustering was performed on the random samples (see Hennig, 2007; Hennig & Liao, 2013, pp. 325–330). The result was evaluated using the Jaccard similarity index, expressed in (4) below (see Arnaboldi et al., 2015, pp. 87–88), which reflects the similarity between the clustering solution for the random sample under inspection W_1 and the original clustering solution W_2, by dividing the intersection $\cap$ of both sets (i.e., members of the same cluster in both sets) by their union $\cup$ (i.e., all members of both sets). In other words, the coefficient reflects the proportion of observations from the sample that belong to the cluster that the matches the same cluster in the originally found or 'true' clusters (cf. Hennig, 2007, p. 261).

(4) $J(W_1, W_2) = \frac{|W_1 \cap W_2|}{|W_1 \cup W_2|}$

The Jaccard coefficient is a value between 0 and 1 and the higher the value, the greater the overlap between the current sample and the clustering solution. For each clustering solution, the mean Jaccard similarity over all 100 samples was used as an index for cluster stability.

D.3 Evaluation of hierarchical cluster solutions

Agglomerative clustering starts with k clusters, where k is equal to the number of observations – in this case, 4109 conditionals. It goes without saying that a 'solution' of 4109 clusters does not provide any insights. Therefore, at each *run* or *iteration*, the algorithm merges the closest clusters, until the number of clusters is 2 (as a 'solution' of one cluster is as uninformative as a $k{=}n$ 'solution'). Now, the question is how the algorithm determines which clusters are closest. With the initial clusters, each cluster holds one observation, and the distance between those 'clusters' coincides with the calculated distance between the two observations. This becomes problematic, however, for each subsequent step, as clusters now contain more than one observation. The parameter *linkage* determines how the algorithm calculates the distance between two clusters, i.e., how the 'closeness' of two clusters is defined (see Kaufman & Rousseeuw, 1990, pp. 45–48). In *single linkage* or the *nearest neighbour rule* the similarity between two clusters is defined as the distance between their two most similar members, and consequently, the two clusters with the smallest distance between their most similar members are merged. This linkage criterion is said to be *local*, because it only considers the areas of clusters that are closest to each other. Next, in *complete linkage* or the *furthest neighbour rule* the similarity of two clusters is defined as the distance between their two most dissimilar members. The complete linkage criterion is non-local, as it is influenced by complete clusters, which lie in between the most dissimilar members of each cluster, instead of only their closest areas. While this can be seen as an advantage, it

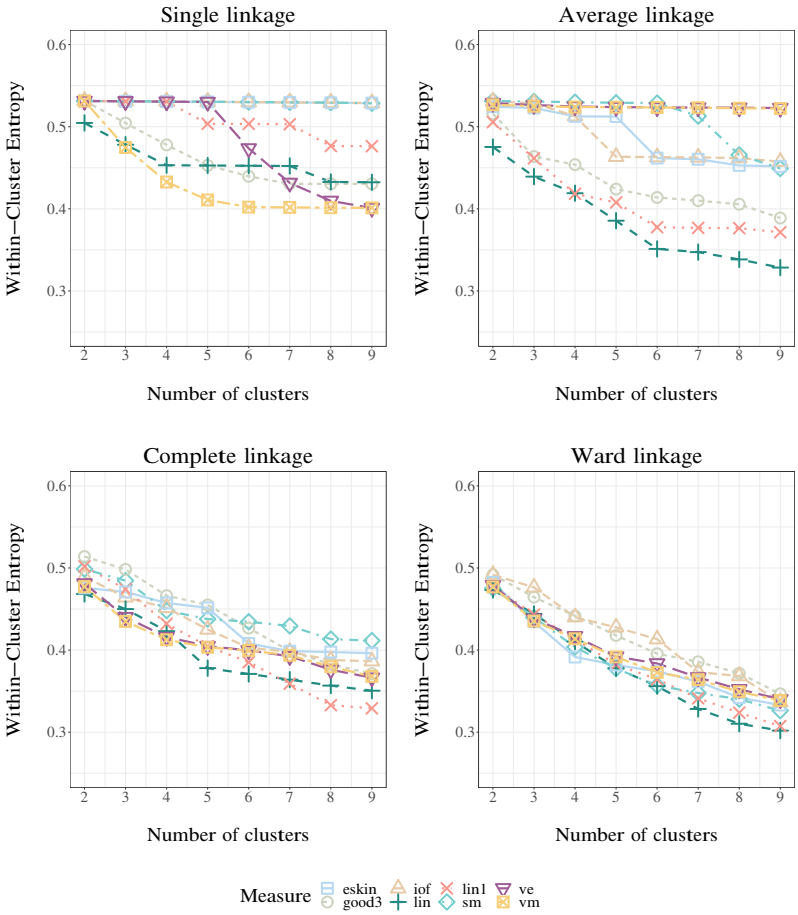
also means that complete linkage is more sensitive to outliers (see Kaufman & Rousseeuw, 1990, p. 227; Cibulková et al., 2019, p. 37). *Average linkage* is, as the name implies, a compromise between single and complete linkage in that it measures the distance between two clusters in terms of the difference between the average of the dissimilarities of all their respective members. Finally, there is Ward's *Minimum Variance Method* (cf. Ward, 1963; see also Anderberg, 1973, pp. 42–44; Kaufman & Rousseeuw, 1990, pp. 230–234; Legendre & Legendre, 1998, pp. 329–333). Ward linkage calculates the distance from each observation to the centroid (the mean distance) of the cluster it is assigned to.[1] At the start, all clusters contain only one observation, so the centroid and observation coincide, hence the sum of all distances is 0. At each next step, however, the distance between observations centroids increases, and the sum of the squared distances also increases. At each step, Ward linkage forms clusters based on the combination of observations or clusters that increase the squared distance from the centroids the least. As the optimum linkage depends on the specific dataset used, and the optimum can be operationalised by information-theoretic notions, linkage was chosen by comparison of evaluations.

The number of clusters or k depends, as one may expect, mostly on the inherent structure in the dataset, and on theoretical preconceptions, as for some datasets, there may be theoretically informed choices for the number of clusters. It also depends on what was called 'simplicity' in section 6.2.5, as a classification with a large number of types may miss important generalisations, whereas a classification with a small number of types may risk overgeneralisation (see also the discussion in section 7.3). For each distance matrix discussed in section 6.3, a clustering solution using each combination of linkage parameter (*single*, *average*, *complete*) and number of clusters k (2 to 9) was generated. For each of these solutions, the five evaluation indices discussed in the previous section were calculated. Each criterion is discussed below.

Figure D.1 below present the 'Within-Cluster Entropy' (WCE) of clustering solutions using each of the linkages discussed above for 2 to 9 clusters. As discussed before, a lower WCE-value indicates more homogeneity within the clusters.

[1] There are two algorithms implementing Ward's *Minimum Variance Method*, which may lead to confusion. See Murtagh and Legendre (2014) for a clear explanation and overview. In this study, Ward's (1963) original criterion is used, which is implemented as 'ward.D2' in the (base) R-function *hclust* and as 'ward' in the Agnes-function of the R-package *cluster* (Maechler et al., 2019).

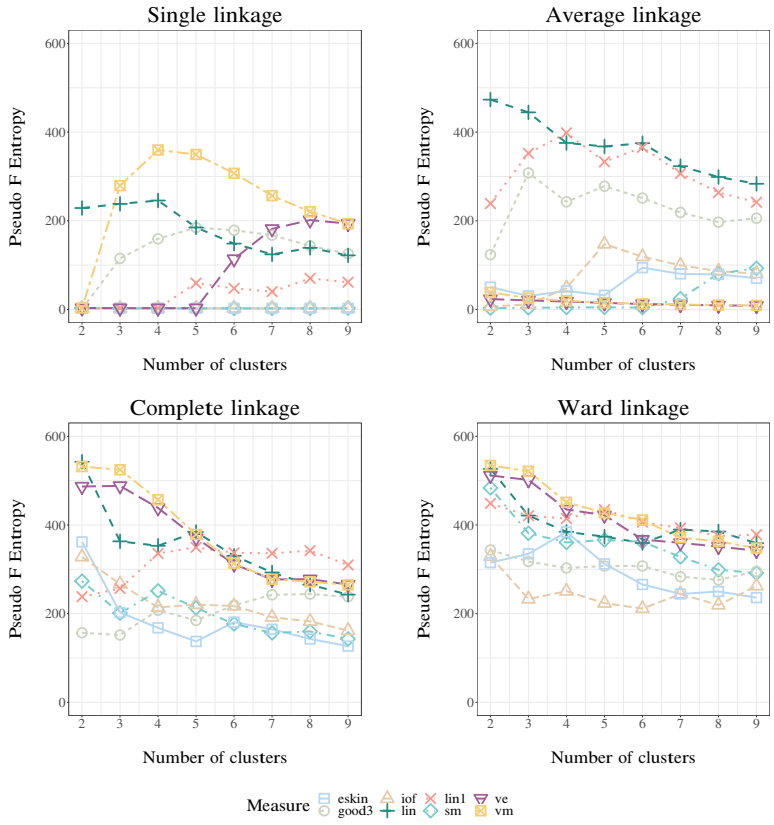
Figure D.1:
Evaluation of optimal linkage and number of clusters by Within-Cluster Entropy (WCE)



What can be seen in Figure D.1, is that the WCE-values for single-linkage solutions are higher on average than for the other linkages. For average linkage, especially the Goodall, Lin, and Lin1 measures perform better, and for complete and Ward linkage, it can be seen that the Lin and Lin1 measure produce the most homogeneous clusters, with WCE-values decreasing with increasing number of clusters, especially between 2 and 6 clusters with complete linkage.

In Figure D.2 below, the 'Pseudo F Entropy' (PsFE) of clustering solutions using each of the linkages discussed above for 2 to 9 clusters is presented. As discussed before, a higher PsFE-value indicates more heterogeneity between the clusters (i.e., better separated clusters).

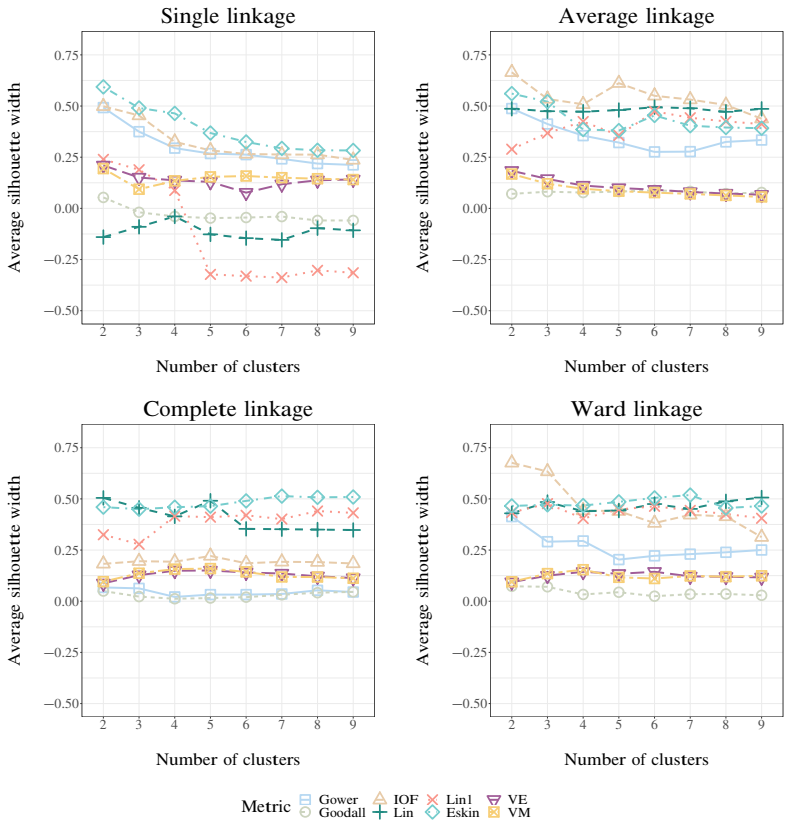
Figure D.2:
Evaluation of optimal linkage and number of clusters by Pseudo F Entropy (PsFE)



What can be seen in Figure D.2, is that a number of PsFE-values for single and, to a lesser extent, average linkage are extremely low in comparison to complete and Ward linkage. For single linkage, this is comparable to the results on clustering of categorical data reported by Ladds et al. (2018, p. 13). For average linkage, high PsFE-values were found for the Lin measure, especially between two and five clusters, and for the Lin1 measure, which is more stable, especially between two and six clusters. As these values, especially for the low-cluster solutions, strongly deviate from the other measures, dispersion within these solutions must be critically assessed. All the other measures have low separation values, although some measures show clear increases from 4 clusters and up, such as solutions using the Goodall measure and IOF measure. Complete linkage shows a somewhat different picture, with high values for the VE and VM measures, which however decrease rapidly after two-cluster solutions. Again, Lin and Lin1 are relatively high, with a peak for the Lin1 measure at four clusters. Ward linkage also produces high values for the VE and VM measure, which, like with complete linkage decrease rapidly after two-cluster solutions. Given the stability of separation values, this figure suggest a solution of two to four clusters using the Lin1 measure with average, complete or Ward linkage, or the VE or VM measure with complete or Ward linkage.

The consistency of cluster membership is visualised by means of Silhouette Coefficient, i.e., the maximal average silhouette width for the complete dataset, in Figure D.3 below.

Figure D.3:
Evaluation of optimal linkage and number of hierarchical clusters by Silhouette Coefficient (SC)

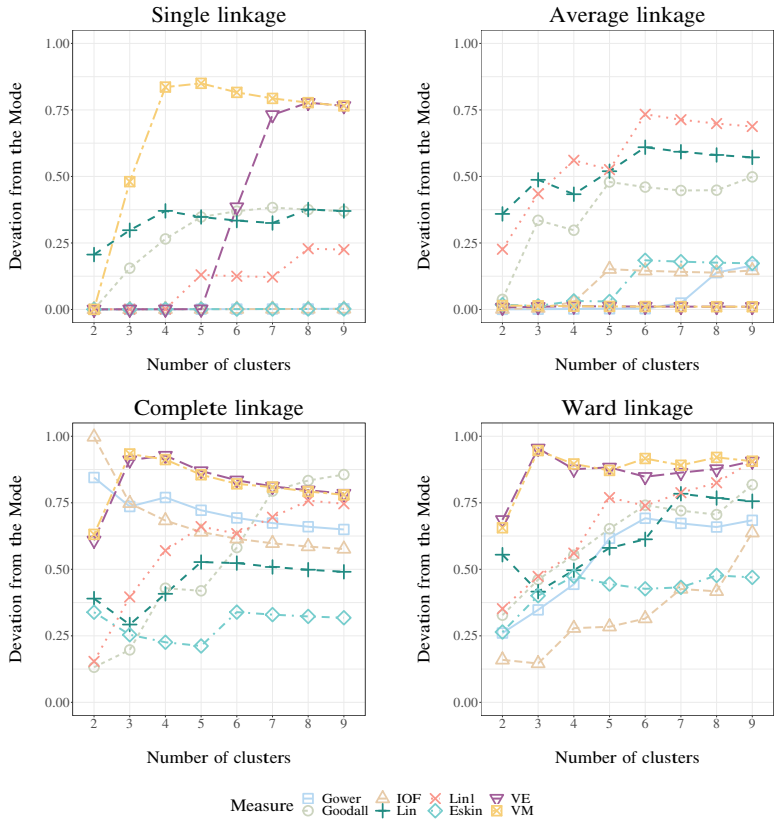


Before looking at the values in this figure, it is important to stress here that no strong conclusions should be drawn yet, as silhouette coefficients can be used especially to estimate the correct number of clusters, but they only paint part of the picture of cluster quality. For the single-linkage solutions, high coefficients were found for the Eskin and IOF measures. Lin and Lin1, however, produce negative silhouette coefficients, indicating that, on average, conditionals seem to be placed in the wrong clusters. For average linkage, IOF seems to perform best especially in 2- and 5-cluster solutions. For complete linkage, values are lower, which can be explained by the fact that complete linkage tends to produce very compact clusters, often resulting in less separation between clusters (Kaufman & Rousseeuw, 1990, pp. 7–48). The Lin, Lin1 and Eskin measures perform around the lower 'reasonable structure'-bound. For Ward

linkage, high values were obtained for the IOF measure, which, however, decrease rapidly after three-cluster solutions. It can also be seen that the Lin, Lin1 and Eskin measures perform reasonably well, as with complete linkage. None of the coefficients suggests a 'strong structure', however.

As discussed before, it is particularly useful to interpret silhouette coefficients in relation to measures of dispersion, because clusters with extreme size differences may still yield high silhouette widths, while they are not very informative. As can be seen in Figure D.4 below, there are indeed solutions, especially produced with single and average linkage and low cluster numbers, that suffer from this issue.

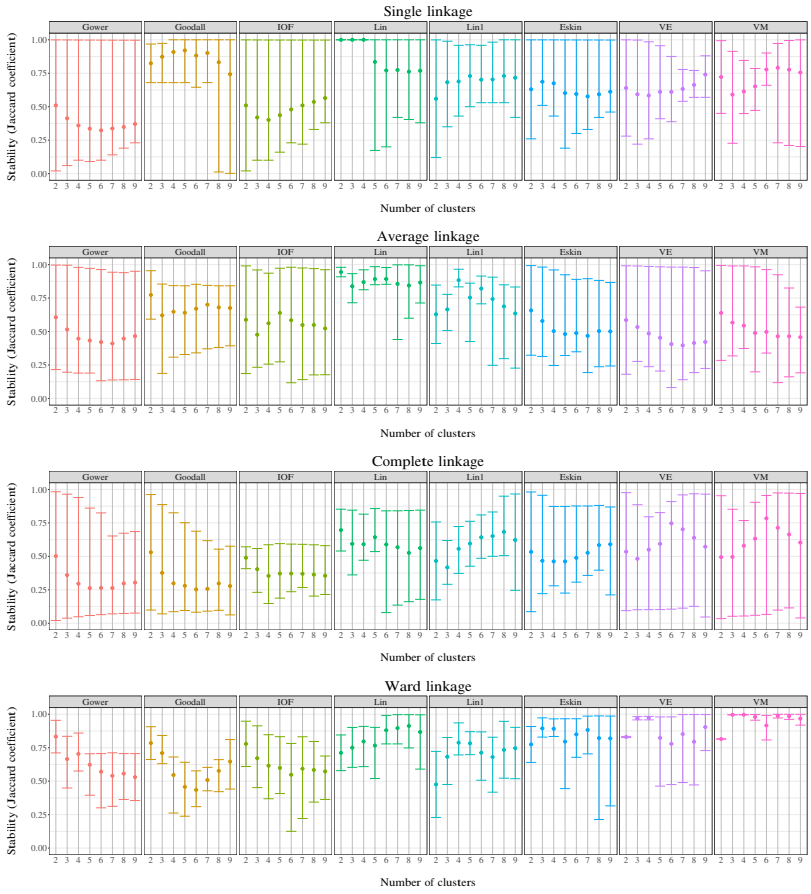
Figure D.4:
Evaluation of dispersion of hierarchical clustering solutions by Deviation from the Mode (DM)



As dispersion values, implemented as *deviation from the mode* (DM), were only used here to check for very skewed membership distributions, not as an absolute indication of quality, the figure above is useful mostly for identifying solutions with extremely low dispersion values, i.e., it would be undesirable to select a solution simply on the basis of high a high DM-value, because it does not make sense to claim that a clustering in which the memberships are balanced evenly is a good solution per se. In fact, given what we have discussed in chapter 3, it is probable that a prototypical type of conditional is more frequent than non-prototypical types. As a clear example of the importance of balancing silhouette widths and (some measure of) dispersion, take the Eskin 2-cluster single-linkage solution: it has one of the highest average silhouette widths, but its dispersion is extremely low (0.0004867364). Upon inspection of the solution itself, it turned out that this solution consists of two clusters, of which one cluster consists of only one conditional, while the other cluster holds the rest. This is of course also reflected in high within-cluster variation and low between-cluster separation.

Finally, the stability of the solutions was evaluated, because it is important to assess whether a particular solution is reproducible and stable. Figure D.5 below presents the stability values in terms of the average Jaccard coefficients based on a bootstrapping procedure of 100 samples per solution (see Hennig, 2007; Zumel & Mount, 2020, pp. 323–325).

Figure D.5:
Evaluation of stability of hierarchical clustering solutions by Jaccard coefficient



Note. Evaluations are generated by bootstrapping (n=100). Dots represent the mean Jaccard coefficient; error bars represent standard deviation.

It can be seen that the stability for most single-linkage solutions are low and/or show more deviation compared to average, complete and Ward's linkage.[2] Especially the latter shows higher mean Jaccard coefficients with less deviation overall. It can also be seen that the Lin and Lin1 measures score high on stability, in terms of both a high mean Jaccard coefficient and relatively low deviation. All other measures seem to be less stable, especially the VE and VM measure for complete linkage, while, remarkly they score high on stability for Ward linkage. Given their low silhouette coefficients, however, these should be interpreted with caution.

D.4 Evaluation of partitional cluster solutions

In partitional clustering, the first parameter, or rather choice of algorithms, depends on the partitioning approach. First, the *Partitioning Around Medoids* (PAM) described in section 6.4.2 was selected because of its widespread application, also to categorical datasets (see e.g., Ladds et al., 2018; for linguistics-oriented studies using *PAM*, see Douven, 2017a; Wälchli, 2018). As discussed, this algorithm works in two steps. First, in the so-called 'build phase', in which the algorithm selects k medoids (i.e., most representative points), allocating each observation to the nearest medoid. Second, in the 'swap phase', changes are made to the allocation of observations to medoids and the average dissimilarity per cluster is calculated. This is done until the average dissimilarities no longer decrease. As an observation can only be member of one cluster, this is a form of hard-clustering. Second, Fuzzy Analysis or *Fanny* was used, which is a form of soft-clustering, as it assigns to each object not a definitive cluster choice, but a membership coefficient, indicating how well that particular objects fits within each cluster. In contrast to the use of representative objects by PAM, this approach does not choose representative observations as medoids, but it minimises the dispersion over all clusters for each observation, as memberships of individual objects should be as large for its most appropriate cluster as possible, while being as low as possible for the other clusters formed (for more details, see Kaufman & Rousseeuw, 1990, p. 171). The algorithm is also capable of hard-clustering, however, simply by selecting the cluster with the highest membership coefficient for each object. This is important, as it allows for applying the same evaluation measurements as for the other clustering solutions.

[2] As can be seen, the 2- to 4-cluster Lin single-linkage solutions, and a number of VE and VM solutions using Ward linkage are maximised and show no deviation. In the latter cases, this is because of very low dispersion, resulting in one large and one small, but very stable cluster. Inspection of the single linkage Lin-clusters however show that this is not the case. Another cause may be a number of very similar clusters merged in an early step, the so-called *chaining effect* (see Kaufman & Rousseeuw, 1990, p. 48). As a number of evaluation measures were used, however, for which single linkage performs worst, this point will not be inspected further.

The second parameter is the number of clusters k, which needs to be set before clustering. To evaluate the optimal number of clusters, solutions with k ranging from 2 to 9 are generated and evaluated. As mentioned before, this is a usual practice in studies developing and applying clustering algorithms.

In Figure D.6 below, the homogeneity within clusters expressed as WCE is presented for both the partitioning around medoids algorithm (PAM) and the fuzzy analysis (FANNY) algorithm.

Figure D.6:
Evaluation of PAM and FANNY solutions by Within-Cluster Entropy coefficient (WCE)

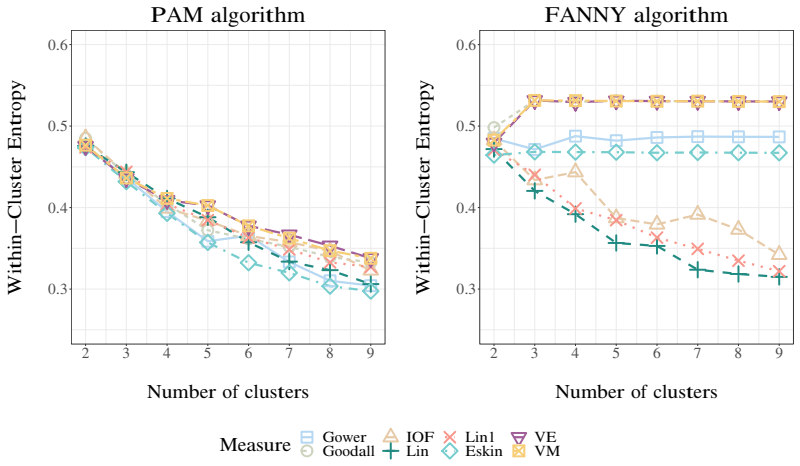
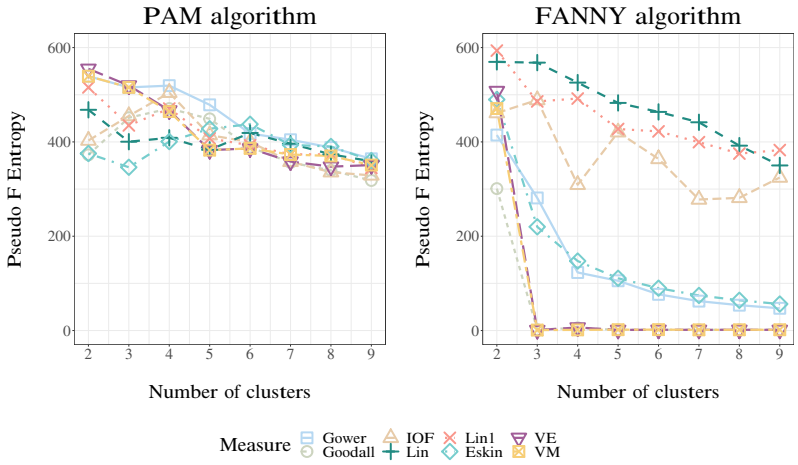


Figure D.1 shows that the within-cluster variability decreases for higher cluster numbers for all measures with the PAM algorithm, comparable to the decrease of within-cluster variability for hierarchical clustering using complete and Ward linkage (see Figure D.1 on page 515). For the FANNY algorithm, however, a relatively high within-cluster variability was found, with only the distances from the Lin, Lin1 and VE measure decreasing with an increasing number of clusters.

In Figure D.7 below, cluster heterogeneity measured in terms of PsFE is presented. As with hierarchical clustering in the previous section, higher PsFE values indicate more better separation between clusters.

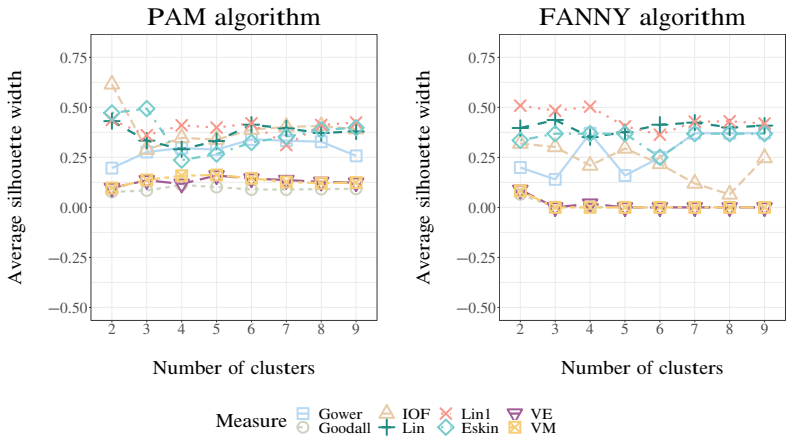
Figure D.7:
Evaluation of PAM and FANNY solutions by Pseudo F Entropy (PsFE)



Again, Figure D.7 presents more stable results for all measures using the PAM algorithm, whereas the FANNY algorithm provides relatively high separation values for the Lin and Lin1, and to a lesser extent the IOF measures only. For the other measures, we see a drastic drop in between-cluster variance. Please note that a contributing factor may be the fact that the soft-clustering has to be converted into hard (or 'crisp') cluster assignments.

The consistency of cluster membership is visualised by means of silhouette coefficients, as presented in Figure D.8 below.

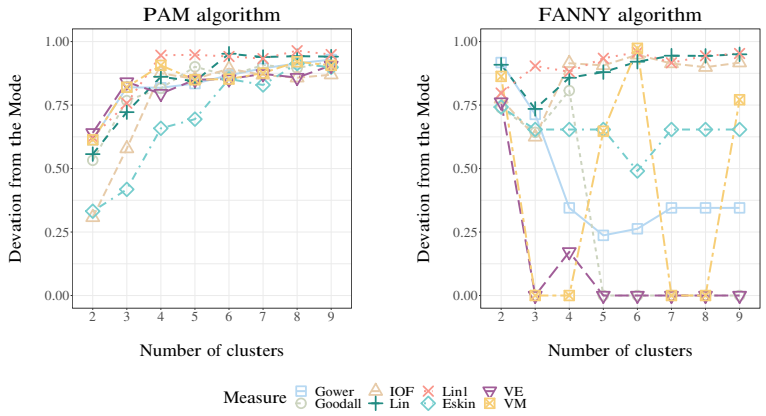
Figure D.8:
Evaluation of optimal algorithm and number of partitional clusters by Silhouette Coefficient (SC)



Using the same guidelines as for hierarchical clustering in the previous section, it can be seen that none of the solutions achieve high silhouette widths. In contrast to the single-linkage solutions evaluated in the previous section, none of the solutions had negative silhouette coefficients. In line with the previous measures, very low silhouette coefficients were found for fuzzy clustering for a number of measures, and the same trend can be observed, albeit less dramatically so, for the partitional algorithm. As for the results of hierarchical clustering, a peak in the PAM results for the two-cluster IOF-solution was found, but this result has to be evaluated in terms of dispersion too, as for hierarchical clustering, two-cluster solutions sometimes had extremely low dispersion values. What can also be seen in Figure D.8, is that the Lin and Lin1 measure reach relatively high and stable SC's in fuzzy clustering, peaking at just above 0.50 for the four-cluster Lin1 solution.

As mentioned before, it is important to uncover potential problems with extreme cluster sizes. The deviations from the mode (DM) per clustering solution are presented in Figure D.9 below.

Figure D.9:
Evaluation of dispersion of partitional clustering solutions by Deviation from the Mode (DM)



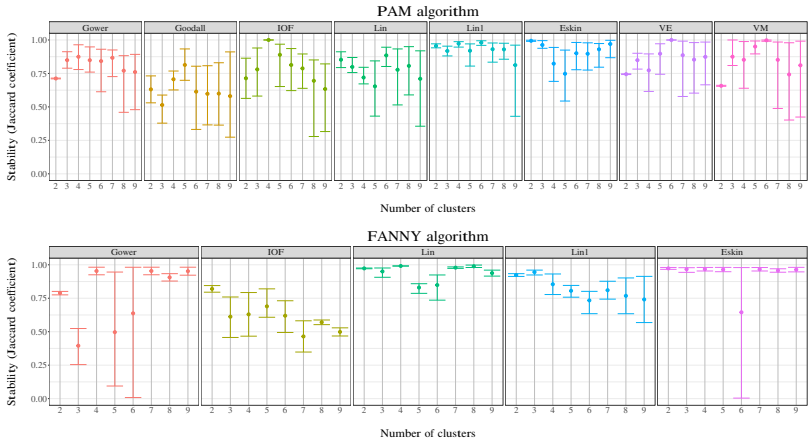
Again, higher dispersion values are not better per se – these figures are used to find extremely low dispersions. The lineplots for the FANNY results may look strange due to the sharp decreases and increases in dispersion values, but upon closer inspection, this can be explained by the fact that a number of solutions effectively clustered all conditionals into one cluster, rendering zero-dispersion and unusable results. The reason for this, in turn, is that these solutions suffered from evenly-spread cluster memberships, i.e., each observation was assigned the same probability for each of the clusters. Inspection of, for instance, the fuzzy five-cluster Goodall solution indeed revealed that each observation had membership coefficients for each cluster very close to 0.2, which simply is $\frac{1}{k=5}$.[3] Forcing a crisp-clustering, then, forces the algorithm to choose the highest probability, while in fact all probabilities are the same. The algorithm then assigns all conditionals to the same cluster. This was the case for a number of solutions based on the Goodall, VE and VM measures. Therefore, these measures were not included in the bootstrapping procedure of which I will present the results below. Discarding the Goodall, VE and VM measures, lower values were found for the IOF and Eskin measure for solutions of four or less clusters resulting from the PAM algorithm.

Finally, the stability measures for the solutions were generated and inspected. In Figure D.10 below, the stability values in terms of the average Jaccard coefficients are presented, which are the result from the same procedure as re-

[3] As an illustration, the mean of the cluster coefficients for cluster 1 in this solution was 0.2 with a standard deviation of 0.00000002181074.

ported in the previous section. Please note that, given the unusable clusters resulting from a number of Goodall, VE and VM solutions, these measures were removed from the stability plot for the FANNY algorithm.

Figure D.10:
Evaluation of stability of partitional clustering solutions using Jaccard coefficient (SC)



Note. Evaluations are generated by bootstrapping (n=100). Dots represent the mean Jaccard coefficient; error bars represent standard deviation.

In Figure D.10, high stabilities with low variance for Lin and especially Lin1 can be seen, although stability decreases for those measures with an increase of clusters. High stability values were also found for the two-cluster Eskin solution, and the six-cluster VE and VM solutions.

As fuzzy clustering is a soft-clustering algorithm, it produces membership coefficients (MC). These are presented in Table D.1 below, which allows inspection of the coefficients to see how the probabilities of those cluster memberships are distributed.

Table D.1:
Membership coefficients (MC) of Lin1 FANNY solutions (2-4 clusters)

	Cluster 1			Cluster 2			Cluster 3			Cluster 4		
#	MC	*sd*	*max*	MC	*sd*	*max*	MC	*sd*	*max*	MC	*sd*	*max*
2 cl.	0.43	*0.29*	*0.98*	0.57	*0.29*	*0.99*						
3 cl.	0.30	*0.26*	*0.94*	0.38	*0.29*	*0.92*	0.32	*0.27*	*0.92*			
4 cl.	0.24	*0.26*	*0.93*	0.17	*0.21*	*1.00*	0.32	*0.30*	*0.92*	0.27	*0.27*	*0.92*

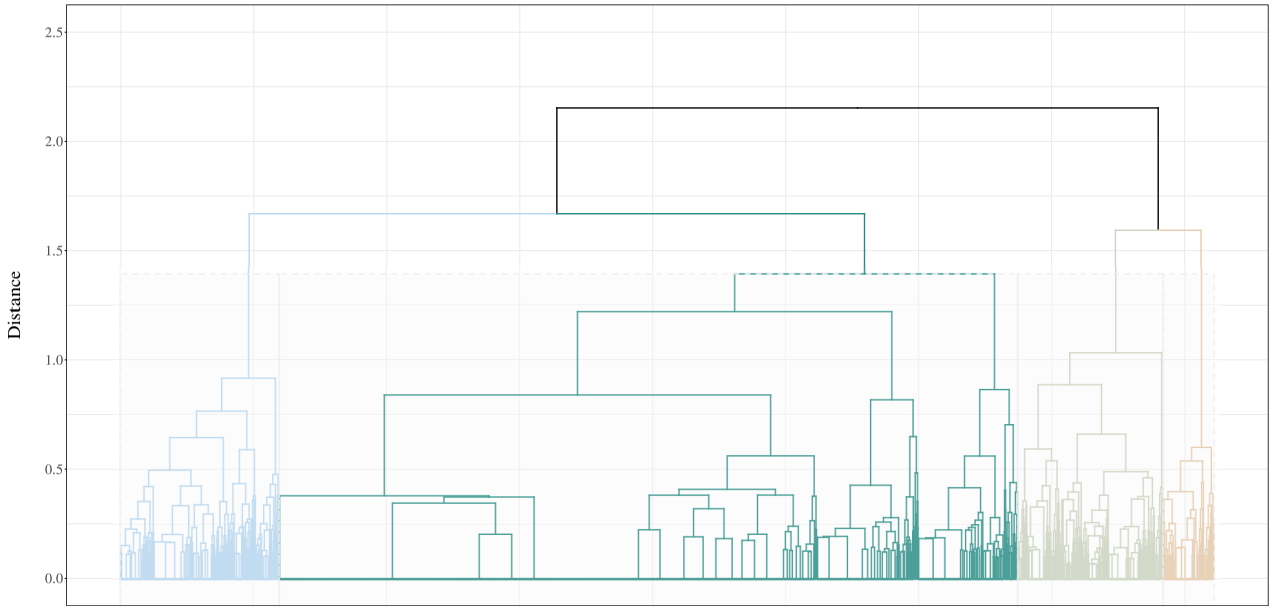
These numbers reflect that most conditionals have been assigned to clusters with probabilities well outside the problematic figures of evenly-spread membership assignments that were found for the fuzzy clustering based on the Goodall, VE and VM measures.

D.5 Dendrogram of hierarchical clustering

In additional to the cluster visualisation presented in Figure 6.5 in section 6.5, Figure D.11 presents the results of the selected hierarchical clustering as a dendrogram. The four clusters are indicated by borders. Please note that, due to the large dataset, only the top of the dendrogram, at which the main clusters discussed are formed, is displayed.

The horizontal axis at the bottom of the dendrogram accommodates all conditionals. The width of each cluster therefore approximates its size. The height or distance at which two objects or clusters join indicates their similarity, i.e., the smaller the vertical distance, the more similar two objects or clusters are. In Figure D.11, it can be observed that the sub-clusters in the left-most cluster are joined at a lower point in the dendrogram, meaning that they are more similar than, for instance, the sub-clusters of the second cluster from the left, as the former sub-clusters join at a height below 1, whereas the latter join at a height well above 1.

Figure D.11:
Dendrogram of hierarchical clustering (Lin, average, 4 clusters)



Note. From left to right, clusters 1, 3, 2, and 4 are presented. The colours used correspond to the colours used in section 6.5.

APPENDIX E

Annotation guidelines and experimental materials (classifications)

E.1 Introduction

In this appendix, the annotation guidelines for applying classifications of conditionals to corpus data are presented (see sections E.2 to E.4). Furthermore, in section E.5, the results of the calculation of rater reliability on corpus items are presented. See chapter 3 for discussion of these classifications, and section 4.2 for discussion on applying classifications to corpus data. Note furthermore that the annotation guidelines below are presented in English, whereas they were presented in Dutch for the experiment reported on in section 4.2.

E.2 Quirk et al.'s (1985) classification

E.2.1 Introduction

Quirk et al. (1985) present a classification of conditionals based on the type of connection between the antecedent and consequent. Before annotating, please read the section 'Conditional clauses' in Quirk et al. (1985, pp. 1089–1099) carefully.

The three main types of conditionals in this classification are direct conditionals, which are further divided into open and hypothetical conditionals, indirect conditionals, which are further divided into politeness, uncertainty, metalinguistic, and speech-act conditionals, and rhetorical conditionals. All types and subtypes are exemplified in (1) to (7) respectively.

(1) **Direct open conditional**
If you put the baby down, she'll scream. (Quirk et al., 1985, p. 1088)
Als je de baby neerzet, gaat ze schreeuwen.

(2) **Direct hypothetical conditional**
If you had listened to me, you wouldn't have made so many mistakes. (Quirk et al., 1985, p. 1091)
Als je naar me geluisterd zou hebben, zou je niet zo veel fouten hebben gemaakt.

(3) **Indirect politeness conditional**
If may be quite frank with you, I don't approve of any concessions to ignorance. (Quirk et al., 1985, p. 1095)
Als ik heel eerlijk mag zijn, keur ik geen enkele vorm van onwetendheid goed.

(4) **Indirect uncertainty conditional**
If I understand you correctly, the theory is heavily outdated.
Als ik je goed begrijp, is de theorie zwaar verouderd.

(5) **Indirect metalinguistic conditional**
She thinks she is more 'zen' than we are, if that's the right way of phrasing it.
Ze denkt dat ze meer 'zen' is dan wij zijn, als dat de juiste verwoording is.

(6) **Indirect speech-act conditional**
If you're interested, there's a flyer in my bag.
Als je interesse hebt, er zit een folder in mijn tas.

(7) **Rhetorical conditional**
If they're Irish, I'm the Pope. (Quirk et al., 1985, p. 1091)
Als zij Iers zijn, ben ik de Paus.

E.2.2 Instructions

Determine the most suiting type. Please note that grammatical features of types can be of help in determining the type, but they will most likely not exhaustively determine the type. As the type reflects a connection between antecedent and consequence, determining its type is at least partly interpretative, i.e. world

knowledge is needed. Annotate the according label. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the letters after the semicolon. They will be converted to their full counterparts after you are done annotating.

Direct-open (dio; 1)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent. The speaker is neutral about the likelihood of the situations.

Examples

(a) Als je de baby neerzet, gaat ze schreeuwen.
If you put the baby down, she'll scream.

(b) Als Colin in Londen is, dan verblijft hij in het Hilton.
If Colin is in London, he {stays/will stay} at the Hilton.

Tests

- The conditional can be embedded into a matrix clause, as in 'I think that if you put the baby down, she'll scream.'
- The conditional can be converted into a direct-hypothetical conditional by backshifting the tense of the finite verb, as in 'If you would put the put the baby down, she'd scream.'
- The integrative and resumptive word-order patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *Als Colin in Londen is, hij verblijft in het Hilton.* 'If Colin is in London, he {stays/will stay} at the Hilton.'

Direct-hypothetical (dih; 2)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent. The speaker expresses disbelief that the condition will be, is or was fulfilled and conveys the unlikelihood of the consequent being true.

Examples

(a) Als je naar me geluisterd zou hebben, zou je niet zo veel fouten hebben gemaakt.
If you would have listened to me, you wouldn't have made so many mistakes.

(b) Als ze harder zou praten, zou ze beter te verstaan zijn.
If she would speak louder, she would be easier to hear.

Tests

- The conditional can be embedded into a matrix clause, as in 'I think that if you had listened to me, you wouldn't have made so many mistakes.'
- The conditional can be converted into a direct-open conditional by removing the modal auxiliaries and adverbs and changing the tense of the finite verb into present tense, as in 'If you listen to me, you will not make so many mistakes.'
- The integrative and resumptive word-order patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *Als je naar me geluisterd zou hebben, je zou niet zo veel fouten hebben gemaakt.* 'If you had listened to me, you wouldn't have made so many mistakes.'

Indirect-politeness (inp; 3)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent expresses courtesy to soften the speech act in the consequent.

Examples

(a) Als ik heel eerlijk mag zijn, ben ik het niet met je eens.
If I may be honest, I do not agree with you.

(b) Als je het niet erg vindt dat ik het zeg, je trui is te kort.
If you do not mind me saying, your sweater is too short.

Tests

- The conditional cannot be embedded into a matrix clause, as in 'I think that if I may be honest, I do not agree with you.'
- Verb tense cannot be used to indicate epistemic uncertainty, as in *Als ik heel eerlijk zou mogen zijn, zou ik dat niet met je eens zijn.* 'If I would be permitted to be honest, I would not agree with you.'
- The integrative, resumptive and non-integrative word-order patterns are possible.

Indirect-uncertainty (inu; 4)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent expresses the uncertainty of the speaker. Mind that this may also be a politeness strategy.

Examples

(a) Als ik het goed heb begrepen, is de theorie verouderd.
If I have understood correctly, the theory is outdated.

Tests

- Verb tense cannot be used to indicate epistemic uncertainty, as in *Als ik het goed begrepen zou hebben, zou de theorie verouderd zijn.* 'If I would have understood correctly, the theory would be outdated.'
- The conditional cannot be embedded into a matrix clause, as in 'I think that if I understood correctly, the theory is outdated.'
- The integrative and resumptive patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *Als ik het goed heb begrepen, de theorie is verouderd.* 'If I have understood correctly, the theory is outdated.'

Indirect-metalinguistic (inm; 5)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent is used to comment on the linguistic form of (a part of) the consequent.

Examples

(a) Zijn stijl is bloemig, als dat het juiste woord is.
His style is florid, if that's the right word

Tests

- Verb tense cannot be used to indicate epistemic uncertainty, as in 'His style would be florid, if that would be the right word.'
- The conditional cannot be embedded into a matrix clause, as in 'I think that his style is florid, if that's the right word.'
- Only the sentence-medial and sentence-final clause order can be used. Sentence-initial order leads less acceptable result, as in *Als dat het juiste woord is, zijn stijl is bloemig.* 'If that's the right word, his style is florid.'

Indirect-speech act (ins; 6)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent presents the condition under which the speaker utters the consequent.

Examples

(a) Als je mijn kant op gaat, ik kan wel een lift gebruiken.
If you're going my way, I need a lift back.

(b) Als je trek hebt, er staan koekjes op tafel.
If you're hungry, there are cookies on the table.

Tests

- Verb tense cannot be used to indicate epistemic uncertainty, as in 'If you would go/be going my way, I would need a lift back.'

- The conditional cannot be embedded into a matrix clause, as in 'I think that if you're going my way, I need a lift back.'
- The resumptive syntactic pattern is not possible, as in *Als je trek hebt, dan staan er koekjes op tafel.* 'If you're hungry, then there are cookies on the table.'

Rhetorical conditional (rhe; 7)

The rhetorical conditional has the form of an open conditional, but presents an absurd situation in either the antecedent or the consequent. This absurdity licenses the falsehood of the other clause.

Examples

(a) Als zij Iers zijn, ben ik de Paus.
If they're Irish, I'm the Pope.

(b) Hij is negentig als hij al een dag oud is.
He's ninety if he's a day.

Notice that the subtype in which the antecedent is absurd ('He is a day'), as in (b), does not seem to be used in Dutch as opposed to English.

Tests

- Verb tense cannot be used to indicate epistemic uncertainty without losing the rhetorical function, as in 'If they were Irish, I would be the pope'.
- The conditional cannot be embedded into a matrix clause without losing the rhetorical function, as in 'I think that if they are Irish, I am the pope.'
- The integrative and resumptive word-order patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *Als zij Iers zijn, ik ben de Paus.* 'If they're Irish, I'm the Pope.'
- Only the sentence-initial clause order is possible (e.g., ? *Ik ben de paus(,) als zij Iers zijn.* 'I am the pope(,) if they're Irish.'

E.2.3 Problem cases

Please take note of the following known problem cases and code accordingly.

Incomplete utterance (*NA*)

If the utterance is incomplete, use this label. For instance, the antecedent or consequent does not have a finite verb or, in case of an imperative, does not have an overt subject.

Examples

(a) Als niet, dan toch.
If not, then still.

Evaluative conditionals (evo; 8, evh; 9)

Especially in Dutch conditionals, there seems to be a use in which the consequent present an evaluative remark on the situation presented in the antecedent. Use the labels 'evo' (8) for open evaluative conditionals (the speaker is neutral about the likelihood of the situations) and 'evh' (9) for hypothetical evaluative conditionals (the speaker expresses disbelief about the truth of the condition).

Examples

(a) Als dat mogelijk is, is dat geweldig.
If that is possible, that's great.

(b) Als dat zou kunnen, zou ik heel gelukkig zijn.
If that would be possible, I would be very happy.

E.3 Athanasiadou and Dirven's (1996) classification

E.3.1 Introduction

Athanasiadou and Dirven (1996) present a classification of conditionals based on the type of connection between the antecedent and consequent. Before annotating, please read Athanasiadou and Dirven's 'Typology of *if*-Clauses' (1996), 'Conditionality, Hypotheticality, Counterfactuality' (1997a), and 'Pragmatic Conditionals' (2000) carefully. The three main types of conditionals in this classification are neutral and non-neutral hypothetical conditionals, course-of-event conditionals and pragmatic conditionals. Further divisions into subtypes are not considered here. All types are exemplified below.

(8) **Unmarked hypothetical conditional**

(a) If the weather is fine, we'll go for a swim.

(b) If there is no water in your radiator, your engine will overheat immediately.

(c) If I go bald I'll shoot myself

(9) **Marked hypothetical conditional**

(a) If the weather would be fine, we would go for a swim.

(b) If there were a beast, I'd have seen it.

(10) **Course-of-event conditional**

(a) If there is a drought like this year, the eggs remain dormant.

(b) If you heat water to 100 degrees, it boils.

(c) He looked at his watch; if the soldier was coming, it was nearly time.

(11) **Pragmatic conditional**

(a) If there's one human species that ought to be put out to pasture, it's Presidents and Prime Ministers.

(b) If the super-organism created by a colony of termites can be compared to an antelope, then the disciplined aggressive columns of the army ants must be reckoned to be the insect equivalent of a beast of prey.

(c) What about the parents demonstrating, if there are no friends?

(d) I've come to offer my congratulations, if that's the right word.

E.3.2 Instructions

Determine the most suiting type. Please note that grammatical features of types can be of help in determining the type, but they will most likely not exhaustively determine the type. As the type reflects the coherence relation between antecedent and consequence, determining its type is at least partly interpretative, i.e. world knowledge is needed. Annotate the according label. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in your annotation. If you prefer shorter labels, you can use the letters after the semicolon. They will be converted to their full counterparts after you are done annotating.

Unmarked hypothetical conditional (hyn; 1)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent and the situations are hypothetical and not marked for (un)certainty, i.e., it is *neutral*. The speaker does not commit herself to the actual occurrence of the antecedent nor of the consequent. The hypothetical character relates to the occurrence of the antecedent and consequent, not to the relation between the two. The antecedent may present a cause of the consequent, a condition or a supposition.

Examples

(a) Als het goed weer is, gaan we zwemmen.
If the weather is fine, we'll go for a swim.

(b) Als er geen water in je radiator zit, oververhit je motor direct.
If there is no water in your radiator, your engine will overheat immediately.

Tests

- The conditional can be paraphrased with 'on condition that' or 'supposing that', as in 'We're going for a swim, on condition that the weather is fine'. If not, it can be paraphrased with 'because': 'He will shoot himself because he goes bald'.
- Negating the antecedent allows for the negated and non-negated consequent, as in *Als het geen mooi weer is, gaan we (toch) zwemmen.* 'If the weather is not nice, we will go for a swim (anyway).' and *Als het geen mooi weer is, gaan we niet zwemmen.* 'If the weather is not nice, we will not go for a swim.'

Marked hypothetical conditional (hym; 2)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent and the situations are hypothetical and, in contrast to the neutral version above, marked by verb tense and/or adverbial modification for a degree of (un)certainty. The speaker commits or distances herself from the actual occurrence of the antecedent and consequent. The hypothetical character relates to the occurrence of the antecedent and consequent, not to the relation between the two. The antecedent may present a cause of the consequent, a condition or a supposition.

Examples

(a) Als het goed weer zou zijn geweest, gingen we zwemmen.
If the weather would have be fine, we would go for a swim.

(b) Als er geen water in je radiator zou zitten, oververhitte je motor direct.
If there would be no water in your radiator, your engine would overheat immediately.

Tests

- The conditional can be paraphrased with 'on condition that' or 'supposing that', as in 'We would go for a swim, on condition that the weather was fine'. If not, it can be paraphrased with 'because': 'He would shoot himself because he went bald'.
- In most cases, the finite verb is backshifted to indicate uncertainty, as in the examples above.
- Negating the antecedent allows for the negated and non-negated consequent, as in *Als het geen mooi weer is, gaan we (toch) zwemmen.* 'If the weather is not nice, we will go for a swim (anyway).' and *Als het geen mooi weer is, gaan we niet zwemmen.* 'If the weather is not nice, we will not go for a swim.'

Course-of-event conditional (cec; 3)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent. The clauses refer to two different events which are seen as being in a relation of mutual dependency, i.e. a 'whenever' relationship. There is a suggestion of a real occurrence of the two events such that whenever the first occurs, the second occurs, too, but the second is not seen as being triggered by the first. There is hypotheticality, but re-occurrence. There is no marking of uncertainty. The relation can be either co-occurrence or recurring inference.

Examples

(a) Als er brand is, kun je de brandweer bellen.
If there is a fire, you can call the fire department.

(b) Als hij er weer aankomt, is het etenstijd.
If he's coming again, it's time for dinner.

Tests

- The focus particle *altijd* 'always' or *elke keer* 'whenever' can be added, as in *Elke keer als hij aankomt, is het etenstijd.* 'Every time {if/when} he comes, it's time for dinner.'
- Negating the antecedent implicates the negation of the consequent, as in *Als het niet waait, wappert onze vlag niet.* 'If there is no wind, our flag doesn't wave.' and not *Als het niet waait, wappert onze vlag (toch).* 'If there is no wind, our flag waves (anyway).'

Pragmatic conditional (pra; 4)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent expresses a felicity condition for the speech act carried out in the consequent – either the antecedent expresses identification of, provides context for or comments on the consequent. Verb tense can, but is mostly not used to express unlikelihood of the situation in the antecedent. This type does not make a prediction involving an alternative scenario, i.e., nothing is said or implicated with respect to the non-occurrence of the situation in the antecedent. The consequent may be non-declarative, i.e. the consequent may be interrogative, imperative or exclamative. There can also be an inferential relation between antecedent and consequent. Mind that this may also be a politeness strategy.

Examples

(a) Als er een goed is in fietsen, dan is het Dumoulin.
If there is one who is good at cycling, it's Dumoulin.

(b) Ik wil je graag feliciteren, als dat het juiste woord is.
I'd like to congratulate you, if that is the right word.

(c) Als iemand me nodig heeft, ik ben boven.
If anyone needs me, I'm upstairs.

(d) Als hij trek heeft, wat mag hij dan eten?
If he is hungry, what can he eat?

(e) Als ik een man gras zag eten, zou ik zeggen dat hij trek had.
If I would see a man eating grass, I'd say he's hungry.

(f) Als ze gescheiden is, moet ze getrouwd zijn geweest.
If she's divorced, then she must have been married before.

Tests

- Verb tense cannot easily be used to indicate epistemic uncertainty, as in *Als er een goed zou zijn in fietsen, dan zou het Dumoulin zijn.* 'If there would be one who is good at cycling, it would be Dumoulin.'
- The conditional cannot be embedded into a matrix clause, as in 'I think that if you need me, I'm upstairs.'
- The negation of the antecedent does not trigger an alternative scenario, i.e. 'If you don't need me, I'm not upstairs'.'
- The sentence-medial or sentence-final pattern cannot (easily) be used in this type, as in 'My name, if you need help, is Anne'.

E.3.3 Problem cases

Please take note of the following known problem cases and code accordingly.

Incomplete utterance (*NA*)
If the utterance is incomplete, use this label. For instance, the antecedent or consequent does not have a finite verb or, in case of an imperative, does not have an overt subject.

Examples

(a) Als niet, dan toch.
If not, then still.

E.4 Dancygier and Sweetser's (2005) classification

E.4.1 Introduction

Dancygier and Sweetser (2005) present a classification of the type of connection between the antecedent and consequent in conditionals. Before annotating, please read (at least) paragraphs 2.1, 2.2, 5.1 to 5.3 and 5.6 to 5.7 from Dancygier and Sweetser's *Mental Spaces in Grammar: Conditional Constructions*. The two main types of conditionals in this classification are predictive

conditionals, functioning in the content domain, and non-predictive conditionals. The latter type is further divided into inferential conditionals, speech-act conditionals and metalinguistic conditionals, functioning in the epistemic, pragmatic and metatextual domain respectively. All types and subtypes are exemplified below.

(12) **Specific predictive conditional**

(a) If you mow the lawn, I'll give you ten dollars.

(b) If Hiro reaches out and takes the hypercard, then the data it represents will be transferred from this guy's system into Hiro's computer.

(13) **Generic predictive conditional**

(a) If I leave the house, he gets angry.

(b) If you heat water to 100 degrees, it boils.

(14) **(Non-predictive) inferential conditional conditional**

(a) If he typed her thesis, he loves her.

(b) If the lights are on, they must be home.

(15) **(Non-predictive) speech-act conditional**

(a) If you need any help, my name is Ann.

(b) If you are hungry, there are biscuits on the sideboard.

(16) **(Non-predictive) metalinguistic conditional**

(a) My ex-husband, if that's the right word, hates onion soup.

(b) That's what we're in business to do, get this cocksucker nailed, if you'll excuse my Greek.

(17) **(Non-predictive) meta-metaphoric conditional**

(a) If the beautiful Golden Gate is the thoroughbred of bridges, the Bay Bridge is the workhorse.

E.4.2 Instructions

Determine the most suiting type. Please note that grammatical features of types can be of help in determining the type, but they will most likely not exhaustively determine the type. As the type reflects the coherence relation between antecedent and consequence, determining its type is at least partly interpretative, i.e. world knowledge is needed. Annotate the according label. Below the coding instructions are presented, together with examples. The labels are presented between parentheses. The parentheses are not to be included in

your annotation. If you prefer shorter labels, you can use the numbers after the semicolon. They will be converted to their full counterparts after you are done annotating.

Specific-predictive (content) (spr; 1)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent. There is a reference to a specific situation, not to the general occurrence of a situation. The speaker can be neutral about the likelihood of the situations, or use verb tense and modal adverbs to express less likelihood of occurrence of the situation. In both cases, a prediction is made and the alternative scenario is triggered, i.e., if the situation in the antecedent does not occur, the situation in the consequent will most likely also not occur.

Examples

(a) Als je het gras maait, krijg je tien dollar.
If you mow the lawn, I'll give you ten dollars.

(b) Als Hiro zijn arm uitstrekt en de hypercard pakt, zullen de gegevens worden overgezet van het systeem van deze vent naar Hiro's systeem.
If Hiro reaches out and takes the hypercard, then the data it represents will be transferred from this guy's system into Hiro's computer.

Tests

- The conditional can be embedded into a matrix clause, as in 'I think that if he takes the card, the data will be transferred.'
- The finite verb is backshifted to indicate uncertainty, i.e., *Als Hiro de kaart zal pakken* 'If Hiro will take the card' becomes *Als Hiro de kaart pakt* 'If Hiro takes the card' and *Als ik won* 'If I won' becomes *Als ik had gewonnen* 'If I had won'.
- An prediction is made and an alternative scenario is triggered, i.e. 'If you mow the lawn' triggers both the situation of mowing the lawn and of not mowing the lawn and its consequences.
- The integrative and resumptive patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *Als je het gras maait, ik geef je tien dollar.* 'If you mow the lawn, I will give you ten Dollar.'

Generic-predictive (content) conditional (gpr; 2)

The (truth of) the situation in the consequent is dependent on the (truth of) the situation in the antecedent. There is a reference to a generic, reoccurring, sometimes (natural) law-like pattern, indicated by the simple present (or 'generic present') in both clauses. In a minority of cases, also the simple past can be used, as exemplified below. A prediction is made

and the alternative scenario is triggered, i.e., if the situation in the antecedent does not occur, the situation in the consequent will most likely also not occur.

Examples

(a) Hij wordt (altijd) boos als ik het huis verlaat.
He gets angry if I leave the house

(b) Als je water opwarmt tot 100 graden, kookt het.
If you heat water to 100 degrees, it boils.

(c) Als Mrs. Dugan de telefoon niet kon opnemen (wat vaak gebeurde), sprak Muriel met Claire.
If Mrs. Dugan couldn't come to the phone (which was often the case), Muriel talked to Claire instead.

Tests

- The conditional can be embedded into a matrix clause, as in 'I think that if you heat water to 100 degrees, it will boil.'
- When a temporal adverbial and/or a modal auxiliary are added, the conditional loses its generic meaning, as in 'He gets angry/will get angry if I leave the house today'.
- It is possible to add the focus particle *altijd* 'always' before the conditional conjunction without changing the meaning, as in *Altijd als ik het huis verlaat, wordt hij boos.* 'Always if .'
- The integrative and resumptive patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *? Als ik het huis verlaat, hij wordt boos.* 'If I leave the house, he gets angry.'

Inferential conditional (inf; 3)

The consequent presents a conclusion based on the argument presented in the antecedent. With respect to the predictive conditional, cause and effect appear reversed in this type. For example, the 'epistemic version' of 'If you heat water to 100 degrees, it boils' is 'If water boils, it is heated to 100 degrees'. The situations represented can be both specific and generic. The speaker can be neutral about the likelihood of the situations, or use verb tense and modal adverbs to express less likelihood of occurrence of the situation. In both cases, a prediction is made and the alternative scenario is triggered, i.e., if the situation in the antecedent does not occur, the situation in the consequent will most likely also not occur.

Examples

(a) Als het universum oneindig is, moet er elders leven zijn.
If the universe is infinite, there must be life somewhere else.

(b) Als het licht aan is, zijn ze thuis.
If the lights are on, the are home.

Tests

- The conditional can be embedded into a matrix clause, as in 'I think that if the lights are on, they are home.'
- Verb tense can be used to indicate epistemic uncertainty, as in 'If the lights would be on, they would be at home'.
- The integrative and resumptive word-order patterns are possible, but applying the non-integrative pattern leads to less acceptable results, as in *? Als het licht aan is, ze zijn thuis.* 'If the light is on, they are home.'
- The epistemic modal verb *moeten* 'must' can be added to the consequent, as in *Als het licht aan is, moeten ze thuis zijn.* 'If the light is on, they must be home.'

Speech-act conditional (spa; 4)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent expresses a felicity condition for the speech act carried out in the consequent. The situations represented can be both specific and generic. Verb tense can, but is most likely not used to express unlikelihood of the situation in the antecedent. This type does not make a prediction involving an alternative scenario, i.e., nothing is said or implicated with respect to the non-occurrence of the situation in the antecedent. The consequent may be non-declarative, i.e. the consequent may be interrogative, imperative or exclamative. Mind that this may also be a politeness strategy.

Examples

(a) Als ik het mag vragen, wat vind je van mijn trui?
If I may ask, what do you think of my sweater?

(b) Als hulp nodig hebt, mijn naam is Anne.
If you need help, my name is Ann.

Tests

- Verb tense cannot easily be used to indicate epistemic uncertainty, as in *Als ik het zou mogen vragen, wat zou je van mijn trui vinden?* 'If I would be permitted to ask, what would you say of my sweater?'
- The conditional cannot be embedded into a matrix clause, as in 'I think that if you need help, my name is Ann.'
- The negation of the antecedent does not trigger an alternative scenario, i.e. 'If you don't need help, my name is not Ann'.'
- The sentence-medial or sentence-final pattern cannot (easily) be used in this type, as in 'My name is Anne, if you need help'.

- The non-intergative word-order pattern is preferred for this type. Integrative and resumptive word-order patterns lead to less acceptable results, as in *? Als je hulp nodig hebt, is mijn naam Anne.* 'If you need help, my name is Ann.' and *? Als je hulp nodig hebt, dan is mijn naam Anne.* 'If you need help, then my name is Ann.'

Metalinguistic conditional (mel; 5)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent is used to comment on the linguistic form of (a part of) the consequent. The antecedent must follow the consequent or intercalate the consequent, but cannot precede it, as in *Als dat het juiste woord is, zijn stijl is bloemig.* 'If that's the right word, his style is florid.'

Examples

(a) Mijn ex-man, als dat het juiste woord is, haat soep.
My ex-husband, if that's the right word, hates soup.

(b) Oma voelt zich beroerd, als ik dat zo mag zeggen.
Grandma is feeling lousy, if you'll allow me to put it that way.

Tests

- Verb tense cannot be used to indicate epistemic uncertainty, as in 'My ex-husband, if that would be the right word, hates soup.'
- Only the sentence-medial and sentence-final clause order can be used. Sentence-initial order leads less acceptable result, as in *Als dat het juiste woord is, haat mijn ex-man soep.* 'If that's the right word, my ex-husband hates soup.'

Meta-metaphoric conditional (mem; 6)

The (truth of) the situation in the consequent is not dependent on the (truth of) the situation in the antecedent. Rather, the antecedent presents a metaphor that is continued in the consequent.

Examples

(a) Als de Golden Gate de volbloed van de bruggen is, dan is de Bay Bridge het werkpaard.
If the beautiful Golden Gate is the thoroughbred of bridges, the Bay Bridge is the workhorse.

(b) Als het leven kaarslicht is, dan zijn mensen motten die erin verbranden.
If life is a candle-flame, then people are moths burned on the flame.

Tests

- Verb tense can be used to indicate epistemic uncertainty, as in 'If Moriarty would be the Napoleon of crime, then Holmes would be a civilian Wellington.'
- The conditional can be embedded into a matrix clause, as in 'I think that if Moriarty is the Napoleon of crime, then Holmes is a civilian Wellington.'
- The sentence-medial or sentence-final pattern cannot (easily) be used in this type, as in 'People are moths burned on the flame, if life is a candle-flame.'.
- The intergative and resumptive word-order patterns are preferred for this type. The non-integrative pattern leads to less acceptable results, as in *? Als de Golden Gate de volbloed van de bruggen is, de Bay Bridge is het werkpaard.* 'If the beautiful Golden Gate is the thoroughbred of bridges, the Bay Bridge is the workhorse.'

E.4.3 Problem cases

Please take note of the following known problem cases and code accordingly.

Incomplete utterance (*NA*)
If the utterance is incomplete, use this label. For instance, the antecedent or consequent does not have a finite verb or, in case of an imperative, does not have an overt subject.

Examples

(a) Als niet, dan toch.
If not, then still.

Deciding between predictive and inferential reading
Although in theory and most examples, the difference between the content and epistemic conditionals below are clear, in practice, it is sometimes hard to distinguish between the two. Use the test above to choose the most suitable type.

Examples

(a) If he loves her, he'll type her thesis.
Content-level predictive conditional: The loving is a precondition for the typing. (Dancygier & Sweetser, 2005, p. 117)

(b) If he typed her thesis, he loves her.
My knowledge that the typing happened is a precondition for my conclusion about the loving. (Dancygier & Sweetser, 2005, p. 117)

E.5 Average agreement per corpus item

Table E.1:
Average agreement per item

Corpus item	Quirk et al.	Athanasiadou and Dirven	Dancygier and Sweetser	*mean*	*sd*
1	0.61	0.69	0.56	*0.62*	*0.06*
2	0.43	0.24	0.31	*0.33*	*0.08*
3	0.12	-0.06	0.35	*0.14*	*0.17*
4	0.79	-0.02	0.18	*0.32*	*0.35*
5	1.00	0.69	0.18	*0.62*	*0.34*
6	0.79	0.43	0.69	*0.64*	*0.15*
7	0.06	0.00	-0.07	*0.00*	*0.05*
8	0.43	0.04	0.31	*0.23*	*0.20*
9	0.11	0.15	-0.15	*0.03*	*0.13*
10	1.00	0.28	0.56	*0.61*	*0.30*
11	0.14	0.35	0.24	*0.24*	*0.09*
12	0.28	0.43	0.18	*0.30*	*0.10*
13	0.43	0.18	0.22	*0.28*	*0.11*
14	0.32	0.16	-0.05	*0.14*	*0.15*
15	0.45	0.02	0.09	*0.19*	*0.19*
16	0.43	0.02	-0.07	*0.13*	*0.22*
17	1.00	0.84	0.57	*0.81*	*0.18*
18	1.00	0.22	0.84	*0.69*	*0.34*
19	-0.18	0.44	-0.17	*0.03*	*0.29*
20	0.10	0.70	0.42	*0.41*	*0.25*
21	-0.01	0.05	0.01	*0.02*	*0.03*
22	-0.07	0.05	0.18	*0.05*	*0.10*
23	0.45	0.56	0.84	*0.62*	*0.16*

Note. Average agreement scores per item are reported in terms of O'Connell-Dobson-Schouten coefficients (see O'Connell & Dobson, 1984).

E.6 Materials

Below the materials used in the experiment are presented. Note that the conditionals in focus were presented in bold in the experiment as well.

Table E.2:
Experimental materials (corpus, control, and test items)

No.	Item	Corpus	Mode	Genre	Source
1	Een hernia kan zomaar terugkomen. Maar Schultz worstelt nog met haar forehand sinds ze haar grip op advies van Franker bij een zekere Belser heeft laten veranderen. **Telkens als ze geestelijk in nood zat, omdat die forehand niet werkte, moest ik haar komen redden.** En Brenda had als meisje van zeventien echt een van de beste forehands in het circuit.	Condiv	Written	News	nrc/ nieuws6
2	Sturing en Bos waren met hun mentaliteit en instelling gouden jongens voor de trainer. **Je kon ze 's nachts om hulp vragen als er stront aan de knikker was...**	Condiv	Written	News	tele/ nie_s12
3	Mensen die nu in de nachtopvang zitten, zien dat lotgenoten uit het eigen circuit toch onder de pannen komen. En dan krijg je de drang van: **als Jan het kan, moet ik dat ook kunnen.** Het is een kwestie van lange adem. Maar het lukt.	Condiv	Written	News	limburg/ div06

Note. Due to size, this page presents part 1 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
4	Maar dat kan niet want de ZCTU beschikt niet over de kwaliteiten van een president, aldus Moegabe, die er voor de goede orde aan toevoegde: "**De vakbonden vergissen zich als ze geloven dat ze sterker zijn dan mijn regering.** Ik waarschuw de ZCTU. Ik maak geen grapjes, ik ben bloedserieus."	Condiv	Written	News	tele/ nie_s5
5	Ja, maar toen was er in feite slechts één tegenstander met wie rekening moest worden gehouden, de Sovjet-Unie, die zowel door Amerikanen als Europeanen als een potentiële bedreiging van hun veiligheid werd beschouwd. **Nu is het beeld, als u wilt: het vijandbeeld, veel diffuser geworden en de Amerikaanse bereidheid in te grijpen navenant onzekerder.**	Condiv	Written	News	nrc/ varia5
6	Het kan zijn dat men daarom al snel tevreden is met de beperkte frequentieweergave van de cd. **Maar als de cd-kopers vaker muziek zouden horen in een goede concertzaal, zou men zich wel eens achter de oren krabben.**	Condiv	Written	News	nrc/ nieuws4
7	We moeten oppassen dat de toeloop op de opleidingen in Limburg niet te groot wordt. **Het is gevaarlijk als 'genoeg werk' het enige argument is om aan de Pabo te gaan studeren.**	Condiv	Written	News	limburg/ nieuws04

Note. Due to size, this page presents part 2 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
8	Met dit oordeel kan een einde komen aan de monopoliepositie van KPN's CD-foongids, het telefoonboek op cd-rom. **KPN mag ook geen beperkende voorwaarden meer opleggen als het de gegevens verkoopt.**	Condiv	Written	News	nrc/ nieuws6
9	Khoury is de enige die vaak naar Israël reist, ook al vinden collega's dat ongepast. Omdat hij niemand kan vinden die dat voor The Arab Daily wil doen, zoekt hij nu een Israëlische correspondent. '**OK, als Israël nog onze vijand is voor wat het de Palestijnen aandoet, lees erover, van binnenuit!** Het belangrijkste nieuws is dat van de vijand."	Condiv	Written	News	nrc/ nieuws9
10	Veel mensen doen dingen waarvan ze best weten dat het niet mag. **Maar als het niet wordt bestraft, gaat zoiets wennen en gaat het steeds een stukje verder.** Daar gaan we nu een stokje voor steken.	Condiv	Written	News	tele/ nie_sp1
11	Het was hem in zijn praktijk opgevallen dat hoeren vaak frigide zijn bij hun klant en dan de smoes verzinnen dat **het onprofessioneel zou zijn als ze wel tot een orgasme zouden geraken.** Het is niet eens een smoes.	Condiv	Written	News	tele/ verstr4.txt

Note. Due to size, this page presents part 3 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
12	'Ik geef toe, we hadden de kunstenaar kunnen informeren. Het beeld ligt nu tijdelijk boven in het magazijn en wordt opgepoetst. **In volle glorie kan meneer Huisman straks de heronthulling verrichten, als het aan mij ligt**', stelt Closset.	Condiv	Written	News	limburg/ div02.txt
13	Of ik tevreden ben over mijn eigen spel in de film? Ja, want de film is goed. **En als een film goed is, dan ben jij vanzelf ook goed.**	Condiv	Written	News	limburg/ div05.txt
14	Zij koesterden een droom: het uit de ruimte bijlichten van donkere plekken op aarde in Arctische gebieden met poolnacht. Maar ook streken waar een natuurramp is gebeurd, teneinde nachtelijk reddingswerk te vereenvoudigen. De verborgen doelstelling is natuurlijk een militaire: want **als je een poolstad kan verlichten, kan dat ook met een slagveld.**	Condiv	Written	News	tele/ nie_sp9.txt
15	Oh dus dat is gewoon voor de arbeids-uh-vriendelijkheid dat jullie daar uh zorgen dat er wat daglicht binnen kan komen. Ja, en **als je midden in de fabriek zit ja je weet niet of 't buiten onweert of dondert of regent of de zon schijnt.** dat weet je pas als je naar buiten stapt . nee .	CGN	Spoken	Face-to-face	fn000400

Note. Due to size, this page presents part 4 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
16	Mmm? **Als je 't niet zou weten dan hoor je niet dat de radio aan staat. Nee, maar was trouwens wel gaaf dat concert.**	Corpus	Spoken	Face-to- face	fn000411
17	Ja een maand geleden vroor 't nog 's nachts en uh hij start gelijk absoluut direct. Maar **als 'k dan 't gas losliet dan sloeg ie af.** dus die automatische choke vind ik half.	CGN	Spoken	Face-to- face	fn000948
18	Ja maar ik bedoel als dat uh... **als je huis een meter in 't water staan dan loopt 't uh toch wel onder.** Mmm ja. Nou kijk, d'r is een verschil tussen een meter en twintig centimeter he.	CGN	Spoken	Face-to- face	fn007816
19	Oh. Ja ik vind dat wel heel interessant en ik **als ik dat zo zie dan denk ik ook dat die man 't heeft gedaan.** ook als je dat hoort maar van de andere kant ja.	CGN	Spoken	Face-to- face	fn000458

Note. Due to size, this page presents part 5 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
20	Ja. **Als ik die morgen die floppy's niet kan vinden, waar liggen die dan?**	CGN	Spoken	Face-to- face	fn000646
21	Zat ik ook aan te denken. **Als wij nou zaterdag Wietske op gaan halen zaterdagmiddag dan kun jij zaterdagmorgen 't huisje in orde maken.** Uh .	CGN	Spoken	Face-to- face	fn007817
22	Ik moet 't allemaal regelen zegt ie. Dus als ik uh ja 't kwam er eigenlijk op neer van ik moet leren delegeren dus **als ik iemand de opdracht geef om mijn verslagen te tikken dan delegeer ik toch?**	CGN	Spoken	Face-to- face	fn000363
23	Maar heb je meer van die bossen hier in de buurt? **Als je naar Woerleit gaat, aan de overkant van de snelweg is ook een groot bos.**	CGN	Spoken	Face-to- face	fn000676
24	Er is een hoop geklieder en geklooi in de postmoderne kunstscene. **Als dat kunst is, dan ben ik ook een kunstenaar!**	Control	*NA*	*NA*	*NA*

Note. Due to size, this page presents part 6 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
25	Het is inmiddels juli, he... **als Oscar wordt ingeloot bij, ehm, bij geneeskunde, dan wacht hem een gouden toekomst.** Ik hoop echt voor ik hoop voor hem dat het gaat lukken.	Control	*NA*	*NA*	*NA*
26	Treurig als het is, hoort de dood ook het bij het leven. **Als het leven een kaars is, zijn mensen de motten die op de vlam afvliegen**.	Control	*NA*	*NA*	*NA*
27	Dus **als ik het goed begrijp, heeft de zanger van REM de band niet verlaten?** De platenbaas knikte en duwde me de nieuwe cd in handen. Of ik die even wilde recenseren.	Control	*NA*	*NA*	*NA*
28	Het is een genot om naar te luisteren. Dus: **als je zin hebt en nog geen avondplannen, het Nederlands Philharmonisch Orkest speelt vanavond Tsjaikovski.**	Control	*NA*	*NA*	*NA*

Note. Due to size, this page presents part 7 of 8 parts of the full table.

No.	Item	Corpus	Mode	Genre	Source
29	**Als er geen koffie zou zijn, zouden de gasten niet zo lang blijven hangen.**	Control	*NA*	*NA*	*NA*
30	Jij bent uh ook zo iemand die veel sport. En **steeds als je gesport hebt, is je energieniveau gezakt.** Eet jij dan wel genoeg?	Control	*NA*	*NA*	*NA*
31	Zoals de kalender aangeeft, is het weer oktober en **als de O weer in de maand is, wordt het snel koud.**	Control	*NA*	*NA*	*NA*
32	Met deze uitspraak kan een einde komen aan de marktpositie van Google. **De zoekmachinegigant mag ook geen privacyrestricties meer opleggen als het de zoekgegevens verkoopt.**	Re-test	*NA*	*NA*	*NA*
33	Wellicht vinden mensen de beperkte beeldkwaliteit van analoge tv daarom al snel voldoende. Maar **als de tv-kijkers vaker zouden kijken naar echte HD-zenders, zou men wel aan het twijfelen slaan.**	Re-test	*NA*	*NA*	*NA*

Note. Due to size, this page presents part 8 of 8 parts of the full table.

APPENDIX F

Supplementary Materials

The supplementary materials for this dissertation, such as data, programming code, and annotation guidelines, can be found online on Dataverse at https://doi.org/10.34894/3QTEKH (for details, see Reuneker, 2022a), and at https://www.reuneker.nl/dissertation (for details, see Reuneker, 2022b). In case a password is required, use '5%*uGnP$5DF3' (without quotation marks).

Bibliography

Aarts, F. G. A. M. (1988). 'Review of 'A Comprehensive Grammar of the English Language by Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik: The Great Tradition Continued'. In: *English Studies* 69.2, pp. 163–173. DOI: https://doi.org/10.1080/00138388808598565.

Aarts, F. and H. Wekker (1987). *A Contrastive Grammar of English and Dutch*. Leiden: Nijhoff.

Ackerman, M., A. Adolfsson and N. Brownstein (2016). *An Effective and Efficient Approach for Clusterability Evaluation*. URL: https://arxiv.org/abs/1602.06687.

Acock, A. C. and G. R. Stavig (1979). 'A Measure of Association for Nonparametric Statistics'. In: *Social Forces* 57.4, pp. 1381–1386. DOI: https://doi.org/10.2307/2577276.

Adams, E. W. (1966). 'Probability and the Logic of Conditionals'. In: *Studies in Logic and the Foundations of Mathematics*. Vol. 43. Elsevier, pp. 265–316.

— (1970). 'Subjunctive and Indicative Conditionals'. In: *Foundations of Language* 6.1, pp. 89–94.

— (1975). *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Dordrecht: Springer.

Adolfsson, A., M. Ackerman and N. C. Brownstein (2019). 'To Cluster, or not to Cluster: An Analysis of Clusterability Methods'. In: *Pattern Recognition* 88, pp. 13–26. DOI: https://doi.org/10.1016/j.patcog.2018.10.026.

Aggarwal, C. C. (2014). 'An Introduction to Cluster Analysis'. In: *Data Clustering: Algorithms and Applications*. Ed. by C. C. Aggarwal and C. K. Reddy. Boca Raton, Florida: Chapman & Hall/CRC, pp. 1–28.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Akaike, H. (1969). 'Fitting Autoregressive Models for Prediction'. In: *Annals of the institute of Statistical Mathematics* 21.1, pp. 243–247.

Akaike, H. (1974). 'A new Look at the Statistical Model Identification'. In: *IEEE Transactions on Automatic Control* 19.6, pp. 716–723.

Akatsuka, N. (1985). 'Conditionals and the Epistemic Scale'. In: *Language* 61.3, pp. 625–639. DOI: https://doi.org/10.2307/414388.

— (1986). 'Conditionals are Discourse-Bound'. In: *On Conditionals*. Ed. by E. C. Traugott, A. G. B. ter Meulen, J. S. Reilly and C. A. Ferguson. Cambridge University Press, pp. 333–352. DOI: https://doi.org/10.1017/cbo9780511753466.

— (1991). 'Dracula Conditionals and Discourse'. In: *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*. Ed. by C. Georgopoulos and R. Ishihara, pp. 25–37. DOI: https://doi.org/10.1007/978-94-011-3818-5_2.

— (1997a). 'Negative Conditionality, Subjectification, and Conditional Reasoning'. In: *On Conditionals Again*. Ed. by A. Athanasiadou and R. Dirven. Amsterdam: John Benjamins Publishing Company, pp. 323–354. DOI: https://doi.org/10.1075/cilt.143.19aka.

— (1997b). 'On the co-Construction of Counterfactual Reasoning'. In: *Journal of Pragmatics* 28.6, pp. 781–794.

— (1999). 'Towards a Theory of Desirability in Conditional Reasoning'. In: *Function and Structure: In Honor of Susumu Kuno*. Ed. by A. Kamio and K.-i. Takami. Amsterdam: John Benjamins Publishing Company, pp. 195–213. DOI: https://doi.org/10.1075/pbns.59.11aka.

Akmajian, A. (1984). 'Sentence Types and the Form-Function Fit'. In: *Natural Language & Linguistic Theory* 2.1, pp. 1–23.

Albert-Balázsi, J. (2018). 'Negatie in het Nederlands en in het Hongaars'. In: *Internationale Neerlandistiek* 56.1, pp. 43–56. DOI: https://doi.org/10.5117/in2018.1.albe.

Alelyani, S., J. Tang and H. Liu (2013). 'Feature Selection for Clustering: A Review'. In: *Data clustering: Algorithms and Applications* 29.1.

Alishahi, A. and S. Stevenson (2008). 'A Computational Model of Early Argument Structure Acquisition'. In: *Cognitive Science* 32.5, pp. 789–834. DOI: https://doi.org/10.1080/03640210801929287.

Almeida, J. A. S., L. M. S. Barbosa, A. A. C. C. Pais and S. J. Formosinho (2007). 'Improving Hierarchical Cluster Analysis: A new Method With Outlier Detection and Automatic Clustering'. In: *Chemometrics and Intelligent Laboratory Systems* 87.2, pp. 208–217. DOI: https://doi.org/10.1016/j.chemolab.2007.01.005.

Anderberg, M. R. (1973). *Cluster Analysis for Applications*. Burlington: Academic Press/Elsevier Science. DOI: https://doi.org/10.1016/c2013-0-06161-0.

Anderson, A. R. (1951). 'A Note on Subjunctive and Counterfactual Conditionals'. In: *Analysis* 12.2, pp. 35–38.

Anderson, A. R. and N. D. Belnap (1975). *Entailment: the Logic of Relevance and Necessity*. 2nd ed. Princeton: Princeton University Press.

Andor, M. (2015). 'Natural Language Conditionals'. PhD thesis. Santiago de Compostela: Universidad de Santiago de Compostela.

Andritsos, P. and P. Tsaparas (2010). 'Categorical Data Clustering'. In: *Encyclopedia of Machine Learning*. Ed. by C. Sammut and G. I. Webb. Boston, MA: Springer US, pp. 154–159. DOI: https://doi.org/10.1007/978-0-387-30164-8_99.

Anscombre, J.-C. and O. Ducrot (1989). 'Argumentativity and Informativity'. In: *From Metaphysics to Rhetoric*. Dordrecht: Kluwer Academic Publishers Group, pp. 71–87.

Ariel, M. (2008). *Pragmatics and Grammar*. Cambridge: Cambridge University Press.

— (2010). *Defining Pragmatics*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511777912.

Arnaboldi, V., A. Passarella, M. Conti and R. I. M. Dunbar (2015). *Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs*. Amsterdam; Boston: Elsevier.

Arregui, A. (2007). 'When Aspect Matters: the Case of Would-Conditionals'. In: *Natural Language Semantics* 15.3, pp. 221–264. DOI: https://doi.org/10.1007/s11050-007-9019-6.

— (2020). 'Counterfactuals: "If Kangaroos Had No Tails..."' In: *The Wiley Blackwell Companion to Semantics*. Ed. by D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann and T. Zimmermann. New York: John Wiley & Sons. DOI: https://doi.org/10.1002/9781118788516.

Artstein, R. and M. Poesio (2008). 'Inter-Coder Agreement for Computational Linguistics'. In: *Computational Linguistics* 34.4, pp. 555–596.

Athanasiadou, A. and R. Dirven (1995). 'Conditional Conjunctions and Conceptual Companions'. In: *Current Approaches to the Lexicon*. Ed. by R. Dirven and J. Vanparys. Bern: Peter Lang, pp. 225–249.

— (1996). 'Typology of *if*-Clauses'. In: *Cognitive Linguistics in the Redwoods*. Ed. by E. H. Casad. Berlin; New York: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110811421.609.

— (1997a). 'Conditionality, Hypotheticality, Counterfactuality'. In: *On Conditionals Again*. Ed. by A. Athanasiadou and R. Dirven. Amsterdam: John Benjamins Publishing Company, pp. 61–96. DOI: https://doi.org/10.1075/cilt.143.05ath.

— (1997b). *On Conditionals Again*. Ed. by A. Athanasiadou and R. Dirven. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/cilt.143.

— (2000). 'Pragmatic Conditionals'. In: *Constructions in Cognitive Linguistics: Selected Papers From the International Cognitive Linguistics Conference, Amsterdam, 1997*. Ed. by A. Foolen and F. van der Leek. Amsterdam; Philadelphia: John Benjamins Publishing Company, pp. 1–26.

Atkins, S., J. Clear and N. Ostler (1992). 'Corpus Design Criteria'. In: *Literary and Linguistic Computing* 7.1, pp. 1–16. DOI: https://doi.org/10.1093/llc/7.1.1.

Audring, J. (2016). 'Een Wachtende Partner is een Chagrijnige Partner'. In: *Aries Netwerk: Een Constructicon*. Ed. by A. Reuneker, R. J. U. Boogaart and S. E. Lensink. Leiden: Amsterdam University Press, pp. 16–19.

Auer, P. (2000). 'Pre-and Post-Positioning of *wenn*-clauses in Spoken and Written German'. In: *Cause, Condition, Concession, Contrast. Cognitive and Discourse Perspectives*. Ed. by E. Couper-Kuhlen and B. Kortmann. Berlin; Boston: De Gruyter Mouton, pp. 173–204.

Austin, J. L. (1962). *How to Do Things With Words*. Oxford: Clarendon Press.

— (1970). 'Ifs and Cans'. In: *Philosophical Papers*. Ed. by J. L. Austin. 2nd ed. Oxford: Clarendon Press, pp. 205–232.

Bach, K. (1994). 'Conversational Impliciture'. In: *Mind & Language* 9.2, pp. 124–162. DOI: https://doi.org/10.1111/j.1468-0017.1994.tb00220.x.

— (1998). 'Standardisation Revisited'. In: *Presupposition, Implicature, and Indirect Speech Acts*. Ed. by A. Kasher. Vol. 4. Pragmatics: Critical Concepts. London: Routledge.

— (1999). 'The Myth of Conventional Implicature'. In: *Linguistics and Philosophy* 22.4, pp. 327–366.

— (2001). 'You don't say?' In: *Synthese* 128.1, pp. 15–44.

Bach, K. and R. M. Harnish (1979). *Communication and Speech Acts*. Cambridge: MIT Press.

Bailey, K. D. (1994). *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, California: Sage Publications.

Baker, P., A. Hardie and T. McEnery (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Balk-Smit Duyzentkunst, F. (1963). *De grammatische functie. Methode van grammaticale analyse, aan het Nederlands gedemonstreerd*. Groningen: J. B. Wolters.

Banerjee, A. and S. Chaudhury (2010). 'Statistics Without Tears: Populations and Samples'. In: *Industrial Psychiatry Journal* 19.1, pp. 60–65. DOI: https://doi.org/10.4103/0972-6748.77642.

Banerjee, M., M. Capozzoli, L. McSweeney and D. Sinha (1999). 'Beyond Kappa: A Review of Interrater Agreement Measures'. In: *Canadian Journal of Statistics* 27.1, pp. 3–23. DOI: https://doi.org/10.2307/3315487.

Baptista, L. (2011). 'Say What? On Grice On What Is Said'. In: *European Journal of Philosophy* 22.1, pp. 1–19. DOI: https://doi.org/10.1111/j.1468-0378.2011.00467.x.

Barbiers, S. (2007). 'On the Periphery of Imperative and Declarative Clauses in Dutch and German'. In: *Imperative Clauses in Generative Grammar: Studies in Honour of Frits Beukema*. Ed. by W. van der Wurff. Amsterdam: John Benjamins Publishing Company, pp. 95–112. DOI: https://doi.org/10.1075/la.103.03bar.

— (2018). 'Adverbs in Strange Places: On the Syntax of Adverbs in Dutch'. In: *Nederlandse Taalkunde* 23.1, pp. 57–88. DOI: https://doi.org/10.5117/nedtaa2018.1.barb.

Barwise, J. (1986). 'Conditionals and Conditional Information'. In: *On Conditionals*. Ed. by E. C. Traugott, A. G. B. ter Meulen, J. S. Reilly and C. A. Ferguson. Cambridge University Press, pp. 21–54.

Beaver, D. (1997). 'Presupposition'. In: *Handbook of Logic and Language*. Ed. by J. van Benthem and A. G. B. ter Meulen. Amsterdam; New York: MIT Press, pp. 939–1008.

Beaver, D. I. and B. Geurts (2014). 'Presupposition'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Metaphysics Research Lab, Stanford University.

Beekhuizen, B. F. (2015). 'Constructions Emerging: A Usage-Based Model of the Acquisition of Grammar'. PhD thesis. Leiden: Leiden University.

Beekhuizen, B. F., J. Watson and S. Stevenson (2017). 'Semantic Typology and Parallel Corpora: Something About Indefinite Pronouns'. In: *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Ed. by G. Gunzelmann, A. Howes, T. Tenbrink and E. J. Davelaar. Austin, Texas: Cognitive Science Society, pp. 112–117.

Beller, S. (2002). 'Conditional Promises and Threats – Cognition and Emotion'. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 24.

Bennett, D. J. (2004). *Logic Made Easy: How to Know When Language Deceives you*. New York; London: W. W. Norton & Company.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/0199258872.001.0001.

Bennis, H. J. (2007). 'Featuring the Subject in Dutch Imperatives'. In: *Imperative Clauses in Generative Grammar: Studies in Honour of Frits Beukema*. Ed. by W. van der Wurff. Amsterdam: John Benjamins Publishing Company, pp. 113–134. DOI: https://doi.org/10.1075/la.103.04ben.

Berez, A. L. and S. T. Gries (2008). 'In Defense of Corpus-Based Methods: A Behavioral Profile Analysis of Polysemous *get* in English'. In: *Proceedings of the 24th Northwest Linguistics Conference*. Ed. by S. Moran, D. S. Tanner and M. Scanlon. Washington: University of Washington Working Papers in Linguistics, pp. 157–166.

Berry, M. W., A. Mohamed and B. W. Yap (2019). *Supervised and Unsupervised Learning for Data Science*. Springer.

Bezdek, J. C., R. Ehrlich and W. Full (1984). 'FCM: The Fuzzy c-Means Clustering Algorithm'. In: *Computers & Geosciences* 10.2-3, pp. 191–203.

Bhatt, R. and R. Pancheva (2005). 'Conditionals'. In: *The Wiley-Blackwell Companion to Syntax*. Ed. by M. Everaert and H. C. van Riemsdijk. Malden, MA: Wiley-Blackwell; John Wiley & Sons, pp. 638–687.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511621024.

— (1993). 'Representativeness in Corpus Design'. In: *Literary and Linguistic Computing* 8.4, pp. 243–257. DOI: https://doi.org/10.1093/llc/8.4.243.

— (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511519871.

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/scl.23.

Biber, D. and S. Conrad (2009). *Register, Genre, and Style*. Cambridge; New York: Cambridge University Press.

Biber, D., S. Conrad and R. Reppen (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511804489.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow, England: Longman.

Binnick, R. I. (2006). 'Aspect and Aspectuality'. In: *The Handbook of English Linguistics*. Ed. by B. Aarts and A. McMahon. Malden: Blackwell Publishing, pp. 244–268. DOI: https://doi.org/10.1002/9780470753002.ch11.

— (2020). 'Aspect and Aspectuality'. In: *The Handbook of English Linguistics*. Ed. by B. Aarts, A. McMahon and L. Hinrichs. 2nd ed. Chichester: John Wiley & Sons, pp. 183–205. DOI: https://doi.org/10.1002/9781119540618.ch11.

Birner, B. J. (2013). *Introduction to Pragmatics*. Chichester: Wiley-Blackwell; John Wiley & Sons.

Blakemore, D. (2002). *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge; New York: Cambridge University Press.

— (2004). 'Discourse Markers'. In: *The Handbook of Pragmatics*. Ed. by L. R. Horn and G. Ward. Oxford: Blackwell Publishing Ltd, pp. 221–240. DOI: https://doi.org/10.1002/9780470756959.ch10.

Blakemore, D. and R. Carston (1999). 'The Pragmatics of *and*-conjunctions: The Non-narrative Cases'. In: *UCL Working Papers in Linguistics*. Vol. 11, pp. 1–20.

Blattberg, R. C., B.-D. Kim and S. A. Neslin (2008). 'Cluster Analysis'. In: *Database Marketing*. Vol. 18. International series in Quantitative Marketing. New York: Springer, pp. 401–422. DOI: https://doi.org/10.1007/978-0-387-72579-6_16.

Bloomfield, L. (1984). *Language*. Chicago: University of Chicago Press.

Bogaards, M. (2019). 'A Mandarin map for Dutch Durativity'. In: *Nederlandse Taalkunde* 24.2, pp. 157–193. DOI: https://doi.org/10.5117/nedtaa2019.2.002.boga.

Bollens, S. M., J. K. Breckenridge, J. R. Cordell, C. A. Simenstad and O. Kalata (2014). 'Zooplankton of Tidal Marsh Channels in Relation to Environmental Variables in the Upper San Francisco Estuary'. In: *Aquatic Biology* 21.3, pp. 205–219. DOI: https://doi.org/10.3354/ab00589.

Bolognesi, M., R. Pilgram and R. van den Heerik (2017). 'Reliability in Content Analysis: The Case of Semantic Feature Norms Classification'. In: *Behavior Research Methods* 49.6, pp. 1984–2001. DOI: https://doi.org/10.3758/s13428-016-0838-6.

Bonnefon, J.-F. (2009). 'A Theory of Utility Conditionals: Paralogical Reasoning From Decision-Theoretic Leakage'. In: *Psychological Review* 116.4, pp. 888–907. DOI: https://doi.org/10.1037/a0017186.

Boogaart, R. J. U. (1999). 'Aspect and Temporal Ordering: A Contrastive Analysis of Dutch and English'. PhD thesis. Amsterdam: Vrije Universiteit Amsterdam.

— (2004). 'Aspect and Aktionsart'. In: *Morphologie: Ein Internationales Handbuch zur Flexion und Wortbildung*. Ed. by G. Booij, C. Lehmann and J. Mugdan. Berlin; New York: Walter de Gruyter.

— (2007a). 'Conditionele constructies met *moest(en)* en *mocht(en)* in Belgisch-Nederlands en Nederlands-Nederlands'. In: *Neerlandistiek.nl* 7.5.

— (2007b). 'The Past and Perfect of Epistemic Modals'. In: *Recent Advances in the Syntax and Semantics of Tense, Aspect and Modality*. Ed. by L. de Saussure, J. Moeschler and G. Puskás. Berlin; New York: Mouton de Gruyter, pp. 47–69. DOI: https://doi.org/10.1515/9783110198768.47.

— (2009). 'Semantics and Pragmatics in Construction Grammar: The Case of Modal Verbs'. In: *Constructional Approaches to Language*. Ed. by A. Bergs and G. Diewald. Amsterdam: John Benjamins Publishing Company, pp. 213–241. DOI: https://doi.org/10.1075/cal.9.09boo.

— (2013). 'De Modaliteit van temporaliteit'. In: *Nederlandse Taalkunde* 18.3, pp. 324–338. DOI: https://doi.org/10.5117/nedtaa2013.3.boog.

Boogaart, R. J. U. and E. L. J. Fortuin (2016). 'Modality and Mood in Cognitive Linguistics and Construction Grammars'. In: *The Oxford Handbook of Modality and Mood*. Ed. by J. Nuyts and J. van der Auwera. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199591435.013.23.

Boogaart, R. J. U., H. Jansen and M. van Leeuwen (2021). 'Introduction: From Argumentation to Language – And Vice Versa'. In: *The Language of Argumentation*. Ed. by R. J. U. Boogaart, H. Jansen and M. van Leeuwen. Cham: Springer International Publishing, pp. 1–14. DOI: https://doi.org/10.1007/978-3-030-52907-9_1.

Boogaart, R. J. U. and T. A. J. M. Janssen (2007). 'Tense and Aspect'. In: *The Oxford Handbook of Cognitive Linguistics*. Ed. by D. Geeraerts and H. Cuyckens. Oxford University Press, pp. 803–828. DOI: https://doi.org/10.1093/oxfordhb/9780199738632.013.0031.

— (2010). 'Mood in Dutch'. In: *Mood in the Languages of Europe. Germanic, Romance, Celtic Languages and Basque*. Ed. by B. Rothstein and R. Thieroff. Vol. 120. Studies in Language Companion Series. John Benjamins Publishing Company, pp. 117–132.

Boogaart, R. J. U. and A. Reuneker (2017). 'Intersubjectivity and Grammar'. In: *The Cambridge Handbook of Cognitive Linguistics*. Ed. by B. Dancygier. Cambridge: Cambridge University Press, pp. 188–206. DOI: https://doi.org/10.1017/9781316339732.013.

Boogaart, R. J. U. and R. Trnavac (2011). 'Imperfective Aspect and Epistemic Modality'. In: *Cognitive Approaches to Tense, Aspect, and Epistemic Modality*. Ed. by A. Patard and F. Brisard. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 217–248.

Boogaart, R. J. U. and K. Verheij (2013). 'Als dát Geen Insubordinatie is! De Pragmatiek van Zelfstandige Conditionele Zinnen'. In: *Honderd Jaar Taalwetenschap. Artikelen Aangeboden aan Saskia Daalder bij Haar Afscheid van de Vrije Universiteit*. Ed. by T. A. J. M. Janssen and J. Noordegraaf. Amsterdam/Münster: Stichting Neerlandistiek VU, Nodus Publikationen, pp. 13–28.

Booij, G. (2002). 'Constructional Idioms, Morphology, and the Dutch Lexicon'. In: *Journal of Germanic Linguistics* 14.04. DOI: https://doi.org/10.1017/s1470542702000168.

— (2018). 'The Construction of Words: Introduction and Overview'. In: *The Construction of Words*. Ed. by G. Booij. Cham: Springer International Publishing, pp. 3–16. DOI: https://doi.org/10.1007/978-3-319-74394-3_1.

Boole, G. (1854). *An Investigation of The Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. London: Macmillan.

Borg, I. and P. J. F. Groenen (2005). *Modern Multidimensional Scaling: Theory and Applications*. New York, London: Springer.

Boriah, S., V. Chandola and V. Kumar (2008). 'Similarity Measures for Categorical Data: A Comparative Evaluation'. In: *Proceedings of the 2008 Siam International Conference on Data Mining*. Society for Industrial and Applied Mathematics, pp. 243–254. DOI: https://doi.org/10.1137/1.9781611972788.22.

Bouhamed, H., T. Lecroq and A. Rebaï (2012). 'New Filter Method for Categorical Variables' Selection'. In: *International Journal of Computer Science Issues* 9.3, pp. 10–19.

Bozdogan, H. (1987). 'Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions'. In: *Psychometrika* 52.3, pp. 345–370. DOI: https://doi.org/10.1007/bf02294361.

Bracht, G. H. and G. V. Glass (1968). 'The External Validity of Experiments'. In: *American Educational Research Journal* 5.4, pp. 437–474. DOI: https://doi.org/10.3102/00028312005004437.

Braine, M. D. (1978). 'On the Relation Between the Natural Logic of Reasoning and Standard Logic.' In: *Psychological Review* 85.1, pp. 1–21. DOI: https://doi.org/10.1037/0033-295x.85.1.1.

Breitbarth, A., S. Delva and T. Leuschner (2016). 'A (Very) Imperfect Sandwich: English *should*, German *sollte*, Dutch *mocht/moest* as Grammaticalizing Markers of Conditionality'. In: *Journal of Germanic Linguistics* 28.4, pp. 282–316. DOI: https://doi.org/10.1017/s1470542716000143.

Broekhuis, H. and N. Corver (2016). *Syntax of Dutch: Verbs and Verb Phrases*. Vol. 3. Comprehensive Grammar Resources. Amsterdam: Amsterdam University Press.

Broekhuis, H., N. Corver and R. Vos (2015a). *Syntax of Dutch: Verbs and Verb Phrases*. Vol. 1. Comprehensive Grammar Resources. Amsterdam: Amsterdam University Press.

— (2015b). *Syntax of Dutch: Verbs and Verb Phrases*. Vol. 2. Comprehensive Grammar Resources. Amsterdam: Amsterdam University Press.

Broekhuis, H., H. den Besten, K. Hoekstra and J. Rutten (1995). 'Infinitival Complementation: On Remnant Extraposition'. In: *Linguistic Review* 12, pp. 93–122.

Broekhuis, H. and E. Keizer (2012). *Syntax of Dutch: Nouns and Noun Phrases*. Ed. by M. d. Dikken and H. J. Bennis. Vol. 1. Comprehensive Grammar Resources. Amsterdam: Amsterdam University Press.

Brown, P. and S. C. Levinson (1987). *Politeness: Some Universals in Language Usage*. Studies in Interactional Sociolinguistics. Cambridge; New York: Cambridge University Press.

Buchstaller, I. and G. Khattab (2014). 'Population Samples'. In: *Research Methods in Linguistics*. Ed. by R. J. Podesva and D. Sharma. Cambridge University Press, pp. 74–95. DOI: https://doi.org/10.1017/cbo9781139013734.006.

Bücker, J. (2016). 'Focus, Scales, and Concessive Conditionals in German'. In: *Language Sciences* 58, pp. 163–177. DOI: https://doi.org/10.1016/j.langsci.2016.03.005.

Bueno, O. and S. A. Shalkowski (2021). 'Introduction. Modality Matters: Philosophical Significance'. In: *The Routledge Handbook of Modality*. Ed. by O. Bueno and S. A. Shalkowski. New York: Routledge, pp. 1–7.

Bunt, H. C. and R. A. Muskens (1999). 'Computational Semantics'. In: *Computing Meaning*. Ed. by H. C. Bunt and R. A. Muskens. Vol. 1. Studies in Linguistics and Philosophy 73. Kluwer Academic Publishers, pp. 1–32.

Burke, M. B. (1994). 'Denying the Antecedent: A Common Fallacy?' In: *Informal Logic* 16.1. DOI: https://doi.org/10.22329/il.v16i1.2432.

Burton, G. (2021). 'Rehabilitating the ELT Conditional System'. In: *ELT Journal*. DOI: 10.1093/elt/ccab064.

Bybee, J. L. (2013). 'Usage-Based Theory and Exemplar Representations of Constructions'. In: *The Oxford Handbook of Construction Grammar*. Ed. by T. Hoffmann and G. Trousdale. Oxford: Oxford University Press, pp. 49–69. DOI: https://doi.org/10.1093/oxfordhb/9780195396683.013.0004.

Bybee, J. L., R. D. Perkins and W. Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Cabin, R. and R. Mitchell (2000). 'To Bonferroni or Not to Bonferroni: When and How Are the Questions'. In: *Bulletin of the Ecological Society of America* 81, pp. 246–248. DOI: https://doi.org/10.2307/20168454.

Cantwell, J. (2018). 'Conditionals'. In: *Introduction to Formal Philosophy*. Ed. by S. O. Hansson and V. F. Hendricks. Cham: Springer International Publishing, pp. 131–146. DOI: https://doi.org/10.1007/978-3-319-77434-3_6.

Capecchi, V. and F. Möller (1968). 'Some Applications of the Entropy to the Problems of Classification'. In: *Quality & Quantity* 2.1-2, pp. 63–84.

Cappelle, B. (2017). 'What's Pragmatics Doing Outside Constructions?' In: *Semantics and Pragmatics: Drawing a Line*. Ed. by I. Depraetere and R. Salkie. Cham: Springer International Publishing, pp. 115–151. DOI: https://doi.org/10.1007/978-3-319-32247-6_8.

Carr, E. H. (1986). *What is History?* 2nd ed. London: Macmillan.

Carston, R. (1998). 'The Semantics/Pragmatics Distinction: A View From Relevance Theory'. In: *UCL Working Papers in Linguistics*. Vol. 10. London: University College London, pp. 1–30.

— (2002). *Thoughts and Utterances*. Malden, MA: Blackwell Publishing. DOI: https://doi.org/10.1002/9780470754603.

— (2006). 'Relevance Theory and the Saying/Implicating Distinction'. In: *The Handbook of Pragmatics*. Ed. by L. R. Horn and G. Ward. Oxford: Blackwell, pp. 633–656. DOI: https://doi.org/10.1002/9780470756959.ch28.

— (2011). 'Truth-conditional Semantics'. In: *Philosophical Perspectives for Pragmatics*. Ed. by M. Sbisà, J.-O. Östman and J. Verschueren. Amsterdam: John Benjamins Publishing Company, pp. 280–288. DOI: https://doi.org/10.1075/hoph.10.24car.

Carter, S. (2021). 'A Suppositional Theory of Conditionals'. In: *Mind*. DOI: https://doi.org/10.1093/mind/fzaa071.

Carter-Thomas, S. (2007). 'The "Iffiness" of Medical Research Articles: A Comparison of English 'if' and French 'si''. In: *Language and Discipline Perspectives on Academic Discourse*. Ed. by K. Flottum. Cambridge Scholars Press, pp. 150–175.

Carter-Thomas, S. and E. Rowley-Jolivet (2008). '*If*-Conditionals in Medical Discourse: From Theory to Disciplinary Practice'. In: *Journal of English for Academic Purposes* 7.3, pp. 191–205. DOI: https://doi.org/10.1016/j.jeap.2008.03.004.

Castroviejo Miró, E. (2008). 'Deconstructing Exclamations'. In: *Catalan Journal of Linguistics* 7.1, pp. 41–90. DOI: https://doi.org/10.5565/rev/catjl.132.

Celce-Murcia, M. and D. Larsen-Freeman (1999). *The Grammar Book: An ESL/EFL Teacher's Course*. 2nd ed. Boston: Heinle & Heinle.

Chafe, W. (1984). 'How People Use Adverbial Clauses'. In: *Annual Meeting of the Berkeley Linguistics Society* 10, pp. 437–449. DOI: https://doi.org/10.3765/bls.v10i0.1936.

Chafe, W. and J. Nichols (1986). *Evidentiality: The Linguistic Coding of Epistemology*. Norwood: Ablex Publishing Corporation.

Chafe, W. and D. Tannen (1987). 'The Relation Between Written and Spoken Language'. In: *Annual Review of Anthropology* 16.1, pp. 383–407. DOI: https://doi.org/10.1146/annurev.an.16.100187.002123.

Chakraborty, C. (1997). 'Grice's Theory of Ordinary Conditionals'. In: *Indian Philosophical Quarterly* 24.4, pp. 543–560.

Chamalis, T. and A. Likas (2018). 'The Projected Dip-Means Clustering Algorithm'. In: *Proceedings of the 10th Hellenic Conference on Artificial Intelligence - Setn '18*. Patras, Greece: ACM Press, pp. 1–7. DOI: https://doi.org/10.1145/3200947.3201008.

Chambers, C. D. (1895). 'The Classification of Conditional Sentences'. In: *The Classical Review* 9.6, pp. 293–294.

Chapelle, O., B. Schölkopf and A. Zien (2006). 'Introduction to Semi-Supervised Learning'. In: *Semi-Supervised Learning*. Ed. by O. Chapelle, B. Schölkopf and A. Zien. Cambridge, Massachusetts: MIT Press, pp. 1–12.

Chapman, S. (2011). *Pragmatics*. Modern Linguistics Series. Houndmills, Hampshire; New York: Palgrave Macmillan.

Charivarius (1943). *Is dat Goed Nederlands?* 5th ed. Amsterdam: Uitgeverij Ploegsma.

Chierchia, G. and S. McConnell-Ginet (1990). *Meaning and Grammar: An Introduction to Semantics*. Massachusetts: MIT Press.

Christie, A. (1926). *The Murder of Roger Ackroyd*. Glasgow: William Collins, Sons.

Ciardelli, I. (2021). 'The Restrictor View, Without Covert Modals'. In: *Linguistics and Philosophy*. DOI: https://doi.org/10.1007/s10988-021-09332-z.

Cibulková, J., Z. Šulc, S. Sirota and H. Řezanková (2019). 'The Effect of Binary Data Transformation in Categorical Data Clustering'. In: *Statistics in Transition New Series* 20.2, pp. 33–47. DOI: https://doi.org/10.21307/stattrans-2019-013.

Cicchetti, D. V. and A. R. Feinstein (1990). 'High Agreement but low Kappa: II. Resolving the Paradoxes'. In: *Journal of Clinical Epidemiology* 43.6, pp. 551–558. DOI: https://doi.org/10.1016/0895-4356(90)90159-m.

Cichosz, P. (2015). '(Dis)Similarity Measures'. In: *Data Mining Algorithms*. Chichester, UK: John Wiley & Sons, pp. 311–327. DOI: https://doi.org/10.1002/9781118950951.ch11.

Cirillo, F. (2009). *The Pomodoro Technique*. San Francisco, California: Lulu.

Clark, B. (1993). 'Relevance and "Pseudo-Imperatives"'. In: *Linguistics and Philosophy* 16.1, pp. 79–121.

Clark, H. H. (1996). *Using Language*. Cambridge, England; New York: Cambridge University Press.

Clarke, R. K. (1993). 'Non-Parametric Multivariate Analyses of Changes in Community Structure'. In: *Australian Journal of Ecology* 18.1, pp. 117–143.

Cleveland, W. S. (1993). 'A Model for Studying Display Methods of Statistical Graphics'. In: *Journal of Computational and Graphical Statistics* 2.4, pp. 323–243. DOI: https://doi.org/10.2307/1390686.

Cohen, J. (1960). 'A Coefficient of Agreement for Nominal Scales'. In: *Educational and Psychological Measurement* 20.1, pp. 37–46.

— (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: L. Erlbaum Associates.

— (1992). 'A Power Primer'. In: *Psychological Bulletin* 112.1, pp. 155–159.

Cohen, J. L. (1971). 'The Logical Particles of Natural Language'. In: *Pragmatics of Natural Language*. Ed. by Y. Bar-Hillel. Dordrecht: Reidel, pp. 50–68.

Colleman, T. (2009). 'Verb Disposition in Argument Structure Alternations: A Corpus Study of the Dative Alternation in Dutch'. In: *Language Sciences* 31.5, pp. 593–611. DOI: https://doi.org/10.1016/j.langsci.2008.01.001.

Comrie, B. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.

— (1986). 'Conditionals: A Typology'. In: *On Conditionals*. Ed. by E. C. Traugott, A. T. Meulen, J. S. Reilly and C. A. Ferguson. Cambridge University Press, pp. 77–100. DOI: https://doi.org/10.1017/cbo9780511753466.

— (1989). 'On Identifying Future Tenses'. In: *Tempus - Aspekt - Modus*. Ed. by W. Abraham and T. A. J. M. Janssen. Berlin; New York: De Gruyter.

— (1996). 'Markedness'. In: *Handbook of Pragmatics*. Ed. by J. Verschueren, J.-O. Östman, J. Blommaert and C. Bulcaen. Amsterdam: John Benjamins Publishing Company, pp. 1–13. DOI: https://doi.org/10.1075/hop.2.mar2.

Cook, R. T. (2009). *A Dictionary of Philosophical Logic*. Edinburgh: Edinburgh University Press.

Cooper, W. S. (1978). ''If-Then': A Case Study in Logico-Linguistic Analysis'. In: *Foundations of Logico-Linguistics*. Dordrecht: Springer, pp. 158–211. DOI: https://doi.org/10.1007/978-94-009-9820-9_8.

Copi, I. (1973). *Symbolic Logic*. 4th ed. London: Macmillan Company.

Cornillie, B. (2009). 'Evidentiality and Epistemic Modality: On the Close Relationship Between two Different Categories'. In: *Functions of Language* 16.1, pp. 44–62. DOI: https://doi.org/10.1075/fol.16.1.04cor.

Coussé, E. and A. Oosterhof (2012). 'Het Imperativische Participium in het Nederlands. Vorm, Betekenis en Gebruik'. In: *Nederlandse taalkunde* 17.1, pp. 26–55. DOI: https://doi.org/10.5117/nedtaa2012.1.het_515.

Cox, T. F. and M. A. A. Cox (2001). *Multidimensional Scaling*. 2nd ed. Monographs on Statistics and Applied Probability. Boca Raton: Chapman & Hall/CRC.

Cramer, M., S. Hölldobler and M. Ragni (2021). 'When are Humans Reasoning With Modus Tollens?' In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 43.43.

Crocco, G., L. Fariñas del Cerro and A. Herzig (1995). 'Introduction'. In: *Conditionals: From Philosophy to Computer Science*. Ed. by G. Crocco, L. Fariñas del Cerro and A. Herzig. Studies in Logic and Computation. Oxford; New York: Clarendon Press; Oxford University Press, pp. 1–12.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford; New York: Oxford University Press.

Croft, W. and D. A. Cruse (2004). *Cognitive Linguistics*. Cambridge Textbooks in Linguistics. Cambridge, United Kingdom; New York: Cambridge University Press.

Crouch, R. (1994). 'Tense in Simple Conditionals'. In: *Proceedings of the Ninth Amsterdam Colloquium*. Ed. by P. Dekker and M. Stokhof. Amsterdam: ILLC, University of Amsterdam.

Crupi, V. and A. Iacona (2021). 'Probability, Evidential Support, and the Logic of Conditionals'. In: *Argumenta* 6.2, pp. 211–222. DOI: 10.14275/2465-2334/202112.CRU.

Cruse, D. A. (2000). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Cuypere, L. de (2008). *Limiting the Iconic: From the Metatheoretical Foundations to the Creative Possibilities of Iconicity in Language*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

D'Hertefelt, S. (2015). 'Insubordination in Germanic: A Typology of Complement and Conditional Constructions'. PhD thesis. Leuven: KU Leuven.

Daalder, S. (1994). '*Tenzij* in irreëel conditioneel verband'. In: *Nauwe Betrekkingen: Voor Theo Janssen bij Zijn Vijftigste Verjaardag*. Ed. by R. J. U. Boogaart and J. Noordegraaf. Amsterdam; Münster: Stichting Neerlandistiek VU; Nodus Publikationen, pp. 31–38.

— (2006). 'Regelmaat en interpretatie bij *mits*-constructies in het moderne Nederlands'. In: *Tussen Semantiek en Pragmatiek: Aangeboden aan Theo Jansen bij Zijn Afscheid als Hoogleraar Nederlandse Taalkunde en Taalbeheersing aan de Vrije Universiteit Amsterdam*. Ed. by R. J. U. Boogaart and T. A. J. M. Janssen. Voortgang: Jaarboek voor de neerlandistiek. Amsterdam: Stichting Neerlandistiek, pp. 51–60.

— (2009). 'Conditional Constructions: The Special Case of Modern Dutch *mits*'. In: *Journal of Germanic Linguistics* 21.2, pp. 231–248. DOI: https://doi.org/10.1017/s1470542709000270.

Dąbrowska, E. (2014). 'Recycling Utterances: A Speaker's Guide to Sentence Processing'. In: *Cognitive Linguistics* 25.4, pp. 617–653. DOI: https://doi.org/10.1515/cog-2014-0057.

Dancygier, B. (1985). '*If*, *unless*, and their Polish Equivalents'. In: *Papers and Studies in Contrastive Linguistics* 20, pp. 64–72.

— (1992). 'Two Metatextual Operators: Negation and Conditionality in English and Polish'. In: *Annual Meeting of the Berkeley Linguistics Society* 18.1, pp. 61–75. DOI: https://doi.org/10.3765/bls.v18i1.1601.

— (1993). 'Interpreting Conditionals: Time, Knowledge, and Causation'. In: *Journal of Pragmatics* 19.5, pp. 403–434. DOI: https://doi.org/10.1016/0378-2166(93)90003-8.

— (1998). *Conditionals and Prediction: Time, Knowledge, and Causation in Conditional Constructions*. Cambridge Studies in Linguistics. Cambridge; New York: Cambridge University Press.

— (2002). 'Mental Space Embeddings, Counterfactuality, and the use of *unless*'. In: *English Language and Linguistics* 6.2, pp. 347–377. DOI: https://doi.org/10.1017/s1360674302000278.

— (2003). 'Classifying Conditionals: Form and Function. Review of 'Conditionals' by Renaat Declerck and Susan Reed''. In: *English Language and Linguistics* 7.2, pp. 309–323. DOI: https://doi.org/10.1017/s1360674303001126.

Dancygier, B. and E. Mioduszewska (1984). 'Semanto-Pragmatic Classification of Conditionals'. In: *Studia Anglica Posnaniensia* 17, pp. 121–134.

Dancygier, B. and E. E. Sweetser (1997). '*Then* in Conditional Constructions'. In: *Cognitive Linguistics* 8.2, pp. 109–136. DOI: https://doi.org/10.1515/cogl.1997.8.2.109.

— (2000). 'Constructions with *if*, *since*, and *because*: Causality, Epistemic Stance, and Clause Order'. In: *Cause - Condition - Concession - Contrast*. Ed. by E. Couper-Kuhlen and B. Kortmann. 2000th ed. Berlin; New York: Mouton de Gruyter, pp. 111–142. DOI: https://doi.org/10.1515/9783110219043.2.111.

— (2005). *Mental Spaces in Grammar: Conditional Constructions*. Cambridge Studies in Linguistics. New York: Cambridge University Press.

Davies, E. C. (1979). *On the Semantics of Syntax*. London: Croom Helm.

De, M. (2021). 'Possible Worlds'. In: *The Routledge Handbook of Modality*. New York: Routledge, pp. 10–20.

de Finetti, B. (1936). 'La Logique de la Probabilité'. In: *Actes du CongrèS International de Philosophie Scientifique* 4, pp. 1–9.

— (1995). 'The Logic of Probability'. In: *Philosophical Studies* 77.1, pp. 181–190. DOI: https://doi.org/10.1007/bf00996317.

de Haan, F. (2006). 'Typological Approaches to Modality'. In: *The Expression of Modality*. Ed. by W. Frawley. Berlin; Boston: De Gruyter Mouton, pp. 27–70.

— (2009). 'On the Status of 'Epistemic' *must*'. In: *Studies on English Modality: In Honour of Frank Palmer*. Ed. by A. Tsangalidis and R. Facchinetti. Bern: Peter Lang, pp. 261–284.

de Haan, S. (1986). 'Gebruiksmogelijkheden van de Nederlandse imperatief'. In: *TABU* 16, pp. 250–260.

— (1991). 'Meaning and Use of the Dutch Perfect'. In: *The Function of Tense in Texts*. Ed. by J. Gvozdanovicć and T. A. J. M. Janssen. Amsterdam: Elsevier, pp. 143–156.

de Hoop, H. and S. Tarenskeen (2015). 'It's all about *you* in Dutch'. In: *Journal of Pragmatics* 88, pp. 163–175. DOI: https://doi.org/10.1016/j.pragma.2014.07.001.

de Jong, W. (2021). 'Superkrachten voor je hoofd'. In: *NRC (03-27-2021)*, pp. 10–11.

de Raadt, A., M. J. Warrens, R. J. Bosker and H. A. L. Kiers (2019). 'Kappa Coefficients for Missing Data'. In: *Educational and Psychological Measurement* 79.3, pp. 558–576. DOI: https://doi.org/10.1177/0013164418823249.

de Rooy, J. (1965). 'Als-of-dat: een semantisch-onomasiologische studie over enkele subordinerende conjuncties in het ABN, de Nederlandse dialecten en het Fries, vergelijkend-synchronisch beschouwd'. PhD thesis. Assen: Katholieke Universiteit Nijmegen.

de Schepper, K., G. van Bergen, S. Lestrade and W. Stoop (2014). 'Prag-Raising Versus Neg-Raising'. In: *Nederlandse Taalkunde* 19.1, pp. 105–117. DOI: https://doi.org/10.5117/nedtaa2014.1.sche.

de Smedt, T. and W. Daelemans (2012). 'Pattern for Python'. In: *Journal of Machine Learning Research* 13.66, pp. 2063–2067.

de Sutter, G. (2005). 'Rood, Groen, Corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen'. PhD thesis. Leuven: KU Leuven.

de Vuyst, J. (1983). 'Situation-Descriptions: Temporal and Aspectual Semantics'. In: *Studies in Model Theoretic Semantics*. Ed. by A. G. B. ter Meulen. Dordrecht: Foris, pp. 161–176.

Dean, R. B. and W. J. Dixon (1951). 'Simplified Statistics for Small Numbers of Observations'. In: *Analytical chemistry* 23.4, pp. 636–638.

Declerck, R. (1984). ''Pure Future' *will* in *if*-clauses'. In: *Lingua* 63.3-4, pp. 279–312.

— (1988). 'Restrictive *when*-clauses'. In: *Linguistics and Philosophy* 11.2, pp. 131–168. DOI: https://doi.org/10.1007/bf00632459.

— (1994). 'The *only*/*already* Puzzle: A Question of Perspective'. In: *Cognitive Linguistics* 5.4, pp. 307–350. DOI: https://doi.org/10.1515/cogl.1994.5.4.307.

— (2011). 'The Definition of Modality'. In: *Human Cognitive Processing*. Ed. by A. Patard and F. Brisard. Vol. 29. Amsterdam: John Benjamins Publishing Company, pp. 21–44. DOI: https://doi.org/10.1075/hcp.29.05dec.

Declerck, R. and S. Reed (2000). 'The Semantics and Pragmatics of *unless*'. In: *English Language and Linguistics* 4.2, pp. 205–241. DOI: https://doi.org/10.1017/s136067430000023x.

— (2001). *Conditionals: A Comprehensive Empirical Analysis*. Topics in English Linguistics. Berlin; New York: Mouton de Gruyter.

Delucchi, K. L. (1976). 'On the use and misuse of Chi-Square'. In: *A Handbook for Data Analysis in the Behavioral Sciences: Statistical issues*. Ed. by G. Keren and C. Lewis, pp. 295–319.

den Hertog, C. H. (1903). *Nederlandsche spraakkunst: Handleiding ten dienste van aanstaande (taal)onderwijzers*. Amsterdam: Versluys.

Depraetere, I. (2019). 'Meaning in Context and Contextual Meaning: A Perspective on the Semantics-Pragmatics Interface Applied to Modal Verbs'. In: *Anglophonia* 28. DOI: https://doi.org/10.4000/anglophonia.2453.

DeRose, K. and R. E. Grandy (1999). 'Conditional Assertions and "Biscuit" Conditionals'. In: *Noûs* 33.3, pp. 405–420. DOI: https://doi.org/10.1111/0029-4624.00161.

Desarbo, W. S. and D. K. Hildebrand (1980). 'A Marketer's Guide to Log-Linear Models for Qualitative Data Analysis'. In: *Journal of Marketing* 44.3, pp. 40–51. DOI: https://doi.org/10.2307/1251110.

Dexter, E., G. Rollwagen-Bollens and S. M. Bollens (2018). 'The Trouble With Stress: A Flexible Method for the Evaluation of Nonmetric Multidimensional Scaling'. In: *Limnology and Oceanography: Methods* 16.7, pp. 434–443.

Deygers, K., V. van Den Heede, S. Grondelaers, D. Speelman and H. van Aken (2000). 'Het Condiv-Corpus Geschreven Nederlands'. In: *Nederlandse Taalkunde* 5.4, pp. 356–363.

Deza, M. and E. Deza (2013). *Encyclopedia of Distances*. 2nd ed. Berlin: Springer.

Diepeveen, J., R. J. U. Boogaart, J. Brantjes, P. Byloo, T. A. J. M. Janssen and J. Nuyts (2006). *Modale Uitdrukkingen in Belgisch-Nederlands en Nederlands-Nederlands: Corpusonderzoek en Enquête*. Münster: Nodus Publikationen; Stichting Neerlandistiek VU Amsterdam.

Diessel, H. (2005). 'Competing Motivations for the Ordering of Main and Adverbial Clauses'. In: *Linguistics* 43.3. DOI: https://doi.org/10.1515/ling.2005.43.3.449.

— (2013). 'Adverbial Subordination'. In: *Bloomsbury Companion to Syntax*. Ed. by S. Luraghi and C. Parodi. London: Continuum, pp. 341–354.

— (2015). 'Usage-Based Construction Grammar'. In: *Handbook of Cognitive Linguistics*. Ed. by E. Dąbrowska and D. Divjak. Handbooks of Linguistics and Communication Science. Berlin; Boston: De Gruyter Mouton, pp. 296–322. DOI: https://doi.org/10.1515/9783110292022-015.

Dietz, E.-A., S. Hölldobler and L. M. Pereira (2015). 'On Conditionals.' In: *Gcai 2015*. Vol. 36. Epic series in Computer Science, pp. 79–92.

Dirikx, A. and D. Gelders (2010). 'To Frame is to Explain: A Deductive Frame-Analysis of Dutch and French Climate Change Coverage During the Annual UN Conferences of the Parties'. In: *Public Understanding of Science* 19.6, pp. 732–742. DOI: https://doi.org/10.1177/0963662509352044.

Divjak, D. (2010). *Structuring the Lexicon: A Clustered Model for Near-synonymy*. Berlin; New York: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110220599.

Divjak, D. and N. Fieller (2014). 'Cluster Analysis: Finding Structure in Linguistic Data'. In: *Human Cognitive Processing*. Ed. by D. Glynn and J. A. Robinson. Vol. 43. Amsterdam: John Benjamins Publishing Company, pp. 405–441. DOI: https://doi.org/10.1075/hcp.43.16div.

Divjak, D. and S. T. Gries (2006). 'Ways of Trying in Russian: Clustering Behavioral Profiles'. In: *Corpus Linguistics and Linguistic Theory* 2.1, pp. 23–60. DOI: https://doi.org/10.1515/cllt.2006.002.

— (2008). 'Clusters in the Mind?: Converging Evidence From Near Synonymy in Russian'. In: *The Mental Lexicon* 3.2, pp. 188–213.

Dolgova Jacobsen, N. (2016). 'The Best of Both Worlds: Combining Cognitive Linguistics and Pedagogic Tasks to Teach English Conditionals'. In: *Applied Linguistics* 39.5, pp. 668–693. DOI: https://doi.org/10.1093/applin/amw030.

Donovan, J. (1895). 'Sonnenschein's Greek Grammar'. In: *The Classical Review* 9.1, pp. 60–67.

Douven, I. (2016). *The Epistemology of Indicative Conditionals: Formal and Empirical Approaches*. Cambridge: Cambridge University Press.

— (2017a). 'Clustering Colors'. In: *Cognitive Systems Research* 45, pp. 70–81. DOI: https://doi.org/10.1016/j.cogsys.2017.05.004.

— (2017b). 'How to Account for the Oddness of Missing-Link Conditionals'. In: *Synthese* 194.5, pp. 1541–1554. DOI: https://doi.org/10.1007/s11229-015-0756-7.

Douven, I., S. Elqayam and K. Krzyżanowska (2021). 'Inferentialism: A Manifesto'. In: *Conditionals: Logic, Linguistics, and Psychology*. Palgrave Macmillan. in press.

Douven, I., S. Elqayam, H. Singmann and J. van Wijnbergen-Huitink (2018). 'Conditionals and Inferential Connections: A Hypothetical Inferential Theory'. In: *Cognitive Psychology* 101, pp. 50–81. DOI: https://doi.org/10.1016/j.cogpsych.2017.09.002.

Dowty, D. R. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht; Boston: D. Reidel Publishing Company.

— (1986). 'The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics?' In: *Linguistics and Philosophy* 9.1, pp. 37–61.

Dowty, D. R., R. E. Wall and S. Peters (1981). *Introduction to Montague Semantics*. Dordrecht: Springer.

Ducrot, O. (1996). *Slovenian Lectures. ConféRences Slovènes*. Ljubljana: Institut za humanisticne studije.

Dudman, V. H. (1984). 'Conditional Interpretations of *if*-Sentences'. In: *Australian Journal of Linguistics* 4.2, pp. 143–204.

Duin, M. (2011). 'Als deze zin ook met *wanneer* kan, wat is het verschil dan? Het verschil tussen *wanneer* en *als*'. MA thesis. Leiden: Leiden University.

Duinhoven, A. M. (1984). 'Ban de bom! Over Vorm en Betekenis van de Imperatief'. In: *Nieuwe Taalgids* 77.2, pp. 148–156.

— (1995). 'Had Gebeld! De Irreële Imperatief'. In: *Tijdschrift voor Nederlandse Taal- en Letterkunde* 111.

— (1997). *Middelnederlandse syntaxis: Synchroon en diachroon. Deel 2: De werkwoordgroep*. Leiden: Nijhoff.

Dummett, M. (1973). *Frege: Philosophy of Language*. London: Gerald Duckworth.

Dy, J. G. and C. E. Brodley (2004). 'Feature Selection for Unsupervised Learning'. In: *Journal of Machine Learning Research* 5.Aug, pp. 845–889.

Ebert, C., C. Endriss and S. Hinterwimmer (2008). 'A Unified Analysis of Indicative and Biscuit Conditionals as Topics'. In: *Semantics and Linguistic Theory*. Vol. 18, pp. 266–283.

Edgington, D. (1986). 'Do Conditionals Have Truth Conditions?' In: *Crítica: Revista Hispanoamericana de Filosofía*, pp. 3–39.

— (1995). 'On Conditionals'. In: *Mind* 104.414, pp. 235–329.

— (2008). 'The Presidential Address: Counterfactuals'. In: *Proceedings of the Aristotelian Society* 108, pp. 1–21.

— (2020). 'Counterfactual Conditionals'. In: *The Routledge Handbook of Modality*. Ed. by O. Bueno and S. A. Shalkowski. New York: Routledge, pp. 30–39.

Edgington, D. (2021). 'Suppose and Tell: The Semantics and Heuristics of Conditionals'. In: *History and Philosophy of Logic*, pp. 1–8. DOI: https://doi.org/10.1080/01445340.2021.1958648.

Egre, P. and M. Cozic (2016). 'Conditionals'. In: *The Cambridge Handbook of Formal Semantics*. Ed. by M. Aloni and P. Dekker. Cambridge: Cambridge University Press, pp. 490–524. DOI: https://doi.org/10.1017/cbo9781139236157.018.

Elder, C.-H. (2019a). *Context, Cognition and Conditionals*. Palgrave Studies in Pragmatics, Language and Cognition. Cham: Palgrave Macmillan. DOI: https://doi.org/10.1007/978-3-030-13799-1.

— (2019b). 'Metalinguistic Conditionals and the Role of Explicit Content'. In: *Linguistics* 57.6, pp. 1337–1365. DOI: https://doi.org/10.1515/ling-2019-0029.

Elder, C.-H. and K. M. Jaszczolt (2016). 'Towards a Pragmatic Category of Conditionals'. In: *Journal of Pragmatics* 98, pp. 36–53. DOI: https://doi.org/10.1016/j.pragma.2016.04.013.

Elder, C.-H. and E. Savva (2018). 'Incomplete Conditionals and the Syntax–Pragmatics Interface'. In: *Journal of Pragmatics* 138, pp. 45–59. DOI: https://doi.org/10.1016/j.pragma.2018.09.015.

Elliott, W. E. (1981). 'Conditional Sentences in the New Testament'. PhD thesis. Winona Lake: Grace Theological Seminary.

Enders, C. K. (2010). *Applied Missing Data Analysis*. Methodology in the Social Sciences. New York: Guilford Press.

Eskin, E., A. Arnold, M. Prerau, L. Portnoy and S. Stolfo (2002). 'A Geometric Framework for Unsupervised Anomaly Detection'. In: *Applications of Data Mining in Computer Security*. Springer, pp. 77–101.

Espino, O. and R. M. J. Byrne (2012). 'It is not the Case That if you Understand a Conditional you Know how to Negate it'. In: *Journal of Cognitive Psychology* 24.3, pp. 329–334. DOI: https://doi.org/10.1080/20445911.2011.639759.

Espino, O., R. M. J. Byrne and P. N. Johnson-Laird (2020). 'Possibilities and the Parallel Meanings of Factual and Counterfactual Conditionals'. In: *Memory & Cognition* 48.7, pp. 1263–1280. DOI: https://doi.org/10.3758/s13421-020-01040-6.

Ester, M. (2014). 'Density-Based Clustering'. In: *Data Clustering: Algorithms and Applications*. Ed. by C. C. Aggarwal and C. K. Reddy. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Boca Raton: Chapman & Hall/CRC, pp. 111–126.

Estiri, H., J. G. Klann and S. N. Murphy (2019). 'A Clustering Approach for Detecting Implausible Observation Values in Electronic Health Records Data'. In: *BMC Medical Informatics and Decision Making* 19.142. DOI: https://doi.org/10.1186/s12911-019-0852-6.

Evans, J. St. B. T. (1972). 'Reasoning With Negatives'. In: *British Journal of Psychology* 63.2, pp. 213–219. DOI: https://doi.org/10.1111/j.2044-8295.1972.tb02102.x.

— (2005). 'The Social and Communicative Function of Conditional Statements'. In: *Mind & Society* 4.1, pp. 97–113. DOI: https://doi.org/10.1007/s11299-005-0003-x.

Evans, J. S. B. T., J. Clibbens and B. Rood (1996). 'The Role of Implicit and Explicit Negation in Conditional Reasoning Bias'. In: *Journal of Memory and Language* 35.3, pp. 392–409.

Evans, J. S. B. T. and S. J. Handley (1999). 'The Role of Negation in Conditional Inference'. In: *The Quarterly Journal of Experimental Psychology Section A* 52.3, pp. 739–769. DOI: https://doi.org/10.1080/713755834.

Evans, J. S. B. T., S. J. Handley and D. E. Over (2003). 'Conditionals and Conditional Probability'. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.2, pp. 321–335. DOI: https://doi.org/10.1037/0278-7393.29.2.321.

Evans, J. S. B. T. and D. E. Over (2004). If*: Supposition, Pragmatics, and Dual Processes*. Oxford: Oxford University Press.

Evans, N. (2007). 'Insubordination and its Uses'. In: *Finiteness: Theoretical and Empirical Foundations*. Ed. by I. Nikolaeva. Oxford; New York: Oxford University Press, pp. 366–431.

Everitt, B. (1977). *The Analysis of Contingency Tables*. Monographs on Applied Probability and Statistics. London; New York: Chapman & Hall; John Wiley & Sons.

Evers-Vermeul, J., J. Hoek and M. C. Scholman (2017). 'On Temporality in Discourse Annotation: Theoretical and Practical Considerations'. In: *Dialogue & Discourse* 8.2, pp. 1–20.

Fauconnier, G. (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge; New York, NY, USA: Cambridge University Press.

Faure, D. and C. Nédellec (1998). 'A Corpus-Based Conceptual Clustering Method for Verb Frames and Ontology Acquisition'. In: *Lrec Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Applications*. Vol. 707. 728.

Feger, H. (2001). 'Classification: Conceptions in the Social Sciences'. In: *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, pp. 1966–1973. DOI: https://doi.org/10.1016/b0-08-043076-7/00699-9.

Feinstein, A. R. and D. V. Cicchetti (1990). 'High Agreement but low Kappa: I. the Problems of two Paradoxes'. In: *Journal of Clinical Epidemiology* 43.6, pp. 543–549. DOI: https://doi.org/10.1016/0895-4356(90)90158-l.

Ferguson, G. (2001). 'If you pop Over There: A Corpus-Based Study of Conditionals in Medical Discourse'. In: *English for Specific Purposes* 20.1, pp. 61–82. DOI: https://doi.org/10.1016/s0889-4906(99)00027-7.

Field, A. P., J. Miles and Z. Field (2012). *Discovering Statistics Using R*. London; Thousand Oaks, California: Sage.

Fillenbaum, S. (1975). 'If: Some Uses'. In: *Psychological Research* 37.3, pp. 245–260. DOI: https://doi.org/10.1007/bf00309037.

Fillenbaum, S. (1986). 'The Use of Conditionals in Inducements and Deterrents'. In: *On Conditionals*. Ed. by E. C. Traugott, A. T. Meulen, J. S. Reilly and C. A. Ferguson. Cambridge: Cambridge University Press, pp. 179–195.

Fillmore, C. J. (1975). 'An Alternative to Checklist Theories of Meaning'. In: *Proceedings of the First Annual Linguistics Society, Berkeley*. Berkeley: Berkeley Linguistics Society, pp. 123–131.

— (1985). 'Frames and the Semantics of Understanding'. In: *Quaderni di Semantica* 6.2, pp. 222–254.

— (1986). 'Varieties of Conditional Sentences'. In: *Eastern States Conference on Linguistics*. Vol. 3. Ithica: DMLL Publications, pp. 163–182.

— (1992). 'Epistemic Stance and Grammatical Form in English Conditional Sentences'. In: *The 26th Regional Meeting of the Chicago Linguistic Society*. Vol. 26, pp. 137–161.

Fillmore, C. J., P. Kay and M. C. O'Connor (1988). 'Regularity and Idiomaticity in Grammatical Constructions: The Case of *let alone*'. In: *Language* 64.3, pp. 501–538. DOI: https://doi.org/10.2307/414531.

Finkbeiner, R. (2019). 'Reflections on the Role of Pragmatics in Construction Grammar'. In: *Constructions and Frames* 11.2, pp. 171–192. DOI: 10.1075/cf.00027.fin.

Fleischman, S. (1995). 'Imperfective and Irrealis'. In: *Typological Studies in Language*. Ed. by J. L. Bybee and S. Fleischman. Vol. 32. Amsterdam: John Benjamins Publishing Company, pp. 519–551. DOI: https://doi.org/10.1075/tsl.32.23fle.

Fleiss, J. and J. Cohen (1973). 'Kappa/Weighted Kappa'. In: *Educational and Psychological Measurement* 33, pp. 613–619.

Foolen, A. (1993). 'De Betekenis van partikels: een dokumentatie van de stand van het onderzoek, met bijzondere aandacht voor 'maar''. PhD thesis. Nijmegen: Katholieke Universiteit Nijmegen.

Fop, M. and T. B. Murphy (2018). 'Variable Selection Methods for Model-Based Clustering'. In: *Statistics Surveys* 12.0, pp. 18–65. DOI: https://doi.org/10.1214/18-ss119.

Forbes, G. (1994). *Modern Logic: A Text in Elementary Symbolic Logic*. New York: Oxford University Press.

Ford, C. E. (1993). *Grammar in Interaction: Adverbial Clauses in American English Conversations*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511554278.

— (1997). 'Speaking Conditionally: Some Contexts for *if*-clauses in Conversation'. In: *On Conditionals Again*. Ed. by A. Athanasiadou and R. Dirven. Amsterdam: John Benjamins Publishing Company, pp. 387–414. DOI: https://doi.org/10.1075/cilt.143.21for.

Ford, C. E. and S. A. Thompson (1986). 'Conditionals in Discourse: A Text-Based Study From English'. In: *On Conditionals*. Ed. by E. C. Traugott, A. T. Meulen, J. S. Reilly and C. A. Ferguson. Cambridge: Cambridge University Press, pp. 353–372. DOI: https://doi.org/10.1017/cbo9780511753466.019.

Fortuin, E. L. J. (2004). 'De syntaxis van imperatiefsubjecten en modale Partikels: een pragma-semantische benadering'. In: *Nederlandse Taalkunde* 9, pp. 355–375.

— (2011). 'Iconicity, Economy and Frequency: the Paratactic Conditional Perfective Present Construction in Russian'. In: *Russian Linguistics* 35.1, pp. 89–123. DOI: https://doi.org/10.1007/s11185-010-9071-6.

— (2019). 'Universality and Language-Dependency of Tense and Aspect: Performatives from a Crosslinguistic Perspective'. In: *Linguistic Typology* 23.1, pp. 1–58. DOI: https://doi.org/10.1515/lingty-2018-0018.

Fortuin, E. L. J. and R. J. U. Boogaart (2009). 'Imperative as Conditional: From Constructional to Compositional Semantics'. In: *Cognitive Linguistics* 20.4. DOI: https://doi.org/10.1515/cogl.2009.028.

Fowlkes, E. B., R. Gnanadesikan and J. R. Kettenring (1988). 'Variable Selection in Clustering'. In: *Journal of classification* 5.2, pp. 205–228.

Fraley, C. and A. E. Raftery (1998). 'How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis'. In: *The Computer Journal* 41.8, pp. 578–588.

— (2002). 'Model-Based Clustering, Discriminant Analysis, and Density Estimation'. In: *Journal of the American Statistical Association* 97.458, pp. 611–631. DOI: https://doi.org/10.1198/016214502760047131.

Frege, G. (1879). *Begriffsschrift, Eine der Arithmetischen Nachgebildete Formelsprache des Reinen Denkens*. Verlag von Louis Nebert.

— (1948). 'Sense and Reference'. In: *The Philosophical Review* 57.3, pp. 209–230. DOI: https://doi.org/10.2307/2181485.

Fugard, A., N. Pfeifer, M. B and G. Kleiter (2011). 'How People Interpret Conditionals: Shifts Toward the Conditional Event'. In: *Journal of Experimental Psychology Learning Memory and Cognition* 37, pp. 635–648.

Funk, W.-P. (1985). 'On a Semantic Typology of Conditional Sentences'. In: *Folia Linguistica* 19.3-4, pp. 365–413. DOI: https://doi.org/10.1515/flin.1985.19.3-4.365.

Gabrielatos, C. (2010). 'A Corpus-Based Examination of English if-Conditionals Through the Lens of Modality: Nature and Types'. PhD thesis. Lancaster University.

— (2019). '*If*-Conditionals and Modality: Frequency Patterns and Theoretical Explanations'. In: *Journal of English Linguistics* 47.4, pp. 301–334. DOI: https://doi.org/10.1177/0075424219875994.

— (2020). 'Conditionals, Modality, and SchröDinger's cat: Conditionals as a Family of Linguistic Qubits'. In: *Re-Assessing Modalising Expressions: Categories, Co-Text, and Context*. Ed. by P. Hohaus and R. Schulze. Amsterdam: John Benjamins Publishing Company, pp. 141–172.

Gabrielatos, C. (2021). '*If*-conditionals: Corpus-based Classification and Frequency Distribution'. In: *ICAME Journal* 45.1, pp. 87–124. DOI: 10.2478/icame-2021-0003.

Gamut, L. T. F. (1991). *Language, Logic, and Meaning: Introduction to Logic*. Vol. 1.

Gastwirth, J. L. (1972). 'The Estimation of the Lorenz Curve and Gini Index'. In: *The Review of Economics and Statistics* 54.3, pp. 306–316. DOI: https://doi.org/10.2307/1937992.

Gazdar, G. (1979). *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.

Gazzo Castañeda, L. E. and M. Knauff (2021). 'Specificity Effects in Reasoning with Counterintuitive and Arbitrary Conditionals'. In: *Memory & Cognition*. DOI: 10.3758/s13421-021-01235-5.

Geach, P. T. (1976). *Reason and Argument*. Berkeley: University of California Press.

Geeraerts, D. (2006). 'Prototype Theory'. In: *Cognitive Linguistics: Basic Readings*. Ed. by D. Geeraerts. Vol. 34. Berlin; New York: Mouton de Gruyter, pp. 141–166. DOI: https://doi.org/10.1515/9783110199901.141.

— (2017). 'Entrenchment as Onomasiological Salience'. In: *Entrenchment and the Psychology of Language Learning*. Ed. by H.-J. Schmid. Berlin; Boston: De Gruyter Mouton, pp. 153–174. DOI: https://doi.org/10.1515/9783110341423-008.

Geerts, G., W. Haeseryn, J. de Rooij and M. C. van den Toorn (1984). *Algemene Nederlandse Spraakkunst*. Groningen: Wolters-Noordhoff.

Geis, M. L. (1973). '*If* and *unless*'. In: *Issues in Linguistics: Papers in Honor of Henry and René Kahane*. Ed. by B. B. Kachru, R. B. Lees, Y. Malkiel, A. Petrangeli and S. Petrangeli. Urbana: University of Illinois Press, pp. 231–253.

Geis, M. L. and W. G. Lycan (1993). 'Nonconditional Conditionals'. In: *Philosophical Topics* 21.2, pp. 35–56. DOI: https://doi.org/10.5840/philtopics199321215.

Geis, M. L. and A. M. Zwicky (1971). 'On Invited Inferences'. In: *Linguistic Inquiry* 2.4, pp. 561–566.

Gerlofs, J. M. (2009). 'The use of Conditionals in Argumentation: A Proposal for the Analysis and Evaluation of Argumentatively Used Conditionals'. PhD thesis. Amsterdam: University of Amsterdam.

Geurts, B. (1999). *Presuppositions and Pronouns*. Amsterdam; New York: Elsevier.

— (2017). 'Presupposition and Givenness'. In: *The Oxford Handbook of Pragmatics*. Ed. by Y. Huang. Vol. 1. Oxford University Press, pp. 180–198. DOI: https://doi.org/10.1093/oxfordhb/9780199697960.013.21.

Giannakidou, A. (1998). *Polarity Sensitivity as (Non)Veridical Dependency*. Linguistik Aktuell. Amsterdam: John Benjamins Publishing Company.

Gildersleeve, B. L. (1882). 'Studies in Pindaric Syntax'. In: *The American Journal of Philology* 3.12, pp. 434–455.

Glasbey, S. (2007). 'Aspectual Composition in Idioms'. In: *Recent Advances in the Syntax and Semantics of Tense, Aspect and Modality*. Ed. by d. Louis, J. Moeschler and G. Puskás. Berlin: Mouton de Gruyter, pp. 71–87.

Godden, D. M. and D. Walton (2004). 'Denying the Antecedent as a Legitimate Argumentative Strategy: A Dialectical Model'. In: *Informal Logic* 24.3, pp. 219–243. DOI: https://doi.org/10.22329/il.v24i3.2146.

Goebel, A. (2017). 'Laws for Biscuits: Independence and Dependence in Conditionals'. In: *Semantics and Linguistic Theory* 27, pp. 377–396. DOI: https://doi.org/10.3765/salt.v27i0.4163.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

— (1996). 'Making One's Way Through the Data'. In: *Grammatical Constructions: Their Form and Meaning*. Ed. by S. A. Thompson and M. Shibatani. Oxford: Oxford University Press, pp. 29–53.

— (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford; New York: Oxford University Press.

— (2013). 'Constructionist Approaches'. In: *The Oxford Handbook of Construction Grammar*. Ed. by T. Hoffmann and G. Trousdale. Oxford University Press, pp. 15–31. DOI: 10.1093/oxfordhb/9780195396683.013.0002.

— (2019). *Explain me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton, NJ: Princeton University Press.

Gomes, G. (2020). 'Concessive Conditionals Without *even if* and Nonconcessive Conditionals with *even if*'. In: *Acta Analytica* 35.1, pp. 1–21. DOI: https://doi.org/10.1007/s12136-019-00396-y.

Goodall, D. W. (1966). 'A New Similarity Index Based on Probability'. In: *Biometrics* 22.4, pp. 882–907. DOI: https://doi.org/10.2307/2528080.

Goodwin, W. W. (1879). *The Greek Verb*. 5th ed. Cambridge: John Wilson & Son.

Gower, J. C. (1971). 'A General Coefficient of Similarity and some of its Properties'. In: *Biometrics* 27.4, pp. 857–871. DOI: https://doi.org/10.2307/2528823.

Greenberg, J. H. (1957). 'The Nature and Uses of Linguistic Typologies'. In: *International Journal of American Linguistics* 23.2, pp. 68–77.

— (1966). 'Some Universals of Grammar With Particular Reference to the Order of Meaningful Elements'. In: *Universals of Language: Report of a Conference Held at Dobbs Ferry, New York, April 13-15, 1961*. Ed. by J. H. Greenberg. 2nd ed. Cambridge, Massachusetts: MIT Press, pp. 73–113.

Grice, H. P. (1957). 'Meaning'. In: *The Philosophical Review* 66.3, pp. 377–388. DOI: https://doi.org/10.2307/2182440.

— (1975). 'Logic and Conversation'. In: *Speech Acts*. Ed. by P. Cole and J. L. Morgan. Leiden: Brill, pp. 41–58. DOI: https://doi.org/10.1163/9789004368811_003.

— (1989). *Studies in the way of Words*. Cambridge, Massachusetts: Harvard University Press.

Grice, H. P. and A. R. White (1961). 'The Causal Theory of Perception'. In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 35, pp. 121–168.

Gries, S. T. and N. Otani (2010). 'Behavioral Profiles: A Corpus-Based Perspective on Synonymy and Antonymy'. In: *ICAME Journal* 34, pp. 121–150.

Gries, S. T. and A. Stefanowitsch (2010). 'Cluster Analysis and the Identification of Collexeme Classes'. In: *Empirical and Experimental Methods in Cognitive/Functional Research*. Ed. by S. Rice and J. Newman. Stanford: CSLI Publications.

Gries, S. Th. (2009). *Quantitative Corpus Linguistics With R: A Practical Introduction*. New York: Routledge.

— (2010). 'Behavioral Profiles: A Fine-Grained and Quantitative Approach in Corpus-Based Lexical Semantics'. In: *The Mental Lexicon* 5.3, pp. 323–346. DOI: https://doi.org/10.1075/ml.5.3.04gri.

— (2013). *Statistics for Linguistics With R: A Practical Introduction*. Berlin: Walter de Gruyter.

Grubbs, F. E. (1969). 'Procedures for Detecting Outlying Observations in Samples'. In: *Technometrics* 11.1, pp. 1–21.

Grusdt, B. and M. Franke (2021). 'Communicating Uncertain Beliefs with Conditionals: Probabilistic Modeling and Experimental Data'. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 43, pp. 1712–1718.

Günthner, S. (2016). 'From Biclausal Constructions to 'Stand-Alone'-Conditionals: 'Syntactically Disintegrated *wenn*-constructions' in Everyday Spoken German'. In: *Sprache Interaktion Arbeitspapierreihe* 58.

Gwet, K. L. (2008). 'Computing Inter-Rater Reliability and its Variance in the Presence of High Agreement'. In: *British Journal of Mathematical and Statistical Psychology* 61.1, pp. 29–48. DOI: https://doi.org/10.1348/000711006x126600.

— (2011). 'On Krippendorff's Alpha Coefficient'. Gaithersburg, MD.

— (2014). *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Advanced Analytics, LLC.

Haegeman, L. M. V. (2003). 'Conditional Clauses: External and Internal Syntax'. In: *Mind and Language* 18.4, pp. 317–339. DOI: https://doi.org/10.1111/1468-0017.00230.

— (2012). *Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery*. Oxford; New York: Oxford University Press.

Haegeman, L. M. V. and H. Wekker (1984). 'The Syntax and Interpretation of Futurate Conditionals in English'. In: *Journal of Linguistics* 20.1, pp. 45–55. DOI: https://doi.org/10.2307/4175718.

Haeseryn, W., K. Romijn, J. Geerts de Rooij and M. C. van den Toorn (1997). *Algemene Nederlandse Spraakkunst*. 2nd ed. Groningen: Wolters-Noordhoff.

Haigh, M., A. J. Stewart, J. S. Wood and L. Connell (2011). 'Conditional Advice and Inducements: Are Readers Sensitive to Implicit Speech Acts During Comprehension?' In: *Acta Psychologica* 136.3, pp. 419–424. DOI: https://doi.org/10.1016/j.actpsy.2011.01.009.

Haiman, J. (1974). 'Concessives, Conditionals, and Verbs of Volition'. In: *Foundations of Language* 11, pp. 341–359. DOI: https://doi.org/10.2307/25000782.

— (1978). 'Conditionals are Topics'. In: *Language* 54.3, pp. 564–589. DOI: https://doi.org/10.1353/lan.1978.0009.

— (1985). *Natural Syntax: Iconicity and Erosion*. Cambridge Studies in Linguistics. Cambridge; New York: Cambridge University Press.

Haiman, J. and S. A. Thompson (1984). '"Subordination" in Universal Grammar'. In: *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*. Vol. 10, pp. 510–523. DOI: https://doi.org/10.3765/bls.v10i0.1973.

Hale, B. (1997). 'Grundlagen §64'. In: *Proceedings of the Aristotelian Society*. Vol. 97. 3, pp. 243–261. DOI: https://doi.org/10.1111/1467-9264.00015.

Han, J., L. Zhu, M. Kulldorff, S. Hostovich, D. G. Stinchcomb, Z. Tatalovich, D. R. Lewis and E. J. Feuer (2016). 'Using Gini Coefficient to Determining Optimal Cluster Reporting Sizes for Spatial Scan Statistics'. In: *International Journal of Health Geographics* 15.1. DOI: https://doi.org/10.1186/s12942-016-0056-6.

Handley, S. J., J. S. B. T. Evans and V. A. Thompson (2006). 'The Negated Conditional: A Litmus Test for the Suppositional Conditional?' In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32.3, pp. 559–569. DOI: https://doi.org/10.1037/0278-7393.32.3.559.

Harnad, S. (2017). 'To Cognize is to Categorize: Cognition is Categorization'. In: *Handbook of Categorization in Cognitive Science*. Elsevier, pp. 21–54.

Harris, R. J. (2001). *A Primer of Multivariate Statistics*. 3rd ed. Mahwah: Lawrence Erlbaum Associates.

Hartigan, J. A. and P. M. Hartigan (1985). 'The Dip Test of Unimodality'. In: *The Annals of Statistics* 13.1, pp. 70–84. DOI: https://doi.org/10.1214/aos/1176346577.

Hartigan, P. M. (1985). 'Computation of the Dip Statistic to Test for Unimodality'. In: *Applied Statistics* 34.3, pp. 320–325. DOI: https://doi.org/10.2307/2347485.

Hartmann, S. and U. Hahn (2020). 'A New Approach to Testimonial Conditionals'. In: *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Ed. by S. Denison, M. Mack, Y. Xu and B. C. Amstrong, pp. 981–986.

Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford; New York: Clarendon Press; Oxford University Press.

— (2006). 'Against Markedness (and What to Replace it With)'. In: *Journal of Linguistics* 42.1, pp. 25–70.

Hassett, W., S. M. Bollens, T. D. Counihan, G. Rollwagen-Bollens, J. Zimmerman, S. Katz and J. Emerson (2017). 'Veligers of the Invasive Asian Clam Corbicula Fluminea in the Columbia River Basin: Broadscale Distribution, Abundance, and Ecological Associations'. In: *Lake and Reservoir Management* 33.3, pp. 234–248. DOI: https://doi.org/10.1080/10402381.2017.1294218.

Hay, J. and H. Baayen (2003). 'Phonotactics, Parsing and Productivity'. In: *Italian Journal of Linguistics* 15, pp. 99–130.

Hayes, A. F. and K. Krippendorff (2007). 'Answering the Call for a Standard Reliability Measure for Coding Data'. In: *Communication Methods and Measures* 1.1, pp. 77–89.

Heim, I. and A. Kratzer (1998). *Semantics in Generative Grammar*. Oxford: Blackwell Publishers.

Hempel, C. G. and P. Oppenheim (1936). *Der Typusbegriff im Lichte der Neuen Logik (The Concept of Type in the Light of Modern Logic. Scientific-Theoretical Investigations in Constitution Research and Psychology)*. Wissenschaftstheoretische Untersuchungen zur Konstitutionsforschung und Psychologie. Leiden: Sijthoff.

Hempel, C. G. (1965). 'Fundamentals of Taxonomy'. In: *Aspects of Scientific Explanation*. Glencoe: Free Press, pp. 137–154.

Hennig, C. (2007). 'Cluster-Wise Assessment of Cluster Stability'. In: *Computational Statistics & Data Analysis* 52.1, pp. 258–271. DOI: https://doi.org/10.1016/j.csda.2006.11.025.

— (2015). 'What are the True Clusters?' In: *Pattern Recognition Letters* 64, pp. 53–62. DOI: https://doi.org/10.1016/j.patrec.2015.04.009.

Hennig, C. and T. F. Liao (2013). 'How to Find an Appropriate Clustering for Mixed-Type Variables With Application to Socio-Economic Stratification'. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62.3, pp. 309–369.

Heringer, J. T. J. (1972). 'Some Grammatical Correlates of Felicity Conditions and Presuppositions'. PhD thesis. Ohio State University Department of Linguistics.

Hermkens, H. M. (1974). *Verzorgd Nederlands*. 5th ed. Den Bosch: Malmberg.

Hilpert, M. and S. T. Gries (2009). 'Assessing Frequency Changes in Multistage Diachronic Corpora: Applications for Historical Corpus Linguistics and the Study of Language Acquisition'. In: *Literary and Linguistic Computing* 24.4, pp. 385–401. DOI: https://doi.org/10.1093/llc/fqn012.

Hilton, D. (2016). 'Pragmatic Approaches to Deontic Conditionals: From Rule Content to Rule Use'. In: *Cognitive Unconscious and Human Rationality*. Ed. by L. Macchi, M. Bagassi and R. Viale. MIT Press, pp. 183–203. DOI: https://doi.org/10.7551/mitpress/10100.003.0014.

Hobbs, J. R. (1979). 'Coherence and Coreference'. In: *Cognitive Science* 3.1, pp. 67–90. DOI: https://doi.org/10.1207/s15516709cog0301_4.

Hoek, J. (2018). *Making Sense of Discourse: On Discourse Segmentation and the Linguistic Marking of Coherence Relations*. Utrecht: LOT/Netherlands Graduate School of Linguistics.

Hoek, J., J. Evers-Vermeul and T. J. M. Sanders (2019). 'Using the Cognitive Approach to Coherence Relations for Discourse Annotation'. In: *Dialogue & Discourse* 10.2, pp. 1–33.

Hoeksema, J. (2012). 'On the Natural History of Negative Polarity Items'. In: *Linguistic Analysis* 38.1/2, pp. 3–33.

Hoeksema, J. and F. Zwarts (1991). 'Some Remarks on Focus Adverbs'. In: *Journal of Semantics* 8.1-2, pp. 51–70. DOI: https://doi.org/10.1093/jos/8.1-2.51.

Holleman, B. C. and H. L. W. Pander Maat (2009). 'The Pragmatics of Profiling: Framing Effects in Text Interpretation and Text Production'. In: *Journal of Pragmatics* 41.11, pp. 2204–2221. DOI: https://doi.org/10.1016/j.pragma.2008.09.038.

Hopper, P. J. and E. C. Traugott (2003). *Grammaticalization*. Cambridge University Press. DOI: https://doi.org/10.1017/cbo9781139165525.

Horn, L. R. (1972). 'On the Semantic Properties of Logical Operators in English'. PhD thesis. Los Angeles: University of California.

— (1984). 'Toward a New Taxonomy for Pragmatic Inference: Q-Based and R Based Implicature'. In: *Meaning, Form and Use in Context*. Ed. by D. Schiffrin. Washington: Georgetown University Press.

— (1985). 'Metalinguistic Negation and Pragmatic Ambiguity'. In: *Language* 61.1, pp. 121–174. DOI: https://doi.org/10.2307/413423.

— (1989). *A Natural History of Negation*. The David Hume Series. Chicago: University of Chicago Press.

— (2000). 'From *if* to *iff*: Conditional Perfection as Pragmatic Strengthening'. In: *Journal of Pragmatics* 32.3, pp. 289–326. DOI: https://doi.org/10.1016/s0378-2166(99)00053-3.

— (2008). 'Toward a Fregean Pragmatics: Voraussetzung, Nebengedanke, Andeutung'. In: *Explorations in Pragmatics*. Ed. by I. Kecskes and L. R. Horn. Berlin: De Gruyter Mouton, pp. 39–72. DOI: https://doi.org/10.1515/9783110198843.1.39.

— (2016). 'Information Structure and the Landscape of (Non-)at-Issue Meaning'. In: *The Oxford Handbook of Information Structure*. Ed. by C. Féry and S. Ishihara. Vol. 1. Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199642670.013.009.

Horsella, M. and G. Sindermann (1992). 'Aspects of Scientific Discourse: Conditional Argumentation'. In: *English for Specific Purposes* 11.2, pp. 129–139. DOI: https://doi.org/10.1016/s0889-4906(05)80004-3.

Howitt, D. and D. Cramer (2008). *Introduction to Statistics in Psychology*. 4th ed. Harlow: Prentice Hall.

Hsu, H.-M., Y.-F. Wang and K.-M. Hu (2015). 'Direct and Indirect Conditionals: A Corpus-Based Study of Chinese *Yaoshi* and *Yaobushi* in Spoken and Written Discourse'. In: *Taiwan Journal of Linguistics* 13.2, pp. 31–77.

Huang, Y. (2009). 'Neo-Gricean Pragmatics'. In: *Concise Encyclopedia of Pragmatics*. Ed. by J. Mey. 2nd ed. Amsterdam: Elsevier, pp. 676–680.
— ed. (2017). *Introduction*. Oxford: Oxford University Press, pp. 1–12. DOI: https://doi.org/10.1093/oxfordhb/9780199697960.001.0001.
Hubers, F. and H. de Hoop (2013). 'The Effect of Prescriptivism on Comparative Markers in Spoken Dutch'. In: *Linguistics in the Netherlands* 30, pp. 89–101. DOI: https://doi.org/10.1075/avt.30.07hub.
Huddleston, R. (1981). 'Review of 'On the Semantics of Syntax: Mood and Condition in English' by Eirian C. Davies'. In: *Journal of Linguistics* 17.1, pp. 121–124.
— (1988). 'Review of 'A Comprehensive Grammar of the English Language' by Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik'. In: *Language* 64.2, pp. 345–354. DOI: https://doi.org/10.2307/415437.
Huddleston, R. and G. K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press. DOI: https://doi.org/10.1017/9781316423530.
Hwang, M. O. (1979). 'A Semantic and Syntactic Analysis of if-Conditionals'. MA thesis. Los Angeles: University of California.
Iatridou, S. (1991). 'Topics in Conditionals'. PhD thesis. Massachusetts: Massachusetts Institute of Technology.
— (1993). 'On the Contribution of Conditional *Then*'. In: *Natural Language Semantics* 2.3, pp. 171–199. DOI: https://doi.org/10.1007/bf01256742.
— (2000). 'The Grammatical Ingredients of Counterfactuality'. In: *Linguistic Inquiry* 31.2, pp. 231–270.
— (2021). 'Grammar Matters'. In: *Conditionals, Paradox, and Probability*. Ed. by L. Walters and J. Hawthorne. Oxford: Oxford University Press, pp. 122–140. DOI: https://doi.org/10.1093/oso/9780198712732.003.0008.
Ippolito, M. (2003). 'Presuppositions and Implicatures in Counterfactuals'. In: *Natural Language Semantics* 11.2, pp. 145–186. DOI: https://doi.org/10.1023/a:1024411924818.
— (2013). *Subjunctive Conditionals: A Linguistic Analysis*. Linguistic Inquiry Monographs. Cambridge, Massachusetts; London, England: MIT Press.
Israel, M. (1996). 'The *way* Constructions Grow'. In: *Conceptual Structure, Discourse and Language*. Ed. by A. E. Goldberg. Stanford: CSLI, pp. 217–230.
— (2011). *The Grammar of Polarity: Pragmatics, Sensitivity, and the Logic of Scales*. Cambridge University Press.
Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford; New York: Oxford University Press.
Jackson, F. (1979). 'On Assertion and Indicative Conditionals'. In: *The Philosophical Review* 88.4, pp. 565–589. DOI: https://doi.org/10.2307/2184845.
— (1987). *Conditionals*. Oxford; New York: Blackwell.
— (1998). *Mind, Method, and Conditionals: Selected Essays*. London; New York: Routledge.

— (2006). 'Conditionals'. In: *The Blackwell Guide to the Philosophy of Language*. Ed. by M. Devitt and R. Hanley. New Jersey: Wiley-Blackwell; John Wiley & Sons, pp. 212–224.

Jacobsen, N. (2015). 'A Cognitive Linguistic Analysis of English Conditionals in English for Academic Purposes (EAP) Instruction: Implications From Sociocultural Theory'. In: *Cognitive Linguistics and Sociocultural Theory*. Ed. by K. Masuda, C. Arnett and A. Labarca. Berlin: De Gruyter, pp. 103–126. DOI: https://doi.org/10.1515/9781614514442-007.

Jacoby, S. and E. Ochs (1995). 'Co-Construction: An Introduction'. In: *Research on Language and Social Interaction* 28.3, pp. 171–183. DOI: https://doi.org/10.1207/s15327973rlsi2803_1.

Jain, A. K. and R. C. Dubes (1988). *Algorithms for Clustering Data*. New Jersey: Prentice-Hall, Inc.

Jansen, F. (2009). 'Het middenstuk in stukjes: Een onderzoek op basis van een corpus naar de werkzaamheid van grammaticale en communicatieve woordvolgordeprincipes in journalistieke teksten'. In: *Nederlandse Taalkunde* 14.2, pp. 172–191.

Jansen, H. (2003). 'Van omgekeerde strekking. Een pragma-dialectische reconstructie van a contrario-argumentatie in het recht'. PhD thesis. Amsterdam: Universiteit van Amsterdam.

— (2011). '"If That Were True, I Would Never Have..." The Counterfactual Presentation of Arguments That Appeal to Human Behaviour'. In: *Proceedings of the Seventh Conference of the International Society for the Study of Argumentation*. Ed. by F. H. van Eemeren, B. J. Garssen, D. Godden and G. Mitchell. Amsterdam: Rozenberg, pp. 881–889.

Janssen, T. A. J. M. (1989). 'Tempus: Interpretatie en Betekenis'. In: *De Nieuwe Taalgids* 82.312, pp. 323–325.

Jespersen, O. (1940). *A Modern English Grammar on Historical Principles*. London: Allen and Unwin.

Johnson-Laird, P. N. (1982). 'Formal Semantics and the Psychology of Meaning'. In: *Processes, Beliefs, and Questions: Essays on Formal Semantics of Natural Language and Natural Language Processing*. Ed. by S. Peters and E. Saarinen. Synthese Language Library. Dordrecht: Springer, pp. 1–68. DOI: https://doi.org/10.1007/978-94-015-7668-0_1.

— (1986). 'Conditionals and Mental Models'. In: *On Conditionals*. Ed. by E. C. Traugott, A. G. B. ter Meulen, J. S. Reilly and C. A. Ferguson. Cambridge: Cambridge University Press, pp. 55–76. DOI: https://doi.org/10.1017/cbo9780511753466.

— (2012). 'The Truth About Conditionals'. In: *The Science of Reason: A Festschrift for Jonathan St. B. T. Evans*. Ed. by K. Manktelow, D. Over and S. Elqayam. New York: Psychology Press, pp. 135–160.

Johnson-Laird, P. N. and R. M. J. Byrne (2002). 'Conditionals: A Theory of Meaning, Pragmatics, and Inference.' In: *Psychological Review* 109.4, pp. 646–678. DOI: https://doi.org/10.1037/0033-295x.109.4.646.

Joos, M. (1964). *The English Verb: Form and Meanings*. Madison/Milwaukee: University of Wisconsin Press.

Jowett, B., ed. (1892). *The Dialogues of Plato, Translated Into English With Analyses and Introductions*. By Plato. 3rd ed. Vol. 1. Oxford: Oxford University Press.

Juhos, C., A. C. Quelhas and R. M. J. Byrne (2015). 'Reasoning About Intentions: Counterexamples to Reasons for Actions.' In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41.1, pp. 55–76. DOI: https://doi.org/10.1037/a0037274.

Kaegi, A. (1905). *A Short Grammar of Classical Greek*. Ed. by J. A. Kleist. St. Louis: B. Herder.

Kang, N., E. M. van Mulligen and J. A. Kors (2012). 'Training Text Chunkers on a Silver Standard Corpus: Can Silver Replace Gold?' In: *BMC Bioinformatics* 13.17. DOI: https://doi.org/10.1186/1471-2105-13-17.

Kaplan, A. and J. M. Goldsen (1965). 'The Reliability of Content Analysis Categories'. In: *Language of Politics: Studies in Quantitative Semantics*, pp. 83–112.

Kapsner, A. (2020). 'Connexivity and the Pragmatics of Conditionals'. In: *Erkenntnis*. DOI: https://doi.org/10.1007/s10670-020-00325-1.

Karttunen, L. (1971). 'Implicative Verbs'. In: *Language* 47.2, pp. 340–358. DOI: https://doi.org/10.2307/412084.

— (2016). 'Presupposition: What Went Wrong?' In: *Semantics and Linguistic Theory (Salt)*. Ed. by M. Moroney, C.-R. Little, J. Collard and D. Burgdorf. Vol. 26. LSA and CLC Publications, pp. 705–731. DOI: https://doi.org/10.3765/salt.v26i0.3954.

Karttunen, L. and S. Peters (1979a). 'Conventional Implicature'. In: *Syntax and Semantics: Presupposition*. Ed. by C.-K. Oh and D. A. Dinneen. Vol. 11. New York: Academic Press, pp. 1–56.

— (1979b). 'Conventional Implicature'. In: *Syntax and Semantics* 11, pp. 1–56.

Kaufman, L. and P. J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken: John Wiley & Sons.

Kaufmann, M. and S. Kaufmann (2015). 'Conditionals and Modality'. In: *The Handbook of Contemporary Semantic Theory*. Ed. by S. Lappin and C. Fox. Chichester: John Wiley & Sons, pp. 237–270. DOI: https://doi.org/10.1002/9781118882139.ch8.

Kaufmann, W. A. (2013). *Nietzsche: Philosopher, Psychologist, Antichrist*. Princeton: Princeton University Press.

Kay, P. (1990). 'Even'. In: *Linguistics and Philosophy* 13.1, pp. 59–111. DOI: https://doi.org/10.1007/bf00630517.

Kehler, A., L. Kertz, H. Rohde and J. L. Elman (2008). 'Coherence and Coreference Revisited'. In: *Journal of Semantics* 25.1 Spec. Iss. Pp. 1–44. DOI: https://doi.org/10.1093/jos/ffm018.

Kemeny, J. G. (1959). *A Philosopher Looks at Science*. New York: Van Nostrand.

Kemmer, S. and A. Verhagen (1994). 'The Grammar of Causatives and the Conceptual Structure of Events'. In: *Cognitive Linguistics* 5.2, pp. 115–156.

Kempson, R. M. (1975). *Presupposition and the Delimitation of Semantics*. Cambridge: Cambridge University Press.

Khemlani, S., I. Orenes and P. N. Johnson-Laird (2014). 'The Negations of Conjunctions, Conditionals, and Disjunctions'. In: *Acta Psychologica* 151, pp. 1–7. DOI: https://doi.org/10.1016/j.actpsy.2014.05.004.

Kim, H.-Y. (2017). 'Statistical Notes for Clinical Researchers: Chi-Squared Test and Fisher's Exact Test'. In: *Restorative Dentistry & Endodontics* 42.2, pp. 152–155. DOI: https://doi.org/10.5395/rde.2017.42.2.152.

Kirsner, R. S. (1970). 'The Role of *Zullen* in the Grammar of Modern Standard Dutch'. In: *Lingua* 24, pp. 101–154.

Kment, B. (2020). 'Counterfactuals and Causal Reasoning'. In: *Perspectives on Causation: Selected Papers From the Jerusalem 2017 Workshop*. Ed. by E. A. Bar-Asher Siegal and N. Boneh. Cham: Springer International Publishing, pp. 463–482. DOI: https://doi.org/10.1007/978-3-030-34308-8_15.

Kneale, W. C. and M. Kneale (1962). *The Development of Logic*. Oxford: Clarendon Press.

Knott, A. (2001). 'Semantic and Pragmatic Relations and Their Intended Effects'. In: *Human Cognitive Processing*. Ed. by T. J. M. Sanders, J. Schilperoord and W. P. M. S. Spooren. Vol. 8. Amsterdam: John Benjamins Publishing Company, pp. 127–152. DOI: https://doi.org/10.1075/hcp.8.08kno.

Kohavi, R., G. H. John et al. (1997). 'Wrappers for Feature Subset Selection'. In: *Artificial Intelligence* 97.1-2, pp. 273–324.

Kollewijn, R. A. (1892). 'Het Systeem van de Tijden der Werkwoorden'. In: *Taal en Letteren* 2, pp. 141–147.

König, E. (1985). 'On the History of Concessive Connectives in English. Diachronic and Synchronic Evidence'. In: *Lingua* 66.1, pp. 1–19. DOI: https://doi.org/10.1016/s0024-3841(85)90240-2.

— (1991). *The Meaning of Focus Particles: A Comparative Perspective*. London; New York: Routledge.

König, E. and J. van der Auwera (1988). 'Clause Integration in German and Dutch Conditionals, Concessive Conditionals, and Concessives'. In: *Clause Combining in Grammar and Discourse*. Ed. by J. Haiman and S. A. Thompson. Amsterdam: John Benjamins Publishing Company, pp. 101–133. DOI: https://doi.org/10.1075/tsl.18.07kon.

Köpcke, K.-M. and K.-U. Panther (1989). 'On Correlations Between Word Order and Pragmatic Function of Conditional Sentences in German'. In: *Journal of Pragmatics* 13.5, pp. 685–711. DOI: https://doi.org/10.1016/0378-2166(89)90074-x.

Koster, J. (1975). 'Dutch as an SOV Language'. In: *Linguistic Analysis* 1.2, pp. 111–136.

Kratzer, A. (1986). 'Conditionals'. In: *Chicago Linguistics Society* 22.2, pp. 1–15.

— (2012). *Modals and Conditionals*. Oxford; New York: Oxford University Press.

Krifka, M. (1989). 'Nominal Reference, Temporal Constitution and Quantification in Event Semantics'. In: *Semantics and Contextual Expression*. Ed. by R. Bartsch, J. v. Benthem and P. v. Emde Boas. Berlin; Boston: De Gruyter, pp. 75–116. DOI: https://doi.org/10.1515/9783110877335-005.

Kripke, S. A. (1959). 'A Completeness Theorem in Modal Logic'. In: *Journal of Symbolic Logic* 24.1, pp. 1–14. DOI: https://doi.org/10.2307/2964568.

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*. Sage.

Kruskal, J. and M. Wish (1978). *Multidimensional Scaling*. Thousand Oaks, California: Sage Publications.

Kruskal, J. B. (1964). 'Nonmetric Multidimensional Scaling: A Numerical Method'. In: *Psychometrika* 29.2, pp. 115–129.

Krzyżanowska, K., S. Wenmackers and I. Douven (2013). 'Inferential Conditionals and Evidentiality'. In: *Journal of Logic, Language and Information* 22.3, pp. 315–334. DOI: https://doi.org/10.1007/s10849-013-9178-4.

Krzyżanowska, K. (2019). 'What is wrong with False-Link Conditionals?' In: *Linguistics Vanguard* 5.s3, p. 20190006. DOI: https://doi.org/10.1515/lingvan-2019-0006.

Krzyżanowska, K., P. Collins and U. Hahn (2017). 'The Puzzle of Conditionals With True Clauses: Against the Gricean Account'. In: *Cogsci 2017 Proceedings of the 39th Annual Meeting of the Cognitive Science Soxciety, London, UK, 26-29 July 2017*. Ed. by R. Granger, U. Hahn and R. Sutton, pp. 2476–2481.

— (2020). '...That P is Relevant for Q: Indicative Conditionals and Learning From Testimony'. In: *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Ed. by S. Denison, M. Mack, Y. Xu and B. C. Amstrong, pp. 987–993.

Krzyżanowska, K., S. Wenmackers and I. Douven (2014). 'Rethinking Gibbard's Riverboat Argument'. In: *Studia Logica* 102.4, pp. 771–792. DOI: https://doi.org/10.1007/s11225-013-9507-2.

Kuhn, M. (2008). 'Building Predictive Models in R using the Caret Package'. In: *Journal of Statistical Software* 28.5. DOI: https://doi.org/10.18637/jss.v028.i05.

Kuhn, M. and K. Johnson (2013). *Applied Predictive Modeling*. New York: Springer-Verlag.

Kuroda, M. (2007). 'Analysis of Contingency Tables'. In: *Statistical Methods for Biostatistics and Related Fields*. Berlin: Springer, pp. 105–124. DOI: https://doi.org/10.1007/978-3-540-32691-5_6.

Ladds, M. A., N. Sibanda, R. Arnold and M. R. Dunn (2018). 'Creating Functional Groups of Marine Fish from Categorical Traits'. In: *PeerJ* 6. DOI: https://doi.org/10.7717/peerj.5795.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

Landis, J. R. and G. G. Koch (1977). 'The Measurement of Observer Agreement for Categorical Data'. In: *Biometrics* 33.1, pp. 159–174. DOI: https://doi.org/10.2307/2529310.

Langacker, R. W. (1978). 'The Form and Meaning of the English Auxiliary'. In: *Language* 54.4, pp. 853–882. DOI: https://doi.org/10.2307/413237.

— (1987). *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Vol. 1. Stanford, California: Stanford University Press.

— (1988a). 'A Usage-Based Model'. In: *Current Issues in Linguistic Theory*. Ed. by B. Rudzka-Ostyn. Vol. 50. Amsterdam: John Benjamins Publishing Company, pp. 127–161. DOI: https://doi.org/10.1075/cilt.50.06lan.

— (1988b). 'An Overview of Cognitive Grammar'. In: *Current Issues in Linguistic Theory*. Ed. by B. Rudzka-Ostyn. Vol. 50. Amsterdam: John Benjamins Publishing Company, pp. 3–48. DOI: https://doi.org/10.1075/cilt.50.03lan.

— (1991). *Foundations of Cognitive Grammar, Volume II: Descriptive Application*. California: Stanford University Press.

— (2008). *Cognitive Grammar: A Basic Introduction*. Oxford; New York: Oxford University Press.

Lassiter, D. (2020). 'What we can Learn From how Trivalent Conditionals Avoid Triviality'. In: *Inquiry* 63.9–10, pp. 1087–1114. DOI: https://doi.org/10.1080/0020174x.2019.1698457.

— (in press). 'Decomposing Relevance in Conditionals'. In: *Mind & Language*.

Lauerbach, G. (1979). *Form und Funktion Englischer Konditionalsätze mit 'if': Eine Konversationslogische Und Sprechakttheoretische Analyse*. Linguistische Arbeiten; 72. Tübingen: Niemeyer.

Lazarsfeld, P. F. (1966). 'Concept Formation and Measurement in the Behavioral Sciences: Some Historical Observations'. In: *Concepts, Theory and Explanation in the Behavioral Sciences*. Ed. by G. J. Direnzo. New York: Random House, pp. 144–205.

Leclercq, B. (2020). 'Semantics and Pragmatics in Construction Grammar'. In: *Belgian Journal of Linguistics* 34, pp. 225–234. DOI: https://doi.org/10.1075/bjl.00048.lec.

Lee, C.-B. (2001). 'The Information Status of English *if*-clauses in Natural Discourse'. In: *Language Research*, pp. 483–505.

Leech, G. (2007). 'New Resources, or Just Better old Ones? The Holy Grail of Representativeness'. In: *Corpus Linguistics and the Web*. Ed. by M. Hundt, N. Nesselhauf and C. Biewer. Brill, pp. 133–149. DOI: https://doi.org/10.1163/9789401203791_009.

Leech, G. N. (1971). *Meaning and the English Verb*. London: Longman.

Legendre, P. and L. Legendre (1998). *Numerical Ecology*. Amsterdam: Elsevier.

Leuschner, T. (2016). 'Fictive Questions in Conditionals?: Synchronic and Diachronic Evidence From German and English'. In: *Human Cognitive Processing*. Ed. by E. Pascual and S. Sandler. Vol. 55. Amsterdam: John Benjamins Publishing Company, pp. 193–213. DOI: https://doi.org/10.1075/hcp.55.10leu.

Leuschner, T. and D. van den Nest (2015). 'Asynchronous Grammaticalization: V1-Conditionals in Present-day English and German'. In: *Languages in Contrast* 15.1, pp. 34–64. DOI: https://doi.org/10.1075/lic.15.1.03leu.

Levinson, S. C. (1979). 'Pragmatics and Social Deixis: Reclaiming the Notion of Conventional Implicature'. In: *Annual Meeting of the Berkeley Linguistics Society*. Ed. by C. Chiarello, J. Kingston, E. E. Sweetser, J. Collins, H. Kawasaki, J. Manley-Buser, D. W. Marschak, C. O'Connor, D. Shaul, M. Tobey, H. Thompson and K. Turner. Vol. 5. Berkeley, California: Berkeley Linguistics Society, pp. 206–223.

— (1983). *Pragmatics*. Cambridge University Press.

— (2000). *Presumptive Meanings: the Theory of Generalized Conversational Implicature*. Cambridge, Massachusetts: MIT Press.

Levshina, N. (2011). '*Doe wat je niet laten kan:* A Usage-Based Analysis of Dutch Causative Constructions'. PhD thesis. Leuven: Katholieke Universiteit Leuven.

— (2015). *How to do Linguistics With R: Data Exploration and Statistical Analysis*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

— (2016). 'A Geometric Exemplar-based Model of Semantic Structure: The Dutch Causative Construction with *laten*'. In: *Constructional Approaches to Language*. Ed. by J. Yoon and S. T. Gries. Vol. 19. John Benjamins Publishing Company, pp. 241–262. DOI: 10.1075/cal.19.09lev.

Levshina, N. and L. Degand (2017). 'Just *Because*: In Search of Objective Criteria of Subjectivity Expressed by Causal Connectives'. In: *Dialogue & Discourse* 8.1, pp. 132–150. DOI: https://doi.org/10.5087/dad.2017.105.

Lewis, C. I. (1912). 'Implication and the Algebra of Logic'. In: *Mind* 21.84, pp. 522–531.

— (1918). *A Survey of Symbolic Logic*. Berkeley: University of California Press.

Lewis, D. (1973a). 'Causation'. In: *The Journal of Philosophy* 70.17, pp. 556–567. DOI: https://doi.org/10.2307/2025310.

— (1973b). *Counterfactuals*. Oxford: Blackwell.

— (1976). 'Probabilities of Conditionals and Conditional Probabilities'. In: *The Philosophical Review* 85.3, pp. 297–315. DOI: https://doi.org/10.2307/2184045.

Li, J., K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang and H. Liu (2017). 'Feature Selection: A Data Perspective'. In: *ACM Computing Surveys* 50.6, pp. 1–45.

Libbrecht, M. W. and W. S. Noble (2015). 'Machine Learning Applications in Genetics and Genomics'. In: *Nature Reviews Genetics* 16.6, pp. 321–332.

Lin, D. (1998). 'An Information-Theoretic Definition of Similarity'. In: *Proceedings of the 15th International Conference on Machine Learning*. Burlington: Morgan Kaufmann, pp. 296–304.

Linde, C. (1976). 'Constraints on the Ordering of *if*-clauses'. In: *Annual Meeting of the Berkeley Linguistics Society*. Vol. 2, pp. 280–285.

Liu, M. and D. Zhang (2016). 'Feature Selection With Effective Distance'. In: *Neurocomputing* 215, pp. 100–109. DOI: https://doi.org/10.1016/j.neucom.2015.07.155.

Liu, M. (2019a). 'Current Issues in Conditionals'. In: *Linguistics Vanguard* 5.s3. DOI: https://doi.org/10.1515/lingvan-2019-0003.

— (2019b). 'The Elastic Nonveridicality Property of Indicative Conditionals'. In: *Linguistics Vanguard* 5.s3. DOI: https://doi.org/10.1515/lingvan-2019-0007.

Liu, M. and M. Barthel (2021). 'Semantics Processing of Conditional Connectives: German *wenn* 'if' versus *nur wenn* 'only if''. In: *Journal of Psycholinguistic Research*. DOI: 10.1007/s10936-021-09812-0.

Liu, M., S. Rotter and A. Giannakidou (2021). 'Bias and Modality in Conditionals: Experimental Evidence and Theoretical Implications'. In: *Journal of Psycholinguistic Research*. DOI: 10.1007/s10936-021-09813-z.

Löbner, S. (1989). 'German *schon - erst - noch*: An Integrated Analysis'. In: *Linguistics and Philosophy* 12.2, pp. 167–212. DOI: https://doi.org/10.1007/bf00627659.

— (2002). *Understanding Semantics*. London: Arnold.

Lombardi Vallauri, E. (2016). 'Insubordinated Conditionals in Spoken and Non-Spoken Italian'. In: *Insubordination*. Ed. by N. Evans and H. Watanabe. John Benjamins Publishing Company, pp. 145–170.

Łukasiewicz, J. (1970). *Selected Works*. Ed. by L. Borkowski. Amsterdam: North-Holland Publishing Company.

Lycan, W. G. (2001). *Real Conditionals*. Oxford; New York: Clarendon Press; Oxford University Press.

Łyda, A. and M. Zasowska (2021). 'Fictional Threatening: The Case of Disjunctive Conditionals'. In: *Linguistica Silesiana* 42, pp. 207–237. DOI: https://doi.org/10.24425/linsi.2021.137238.

MacFarlane, J. (2020). *Philosophical Logic: A Contemporary Introduction*. Routledge. DOI: https://doi.org/10.4324/9781315185248.

Mackay, J. (2015). 'Actuality and Fake Tense in Conditionals'. In: *Semantics and Pragmatics* 8, pp. 1–12. DOI: https://doi.org/10.3765/sp.8.12.

— (2017). 'Past Tense and Past Times in Subjunctive Conditionals: Past Tense and Past Times in Subjunctive Conditionals'. In: *Pacific Philosophical Quarterly* 98, pp. 520–535. DOI: https://doi.org/10.1111/papq.12195.

Mackie, J. L. (1972). *Truth, Probability and Paradox: Studies in Philosophical Logic*. Clarendon Library of Logic and Philosophy. Oxford: Clarendon Press.

MacQueen, J. (1967). 'Some Methods for Classification and Analysis of Multivariate Observations'. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. 14. Oakland, CA, USA, pp. 281–297.

Maechler, M. (2016). *Diptest: Hartigan's Dip Test Statistic for Unimodality – Corrected*. R package version 0.75-7.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert and K. Hornik (2019). *Cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.0.

Magnus, P. D. (2015). *For all X*. New York: University at Albany.

Mahlberg, M. and C. Smith (2012). 'Dickens, the Suspended Quotation and the Corpus'. In: *Language and Literature: International Journal of Stylistics* 21.1, pp. 51–65. DOI: https://doi.org/10.1177/0963947011432058.

Malamud, S. A. (2012). 'Impersonal Indexicals: one, you, man, and du'. In: *The Journal of Comparative Germanic Linguistics* 15.1, pp. 1–48. DOI: https://doi.org/10.1007/s10828-012-9047-6.

Marradi, A. (1990). 'Classification, Typology, Taxonomy'. In: *Quality & Quantity* 24.2, pp. 129–157.

Marsland, S. (2015). *Machine Learning: An Algorithmic Perspective*. 2nd ed. London: Chapman and Hall/CRC.

Martin, J. R. and P. R. R. White (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Mato Míguez, B. (2016). 'The Expression of Directive Meaning: A Corpus-Based Study on the Variation Between Insubordinated *if*-Clauses, Imperatives and Conditionals in Contemporary Spoken British and American English'. PhD thesis. Santiago de Compostela: Universidad de Santiago de Compostela.

Mauck, S. and P. Portner (2006). 'Review of 'Conditionals: A Comprehensive Empirical Analysis' by Renaat Declerck and Susan Reed'. In: *Lingua* 116.8, pp. 1330–1337. DOI: https://doi.org/10.1016/j.lingua.2005.03.001.

Mauri, C. and J. van der Auwera (2012). 'Connectives'. In: *The Cambridge Handbook of Pragmatics*. Ed. by K. Allan and K. M. Jaszczolt. Cambridge: Cambridge University Press, pp. 377–402. DOI: https://doi.org/10.1017/cbo9781139022453.021.

Mayes, P. (1994). 'Conditionals and the Logic of Desirability: An Interview With Noriko Akatsuka'. In: *Issues in Applied Linguistics* 5.2, pp. 449–461.

McCarthy, M. and A. O'Keeffe (2012). 'What are Corpora and how Have They Evolved?' In: *The Routledge Handbook of Corpus Linguistics*. Ed. by M. McCarthy and A. O'Keeffe. Milton Park, Abingdon; New York: Routledge, pp. 3–13.

McCready, E. (2010). 'Varieties of Conventional Implicature'. In: *Semantics and Pragmatics* 3.8. DOI: https://doi.org/10.3765/sp.3.8.

McCune, B., J. B. Grace and D. L. Urban (2002). *Analysis of Ecological Communities*. 2nd ed. Gleneden Beach, Oregon: MjM Software Design.

McDermott, M. (1999). 'Counterfactuals and Access Points'. In: *Mind* 108.430, pp. 291–334.

McEnery, T. and A. Hardie (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge; New York: Cambridge University Press.

McEvoy, P. and D. Richards (2001). 'Using log-Linear Models to Analyse Categorical Data'. In: *NT Research* 6.5, pp. 867–875.

Ménard, P. A. and A. Mougeot (2019). 'Turning Silver Into Gold: Error-Focused Corpus Reannotation With Active Learning'. In: *Proceedings of Natural Language Processing in a Deep Learning World*. Montreal, Canada: Incoma Limited, Shoumen, Bulgaria, pp. 758–767. DOI: https://doi.org/10.26615/978-954-452-056-4_088.

Mercier, H. and D. Sperber (2011). 'Why do Humans Reason? Arguments for an Argumentative Theory'. In: *Behavioral and Brain Sciences* 34.2, pp. 57–74. DOI: https://doi.org/10.1017/s0140525x10000968.

— (2019). *The Enigma of Reason*. Harvard: Harvard University Press.

Mervis, C. B. and E. Rosch (1981). 'Categorization of Natural Objects'. In: *Annual Review of Psychology* 32.1, pp. 89–115. DOI: https://doi.org/10.1146/annurev.ps.32.020181.000513.

Michaelis, L. A. and K. Lambrecht (1996). 'Toward a Construction-Based Theory of Language Function: The Case of Nominal Extraposition'. In: *Language* 72.2, pp. 215–247. DOI: https://doi.org/10.2307/416650.

Miller, G. A. and P. N. Johnson-Laird (1976). *Language and Perception*. Harvard: Belknap Press.

Miltsakaki, E., R. Prasad, A. Joshi and B. Webber (2004). 'The Penn DIscourse TReebank'. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (Lrec'04)*. Lisbon, Portugal: European Language Resources Association (ELRA), pp. 2237–2240.

Mirolo, C. and C. Izu (2019). 'An Exploration of Novice Programmers' Comprehension of Conditionals in Imperative and Functional Programming'. In: *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. Aberdeen: Acm, pp. 436–442. DOI: https://doi.org/10.1145/3304221.3319746.

Mulken, M. J. P. v. and P. J. Schellens (2012). 'Over Loodzware Bassen en Wapperende Broekspijpen. Gebruik en Perceptie van Taalintensiverende Stijlmiddelen'. In: *Tijdschrift voor Taalbeheersing* 34.1, pp. 26–53.

Murakami, H. (1987a). *Norwegian Wood*. Ed. by J. Rubin. New York: Vintage International.

— (1987b). *Norwegian Wood*. Trans. by E. Fennema. Amsterdam: Olympus, Atlas Contact.

Murtagh, F. and P. Contreras (2017). 'Algorithms for Hierarchical Clustering: An Overview, II'. In: *WIREs Data Mining and Knowledge Discovery* 7.6. DOI: https://doi.org/10.1002/widm.1219.

Murtagh, F. and P. Legendre (2014). 'Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?' In: *Journal of Classification* 31.3, pp. 274–295. DOI: https://doi.org/10.1007/s00357-014-9161-z.

Nall, S. P. and T. M. Nall (2010). 'English '*if*-clauses': A Corpus-Based Genre Analysis of their Incidence, Ordering and Pragmatic Effects.' In: *Journal of Applied Linguistics & Professional Practice* 7.2.

Neuendorf, K. A. (2017). *The Content Analysis Guidebook*. 2nd ed. Los Angeles: Sage.

Newstead, S. E. (1997). 'Conditional Reasoning With Realistic Material'. In: *Thinking & Reasoning* 3.1, pp. 49–76.

Nieuwint, P. J. G. M. (1986). 'Present and Future in Conditional Protases'. In: *Linguistics* 24.2, pp. 371–392. DOI: https://doi.org/10.1515/ling.1986.24.2.371.

— (1992). 'On Conditionals: A Study of Conditional Sentences in English and Dutch'. PhD thesis. Tilburg University.

Nivelle, N. (2008). 'Tegenfeitelijke Voorwaardelijke Zinnen in Argumentatieve Juridische Teksten'. PhD thesis. KU Leuven.

Noakes, T. (1991). *Lore of Running*. 3rd ed. Champaign, Illinois: Leisure Press.

Nolan, D. (2013). 'Why Historians (and Everyone Else) Should Care About Counterfactuals'. In: *Philosophical Studies* 163.2, pp. 317–335. DOI: https://doi.org/10.1007/s11098-011-9817-z.

Noordman, L. G. M. (1979). *Inferring From Language*. Berlin: Springer.

Nuyts, J. (2002). 'Grounding and the System of Epistemic Expressions in Dutch: A Cognitive-Functional View'. In: *Grounding*. Ed. by F. Brisard. Berlin; Boston: De Gruyter. DOI: https://doi.org/10.1515/9783110899801.433.

— (2005). 'The Modal Confusion: On Terminology and the Concepts Behind it'. In: *Modality: Studies in Form and Function*. Ed. by A. Klinge and H. H. Müller. London: Equinox, pp. 5–38.

— (2006). 'Modality: Overview and Linguistic Issues'. In: *The Expression of Modality*. Ed. by W. Frawley. The Expression of Cognitive Categories. Berlin: Mouton de Gruyter, pp. 1–26.

Nuyts, J., P. Byloo and J. Diepeveen (2010). 'On Deontic Modality, Directivity, and Mood: The case of Dutch *mogen* and *moeten*'. In: *Journal of Pragmatics* 42.1, pp. 16–34. DOI: https://doi.org/10.1016/j.pragma.2009.05.012.

Nuyts, J. and W. Vonk (1999). 'Epistemic Modality and Focus in Dutch'. In: *Linguistics* 37.4, pp. 699–737.

O'Connell, D. L. and A. J. Dobson (1984). 'General Observer-Agreement Measures on Individual Subjects and Groups of Subjects'. In: *Biometrics*, pp. 973–983.

Oaksford, M. and N. Chater (2012). *Cognition and Conditionals: Probability and Logic in Human Thinking*. Oxford: Oxford University Press, pp. 1–418.

Ohm, E. and V. A. Thompson (2004). 'Everyday Reasoning With Inducements and Advice'. In: *Thinking & Reasoning* 10.3, pp. 241–272.

Oostdijk, N. (2000). 'The Spoken Dutch Corpus. Overview and First Evaluation.' In: *LREC-2000*. Ed. by M. Gravilidou, G. Carayannis, S. Markantanatou, S. Piperidis and G. Stainhaouer. Vol. 2, pp. 887–894.

Oostdijk, N., M. Reynaert, V. Hoste and I. Schuurman (2013). 'The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch'. In: *Essential Speech and Language Technology for Dutch: Results by the Stevin Programme*. Ed. by P. Spyns and J. Odijk. Berlin: Springer, pp. 219–247. DOI: https://doi.org/10.1007/978-3-642-30910-6_13.

Over, D. E. (2017). 'Causation and the Probability of Causal Conditionals'. In: *The Oxford Handbook of Causal Reasoning*. Ed. by M. R. Waldmann. Vol. 1. Oxford: Oxford University Press, pp. 307–325. DOI: https://doi.org/10.1093/oxfordhb/9780199399550.013.3.

Over, D. E. and N. Cruz (2021). 'The Suppositional Theory of Conditionals and Rationality'. In: *The Handbook of Rationality*. Ed. by M. Knauff and W. Spohn. Cambridge: MIT Press, pp. 395–403.

Over, D. E., I. Douven and S. Verbrugge (2013). 'Scope Ambiguities and Conditionals'. In: *Thinking & Reasoning* 19.3–4, pp. 284–307. DOI: https://doi.org/10.1080/13546783.2013.810172.

Overdiep, G. S. (1937). *Stilistische Grammatica van het Moderne Nederlandsch*. Zwolle: Tjeenk-Willink.

Oyelade, J., I. Isewon, F. Oladipupo, O. Aromolaran, E. Uwoghiren, F. Ameh, M. Achas and E. Adebiyi (2016). 'Clustering Algorithms: Their Application to Gene Expression Data'. In: *Bioinformatics and Biology Insights* 10, pp. 237–253. DOI: https://doi.org/10.4137/bbi.s38316.

Paardekooper, P. C. (1950). '*Als* en *dan* bij vergelijkingen'. In: *De Nieuwe Taalgids* 43, pp. 160–167.

— (1957). 'De tijd als spraakkunstgroep in het ABN'. In: *De Nieuwe Taalgids* 50, pp. 38–45.

— (1970). 'Groter als'. In: *De Nieuwe Taalgids* 63, pp. 329–337.

— (1986). *Beknopte ABN-Syntaksis*. Eindhoven: Uitgave in eigen beheer.

Palmer, F. R. (1974). *The English Verb*. London; New York: Longman.

— (1986). *Mood and Modality*. Cambridge: Cambridge University Press.

— (2001). *Mood and Modality*. 2nd ed. Cambridge Textbooks in Linguistics. Cambridge, UK; New York: Cambridge University Press.

Pander Maat, H. L. W. and T. J. M. Sanders (2006). 'Connectives in Text'. In: *Encyclopedia of Language & Linguistics*. Elsevier, pp. 33–41. DOI: https://doi.org/10.1016/b0-08-044854-2/00500-9.

Panther, K.-U. and L. L. Thornburg (2005). 'Motivation and Convention in Some Speech act Constructions: A Cognitive Linguistic Approach'. In: *Reviewing Linguistic Thought*. Ed. by S. Marmaridou, K. Nikiforidou and E. Antonopoulou. Berlin; New York: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110920826.53.

Partee, B. H. (2007). 'Compositionality and Coercion in Semantics: The Dynamics of Adjective Meaning'. In: *Cognitive Foundations of Interpretation: Proceedings of the Colloquium, Amsterdam, 27-28 October 2004*. Ed. by G. Bouma, M. Aloni and J. Zwarts. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen, pp. 145–162.

Partee, B. H. (2010). 'Formal Semantics: Origins, Issues, Early Impact'. In: *Baltic International Yearbook of Cognition, Logic and Communication* 6.1. DOI: https://doi.org/10.4148/biyclc.v6i0.1580.

Pascual, E. (2014). *Fictive Interaction: The Conversation Frame in Thought, Language, and Discourse*. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/hcp.47.

Pauwels, A. (1953). *De plaats van het hulpwerkwoord, verleden deelwoord en infinitief in de Nederlandse bijzin*. Leuven: M. & L. Symons.

Peirce, C. S. (1933). *The Collected Papers of Charles Sanders Peirce, Vol. III: Exact Logic*. Ed. by C. Hartshorne and P. Weiss. Cambridge: Harvard University Press.

Peugh, J. L. and C. K. Enders (2004). 'Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement'. In: *Review of Educational Research* 74.4, pp. 525–556. DOI: https://doi.org/10.3102/00346543074004525.

Piwek, P. (2000). 'Imperatives, Commitment and Action: Towards a Constraint-Based Model'. In: *LDV Forum*. Vol. 17. 1/2, pp. 92–114.

Pollmann, T. (1975). 'Temporele en conditionele *als*-zinnen: Een terreinverkenning I'. In: *De Nieuwe Taalgids* 86.1, pp. 186–199.

Pollmann, T. and A. Sturm (1977). *Over Zinnen Gesproken: Termen en Begrippen van de Traditionele Grammatica*. Culemborg: Tjeenk Willink.

Poole, M. E. and T. W. Field (1976). 'A Comparison of Oral and Written Code Elaboration'. In: *Language and Speech* 19.4, pp. 305–312. DOI: https://doi.org/10.1177/002383097601900401.

Postma, G. J. (2006). 'Van 'groter dan' naar 'groter als'. Structurele oorzaken voor het verval van het comparatieve voegwoord 'dan'.' In: *Nederlandse Taalkunde* 11.1, pp. 2–22.

Postma, G. J. and H. J. Bennis (2006). 'Variatie in Negatie: Het Gedrag van het Negatieve Cliticum in het Middeldrents van Rond 1400'. In: *Taal & Tongval* 58.1, pp. 148–167.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

— (2007a). 'Into the Conventional-Implicature Dimension'. In: *Philosophy Compass* 2.4, pp. 665–679. DOI: https://doi.org/10.1111/j.1747-9991.2007.00089.x.

— (2007b). 'The Expressive Dimension'. In: *Theoretical Linguistics* 33.2. DOI: https://doi.org/10.1515/tl.2007.011.

Prasad, R., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi and B. Webber (2008). 'The Penn Discourse Treebank 2.0.' In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (Lrec'08)*. Ed. by N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis and D. Tapias. Marrakech, Morocco: European Language Resources Association (ELRA), pp. 2961–2968.

Prasad, R., E. Miltsakaki, N. Dinesh, A. Lee, A. Joshi, L. Robaldo and B. Webber (2017). *The Penn Discourse Treebank 2.0 Annotation Manual*. URL: https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf.

Predelli, S. (2009). 'Towards a Semantics for Biscuit Conditionals'. In: *Philosophical Studies* 142.3, pp. 293–305. DOI: https://doi.org/10.1007/s11098-007-9187-8.

Proeme, H. (1984). 'Over de Nederlandse Imperativus'. In: *Forum der Letteren* 25.1, pp. 241–258.

Puente-Castelo, L. (2021). '"If you Will Take the Trouble to Inquire Into it Rather Closely, I Think you Will Find That it is not Worth Very Much": Authorial Presence Through Conditionals and Citation Sequences in Late Modern English Life Sciences Texts'. In: *"All Families and Genera": Exploring the Corpus of English Life Sciences Texts*. Ed. by I. Moskowich, I. Lareo and G. Camiña. Amsterdam: John Benjamins Publishing Company, pp. 190–208. DOI: https://doi.org/10.1075/z.237.10pue.

Questier, F., B. Walczak, D. L. Massart, C. Boucon and S. de Jong (2002). 'Feature Selection for Hierarchical Clustering'. In: *Analytica Chimica Acta* 466.2, pp. 311–324.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London; New York: Longman.

Raftery, A. E. and N. Dean (2006). 'Variable Selection for Model-Based Clustering'. In: *Journal of the American Statistical Association* 101.473, pp. 168–178. DOI: https://doi.org/10.1198/016214506000000113.

Ramsey, F. P. (1950). 'General Propositions and Causality'. In: *Foundations of Mathematics and Other Logical Essays*. Ed. by F. P. Ramsey. New York, pp. 237–257.

Ramsey, V. (1987). 'The Functional Distribution of Preposed and Postposed 'if' and 'When' Clauses in Written Discourse'. In: *Coherence and Grounding in Discourse*. Ed. by R. S. Tomlin. June 1984, pp. 383–408. DOI: https://doi.org/10.1075/tsl.11.17ram.

Rebholz-Schuhman, D., A. Jimeno-Yepes, v. M., N. Kang, J. Kors, D. Milward, P. Corbett, E. Buyko, K. Tomanek, E. Beisswanger and U. Hahn (2010). 'The Calbc Silver Standard Corpus for Biomedical Named Entities: A Study in Harmonizing the Contributions From Four Independent Named Entity Taggers'. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10)*.

Recanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford: Blackwell.

Rehbein, I., M. Scholman and V. Demberg (2016). 'Annotating Discourse Relations in Spoken Language: A Comparison of the Pdtb and CCR Frameworks'. In: *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). May 23-28, 2016. Portoroŏ, Slov-*

enia. Ed. by N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk and S. Piperidis. Paris: European Language Resources Association, pp. 23–28.

Renkema, J. (2016). *Taal is taal: De interactieve 'basta'-functie van een schijnbare tautologie*. URL: https://www.janrenkema.nl/pdf/artikel/taal_is_taal.pdf.

— (2020). *Schrijfwijzer*. 6th ed. Amsterdam: Boom.

Renmans, B. and W. van Belle (2003). 'The use of the Particle *dan* in Dutch Conditional Sentences'. In: *Leuvense Bijdragen – Leuven Contributions in Linguistics and Philology* 92, pp. 141–158. DOI: https://doi.org/10.2143/lb.92.1.542037.

Rescher, N. (2007). *Conditionals*. Cambridge, Massachusetts: MIT Press.

Reuneker, A. (2016). 'Conditional use of Prepositional Phrases in Dutch: The Case of Zonder ('Without')'. In: *Linguistics in the Netherlands* 33, pp. 121–134. DOI: https://doi.org/10.1075/avt.33.09reu.

— (2017a). *Classifying Conditionals in Discourse. Presentation at* CogLing7*, Nijmegen, The Netherlands*. URL: https://www.reuneker.nl/files/papers/reuneker_2017_cogling_classifying_conditionals_in_discourse.pdf.

— (2017b). 'Sentence-medial *if*-clauses in Dutch'. In: *Nederlandse Taalkunde* 22.1, pp. 137–148. DOI: https://doi.org/10.5117/nedtaa2017.1.reun.

— (2020). 'Clause Order and Syntactic Integration Patterns in Dutch Conditionals'. In: *Linguistics in the Netherlands* 37, pp. 119–134. DOI: https://doi.org/10.1075/avt.00041.reu.

— (2022a). *Data and scripts for* Connecting Conditionals*. DataverseNL*. DOI: https://doi.org/10.34894/3QTEKH.

— (2022b). *File repository for* Connecting Conditionals. URL: https://www.reuneker.nl/dissertation.

Reuneker, A. and R. J. U. Boogaart (2013). 'De argumentatieve oriëntatie van 'het is dat''. In: *Studies in Taalbeheersing 4*. Ed. by R. J. U. Boogaart and H. Jansen. Assen: Van Gorcum, pp. 291–301.

— (2021). '*Dat als*: Een schrijfadvies en de praktijk'. In: *Wat gebeurt er in het Nederlands?! Over taal, frequentie en variatie*. Ed. by N. van der Sijs, L. Fonteyn and M. van der Meulen, pp. 177–182.

Rice, S. (2010). 'Sampling in Geography'. In: *Key Methods in Geography*. Ed. by N. Clifford, S. French and G. Valentine. Los Angeles: Sage, pp. 230–252.

Rieger, A. (2015). 'Defending a Simple Theory of Conditionals'. In: *American Philosophical Quarterly* 52.3, pp. 253–260.

Rijpma, E. and F. G. Schuringa (1972). *Nederlandse Spraakkunst*. van Bakel, Jan. Groningen: Wolters-Noordhoff.

Robertson, A. T. (1919). *A Grammar of the Greek New Testament in the Light of Historical Research*. Cambridge University Press.

Rohlf, F. J. (1970). 'Adaptive Hierarchical Clustering Schemes'. In: *Systematic Biology* 19.1, pp. 58–82.

Rooryck, J. E. C. V. and G. J. Postma (2007). 'On Participial Imperatives'. In: *Imperative Clauses in Generative Grammar: Studies in Honour of Frits Beukema*. Ed. by W. van der Wurff. Amsterdam: John Benjamins Publishing Company, pp. 273–296.

Rosch, E. (1973). 'Natural Categories'. In: *Cognitive Psychology* 4.3, pp. 328–350.

— (1975). 'Cognitive Representations of Semantic Categories'. In: *Journal of Experimental Psychology* 104, pp. 192–233.

— (1978). 'Principles of Categorization'. In: *Cognition and Categorization*. Ed. by B. B. Lloyd, J. Wiles and E. Rosch. Hillsdale, NJ: Erlbaum, pp. 27–48.

Rothschild, D. (2021). 'A Note on Conditionals and Restrictors'. In: *Conditionals, Paradox, and Probability: Themes From the Philosophy of Dorothy Edgington*. Ed. by L. Walters and J. Hawthorne. Oxford University Press, pp. 19–39. DOI: https://doi.org/10.1093/oso/9780198712732.003.0003.

Rowley-Jolivet, E. and S. Carter-Thomas (2008). 'When Practice Belies 'Theory': Form, Function and Frequency of *if*-Conditionals in Specialised Discourse'. In: *ASp* 53-54, pp. 39–61. DOI: https://doi.org/10.4000/asp.343.

Rühlemann, C. and K. Aijmer (2015). 'Corpus Pragmatics: Laying the Foundations'. In: *Corpus Pragmatics*. Ed. by K. Aijmer and C. Rühlemann. Cambridge: Cambridge University Press, pp. 1–26. DOI: https://doi.org/10.1017/cbo9781139057493.001.

Russell, B. (1903). *The Principles of Mathematics*. 2nd ed. Berlin: W. W. Norton & Company.

— (1905). 'On Denoting'. In: *Mind* 14.56, pp. 479–493.

Sadock, J. M. (1978). 'On Testing for Conversational Implicature'. In: *Pragmatics*. Ed. by P. Cole. Syntax and Semantics. New York: Academic Press, pp. 281–297.

Saeed, J. (2011). 'Pragmatics and Semantics'. In: *Foundations of Pragmatics*. Ed. by W. Bublitz and N. R. Norrick. Berlin; New York: De Gruyter Mouton.

Sagan, C. (1977). *The Dragons of Eden: Speculations on the Evolution of Human Intelligence*. New York: Ballantine Books.

Saldanha, E.-A. D., S. Hölldobler and I. L. Rocha (2017). 'Obligation Versus Factual Conditionals Under the Weak Completion Semantics'. In: *Proceedings of the Second Young Scientist's International Workshop on Trends in Information Processing*, pp. 55–64.

Salmon, N. (2005). 'Two Conceptions of Semantics'. In: *Semantics Versus Pragmatics*. Ed. by Z. Szábo. Oxford: Oxford University Press, pp. 317–328.

Sanders, J. and K. van Krieken (2019). 'Traveling Through Narrative Time: How Tense and Temporal Deixis Guide the Representation of Time and Viewpoint in News Narratives'. In: *Cognitive Linguistics* 30.2, pp. 281–304. DOI: https://doi.org/10.1515/cog-2018-0041.

Sanders, T. J. M. and W. P. M. S. Spooren (2007). 'Discourse and Text Structure'. In: *Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, pp. 916–941.

Sanders, T. J. M., W. P. M. S. Spooren and L. G. M. Noordman (1992). 'Toward a Taxonomy of Coherence Relations'. In: *Discourse Processes* 15.1, pp. 1–35. DOI: https://doi.org/10.1080/01638539209544800.

Sanders, T. J. M. and N. Stukker (2012). 'Causal Connectives in Discourse: A Cross-Linguistic Perspective'. In: *Journal of Pragmatics* 44.2, pp. 131–137. DOI: https://doi.org/10.1016/j.pragma.2011.12.007.

Sanders, T. J. M. and E. E. Sweetser (2009). 'Causality in Language and Cognition – What Causal Connectives and Causal Verbs Reveal About the way we Think'. In: *Causal Categories in Discourse and Cognition*. Ed. by T. J. M. Sanders and E. E. Sweetser. Berlin; New York: Walter de Gruyter, pp. 1–18. DOI: https://doi.org/10.1515/9783110224429.

Sandri, G. (1969). 'On the Logic of Classification'. In: *Quality and Quantity* 3.1–2, pp. 80–124. DOI: https://doi.org/10.1007/bf00166506.

Sanford, D. H. (1989). *If P, Then Q: Conditionals and the Foundations of Reasoning*. The Problems of Philosophy. London; New York: Routledge.

Sankoff, G. (1989). 'A Quantitative Paradigm for the Study of Communicative Competence'. In: *Explorations in the Ethnography of Speaking*. Ed. by R. Bauman and J. Sherzer. 2nd ed. Cambridge University Press, pp. 18–49. DOI: https://doi.org/10.1017/cbo9780511611810.005.

Santorio, P. (2019). 'Interventions in Premise Semantics'. In: *Philosophers' Imprint* 19, pp. 1–27.

Scheepers, C. (2017). 'Log-Linear Analysis'. In: *The Sage Encyclopedia of Communication Research Methods*. Thousand Oaks, California: SAGE Publications, Inc, pp. 887–890. DOI: https://doi.org/10.4135/9781483381411.n309.

Scheffler, I. (1982). *Science and Subjectivity*. Indianapolis: Hackett Publishing.

Schelfhout, C., P.-A. Coppen and N. Oostdijk (2003). 'Positions of Parentheticals and Interjections: A Corpus-Based Approach'. In: *Linguistics in the Netherlands*. Ed. by L. Cornips and P. Fikkert. Vol. 20. Amsterdam: John Benjamins Publishing Company, pp. 155–166.

Schiffrin, D. (1992). 'Conditionals as Topics in Discourse'. In: *Linguistics* 30.1, pp. 165–198. DOI: https://doi.org/10.1515/ling.1992.30.1.165.

— (2006). *In Other Words: Variation in Reference and Narrative*. Studies in Interactional Sociolinguistics. Cambridge, UK; New York: Cambridge University Press.

Schilperoord, J. and A. Verhagen (1998). 'Conceptual Dependency and the Clausal Structure of Discourse'. In: *Discourse and Cognition: Bridging the Gap*, pp. 141–163. DOI: https://doi.org/10.1016/j.jaad.2013.03.040.

Schmid, H.-J. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford; New York: Oxford University Press.

Scholman, M. C. J., J. Evers-Vermeul and T. J. M. Sanders (2016). 'Categories of Coherence Relations in Discourse Annotation'. In: *Dialogue & Discourse* 7.2, pp. 1–28. DOI: https://doi.org/10.5087/dad.2016.201.

Schouten, E. M. (2000). *The Role of the Native Language in the non-Native Acquisition of Hypothetical Conditional Structures*. LOT/Netherlands Graduate School of Linguistics. The Hague: Holland Academic Graphics.

Schouten, H. J. A. (1982). 'Measuring Pairwise Interobserver Agreement When all Subjects are Judged by the Same Observers'. In: *Statistica Neerlandica* 36.2, pp. 45–61.

Schroyens, W. and W. Schaeken (2003). 'A Critique of Oaksford, Chater, and Larkin's (2000) Conditional Probability Model of Conditional Reasoning.' In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.1, pp. 140–149. DOI: https://doi.org/10.1037/0278-7393.29.1.140.

Schulz, K. (2011). '"If you'd Wiggled A, Then B Would've Changed": Causality and Counterfactual Conditionals'. In: *Synthese* 179.2, pp. 239–251. DOI: https://doi.org/10.1007/s11229-010-9780-9.

— (2014). 'Fake Tense in Conditional Sentences: A Modal Approach'. In: *Natural Language Semantics* 22.2, pp. 117–144. DOI: https://doi.org/10.1007/s11050-013-9102-0.

Schulze, R. and P. Hohaus (2020). 'Modalising Expressions and Modality: An Overview of Trends and Challenges'. In: *Studies in Language Companion Series*. Ed. by P. Hohaus and R. Schulze. Vol. 216. Amsterdam: John Benjamins Publishing Company, pp. 1–16. DOI: https://doi.org/10.1075/slcs.216.01sch.

Schwarz, G. (1978). 'Estimating the Dimension of a Model'. In: *The Annals of Statistics* 6.2, pp. 461–464. DOI: https://doi.org/10.1214/aos/1176344136.

Schwarz, S. and E. Smitterberg (2020). '"If Anyone Would Have Told me, I Would Have not Believed it": Using Corpora to Question Assumptions About Spoken vs. Written Grammar in EFL Grammars and Other Normative Works'. In: *Studies in Corpus Linguistics*. Ed. by E. Jonsson and T. Larsson. Vol. 97. Amsterdam: John Benjamins Publishing Company, pp. 284–300. DOI: https://doi.org/10.1075/scl.97.17sch.

Schweizer, B. and A. Sklar (1960). 'Statistical Metric Spaces'. In: *Pacific Journal of Mathematics* 10.1, pp. 313–334. DOI: https://doi.org/10.2140/pjm.1960.10.313.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9781139173438.

— (1975). 'Indirect Speech Acts'. In: *Syntax and Semantics*. Ed. by P. Cole and J. L. Morgan. Vol. 3. Brill, pp. 59–82. DOI: https://doi.org/10.1163/9789004368811_004.

— (1976). 'A Classification of Illocutionary Acts'. In: *Language in Society* 5.1, pp. 1–23.

— (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Sebben, S. and J. Ullrich (2021). 'Can Conditionals Explain Explanations? A Modus Ponens Model of B Because A'. In: *Cognition* 215, p. 104812. DOI: https://doi.org/10.1016/j.cognition.2021.104812.

Seuren, P. A. M. (2010). *The Logic of Language: Language from Within*. Oxford; New York: Oxford University Press.

Sevcik, R., H. Rezankova and D. Husek (2011). 'Comparison of Selected Methods for Document Clustering'. In: *Advances in Intelligent Web Mastering–3*. Springer, pp. 101–110.

Sharpe, D. (2015). 'Your chi-Square Test is Statistically Significant: Now What?' In: *Practical Assessment, Research and Evaluation* 20.8, pp. 1–10.

Siegel, M. E. A. (2006). 'Biscuit Conditionals: Quantification Over Potential Literal Acts'. In: *Linguistics and Philosophy* 29.2, pp. 167–203. DOI: https://doi.org/10.1007/s10988-006-0003-2.

Silverman, B. W. (1981). 'Using Kernel Density Estimates to Investigate Multimodality'. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 43.1, pp. 97–99. DOI: https://doi.org/10.1111/j.2517-6161.1981.tb01155.x.

Silvestre, C., M. Cardoso and M. Figueiredo (2013). 'Clustering and Selecting Categorical Features'. In: *Portuguese Conference on Artificial Intelligence*. Springer, pp. 331–342.

Singer, P. (1975). *Animal Liberation: A New Ethics for our Treatment of Animals*. New York: Harper Collins.

Skovgaard-Olsen, N. (2020). 'Relevance and Conditionals: A Synopsis of Open Pragmatic and Semantic Issues'. In: *Logic and Uncertainty in the Human Mind: A Tribute to David E. Over*. Ed. by S. Elgayam, I. Douven, N. Cruz and J. Evans. London; New York: Routledge, pp. 192–206.

Skovgaard-Olsen, N., P. Collins, K. Krzyżanowska, U. Hahn and K. C. Klauer (2019). 'Cancellation, Negation, and Rejection'. In: *Cognitive Psychology* 108, pp. 42–71. DOI: https://doi.org/10.1016/j.cogpsych.2018.11.002.

Skovgaard-Olsen, N., H. Singmann and K. C. Klauer (2016). 'The Relevance Effect and Conditionals'. In: *Cognition* 150, pp. 26–36. DOI: https://doi.org/10.1016/j.cognition.2015.12.017.

Skovgaard-Olsen, N. and P. Collins (2021). 'Indicatives, Subjunctives, and the Falsity of the Antecedent'. In: *Cognitive Science* 45.11. DOI: 10.1111/cogs.13058.

Slater, M. H. and A. Borghini (2011). 'Introduction: Lessons From the Scientific Butchery'. In: *Carving Nature at its Joints: Natural Kinds in Metaphysics and Science*. Ed. by J. K. Campbell, M. O'Rourke and M. H. Slater. Cambridge, Massachusetts: MIT Press, pp. 1–32.

Smedts, W. and W. van Belle (1993). *Taalboek Nederlands*. 4th ed. Kapellen: Pelckmans.

Smessaert, H. (1999). *Perspectief en Vergelijking: Aspectuele Partikels in het Nederlands*. Leuven: Belgisch interuniversitair centrum voor neerlandistiek/Peeters.

Smollett, R. (2005). 'Quantized Direct Objects Don't Delimit After All: A Revised Account of the Role of Quantized Direct Objects in Aspectual Composition'. In: *Perspectives on Aspect*. Ed. by H. J. Verkuyl, H. de Swart and A. van Hout. Dordrecht: Springer, pp. 41–59. DOI: https://doi.org/10.1007/1-4020-3232-3_3.

Smullyan, R. M. (1978). *What is the Name of This Book? The Riddle of Dracula and Other Logical Puzzles*. Englewood Cliffs: Prentice-Hall.

Smyth, H. W. (1920). *A Greek Grammar for Colleges*. New York: American Book Company.

Sneath, P. H. A. and R. R. Sokal (1973). *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. A series of Books in Biology. San Francisco: W. H. Freeman.

Solorio-Fernández, S., J. A. Carrasco-Ochoa and J. F. Martínez-Trinidad (2020). 'A Review of Unsupervised Feature Selection Methods'. In: *Artificial Intelligence Review* 53.2, pp. 907–948. DOI: https://doi.org/10.1007/s10462-019-09682-y.

Sonnenschein, E. A. (1892). *A Greek Grammar for Schools Based on the Principles and Requirements of the Grammatical Society*. London: S. Sonnenschein & Co.

Sorensen, R. (2012). 'Lying With Conditionals'. In: *The Philosophical Quarterly* 62.249, pp. 820–832.

Spärck Jones, K. (1972). 'A Statistical Interpretation of Term Specificity and its Application in Retrieval'. In: *Journal of Documentation* 28, pp. 11–21.

Sperber, D. and D. Wilson (1986). *Relevance: Communication and Cognition*. Cambridge, Massachusetts: Harvard University Press.

Spooren, W. P. M. S. (2004). 'On the use of Discourse Data in Language use Research'. In: *Words in Their Places: A Festschrift for J. Lachlan Mackenzie*. Ed. by M. Hannay, H. Aertsen and R. J. Lyall. Amsterdam: Faculty of Arts, Vrije Universiteit Amsterdam, pp. 381–393.

Spooren, W. P. M. S. and L. Degand (2010). 'Coding Coherence Relations: Reliability and Validity'. In: *Corpus Linguistics and Linguistic Theory* 6.2. DOI: https://doi.org/10.1515/cllt.2010.009.

Stalnaker, R. C. (1968). 'A Theory of Conditionals'. In: *Ifs*. Ed. by W. L. Harper, R. C. Stalnaker and G. Pearce. Dordrecht: Springer, pp. 41–55. DOI: https://doi.org/10.1007/978-94-009-9117-0_2.

— (1972). 'Pragmatics'. In: *Semantics of Natural Language*. Ed. by D. Davidson and G. Harman. Dordrecht: Reidel, pp. 389–408.

— (1974). 'Pragmatic Presuppositions'. In: *Semantics and Philosophy*. Ed. by M. K. Munitz and P. Unger. New York: New York University Press, pp. 197–213.

— (1975). 'Indicative Conditionals'. In: *Philosophia* 5.3, pp. 269–286. DOI: https://doi.org/10.1007/bf02379021.

— (2002). 'Common Ground'. In: *Linguistics and Philosophy* 25.5/6, pp. 701–721.

Stalnaker, R. C. (2019). *Knowledge and Conditionals: Essays on the Structure of Inquiry*. Oxford: Oxford University Press.

Stefanowitsch, A. (2003). 'A Construction-Based Approach to Indirect Speech Acts'. In: *Metonymy and Pragmatic Inferencing*. Ed. by K.-U. Panther and L. L. Thornburg. Amsterdam: John Benjamins Publishing Company, pp. 105–126. DOI: https://doi.org/10.1075/pbns.113.09ste.

Stone, M. A. (2012). 'Denying the Antecedent: Its Effective Use in Argumentation'. In: *Informal Logic* 32.3, pp. 327–356. DOI: https://doi.org/10.22329/il.v32i3.3681.

Strawson, P. F. (1952). *Introduction to Logical Theory*. New York: John Wiley & Sons.

Strijbos, J.-W. and G. Stahl (2007). 'Methodological Issues in Developing a Multi-Dimensional Coding Procedure for Small-Group Chat Communication'. In: *Learning and Instruction* 17.4, pp. 394–404. DOI: https://doi.org/10.1016/j.learninstruc.2007.03.005.

Stroop, J. P. A. (2011). *Hun Hebben de Taal Verkwanseld: Over Polderneder-lands, "Fout" Nederlands en ABN*. Amsterdam: Atheneum.

Stukker, N., T. J. M. Sanders and A. Verhagen (2008). 'Causality in Verbs and in Discourse Connectives: Converging Evidence of Cross-Level Parallels in Dutch Linguistic Categorization'. In: *Journal of Pragmatics* 40.7, pp. 1296–1322.

Stukker, N. and A. Verhagen (2019). *Stijl, Taal en Tekst: Stilistiek op Taalkundige Basis*. Leiden: Leiden University Press.

Šulc, Z. (2016). 'Similarity Measures for Nominal Data in Hierarchical Clustering'. PhD thesis. Prague: University of Economics. DOI: https://doi.org/10.13140/RG.2.2.33118.02882.

Šulc, Z., J. Cibulkova and H. Řezanková (2020). *Nomclust: Hierarchical Cluster Analysis of Nominal Data*. R package version 2.1.4.

Šulc, Z. and H. Řezanková (2015). 'Nomclust: An R Package for Hierarchical Clustering of Objects Characterized by Nominal Variables'. In: *The 9th International Days of Statistics and Economics, Prague, September 10-12, 2015*, pp. 1581–1590.

— (2019). 'Comparison of Similarity Measures for Categorical Data in Hierarchical Clustering'. In: *Journal of Classification* 36.1, pp. 58–72. DOI: https://doi.org/10.1007/s00357-019-09317-5.

Sweetser, E. E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge; New York: Cambridge University Press. DOI: https://doi.org/10.1017/cbo9780511620904.

— (1996). 'Reasoning, Mapping and Meta-Metaphorical Conditionals'. In: *Grammatical Constructions: Their Form and Meaning*. Ed. by S. A. Thompson and S. Masayoshi. Oxford: Oxford University Press, pp. 221–233.

Szabó, Z. G. (2005). *Semantics vs. Pragmatics*. Oxford: New York: Clarendon Press; Oxford University Press.

Sztencel, M. (2018). *Semantics, Pragmatics and Meaning Revisited*. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-69116-9.

Sztencel, M. and L. Clarke (2018). 'Deontic Commitments in Conditional Promises and Threats: Towards an Exemplar Semantics for Conditionals'. In: *Language and Cognition* 10.3, pp. 435–466. DOI: https://doi.org/10.1017/langcog.2018.10.

Sztencel, M. and S. E. Duffy (2019). 'Reporting Conditionals With Modals'. In: *Indirect Reports and Pragmatics in the World Languages*. Ed. by A. Capone, M. García-Carpintero and A. Falzone. Vol. 19. Cham: Springer International Publishing, pp. 201–226.

Tabakhi, S., P. Moradi and F. Akhlaghian (2014). 'An Unsupervised Feature Selection Algorithm Based on ant Colony Optimization'. In: *Engineering Applications of Artificial Intelligence* 32, pp. 112–123.

Talavera, L. (2005). 'An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering'. In: *International Symposium on Intelligent Data Analysis*. Springer, pp. 440–451.

Tang, X. (2017). 'Lexeme-Based Collexeme Analysis With Depcluster'. In: *Corpus Linguistics and Linguistic Theory* 13.1. DOI: https://doi.org/10.1515/cllt-2015-0007.

Taylor, J. R. (2001). 'Prototype Theory in Linguistics'. In: *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, pp. 286–289. DOI: https://doi.org/10.1016/b978-0-08-097086-8.57034-2.

— (2003). *Linguistic Categorization*. 3rd ed. New York: Oxford University Press.

te Winkel, L. A. (1866). 'Over de wijzen en tijden der werkwoorden'. In: *De Taalgids* 8, pp. 66–75.

Tellings, J. L. (2016). 'Counterfactuality in Discourse'. PhD thesis. Los Angeles: University of California Los Angeles.

— (2017). '*Still* as an Additive Particle in Conditionals'. In: *Semantics and Linguistic Theory* 27, pp. 1–21. DOI: https://doi.org/10.3765/salt.v27i0.4117.

— (2020). 'Conditional Answers and the Role of Probabilistic Epistemic Representations'. In: *Proceedings of the Probability and Meaning Conference (PaM 2020)*. Gothenburg: Association for Computational Linguistics, pp. 26–33.

Terkourafi, M. (2009a). 'Conventional Meaning'. In: *Key Ideas in Linguistics and the Philosophy of Language*. Ed. by S. Chapman and C. Routledge. Edinburgh: Edinburgh University Press, pp. 28–29.

— (2009b). 'Conventionality'. In: *The Pragmatics Encyclopedia*. Ed. by L. Cummings. New York: Routledge, pp. 79–81.

Thompson, V. A., J. S. B. T. Evans and S. J. Handley (2005). 'Persuading and Dissuading by Conditional Argument'. In: *Journal of Memory and Language* 53.2, pp. 238–257. DOI: https://doi.org/10.1016/j.jml.2005.03.001.

Thumm, M. (2000). 'The Contextualization of Paratactic Conditionals'. In: *Interaction and Linguistic Structures*. InLiSt 20.

Tietjen, G. L. and R. H. Moore (1972). 'Some Grubbs-Type Statistics for the Detection of Several Outliers'. In: *Technometrics* 14.3, pp. 583–597.

Tiryakian, E. A. (1968). 'Typologies'. In: *International Encyclopedia of the Social Sciences*. Ed. by D. L. Sills. Vol. 16. New York: The Macmillan Company & The Free Press, pp. 177–186.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Studies in Corpus Linguistics. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, Massachusetts; London, England: Harvard University Press.

Toulmin, S. (2003). *The Uses of Argument*. Updated edition. Cambridge; New York: Cambridge University Press.

Traugott, E. C. (1985). 'Conditional Markers'. In: *Proceedings of a Symposium on Iconicity in Syntax, Stanford 1983*. Ed. by J. Haiman. Vol. 6. Amsterdam: John Benjamins Publishing Company, pp. 289–310. DOI: https://doi.org/10.1075/tsl.6.14clo.

— (1989). 'On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change'. In: *Language* 65.1, pp. 31–55. DOI: https://doi.org/10.2307/414841.

— (1999). 'The Role of Pragmatics in a Theory of Semantic Change'. In: *Pragmatics in 1998: Selected Papers From the6th International Pragmatics Conference*. Ed. by J. Verschueren. Vol. 2. Antwerp: International Pragmatics Association, pp. 93–102.

Traugott, E. C. and E. König (1991). 'The Semantics-Pragmatics of Grammaticalization Revisited'. In: *Approaches to Grammaticalization*. Ed. by E. C. Traugott and B. Heine. Amsterdam: John Benjamins Publishing Company, pp. 189–218.

Traugott, E. C., A. G. B. ter Meulen, J. S. Reilly and C. A. Ferguson (1986). *On Conditionals*. Cambridge: Cambridge University Press.

Trnavac, R. and M. Taboada (2021). 'Engagement and Constructiveness in Online News Comments in English and Russian'. In: *Text & Talk* 41.5, p. 000010151520200171. DOI: 10.1515/text-2020-0171.

Tyler, A. and V. Evans (2003). *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition*. 1st ed. Cambridge University Press. DOI: 10.1017/CBO9780511486517.

Tyler, C. L. and M. Kowalewski (2014). 'Utility of Marine Benthic Associations as a Multivariate Proxy of Paleobathymetry: A Direct Test From Recent Coastal Ecosystems of North Carolina'. In: *PLoS ONE* 9.4. Ed. by K. Swadling, e95711. DOI: https://doi.org/10.1371/journal.pone.0095711.

van Bart, P. and A. Sturm (1987). *Zinsanalyse en de termen die daarbij gebruikt worden*. Leiden: M. Nijhoff.

van Belle, W. (2003). *Zwijgen is niet altijd toestemmen: De rol van inferenties bij het interpreteren en argumenteren*. Leuven; Leusden: Acco.

van Belle, W. and I. Canegem-Ardijns (2007). 'Review of 'Mental Spaces in Grammar: Conditional Constructions' by Barbara Dancygier and Eve Sweetser'. In: *Linguistics* 5.4, pp. 824–833.

van Beveren, A., T. Colleman and G. de Sutter (2018). 'De *om*-alternantie: Een verkennende casestudy'. In: *Handelingen: Koninklijke Zuidnederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 71, pp. 191–222.

van den Berge, M. J. C., R. H. Free, R. Arnold, E. de Kleine, R. Hofman, J. M. C. van Dijk and van Dijk, Pim (2017). 'Cluster Analysis to Identify Possible Subgroups in Tinnitus Patients'. In: *Frontiers in Neurology* 8, pp. 1–7. DOI: https://doi.org/10.3389/fneur.2017.00115.

van den Toorn, M. C. (1984). *Nederlandse Grammatica*. 9th ed. Groningen: Wolters-Noordhoff.

van der Auwera, J. (1985). 'Only if'. In: *Logique et Analyse* 28.109, pp. 61–74.

— (1986). 'Conditionals and Speech Acts'. In: *On Conditionals*. Ed. by E. C. Traugott, A. G. B. ter Meulen, J. Snitzer Reilly and C. A. Ferguson. Cambridge: Cambridge University Press, pp. 197–214.

— (1993). '"Already" and "Still": Beyond Duality'. In: *Linguistics and Philosophy* 16.6, pp. 613–653.

— (1997). 'Conditional Perfection'. In: *On Conditionals Again*. Ed. by A. Athanasiadou and R. Dirven. Amsterdam: John Benjamins Publishing Company, pp. 169–190. DOI: https://doi.org/10.1075/cilt.143.10auw.

van der Auwera, J. and V. Gast (2010). 'Categories and Prototypes'. In: *The Oxford Handbook of Linguistic Typology*. Ed. by J. J. Song. Oxford: Oxford University Press, pp. 166–189. DOI: https://doi.org/10.1093/oxfordhb/9780199281251.013.0010.

van der Horst, J. (1995). *Analytische taalkunde*. Groningen: Nijhoff.

— (2010). *Met (het) oog op morgen: Opstellen over taal, taalverandering en standaardtaal*. Leuven University Press. DOI: https://doi.org/10.2307/j.ctt9qdwvp.

van der Wouden, T. (1994). 'Negative Contexts'. PhD thesis. Groningen: University of Groningen.

— (1995). 'De reflectie van logische relaties in woordstructuur: De distributie van het prefix 'on''. In: *TABU* 1, pp. 61–71.

— (1996). 'Hoeven'. In: *TABU* 26, pp. 164–182.

— (1998). 'Dat had niet zo gehoeven: Modaliteit en negatie in de nieuwe ANS'. In: *Nederlandse Taalkunde* 3, pp. 237–252.

— (2000). 'Focus on Appendices in Dutch'. In: *Linguistics in the Netherlands* 17, pp. 233–244. DOI: https://doi.org/10.1075/avt.17.22wou.

— (2002). 'Partikels: Naar een partikelwoordenboek voor het Nederlands'. Dutch. In: *Nederlandse Taalkunde* 1, pp. 20–43.

— (2018). 'Eindelijk alles over *nee*?' In: *Onze Taal* 87.5, p. 17.

van der Wouden, T. and J. Caspers (2010). 'Nederlandse partikelbeschrijving in internationaal perspectief: Waar zijn we eigenlijk en waar moet het toch naar toe?' In: *Internationale Neerlandistiek* 48.2, pp. 52–62. DOI: https://doi.org/10.5117/ivn2010.2.derw.

van der Wouden, T. and A. Foolen (2011). 'Pragmatische partikels in de rechterperiferie'. In: *Nederlandse Taalkunde* 16.3, pp. 307–322.

— (2015). 'Dutch Particles in the Right Periphery'. In: *Final Particles*. Ed. by S. Hancil, A. Haselow and M. Post. Berlin: Mouton de Gruyter, pp. 221–247.

van der Wurff, W. (2007). 'Imperative Clauses in Generative Grammar: An Introduction'. In: *Imperative Clauses in Generative Grammar: Studies in Honour of Frits Beukema*. Ed. by W. van der Wurff. Amsterdam: John Benjamins Publishing Company, pp. 1–94. DOI: https://doi.org/10.1075/la.103.02wur.

van Dijk, T. A. (1979). 'Pragmatic Connectives'. In: *Journal of Pragmatics* 3.5, pp. 447–456. DOI: https://doi.org/10.1016/0378-2166(79)90019-5.

van Duijn, M. and A. Verhagen (2019). 'Recursive Embedding of Viewpoints, Irregularity, and the Role for a Flexible Framework'. In: *Pragmatics* 29.2, pp. 198–225. DOI: https://doi.org/10.1075/prag.18049.van.

van Eemeren, F. H. (2010). *Strategic Maneuvering in Argumentative Discourse: Extending the Pragma-Dialectical Theory of Argumentation*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

van Eemeren, F. H., B. Garssen, E. C. W. Krabbe, A. F. Snoeck Henkemans, B. Verheij and J. H. M. Wagemans (2014). *Handbook of Argumentation Theory*. Dordrecht: Springer.

van Eemeren, F. H., R. Grootendorst and A. F. Snoeck Henkemans (2002). *Argumentation: Analysis, Evaluation, Presentation*. Mahwah: Lawrence Earlbaum Associates.

van Eemeren, F. H. and A. F. Snoeck Henkemans (2017). *Argumentation: Analysis and Evaluation*. Second edition. New York: Routledge.

van Eerten, L. (2007). 'Over het Corpus Gesproken Nederlands'. In: *Nederlandse Taalkunde* 12.3, pp. 194–215.

van Gorp, B. (2007). 'The Constructionist Approach to Framing: Bringing Culture Back in'. In: *Journal of Communication* 57.1, pp. 60–78.

van Haaften, T. (1991). 'De interpretatie van verzwegen subjecten'. PhD thesis. Amsterdam: Vrije Universiteit Amsterdam.

van Haaften, T. and M. van Leeuwen (2021). 'On the Relation Between Argumentative Style and Linguistic Style: Integrating Linguistic-Stylistic Analysis Systematically Into the Analysis of Argumentative Style'. In: *Journal of Argumentation in Context* 10.1, pp. 97–120. DOI: https://doi.org/10.1075/jaic.20014.haa.

van Haeringen, C. B. (1946). *Spanningen in Hedendaags Nederlands*. 's-Gravenhage: D. A. Daamen.

van Leeuwen, M. (2012). 'Rhetorical Effects of Grammar'. In: *Critical Approaches to Discourse Analysis across Disciplines* 5.2, pp. 88–101.

van Olmen, D. (2011). 'The Imperative in English and Dutch: A Functional Analysis in Comparable and Parallel Corpora'. PhD thesis. Universiteit Antwerpen.

van Olmen, D. and S. Heinold (2017). 'Imperatives and Directive Strategies From a Functional-Typological Perspective: An Introduction'. In: *Imperatives and Directive Strategies*. Ed. by D. van Olmen and S. Heinold. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/slcs.184.01van.

van Rooij, R. and K. Schulz (2019). 'Conditionals, Causality and Conditional Probability'. In: *Journal of Logic, Language and Information* 28.1, pp. 55–71. DOI: https://doi.org/10.1007/s10849-018-9275-5.

— (2020). 'Conditionals As Representative Inferences'. In: *Axiomathes*. DOI: https://doi.org/10.1007/s10516-020-09477-9.

van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley: CreateSpace.

Vandergriff, I. (2009). 'Conditionals and Mental Space Set-Up: Evidence from German Word Order'. In: *Constructional Approaches to Language*. Ed. by A. Bergs and G. Diewald. Vol. 9. Amsterdam: John Benjamins Publishing Company, pp. 193–212. DOI: https://doi.org/10.1075/cal.9.08van.

Vandeweghe, W. (1992). *Perspectivische Evaluatie in het Nederlands: De Partikels van de* al/nog/pas-*Groep*. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.

— (2000). *Grammatica van de Nederlandse zin*. Leuven: Garant.

Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.

Verbrugge, S., K. Dieussaert, W. Schaeken, H. Smessaert and W. van Belle (2007). 'Pronounced Inferences: A Study on Inferential Conditionals'. In: *Thinking & Reasoning* 13.2, pp. 105–133. DOI: https://doi.org/10.1080/13546780600713227.

Verbrugge, S., K. Dieussaert, W. Schaeken and W. van Belle (2004). 'Promise is Debt, Threat Another Matter: The Effect of Credibility on the Interpretation of Conditional Promises and Threats'. In: *Canadian Journal of Experimental Psychology* 58.2, pp. 106–112. DOI: https://doi.org/10.1037/h0085801.

Verbrugge, S. and H. Smessaert (2010). 'On the Argumentative Strength of Indirect Inferential Conditionals'. In: *Argumentation* 24.3, pp. 337–362. DOI: https://doi.org/10.1007/s10503-010-9179-2.

— (2011). 'On the Distinction Between Inferential and Meta-Inferential Conditionals in Dutch'. In: *Journal of Pragmatics* 43.14, pp. 3387–3402. DOI: https://doi.org/10.1016/j.pragma.2011.07.008.

Verhagen, A. (1992). 'Praxis of Linguistics: Passives in Dutch'. In: *Cognitive Linguistics* 3.3, pp. 301–342.

— (2000). 'Interpreting Usage: Construing the History of Dutch Causal Verbs'. In: *Usage-Based Models of Language*. Ed. by M. Barlow and S. Kemmer. Stanford, California: CSLI Publications, pp. 261–286.

— (2003). 'The Dutch way'. In: *Usage-Based Approaches to Dutch: Lexicon, Grammar, Discourse*. Ed. by A. Verhagen and J. van de Weijer, pp. 27–56.

Verhagen, A. (2005). *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. New York: Oxford University Press.
— (2007). 'English Constructions From a Dutch Perspective: Where are the Differences?' In: *Structural-Functional Studies in English Grammar: In Honour of Lachlan Mackenzie*. Ed. by M. Hannay and G. J. Steen. Amsterdam: John Benjamins Publishing Company, pp. 257–274. DOI: https://doi.org/10.1075/slcs.83.15ver.
— (2009). 'The Conception of Constructions as Complex Signs: Emergence of Structure and Reduction to Usage'. In: *Constructions and Frames* 1.1, pp. 119–152. DOI: https://doi.org/10.1075/cf.1.1.06ver.
— (2019). '50 Years of Dichotomies: Misunderstandings and the Relevance of Lexical Semantics'. In: *Linguistics in the Netherlands* 36, pp. 60–66. DOI: https://doi.org/10.1075/avt.00023.ver.
— (2021). 'Argumentation in Evolution: On the Relation Between Dyadic and Public Communication'. In: *The Language of Argumentation*. Ed. by R. J. U. Boogaart, H. Jansen and M. van Leeuwen. Cham: Springer International Publishing, pp. 37–60. DOI: https://doi.org/10.1007/978-3-030-52907-9_3.
— (forthcoming). 'Perspective Management and Iconicity: Depiction and Description in Construing Other Minds'.
Verhagen, V. and M. Mos (2016). 'Stability of Familiarity Judgments: Individual Variation and the Invariant Bigger Picture'. In: *Cognitive Linguistics* 27.3, pp. 307–344. DOI: https://doi.org/10.1515/cog-2015-0063.
Verkuyl, H. J. (1972). *On the Compositional Nature of the Aspects*. Dordrecht: Springer. DOI: https://doi.org/10.1007/978-94-017-2478-4.
— (1986). 'Vendler-klassen moeten worden genegeerd'. In: *TABU* 16, pp. 69–84.
— (1989). 'Aspectual Classes and Aspectual Composition'. In: *Linguistics and Philosophy* 12.1, pp. 39–94.
— (1993). *A Theory of Aspectuality: The Interaction Between Temporal and Atemporal Structure*. Cambridge: Cambridge University Press.
— (2008). *Binary Tense*. Csli Lecture Notes. Stanford, CA: CSLI Publications.
Verkuyl, H. J. and H. Broekhuis (2013a). 'Temporaliteit en Modaliteit'. In: *Nederlandse Taalkunde* 18.3, pp. 306–323. DOI: https://doi.org/10.5117/nedtaa2013.3.verk.
— (2013b). 'Temporaliteit én Modaliteit: een Repliek'. In: *Nederlandse Taalkunde* 18.3, pp. 339–345. DOI: https://doi.org/10.5117/nedtaa2013.3.ver2.
Vermunt, J. K. and J. Magidson (2002). 'Latent Class Cluster Analysis'. In: *Applied Latent Class Analysis*. Ed. by J. Hagenaars and A. McCutcheon. Cambridge: Cambridge University Press, pp. 89–106.
von Fintel, K. (1994). 'Restrictions on Quantifier Domains'. PhD thesis. Amherst: University of Massachusetts.

— (2011). 'Conditionals'. In: *Semantics: An International Handbook of Natural Language Meaning*. Ed. by K. von Heusinger, C. Maienborn and P. Portner. Vol. 2. Berlin; Boston: De Gruyter, pp. 1515–1537. DOI: https://doi.org/10.1515/9783110255072.1515.
— (2012). 'Subjunctive Conditionals'. In: *The Routledge Companion to Philosophy of Language*. Ed. by G. Russell and D. Graff Fara. New York: Routledge, pp. 466–477.
von Prince, K. (2019). 'Counterfactuality and Past'. In: *Linguistics and Philosophy* 42.6, pp. 577–615. DOI: https://doi.org/10.1007/s10988-019-09259-6.
von Wright, G. H. (1973). 'On the Logic and Epistemology of the Causal Relation'. In: *Studies in Logic and the Foundations of Mathematics* 74, pp. 293–312. DOI: https://doi.org/10.1016/s0049-237x(09)70366-0.
Wälchli, B. (2018). ''As Long as', 'Until' and 'Before' Clauses'. In: *Baltic Linguistics* 9, pp. 141–236.
Wallwork, A. (2018). *Top 50 Grammar Mistakes*. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-70984-0.
Walters, L. (2021). 'Introduction'. In: *Conditionals, Paradox, and Probability: Themes From the Philosophy of Dorothy Edgington*. Ed. by L. Walters and J. Hawthorne. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198712732.001.0001.
Ward, J. H. J. (1963). 'Hierarchical Grouping to Optimize an Objective Function'. In: *Journal of the American Statistical Association* 58.301, pp. 236–244. DOI: https://doi.org/10.1080/01621459.1963.10500845.
Wason, P. C. (1968). 'Reasoning About a Rule'. In: *Quarterly Journal of Experimental Psychology* 20.3, pp. 273–281. DOI: https://doi.org/10.1080/14640746808400161.
Weber, M. (1949). *The Methodology of the Social Sciences*. New York: Free Press.
Weirich, K. M. (2015). 'Truth and Consequence: Toward a New Account of Indicative Conditionals'. PhD thesis. University of Colorado at Boulder.
Whitehead, A. N. and B. Russell (1910). *Principia Mathematica*. Cambridge: Cambridge University Press.
Wiebe, J., R. Bruce and T. P. O'Hara (1999). 'Development and use of a Gold-Standard Data set for Subjectivity Classifications'. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. Maryland: Association for Computational Linguistics, pp. 246–253.
Wierzbicka, A. (1988). *The Semantics of Grammar*. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/slcs.18.
— (1997). 'Conditionals and Counterfactuals: Conceptual Primitives and Linguistic Universals'. In: *On Conditionals Again*. Ed. by A. Athanasiadou and R. Dirven. Amsterdam: John Benjamins Publishing Company, pp. 15–60. DOI: https://doi.org/10.1075/cilt.143.04wie.
Wilcox, A. R. (1973). 'Indices of Qualitative Variation and Political Measurement'. In: *The Western Political Quarterly* 26.2, pp. 325–343. DOI: https://doi.org/10.2307/446831.

Willer, M. (2022). 'Negating Conditionals'. In: *Oxford Studies in Philosophy of Language*. Ed. by E. Lepore and D. Sosa. Vol. 2. Oxford: Oxford University Press. Forthcoming.

Williamson, T. (2020). *Suppose and Tell: the Semantics and Heuristics of Conditionals*. New York: Oxford University Press.

Wilson, D. (1975). *Presupposition and Non-truth-conditional Semantics*. Cambridge, Massachusetts: Academic Press.

Wittgenstein, L. (1958). *Philosophical Investigations*. Ed. by G. E. M. Anscombe. 3rd ed. Oxford: Basil Blackwell.

Wright, P. and A. J. Hull (1986). 'Answering Questions About Negative Conditionals'. In: *Journal of Memory and Language* 25.6, pp. 691–709. DOI: https://doi.org/10.1016/0749-596x(86)90044-6.

Wright, W. (1888). 'The Empire of the Hittites'. In: *Journal of the Transactions of The Victoria Institute, or Philosophical Society of Great Britain* 21, pp. 55–72.

Xiaofei, H., C. Deng and N. Partha (2005). 'Laplacian Score for Feature Selection'. In: *Advances in Neural Information Processing Systems*, pp. 507–514.

Zakkou, J. (2017). 'Biscuit Conditionals and Prohibited 'Then''. In: *Thought: A Journal of Philosophy* 6.2, pp. 84–92. DOI: https://doi.org/10.1002/tht3.235.

Zevakhina, N. and V. Prigorkina (2020). 'Quantity Inferences in Conditionals: A Pilot Experimental Study'. In: *Proceedings of 11th International Conference of Experimental Linguistics*. ExLing Society. DOI: https://doi.org/10.36505/ExLing-2020/11/0058/000473.

Zufferey, S., J. Moeschler and A. Reboul (2019). *Implicatures*. New York: Cambridge University Press.

Zumel, N. and J. Mount (2020). *Practical Data Science With R*. New York: Manning Publications.

Zwaan, F. L. (1968). 'De "tussenzin"'. In: *Taalkunde in Artikelen*. Ed. by J. Hoogteijling. Groningen: Wolters-Noordhoff, pp. 360–362.

Zwart, J.-W. (2011). *The Syntax of Dutch*. Cambridge; New York: Cambridge University Press.

Zwarts, F. (1981). 'Negatief Polaire Uitdrukkingen I'. In: *Glot* 4, pp. 35–132.