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

Reuneker, A.

CHAPTER 6

Clusters of conditionals

6.1 Introduction

In chapter 2, I discussed two meaning aspects of conditionals, namely one of unassertiveness, and one of connectedness. Although both meanings were analysed as being conventionally attached to the form of conditionals, i.e., as conventional meanings, their more specific interpretations were analysed as (partly) contextually determined, i.e., as conversational implicatures. The literature discussed in chapter 3 provides suggestions for relations between these implicatures and grammatical features of conditionals, such as verb tense, modality, and clause order. Therefore, the types of unassertiveness and connectedness were hypothesised to be generalised to a certain extent, because generalised conversational implicatures are assumed to be default interpretations for a specific grammatical form. As Grice (1989, p. 37) argues, 'the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature'. Such generalised conversational implicatures are thus licensed in the majority of cases in which there is 'an absence of information on the contrary', i.e., a 'default inference associated with specific kinds of linguistic expression'. (Levinson, 2000, p. 59; see also Birner, 2013, p. 37; Ariel, 2010, p. 20). As I argued for briefly already in section 4.3, this is the link between the pragmatic analysis in the first part of this dissertation, and the data-driven, constructional account in the second part. In this chapter, we will move beyond the conjunction *als* 'if' in isolation, and we will attempt to answer the question to what extent the linguistic fea-

tures of conditionals indeed license generalised conversational implicatures of unassertiveness and connectedness, and to which extent they can be viewed as grammatical constructions (i.e., pairings of meaning and form).

In this chapter, I address the question to what extent the linguistic features of conditionals license specific implicatures of unassertiveness and connectedness. To explore and assess systematic relations between grammatical features of conditionals, i.e., their form, and the more specific types of unassertiveness and connectedness, i.e., their meaning, we will subject the feature distributions of Dutch conditionals presented in chapter 5 to a number of bottom-up multivariate analyses. The features are analysed in order to determine whether underlying structures can be found, and if so, to what extent these features can be used to cluster conditionals into groups or classes of conditional constructions licensing specific implicatures of unassertiveness and connectedness. This, then, will not only add a more data-driven answer to the first research question (i.e., which specific implicatures are licensed through unassertiveness of and connectedness in conditionals?), but also to the second research question (to what extent does the grammatical form of conditionals license specific implicatures?).[1]

In section 6.2, I will discuss the goal and types of classification. The reason for doing so is that the clustering approach chosen in this study is aimed at finding groups of conditionals that have similar feature distribution, which means that the nature of the approach is one of classification. In section 6.3, the data preparation for clustering is discussed, as are the necessary calculations and initial tests for assessing the 'clusterability' of the dataset.[2] In section 6.4, I will evaluate the results of the cluster analyses, and in sections 6.5 and 6.6, I will analyse the selected clustering solutions in light of the implicatures discussed throughout this dissertation. Finally, I will offer an interim conclusion in section 6.7, before moving on to the final conclusion and discussion in chapter 7.

6.2 Constructions and classification

6.2.1 Introduction

Before exploring the possibilities of grouping Dutch conditionals based on their features, and evaluating to what extent the resulting groups relate to implicatures of unassertiveness and connectedness, I deem it necessary to elab-

[1] See section 2.7 for research questions.

[2] The approach to data analysis presented in this chapter involves many technical and statistical choices. Although, to some, the term 'algorithm' may have connotations of objectivity, choices by the analyst can have big effects on the results. As I acknowledge that the technical details might not be of interest to all linguists, I have redirected a number of the more detailed technical arguments for the choices made in Appendix C, to which I will refer throughout this chapter.

orate on the construction grammar approach to conditionals opted for here, as discussed preliminarily in section 2.7, and its relation to the classification methodology used in this chapter.

In section 6.2.2, I will discuss the framework of construction grammar in relation to the form and meaning of conditionals central in this dissertation. In section 6.2.3, I will discuss the relation of this approach to classification, both as a scientific endeavour and in everyday life. The two main types of classification, i.e., intensional and extensional classification, will be discussed in section 6.2.4. In section 6.2.5, I will briefly iterate the arguments for the approach chosen in this study, in order to discuss the evaluation of classifications available in this study. In section 6.2.6, I will offer a brief conclusion, before moving on to the preparation for the actual analysis of the data discussed in the previous chapter.

6.2.2 Conditionals as constructions

From the perspective of construction grammar (see e.g., Fillmore, Kay & O'Connor, 1988; Goldberg, 1995; Croft & Cruse, 2004, chapters 9-11; Verhagen, 2005), we are not only interested in the meaning of *if*, whether or not it is considered to be truth-conditional or not (see chapter 2), but also in other types of non-truth-conditional meaning, which we analysed in the chapter mentioned in terms of conventional meanings of unassertiveness and connectedness connected to *als* 'if', and generalised conversational implicatures hypothesised to be attached to the linguistic form of conditionals as grammatical constructions. This means that we will look beyond the conjunction *if* in isolation, and analyse conditionals as grammatical constructions or pairings of form and meaning, i.e., as 'complex signs' (cf. Verhagen, 2009), in which form refers to the grammar of the complete complex conditional sentence.

To be clear on terms, and to reiterate the standpoint defended in chapter 2, I use the term 'meaning' here to include both 'encoded and inferred' meanings (cf. Ariel, 2010, pp. 114–115), including both conventional meaning and context-dependent implicatures. Note, however, that the term 'meaning' itself deserves clarification (see also Cappelle, 2017; Leclercq, 2020). Clark (1996) (cited in Verhagen, 2019, p. 62) offers the following observation.

> It is odd to have to explain the difference between speaker's meaning and signal meaning. In German, they are called *Gemeintes* and *Bedeutung*, in Dutch, *bedoeling* and *betekenis*, and in French, *intention* and *signification*. For theorists working in German, Dutch, and French, they are as different as apples and oranges. (Clark, 1996, p. 127)

Clark continues by arguing that because speaker's meaning and signal meaning are both referred to by the term 'meaning' in English, i.e., the term is used to refer to both encoded (or *conventional*) meaning and to inferred meaning, it remains a 'chronic source of confusion' for theorists (for further discus-

sion and references, see Verhagen, 2019). Encoded or conventional meaning is, by definition, tied to linguistic form, whereas inferred meaning is, at least to some degree, context-dependent. The picture becomes more complex however, as 'meaning' and 'intention' do not coincide with truth-conditional meaning (semantics), and non-truth-conditional meaning (implicatures, pragmatics) respectively, because, as discussed in chapter 2, conventional implicatures are encoded, but non-truth-conditional, and conversational implicatures can be strongly generalised and therefore said to be 'default inferences' (or 'default interpretations'; see Levinson, 2000, pp. 11–12), which further blurs the distinction.[3] As the analyses presented in this chapter are aimed at finding conditional constructions defined as form-meaning pairings, it is important to make explicit that the 'meaning' part of constructions refers to both types of meaning.

It is, in general, hard to draw an exact line between what is encoded and what is inferred meaning, especially when conventionalisation of implicatures is taken into consideration. The moment at which an implicature can be considered conventionalised is hard to define, which we saw already in section 2.5. Although Grice (1989, p. 39) also mentions this issue, he does not analyse it detail when he discusses his distinction between conventional and conversational implicatures.[4] As Croft and Cruse (2004, p. 258) remark in their discussion of different versions of construction grammar, they use the term 'meaning' to 'represent all of the **conventionalized** aspects of a construction's function, which may include not only properties of the situation described by the utterance, but also properties of the discourse in which the utterance is found'. Although I will present analyses which are in line with Dancygier and Sweetser's (2005, p. 8) remark that they include formal aspects of conditionals such as verb tense, clause order and intonational patterns in their analysis of conditionals, I will continue to use the distinctions drawn in the analysis presented in chapter 2. In summary, two conventional, non-truth-conditional meanings of conditionals were distinguished, i.e., their unassertiveness and connectedness, which are general, and further specified by the conversational (non-truth-conditional) implicatures they license in collaboration with the two clauses of a conditional. Distinctions between conventional and conversational aspects of meaning are made not because they are separated easily, or because of a theoretical predisposition on such an account, but, as argued before, for sake of clarity. As we saw earlier chapter 2, the distinction provides clear starting points for further analysis and can, as long as one is explicit about such a choice, clarify the discussion at hand.[5]

[3] See Ariel (2008, chapters 1-4), and Ariel (2010, chapter 4) for overviews and discussion of code-inference distinctions.

[4] On the notion of conventionalisation of implicatures, see Levinson (1979), Levinson (2000, pp. 262–263), Traugott and König (1991), Ariel (2010, p. 164), Schmid (2020, chapter 14).

[5] For a proposal on combining formal-semantic and pragmatic analyses of conditionals using Verhagen's (2005) intersubjectivity approach to grammatical constructions, see Boogaart and Reuneker (2017, p. 204), and the discussions in chapter 7.

Now that we have a clearer picture of what is meant by 'meaning', the question is how to analyse it. I will follow Boogaart (2009, p. 232), who shows, in his analysis of the meaning of the Dutch verb *kunnen* 'can', that there is an option beyond pure monosemy and polysemy by analysing various uses of a linguistic form in its specific grammatical context, i.e., as a grammatical construction or 'pairing of form and meaning'.[6] On the one hand, adopting a monosemous approach, one would analyse conditionals as having one general meaning that is further specified in context, as is done for instance by van der Auwera (1986, p. 200) in his 'Sufficiency Hypothesis', in which the conditionals mean that 'p is a sufficient condition for q' (i.e., p enables q). More specific interpretations, such as causal or inferential connections between antecedents and consequents, are then more specific instances of this meaning. The meaning of *if* is, in these terms, essentially vague and pragmatically enriched by context. Adopting a polysemous approach, on the other hand, one would argue for various distinct meanings of *als* 'if', as is done, for instance, for indicative and subjunctive conditionals (see section 2.5.4 for discussion and references). In this view, the meanings of *if* are distinct 'senses' which are, for instance, related through metaphorical extension, but there is no one 'core meaning' which is common to all those senses. Although the analysis presented in this dissertation is closest to a monosemous approach, I opted here for the approach of construction grammar (see also section 4.3.2), as it explicitly includes the linguistic form in the analysis of, in this case, conditionals. This means that the essentially abstract meanings of unassertiveness and connectedness are indeed enriched pragmatically, but explicitly in terms of the 'grammatical context' that, to a certain degree, licenses these implicatures. In case there is a clear relation between grammatical context and an implicature, we view the implicature as generalised. In case there is no such relation, and no default inference is triggered, the implicature 'remains' particular (see Levinson, 2000, p. 16)[][37]Grice. In the current approach, 'contextual enrichment' is thus substantiated by investigating systematic relations between grammatical features and implicatures. This ties in with the preliminary discussion in chapter 1, in which I described constructions as symbolic units which represent both 'lexical and grammatical structure' (cf. Langacker, 1987, p. 58) as small as morphemes and as large as complete clauses (cf. Croft & Cruse, 2004, p. 257).[7] Constructions require careful and combined analysis of both their formal characteristics and their meaning aspects. This, then, as opposed to a purely monosemous analysis, includes the grammatical features of conditionals beyond the *als* 'if', including the characteristics of the two clauses. Meaning can then be seen as conventionalised 'usage events' (cf. Langacker, 1987, p. 66; Verhagen, 2005, p. 24) tied to differences in the linguistic context in which the conjunction occurs. This is

[6] I will not consider an homonymous approach here. See Boogaart (2009, pp. 215–217) for discussion.

[7] See Boogaart, 2009, p. 230 for a summary of Croft and Cruse's 'essential principles of construction grammar'.

not fundamentally incompatible with the monosemous or the polysemous view, although it leans more towards a monosemous approach. Crucially, however, it adds an explicitly grammatical dimension to the analysis.

When we return to the topic of this dissertation, the benefits of the approach discussed above can be observed. The 'linguistic context', which is important in licensing conversational implicatures, is now defined as the combined grammatical features of conditionals, and different meanings of conditionals as conversational implicatures licensed by the conventional meanings of *als* 'if' and this linguistic context. In case there are patterns of grammatical features that occur frequently and that can be linked to specific implicatures, these can be seen as generalised conversational (or even conventional) implicatures (as opposed to particularised conversational implicatures), and in turn as the meaning part of the form-meaning pairings of construction grammar. This is, however, not uncontroversial, and as we saw in sections 1.3 and 2.4, Leclercq (2020) argues that in construction grammar, it is often unclear what the 'meaning' part precisely consists of. The question at hand here is thus whether (strongly) generalised implicatures can be viewed as part of the meaning of the construction. On the one hand, one can argue, this is the case, as the meaning is clearly linked to linguistic form, and as Goldberg (1995, p. 7) argues, a 'notion rejected by Construction Grammar is that of a strict division between semantics and pragmatics. Information about focused constituents, topicality, and register is represented in constructions alongside semantic information'. Cappelle (2017, p. 145) further argues that 'apart from semantic information, we also make use of pragmatic information in interpreting a construction in use, but not everything that is pragmatic about this interpretation is necessarily to be considered unpredictably context-dependent. There is much pragmatics that is conventionally linked to constructions'. An example of this is Stefanowitsch's (2003) inclusion of the 'pragmatic function' of indirect speech-acts into certain constructions used to perform them. In so-far 'default meanings' (cf. Levinson, 2000) go, however, they can still, given the right context, be cancelled. Once again, this points towards the discussion concerning conventionalisation of implicatures. While I acknowledge that this issue deserves more discussion, in this chapter, I take meaning to include implicatures (see above), and I will follow the line by Goldberg mentioned above, holding open the option that patterns of grammatical features in conditionals may license generalised implicatures, and that their combination may be viewed as constructions. In order to identify such patterns from the bottom up, it is needed to elaborate the methods used, and as the clustering approach is a specific form of classification, I will start by discussing the basic tenets of classification in the next section.

6.2.3 Classification, analysis, and cognition

The main benefit of classification is a reduction of complexity. In case of this dissertation, numerous uses of conditionals will be brought down to a limited number of groups, which may then – to a certain degree – be analysed as homo-

geneous phenomena, i.e., grammatical constructions with identifiable meaning aspects. The basic tenet of classification is the grouping of phenomena in such a way that 'within-group variance' is minimised and 'between-group variance' is maximised. Consider Bailey's example below.

> Imagine that we throw a mixture of 30 knives, forks, and spoons into a pile on a table and ask three people to group them by "similarity." Imagine our surprise when three different classifications result. One person classifies into two groups of utensils, the long and the short. Another classifies into three classes – plastic, wooden and silver. The third person classifies intro three groups – knives, forks and spoons. Whose classification is "best"? (Bailey, 1994, p. 2)

All three people in the example used some criterion to determine similarity between the pieces of cutlery – size, material and use, respectively. By doing so, they reduced the complexity from 30 individual objects to two or three groups. The benefit is that all cutlery can now be understood or described by referring to a limited number of groups. Because of this reduction of complexity, classification is seen by many as a central aspect of science, cognition and general learning processes. As Slater and Borghini (2011, p. 1) reflect, 'Plato famously employed [a] "carving" metaphor as an analogy for the reality of Forms (Phaedrus 265e): like an animal, the world comes to us predivided. Ideally, our best theories will be those which "carve nature at its joints"'.[8] The current use of this metaphor in modern science, however, is to refer to discovering or identifying new species, types or particles. As Slater and Borghini put it: 'we humans love to draw lines around different portions of the world, so there should be no shortage of fascinating possibilities to consider when we ask whether we are, in so doing, carving nature at its joints'. With respect to cognition, Harnad (2017, p. 21) even goes as far as to argue that our categories consist of the different ways we behave towards different *kinds* of things, such as things we do or do not eat, flee from or do not flee from. For Harnad, therefore, 'that is all that cognition is for, and about'. The idea of which classification is best or most natural also has its place in the clustering literature, in which 'realistic clustering' aims to uncover truly existing groups of data, whereas 'constructive clustering' aims to find informative, but not necessarily pre-existing groups of data.[9]

An analyst can use classification as a technique to describe and explain data. As such, classification is seen as a tool 'for conferring organization and stability on our thoughts about reality' (Marradi, 1990, p. 154) and as a 'special kind of

[8] In discussing principles of definition, Socrates offers two principles, of which the following is the relevant principle here.

> The second principle is that of division into species according to the natural formation, where the joint is, not breaking any part as a bad carver might. (Plato's *Phaedrus* (265e), in Jowett's 1892, p. 439 translation)

See Jowett (1892) for the complete dialogue between Socrates and Phaedrus.

[9] For discussion, see Hennig (2015).

scientific concept formation' (Hempel, 1965, p. 139). The analyst may conform to the 'basic rule' of classification: classes formed must be both *exhaustive* and *mutually exclusive* (cf. Greenberg, 1957, p. 69; Lakoff, 1987, p. 162; Marradi, 1990; Bailey, 1994, p. 3). For a classification to be *exhaustive*, the collective of classes must provide room to all objects assumed under the extension of the classification. *Mutual exclusivity* amounts to assigning one type only to an individual object. Sweetser (1990, pp. 124–125), for instance, as we have seen in chapter 3, explicitly mentions this for her analysis of conditionals in the content, epistemic and speech-act domain: 'A given example may be ambiguous between interpretations in two different domains, [...], but no one interpretation of an *if-then* sentence [...] simultaneously expresses conditionality in more than one domain'. The assumption is that a classification is necessarily *monothetic*, meaning that a type is defined by one or a few necessary properties (cf. Marradi, 1990, p. 132).

Many authors emphasise the importance of classification in everyday experience. Bailey (1994, p. 1) argues that 'classification is a very central process in all facets of our lives', for Feger (2001, p. 1967) its 'fundamental purpose [...] is to find structure', and Lakoff (1987, pp. 5–6) argues that, without categorisation, 'we could not function at all'. Kaufman and Rousseeuw (1990, p. 1) argue that 'a child learns to distinguish between cats and dogs, between tables and chairs, between men and women, by means of continuously improving subconscious classification schemes'. Divjak and Fieller (2014, pp. 405–406) argue that categorisation is as fundamental to language as it is to the rest of life. This is not to say that these authors all refer to the classical notion of classification, which has indeed received substantial criticism. Lakoff (1987, pp. 5–7) argues that the rules of exhaustivity and exclusivity do not hold for 'categories of the mind', because groupings are more likely to be *polythetic*, meaning that a member of a class has 'some of the [sufficient] properties of a specified total set, not necessarily the same for every object' (Feger, 2001, p. 1969), i.e., there does not need to be a 'single set of defining attributes that conform to the necessity-cum-sufficiency requirement' (Geeraerts, 2006, p. 143). This criticism has shifted the focus to another model of classification, namely *prototype theory* (cf. Rosch, 1978; Mervis & Rosch, 1981).

In prototype theory, categories are organised around '(cognitively and perceptually) salient representatives' (van der Auwera & Gast, 2010). These representatives are *prototypes* of a *radial category*, with some category members being closer to this representative member than others. As van der Auwera and Gast (2010) exemplify, 'we do not primarily think of a set of features that a bed necessarily exhibits; rather, we associate with that notion specific perceptual experiences like comfort and rest'. The traditional, 'classical', 'objectivist' or 'Aristotelian' model of necessary and sufficient conditions (see Lakoff, 1987; van der Auwera & Gast, 2010) is, then, at most an idealisation, because real objects resist what is called a 'checklist approach' (cf. Fillmore, 1975). In prototype theory, an object is related, as a whole, to experiences with that object,

instead of as a collection of features.[10] In a range of experiments, Rosch (1973; 1975) showed that categories form cognitive structures with an internal organisation based on resemblance. Whereas traditional classes are based on a small set of defining characteristics, Rosch showed that people take into account characteristics that are not necessary and that some objects are better examples of categories than others. In other words, a category is primarily understood in terms of its most representative examples (cf. Taylor, 2001, p. 287). The internal structure of a category has been linked to Wittgenstein's (1958, pp. 31–32) notion of 'family resemblance'. As an example from the domain of conditionals, Athanasiadou and Dirven (1997a, p. 89) show that 'hypothetical conditionals' are the most prototypical in their corpus,[11] because they are the most frequent type of conditional, they have interdependent clauses, internal variation in three sub-types with an internal prototypicality range (the *causal* sub-type being the most prototypical hypothetical conditional, followed by *condition* and *supposition*), they all express epistemic attitudes towards the situations expressed, they can be marked and unmarked, and from hypothetical to other types of conditionals, the causal dependency between the antecedent and the consequent decreases. The idea that class members can be more or less central to the prototype(s) of that class means that there is an internal organisation based on similarity and predicts that there will be border-line cases; those objects that are far away from the prototype will be worse examples of the category, resulting in so-called 'fuzzy boundaries' (see Löbner, 2002, pp. 186–189; cited in van der Auwera & Gast, 2010; see also Mervis & Rosch, 1981, p. 109), providing room for examples that seem to resist clear-cut assignment to one of the classes, because, as Hempel (1965, p. 151) puts it, empirical data often resist 'tidy pigeonholing'. We will take up this point later on in this chapter, and in the discussion in chapter 7. Before doing so, however, we will look at another relevant distinction made in the classification literature, namely that between intensional and extensional classification.

6.2.4 Intensional and extensional classification

Marradi (1990, pp. 130–148) distinguishes between intensional classification and extensional classification.[12] In linguistic classifications of conditionals, these differences are usually not explicitly mentioned. Intensional classification,

[10] While classification based on necessary and sufficient conditions may be preferred by some analysts, and classification based on family resemblance may be considered a more adequate model of cognition by others, the two ways of classifying discussed do not coincide with the claims made about either analysis or cognition. Although proponents of classical classification may project their analyses onto cognition, and proponents of prototype theory may use their theory of cognition to provide analyses, they do not have to do so.

[11] Dancygier (1998, p. 184) argues the same for her closely related class of 'predictive conditionals'.

[12] The activity of *classing* is omitted here, as it comes down to the process of assigning individual observations to *previously defined* classes. The process is, among other terms, also known as categorical assignment (Scheffler, 1982, p. 49) and (class) identification (Capecchi & Möller, 1968, p. 63; Feger, 2001, p. 1967).

also called qualitative classification (Bailey, 1994, p. 6), is the deductive process of forming classes on the basis of a main or fundamental parameter. When this is done at the highest, most general level, a different parameter can be chosen to further differentiate between sub-classes. These sub-classes inherit the properties of their parent classes, except when a more specific parameter 'overwrites' them. This type of classification is most common and, indeed, most classifications of conditionals discussed in chapter 3 are of this type. For instance, Quirk et al. (1985, p. 1091) suggest *directness* as main parameters to define the main types, i.e., direct and indirect conditionals. The sub-classes are based on different parameters. Within the class of direct conditionals, Quirk et al. (1985) use what could be termed 'epistemic distance' as expressed by tense to distinguish between *open* and *hypothetical* conditionals. Because the parameters work on basis of exhausivity and mutual exclusivity, the main risk of this type of classification is that it encourages strict placement in categories for observations that may not belong to one category necessarily, and, related, that it often includes some kind of 'residual category' (Marradi, 1990, p. 141), which, in the best cases, can be made sense of *a posteriori* on theoretical grounds. A clear example is Dancygier and Sweetser's (2005, p. 136) residual class of *meta-spatial conditionals* in which conditionals juxtapose two mental spaces in a domain (their main parameter) that is not *content, epistemic, speech-act* or *metalinguistic* and concerns the very act of comparing two spaces itself (see section 3.3.7). Forming classes on logical grounds may run into problems when applied to empirical findings. The rigid combination of criteria lead to what Weber (1949) calls an 'ideal type', which are idealised examples that may, but do not have to exist in this pure form. We have seen examples of this in previous chapters, such as the tense pattern 'present perfect, past perfect' and the reverse, which are logical possibilities of combining the tenses of the two clauses in conditionals, although they did not occur in the corpus (see section 5.4). Accordingly, Sandri (1969, pp. 86–87) argues that 'those kinds of classification, in which the fundamental requirements are satisfied on purely logical grounds, say very little in the field of the empirical sciences'.

Extensional (*quantitative*, *natural*) classification works on the basis of empirical data, which cannot always be neatly divided into classes and their logical complements. Instead of *deducing* classes from criteria, in an *extensional classification* the classes are *induced* from patterns of properties in an actual population of objects. Classification here is an inductive process and works on basis of perceived similarities between phenomena instead of theoretical constructs.[13] This type of classification works with properties that seem relevant for the study of the phenomena concerned. The multidimensional combination of all properties (the 'logical product', cf. Hempel & Oppenheim, 1936) is called an *attribute space*, *property space* or *feature space* (cf. Greenberg, 1957, pp. 72, 76; Marradi, 1990, p. 143; Bailey, 1994, p. 9). An example of this type of classification is Declerck and Reed's (2001) study of conditionals, in which each

[13]Other common terms are *empirical classification*, *typology* or *numerical taxonomy*.

systematic difference between exemplars results in a new class. A major benefit of extensional classification is the identification of previously unattested types, as can be seen in the remark Mauck and Portner (2006, p. 1336) make in their review of Declerck and Reed (2001): 'the book showed us kinds of conditionals (and conditional-like sentences) which we would not have thought about otherwise'. The main downside of this approach is that the various types are not logically and/or explicitly linked to each other, resulting in a *typology* that is exhaustive, but does not lend itself easily to generalisations, as the concluding remark in Dancygier's (2003, p. 322) review of Declerck and Reed's study makes clear: 'The trees have been described in all their plenitude and variety, but the forest has been overlooked'.

The type of classification I will present in this chapter is of the extensional kind, as the combined features will be used to explore possible structures underlying distributions of grammatical features of conditionals. This avoids the risk involved to intensional classification mentioned above, namely that of forcing conditionals into categories that should theoretically exist, but may not be found empirically due to overlapping boundaries. It is however prone to the risk of ending up with theoretically unmotivated residual categories, and to the main risk discussed with respect to extensional classification, namely that generalisations are not easily made, although they are desirable. We may thus end up with groups of conditionals that lack theoretical importance. Both problems are addressed here. First, most clustering algorithms require, as we will see in the sections to come, a number of clusters as parameter. While this does introduce the risk of forcing all conditionals in a small number of classes, it does eliminate the risk of high numbers of classes resisting generalisation and it decreases the risk of small residual classes. Additionally, in one of the types of clustering presented in section 6.4, the resulting hierarchical structure may preserve differences between conditionals in classes as sub-types of those classes. Furthermore, I will carefully evaluate the optimum number of clusters to address the issue mentioned. Second, the risk of theoretically unmotivated classes was already addressed by carefully using a large body of literature to identify features of importance. We do not wish, however, to eliminate the risk of coming up with classes lacking a clear theoretical motivation, as one of the possible outcomes of this study may be that the features in fact do not support the hypothesis that the grammatical form of conditionals licenses implicatures of unassertiveness and connectedness. In sum, then, we cannot avoid all risks, but I hope to have shown that the relevant risks were identified and anticipated as much as possible.

Before moving on to the data preparation in section 6.3, it is needed to discuss a related and important, but often overlooked aspect of classification, namely its evaluation. We will address this issue in the following section.

6.2.5 Evaluation of classifications

In the field of machine learning, there is particular interest in the extensional type of classification. A large number of algorithms exists which take a collection of variables or 'feature space' and try to determine underlying structures. Although other types of machine learning approaches exist, the two main approaches are *supervised* and *unsupervised machine learning*, and the difference has a large impact on the evaluation of the results, which we will discuss in this section.

As discussed already in chapter 4 (see section 4.6), the term *classification* as used in the computational literature usually refers to *supervised* machine learning. In this type of machine learning, the target labels (or *classes*) for objects are known for at least a number of observations. In contrast, unsupervised algorithms deal with data that lack such labels and are used to identify clusters of features inherent in the data, without any preconception of the nature of these clusters. In summary, Marsland (2015) describes both types of machine learning as follows.

> **Supervised learning** A training set of examples with the correct responses (targets) is provided and, based on this training set, the algorithm generalises to respond correctly to all possible inputs. This is also called learning from exemplars.
> **Unsupervised learning** Correct responses are not provided, but instead the algorithm tries to identify similarities between the inputs so that inputs that have something in common are categorised together. The statistical approach to unsupervised learning is known as density estimation. (Marsland, 2015, pp. 5–6)

Whereas in supervised machine learning an algorithm tries to predict the correct label for an observation based on the distribution of features, aiming at maximum accuracy, in unsupervised machine learning, no such target labels are available, which means that an algorithm has to resort to minimising within-group variance and maximising between-group variance. Although this might be seen as a definite disadvantage of the unsupervised approach, and an advantage of the supervised approach, I provided three arguments against a supervised approach in this study in section 4.3, which I will briefly reiterate here. The first argument was that a supervised approach presupposes that labels can be applied to the data beforehand, which turned out to be highly unreliable for the classifications discussed in chapter 3. The second argument was that a non-trivial selection of classifications used as 'gold standard' has to be made in order evaluate the results, i.e., which types of conditionals is an algorithm supposed to predict based on grammatical features? This would introduce a theoretical bias, which unsupervised machine-learning does not suffer from. The third argument was that an unsupervised approach offers ways of grouping conditionals which can be interpreted along the lines of prototype theory, i.e., these techniques are able to provide detailed insights into category structure and rep-

resentativity of conditionals. I used these arguments so support the choice for an unsupervised approach to the multivariate analysis of the data described in chapter 5. This, however, does leave us with the question of evaluation, which is much more complex for unsupervised algorithms, as the 'true types' are not known *a priori*. In the following sections, I will focus on statistics available to assess the reliability of clustering (i.e., unsupervised classification). First, in the remainder of this section, I will review the ideas behind evaluation and we discuss existing evaluation criteria for classifications in general.

Some scholars argue that classifications should 'faithfully portray the inner structure of reality' (Marradi, 1990, p. 148), while others argue that the objective of classification is instead an increase of our understanding of reality (e.g., Feger, 2001, p. 1972; van der Auwera & Gast, 2010). In this latter sense, classifications are not to be judged *true* or *false*, but more or less *fruitful* for understanding reality (cf. Tiryakian, 1968, p. 5; Kemeny, 1959, p. 195).[14] Given the importance researchers attribute to classification, one would expect its evaluation to be a common theme in research. Surprisingly, Feger (2001, p. 1967) notes that this critical aspect is frequently neglected. Indeed, classifications usually do not offer an explicit measure of quality, beside (implicit) claims of completeness (exhaustivity) and mutual exclusivity. There are, however, heuristics that can be used to evaluate the result of a classification activity. Because the nature of different types of classifications is different, not all heuristics apply to all classifications.

According to Feger (2001, p. 1968), a classification should have a *theoretical foundation* and the parameters provided by this foundation should be central to the purpose of the research (cf. Tiryakian, 1968). The theoretical foundation is the basis for deduction through which classes and their order are formulated. In this study the theoretical foundation is provided by the fact that the features to be included are based on the extensive literature review in chapter 3. Because the feature inventories come from what is known from the literature on conditionals, the analysis presented below forms a test for their ability to discriminate between different implicatures of unassertiveness and connectedness in Dutch conditionals. Next, a classification should be *objective*, in the sense that anyone familiar with the matter should be able to apply it to real data (Feger, 2001, pp. 1968–1971). This means that criteria or dimensions must be explicitly stated. With respect to conditionals, and especially implicatures of connectedness, we already saw that this proved problematic (see section 4.5). Although it may be clear from the previous chapter that I have tried to provide maximal transparency with respect to features that form the basis for the clustering, the algorithm itself involves many choices made by the analyst, which

[14] Apart from this difference, the terms *classification* and *categorisation* are used interchangeably throughout the literature. For instance, van der Auwera and Gast (2010) use the term *category* to refer to 'a set of entities that share one or more properties and that are thus to some extent similar' and trace the term back to Aristotles' *Categories*, who used the term to refer to what is called *classes* here: groupings based on necessary and sufficient conditions. See also Bloomfield (1984, p. 270).

I will elaborate with the same transparency in the remainder of this chapter in order to maximise objectivity. A classification should furthermore be *exhaustive*, i.e., all elements should have a place in the resulting classification. Especially in the case of intensional classifications, this leads, in many cases, to a *residual category* (Marradi, 1990, p. 141) in which all objects that do not fit neatly into one of the theoretically motivated classes are placed. What can be seen in the classifications discussed in chapter 3, is that it sometimes remains unclear whether or not they adhere to what McEnery and Hardie (2012, pp. 14–18) call the 'principle of total accountability': explicitly taking the responsibility to account for all corpus data present in the corpus or sample, including ambiguous, border-line or unclear exemplars. We already saw in section 6.2.4 that some accounts of conditionals include residual categories. Together with *exhaustivity*, *mutual exclusivity* is one of the classic criteria of classifications. When *mutual exclusivity* is used as a criterion, all cases must fall into one class only, as a consequence of the exclusivity of the parameters used. This is similar (although not identical) to for instance Sweetser's (1990) remark cited in in section 6.2.3 that a conditional may be ambiguous between interpretations, but it can only have one interpretation at a time and thus should be assigned to one type, given the specific context is taken into account. As some algorithms are able to provide so-called 'soft' or 'fuzzy' cluster assignments, meaning that one observation can be placed in multiple groups of data by assigning a numerical indication of the fit, I will reflect explicitly on this issue when using this type of clustering algorithm (see section 6.4.5). Next, according to Feger (2001, p. 1968), a classification should be *simple*, in the sense that a 'small amount of information is used to establish the system and identify objects'. A 'minimal set of variables' should be sufficient to discriminate between classes. This is called *parsimony* by Tiryakian (1968), referring to 'the fewest meaningful or significant major types possible to cover the largest number of observations'. Finally, a classification should be able to generate predictions. In Tiryakian's (1968) terms, 'a "good" typological classification would include the criterion of fruitfulness (the typology may have heuristic significance in facilitating the discovery of new empirical entities)'. In section 6.2.4, this quality was addressed directly by Mauck and Portner's (2006) review of Declerck and Reed's (2001) account. Ideally, the results of the clustering should provide insight into the feature-combinations that are most typical for a certain cluster, in turn enabling the placement of new observations into the generated clusters.

6.2.6 Conclusion

In this section, I provided arguments for analysing conditionals as form-meaning pairings, i.e., constructions, in order to investigate relations between grammatical features and implicatures of conditionals. As the features are expected to 'work together' in licensing implicatures of unassertiveness and connectedness, a clustering approach to the data was chosen. Clustering is a type of classification, and in order to explain the choice for and evaluation of this un-

supervised clustering approach, I discussed its advantages and disadvantages in this section. Although no target labels are available for the data, which excludes direct implementation of supervised techniques, a cluster analysis upholds the basic tenet of classification, namely forming groups that exhibit the smallest amount of within-group variance and the largest amount of between-group variance, which will be used to investigate to what extent specific implicatures of unassertiveness and connectedness can be viewed as generalised conversational implicatures. A major benefit of the unsupervised approach taken is that no preconception about these implicatures has to be made beyond the selection of variables. As these variables form the input for the the clustering algorithms, the data preparation needed will be elaborated in the next section.

6.3 Data preparation, variable selection, and distance calculation

6.3.1 Introduction

Before we can select suitable clustering approaches, and subsequently subject the data to the corresponding algorithms, the collective features of conditionals discussed extensively in the previous chapter (the 'feature space') demand a number of preparatory conversions and evaluations. These steps are the subject of this section. As we are dealing with data preparation here mainly, I find it important to remind the reader here why such data preparation and discussion thereof are important. As we are looking for clusters of grammatical features in relation to implicatures of unassertiveness and connectedness, it is vital to assess to what extent these features combined actually indicate the presence of clusters. As remarked throughout chapter 5, it may also be the case that the implicatures are not generalised or conventionalised, and that meaningful clusters cannot be not found. The preparatory steps discussed in this section are meant to enable such assessments, and to enable the clustering algorithms to process the data (see section 6.4).

In section 6.3.2, I will discuss the preliminary variable selection. In section 6.3.3, I will discuss the basics of distance calculation, and more advanced distance calculation will be discussed in section 6.3.4. Then, in in section 6.3.5, the selection and evaluation of distance measures (and resulting matrices) is elaborated. In section 6.3.6, the final variable selection, based on this evaluation, will be presented. In section 6.3.7 I will use the results of the evaluations to identify the most and least representative conditionals in Dutch, and, finally, in section 6.3.8, I will offer a brief conclusion before moving on to the actual clustering in section 6.4.

6.3.2 Initial variable selection

In chapter 5, the features identified in the linguistic literature on conditionals were annotated in a corpus that was balanced on the dimensions *mode* (spoken, written) and *register* (formal, informal). The annotation followed conventional labels in the field of linguistics. This does not, however, necessarily lead to optimal coding for further quantitative data analysis. For example, a feature like *verb tense* is already prone to discussion in linguistics (see section 5.4 on future tenses). I have chosen to use the two-way binary tense system (cf. Broekhuis, Corver & Vos, 2015a) for annotation, and the name already suggests that each possible value or *level* of this feature can be further decomposed into two binary features, namely ±*past* and ±*perfect*. This implies a choice of how to code the feature as a variable for further analysis: either a clause is annotated for verb tense as *simple past* (one feature), or as having the features +*past* and −*perfect*. As one may imagine, there is no simple wrong or right way to go about this. I chose to code the variables with minimal deviation from the levels used in chapter 5, mostly for reasons of interpretability of results. However, variables with skewed distributions and low-frequency values may have negative impact on dimension reduction of the data, and as we saw in the previous chapter, a number of features indeed suffered from this issue. Before performing any analyses, I took a number steps to ensure optimal coding. First, I performed a check on feature independence, resulting in 12 features, for instance, by combining clause order and syntactic integration into one feature (see Appendix C for details).[15] Second, the dispersion of the feature values (i.e., the distribution and possible skewedness) was evaluated, which we will turn to next.

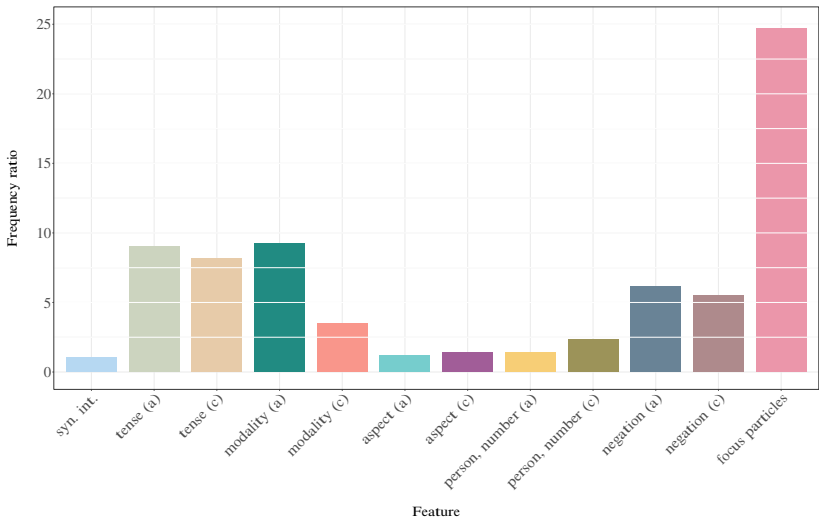
In the clustering literature, *feature* or *variable selection* is gaining attention, but it lags behind the amount of work done in classification literature (cf. Liu & Zhang, 2016, p. 100; Solorio-Fernández, Carrasco-Ochoa & Martínez-Trinidad, 2020). Many studies simply use all variables available, without critically assessing 'clusterability' and the contribution of the individual variables. Levshina (2011, pp. 60–61), for instance, uses no less than 35 variables, which were chosen 'only by practical methodological reasons (although some variables have proven to be useful in the previous studies)'. However, the 'inclusion of unnecessary variables' may have a negative impact on clustering results (see Raftery & Dean, 2006, p. 168). Fowlkes, Gnanadesikan and Kettenring (1988, p. 205) provide clear examples of how clustering algorithms 'can completely fail to identify clear cluster structure if that structure is confined to a subset of the variables'. At the other end of the spectrum, Gabrielatos (2010, pp. 52–53; 2021) uses only two (correlated) variables, 'modal density' (MD) and 'modalisation spread' (MS) in his hierarchical cluster analysis. As was shown in the previous chapter, the

[15] The original features were recoded into the following variables *syntactic integration* (including clause order), *tense (a)* and *tense (c)*, *modality (a)* and *modality (c)*, *aspecgt (a)* and *aspect (c)*, *person and number (a)* and *person and number (c)*, *negation (a)* and *negation (c)*, and *focus particles*.

number of variables in this study lies in between: the 12 variables were chosen explicitly for their relation to implicatures of unassertiveness and connectedness as discussed in the literature in chapter 3. This does not mean, however, that a critical perspective on the informativeness of each of these variables is not needed. In this study, I used 'Deviation from the Mode' or *DM* (cf. Wilcox, 1973, p. 325) as an initial measure of dispersion. DM takes on a value between 0 and 1 and the higher the value, the more evenly spread the values of a variable are, whereas a low value indicates an uneven spread of the values. The calculation and results are discussed in detail in section C.2 of Appendix C. The DM values were only used for the initial inspection of feature distributions, as it enabled identifying variables with skewed distributions. In itself, such a distribution need not be problematic, but a *DM* value near zero may indicate a non-informative distribution. The *DM* values (see Table C.1 on page 485 in Appendix C) indicate that tense in antecedents and consequents, modality in antecedents, negation in antecedents and consequents and particularly focus particles have prevalent values that skew the distributions.

To further assess which variables may have non-informative value distributions, the frequency-ratio of each of the variables was calculated. The idea behind this step is that a large ratio between the frequency of the most frequent value and the second-most frequent value indicates that it may be better to remove the variable from the model (see Kuhn & Johnson, 2013, p. 45). From the results of this analysis, presented in Figure 6.1 below (for technical details, see section C.4 of Appendix C), we can see that the feature *focus particles* has a frequency ratio that is more than twice as high as that of any other feature.

Figure 6.1:
Frequency ratio per feature



This can be explained by the fact that an overwhelming majority of conditionals does not have a focus particle, making the absence of a focus particle a largely uninformative feature in isolation. Furthermore, the results indicate a relatively high frequency ratio for tense in both clauses, because of the prevalence of the simple present, and, interestingly, a higher frequency ratio for modal marking of the antecedent, but not the consequent, as the number of modalised consequents is much higher than the number of modalised antecedents.

Next to these internal measures of dispersion, I will informally rank the variables based on their theoretical relevance, as discussed at length in both chapter 3 and chapter 5. With respect to implicatures of unassertiveness and connectedness, it became clear that tense and modality are the most prominent features in the literature. Note that these classifications concerned English conditionals mainly, whereas in Dutch conditionals other features may prove more informative.[16] For Dutch, however, Reuneker (2016) already showed that modality was of influence on the conditional interpretation of prepositional phrases, which indicates a certain importance of this feature. In section 5.3, I showed that clause order, and, for Dutch, syntactic integration (see Reuneker, 2020) are other important features in relation to implicatures, mainly those of

[16]See also section 4.3.2 and the discussion in chapter 7 on the arguments for and consequences of the choice for a Dutch corpus in relation to the literature that has focused mainly on English conditionals.

connectedness. Next, negation proved an important parameter in the coherency approach to conditionals discussed in section 3.3.8, which deals mainly with implicatures of connectedness, and in Declerck and Reed's classification discussed in section 3.3. As we saw in section 5.9, negation was also suggested to play a role in cooperation with tense in licensing implicatures of unassertiveness, or, more specifically, implicatures of counterfactuality. The feature of aspect and the combined feature of person and number were only weakly linked to conditionals (see sections 3.2.7 and 3.3.9), although, in section 5.7, I discussed how first-person and second-person subjects were related to the use of conditionals to tone down the force of directive speech acts. Focus particles, finally, have been linked to restrictions on connections between antecedents and consequents (see sections 3.3 and 5.10). To conclude, the ranking based on the literature reviewed in chapter 3 suggests the following order of feature importance: verb tense, modal marking, syntactic integration (including clause order and occurrence of resumptive *dan* 'then'), sentence type of the consequent, negation, focus particles, aspect and person and number.

The results in this section suggest that focus particles should be excluded from the final analyses due to a high frequency-ratio, and relatively low theoretical relevance. Theoretical relevance also suggests that special attention needs to be paid to aspect and person and number in subsequent analyses, as the literature review in chapter 3 suggests low theoretical importance. These suggestions will be taken up in the final feature selection discussed in section 6.3.6, but first, as a necessary step, the distance calculation will be discussed.

6.3.3 Basic distance calculation

After coding features as variables, evaluating their dispersion, and composing an initial list of contributing variables, the dataset consisted of 4109 observations (i.e., conditionals) of the 12 variables shown in Figure 6.1, resulting in $4109 * 12 = 49.308$ data points. From this dataset, the next step was the calculation the (dis)similarity between each conditional in terms of its features. This was done because clustering, in basic terms, works on the basic principles of classification discussed in section 6.2 above, as it groups observations in such a way that within-group differences are minimised, while between-group differences are maximised (see Cichosz, 2015, chapter 11 for an introduction on similarity and dissimilarity calculation). 'Distance' in 'distance calculation', then, is the operationalisation of difference.[17]

In some cases, the calculation of distance is straightforward. Two people with different heights have a distance on that variable equal to their difference in height. When we add a variable, like weight, the difference between their weights is incorporated into the distance calculation. As height and weight are measured on different scales, before calculating the distance, such variables should be normalised, for which multiple strategies are available. However,

[17]For technical details of the distance calculations, see section C.5 of Appendix C.

if yet another variable is taken into account, such as gender or eye-colour, distance calculation becomes much more complex, even with normalisation, because numeric variables (height, weight), and categorical variables (gender, eye-colour), are fundamentally different. It is therefore a non-trivial task to use such variables in any distance calculation, yet categorical variables are what most corpus linguists deal with. The use of categorical variables severely limits the choices in distance measures. A common approach to dealing with problem in clustering is to first perform a binary transformation of categorical variables into so-called 'dummy variables', meaning that each variable is coded into variable-value pairs. For instance, the variable 'modality' with values 'epistemic' and 'deontic' is coded into the binary variables 'modality-epistemic', 'modality-deontic', which each then receive 0 for absence and 1 for presence of the value. After this transformation, the more widely available distance measures for binary variables can be applied. However, this can introduce both significant decrease computing speed and a loss of information (for a recent experiment and discussion, see Cibulková et al., 2019). This may be the reason why most corpus-linguistic literature in which clustering techniques are applied use Gower's distance or another, indirect approach, such as clustering through 'behavioural profiles' (see e.g., Divjak & Gries, 2006; Divjak, 2010; Levshina, 2011; Divjak & Fieller, 2014). Although Gower's distance has important limitations, it is readily available for categorical data and can serve for introductory purposes below.

A *distance matrix* reflects the dissimilarity of one observation compared to another in terms of all of its features. In most corpus-linguistic studies, distances are calculated using *Gower's General Similarity Coefficient* (Gower, 1971, p. 861), often abbreviated to *Gower's Distance*, mostly because it is presented as the default option for non-numerical variables. The formula and details for calculation of Gower's distance are presented in section C.5.1 of Appendix C. Here, it will suffice to discuss the measure in more general terms. As, in calculating a distance matrix, all observations are compared to each other, the product rapidly becomes very large. Therefore, I will introduce a small and fictitious data set to clarify the workings of Gower's Distance, and to discuss its importance for this study. The illustrative dataset consists of just four conditionals, exemplified in (397) to (400) below. For each of these conditionals, three features were annotated, namely clause order (sentence-initial, sentence-medial or sentence-final), person and number of the antecedent (1ps, 1pp, 2ps, 2pp, 3ps or 3pp), and modal marking of the consequent (epistemic, evidential, deontic, dynamic or none).

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

(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.

The examples in (397), (398), and (400) have sentence-initial antecedents, whereas the example in (399) has a sentence-final antecedent. The grammatical subject in the antecedent is second-person singular in (397), third-person singular in (398) and (399), and first-person plural in (400). With respect to modality in the consequent, we see that (397) is marked for epistemic modality by the auxiliary *will*, (398) and (399) are not marked for modality, and (400) is marked for deontic modality by means of the modal auxiliary *may*. In the data structure employed in this study, this looks like Table 6.1.

Table 6.1:
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

Assuming no custom weights (see Appendix C.5.1), Gower's distance is the number of features shared between two conditionals, divided by the number of features. The resulting distance matrix is presented in Table 6.2 below.

Table 6.2:
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

Before looking at the distances in Table 6.2, please note that there is a diagonal line of zeros, which is expected, as these numbers represent the distance from one conditional to itself. As the table is symmetrical, the lower-left diagonal is identical to the upper-right diagonal, and by convention only the lower triangle is presented. Looking at the conditionals in (397) and (398), we can see the distance is 0.67, because they share only one out of three features (clause order), as can be seen in Table 6.1. The distance is then simply 1 minus the number of features shared (1) divided by the total number of features (3), i.e., $1-(1/3) = 0.67$. We can also see that (397) and (399) share no features, resulting in a distance of 1, i.e., $1-(0/3) = 1$. The distance between (397) and (400) is $1-(1/3) = 0.67$, because they share only the feature clause order. Looking at

(398) and (399), we see they share person and number, and the non-occurrence of modal marking, resulting in a distance of $1 - (2/3) = 0.33$. The distance between (398) and (400) is 0.67, because they only share clause order. Finally, the distance between (399) and (400) is 1, because no features are shared. We thus see that the more similar conditionals are in terms of their features, the smaller their distance is.[18]

Even with categorical variables only, the calculation of distance is not devoid of problems. Before discussing the problem of missing values, I will discuss another problem, which, as we will see, may provide the key to solving the missing values problem in the first place. In the case of modal marking in the example above, most consequents are not marked for modality. If we were to treat these *no*-values as genuine features of conditionals, their prevalence may introduce problems, as I discussed with respect to focus particles above. While this may look like an isolated problem, it is comparable to another problem encountered already, namely that of highly skewed distribution for features like verb tense. As we saw in section 5.4, around 80% of all clauses in conditionals have simple present verb tense. In default metrics for nominal features, this is not taken into consideration, which means that the similarity measure of two clauses is impacted exactly the same when they both have the highly frequent verb tense *simple present*, as when they have the much less frequent verb tense *simple past*. Are two conditionals that share the simple present tense as equal as two conditionals that share the simple past tense, given that the former tense is much more likely to occur than the latter? As we can see, the relatively straightforward calculation of (dis)similarity has now become a more complex problem involving probability. Intuitively, the following makes sense: the probability of a feature occurring in both conditionals should have as large an impact as the probability of both conditionals not sharing the feature. This idea has already been suggested in very general terms by Anderberg, but at the time computation was too slow to implement such a metric.

> The desire to give rare classes extra weight appears frequently in the biological literature though systematic methods for assigning such weights are not offered. [...] Since rare events have low probabilities, the probability of an event is not a suitable weight; however any inverse function of the probability is potentially interesting. (Anderberg, 1973, pp. 124–125)

In recent years, however, a number of 'probability-based' distance metrics have been implemented, which we will discuss in the next section, and in doing do, we will return to the problems of missing values and skewed distributions.

[18] While similarities may be preferred for interpretation, distances or *dissimilarities* are frequently used in several kinds of machine-learning and dimension-reduction algorithms. The measure for similarity is simply 1 minus distance.

6.3.4 Probability-based distance calculation

The general idea of probability-based distance measures is that the distribution of the variable levels is taken into account in the calculation of distance. Let us look at another simplified example.[19] Suppose we have five conditionals, annotated for two features, as presented in Table 6.3 below.

Table 6.3:
Example data for probability-weighted distance calculation

Example	Clause order	Modality (c)
(1)	initial	epistemic
(2)	initial	epistemic
(3)	initial	no
(4)	final	deontic
(5)	final	deontic

We can see that there are two clause orders present. The probability of a conditional having a sentence-initial antecedent is the number sentence-initial antecedents divided by the total number of conditionals, i.e., 3/5=0.6. The probability of a sentence-final antecedent is 2/5=0.4. Necessarily, the sum of both probabilities is 1. The probabilities of the variables *modality (c)* are 2/5=0.4 for epistemic modality, 1/5=0.2 for no modal marking and 2/5=0.4 for deontic modality. Let us now look at examples (1) and (2). They are identical, and using Gower's distance, as discussed in the previous section, they would receive a dissimilarity of 0. The same goes for examples (4) and (5). They are identical and thus have a dissimilarity of 0. However, the chance of a zero distance is higher for (1) and (2) than for (4) and (5), because the sentence-initial clause order in the former pair is more likely to occur in both conditionals than the sentence-final clause order in the latter pair. Although I endorse the view that examples (1)-(2) and (4)-(5) are both identical in Table 6.3, given the skewedness of a number of features in the dataset (see chapter 5), it would be advantageous to set the weight not per feature (i.e., a constant weight, see Appendix C.5.1), but to make weight dependent on the probability of the feature's values, in order to 'give status to rare classes' (Anderberg, 1973, pp. 124–125). In other words, examples (4)-(5) receive a slightly higher similarity, and examples (1)-(2) a slightly higher dissimilarity, because the probability of a match on sentence-initial clause order is higher than a match on sentence-final clause order.[20]

[19]The technical details for the application of probability-based distances measures to the actual corpus data can be found in Appendix C.

[20]Expressed exclusively in terms of dissimilarity, examples (1)-(2) would receive a slightly higher dissimilarity than examples (4)-(5).

It turns out that this problem has been addressed in the biological and statistical literature on similarity measurements already, although it does not, to my knowledge, seem to have found its way into (corpus) linguistics. Goodall (1966) proposed a measure that captures the exact nature of the probability measure proposed above by adding variable weights based on probability. Such a strategy seems particularly suitable for the current dataset, as a number of features have highly skewed distributions. For instance, as we saw in section 5.4 in the previous chapter, roughly 84% of antecedents, and 88% of consequents have simple present verb tense. Such a similarity should contribute less to the overall similarity between two conditionals, than for instance a correspondence on the simple past verb tense, which occurs in roughly 9% and 11% of antecedents and consequents respectively. In more general cognitive terms this too makes sense, i.e., a low-frequency value is a more informative clue for processing than a high-frequency value, because it is 'marked', somewhat comparable with the argument for Levinson's (2000, p. 39) M-principle discussed in section 2.4. In terms of 'markedness', the simple present in this example would be the unmarked member, whereas other tenses are marked (see Comrie, 1996). Although there is significant criticism on the notion of markedness (see Haspelmath, 2006), here it is used in the same terms as, for instance, Comrie (1996) and Holleman and Pander Maat (2009, p. 2209): when a feature has a skewed distribution, the high-frequent value(s) will be used in a wider variety of contexts (i.e., the *unmarked* values) than the low-frequent value(s) (i.e., the *marked* values). We can see this general idea come to fruition when we calculate the distances between the conditionals in the example data in Table 6.3 using Gower's measure on the one hand and Goodall's measure on the other. First, the distance matrix using Gower's metric was calculated. The results of the calculations using Gower's measure are presented in Table 6.4 below.

Table 6.4:
Distance matrix using Gower's distance on Table 6.3

	Ex. (1)	Ex. (2)	Ex. (3)	Ex. (4)	Ex. (5)
Ex. (1)	0.00				
Ex. (2)	0.00	0.00			
Ex. (3)	0.50	0.50	0.00		
Ex. (4)	1.00	1.00	1.00	0.00	
Ex. (5)	1.00	1.00	1.00	0.00	0.00

As expected, we see here that examples (1) and (2) have a distance of 0, as have examples (4) and (5). Examples (1) and (2) share only clause order with the example in (3), resulting in a distance of 0.5. The other combinations share no

features, resulting in the maximum distance of 1.0. To reiterate, we see that the distributional difference between sentence-initial and sentence-final antecedents is not taken into account.

Now, we will use Goodall's probability-weighted measurement on the same dataset. As the formula can be insightful at this point, I include and discuss it below, instead of referring to the appendices. Goodall's measurement is presented in (401) below.

(401) $S_c(x_{ic}, x_{jc}) = \begin{smallmatrix} 1-p_c^2(x_{ic})\ if\ x_{ic}=x_{jc} \\ 0\ otherwise \end{smallmatrix}$.

Here, x is the dataset, i and j are the individual observations (here conditionals), ranging from 1 to n (the number of observations), and c stands for the feature to be compared, ranging from 1 to m, m being the total number of features. Finally, $p_c(x)$ is the relative frequency of value x for feature c. In case of a match, a similarity based on this relative frequency (the probability of the value) is calculated, whereas in case of a mismatch, *0* is added to the similarity. For each comparison of two conditionals, similarities are summed and subtracted from 1, resulting in their final distance.[21] Applying this measure on the example dataset in 6.3 results in the data presented below in Table 6.5.

Table 6.5:
Distance matrix using Goodall's probability-weighted measure on Table 6.3

	Ex. (1)	Ex. (2)	Ex. (3)	Ex. (4)	Ex. (5)
Ex. (1)	0.00				
Ex. (2)	0.26	0.00			
Ex. (3)	0.68	0.68	0.00		
Ex. (4)	1.00	1.00	1.00	0.00	
Ex. (5)	1.00	1.00	1.00	0.16	0.00

What we see here, is that the examples in (1) and (2) are not as similar as those in (4) and (5), reflected in distances of 0.26 and 0.16 respectively, because the chance of similarity is greater in the former pair than in the latter pair. The advantage of this result is that the distribution of the variables is clearly weighted in the calculation.

There are, however, two main disadvantages. The first, as already mentioned by Anderberg (1973, pp. 124–125), is that the computation of such as probability-weighted metric is highly inefficient. For an example such as the above with only five observations of two variables, this poses no problem, but one can image that the current dataset of more than 4000 observations of 12

[21] See Goodall (1966) for the original probability-based similarity index. The equation here is based on a later implementation, which assigns higher similarity for infrequent matches without using the frequencies of other categories (Šulc & Řezanková, 2015; Šulc, 2016).

variables requires significant calculation time. The second disadvantage is more fundamental, although the algorithm producing the distance matrix in 6.5 takes care of it in a practical way. The problem is that, because the distribution of variables plays a role, the distance between an observation and itself is not necessarily 0. The example in (1), for instance, would, by implication of the metric, have a slightly higher distance from itself than the example in (4) due to the distribution of clause orders. This is unwanted, and while it formally excludes the formula in (401) as a proper metric (see e.g., Deza & Deza, 2013, chapter 1), the algorithm excludes comparisons on the 0-diagonal and simply returns 0 in those cases.[22]

In conclusion, then, the results in Table 6.5 reflect what is required, given the inherent structure of the dataset. The calculation is not biased by *a priori* assumptions used for constant feature weights, but based on the internal distribution of features, which takes into account any skewedness. In the following section, we will evaluate a number of distance measures which implement probabilities, enabling the selection of the most promising calculations for further steps in the clustering approach.

6.3.5 Selection and evaluation of distance measures

The way distributional differences are used for probability-weighting can be implemented in various ways, of which Goodall's (1966) proposal discussed above is an early example. As various datasets tend to respond differently to different distance measures, choosing an appropriate measure varies per dataset (see e.g., Boriah, Chandola & Kumar, 2008, p. 253; Ladds et al., 2018). To deal with this issue, I have selected eight measures, ranging from tested and evaluated measures to state-of-the-art measures to be calculated and compared, in order to evaluate which produces the most promising basis for further data analyses.[23] For the selected measures, I will include a short description of its workings with a focus on the weighting scheme, and I will refer to the publications their published in for details on calculation, considerations and assumptions. Further note that the data visualisations in this section reflect clusterability of the data (i.e., to what extent do the feature distributions indicate *underlying* structures), but not yet actual clustering results (i.e., the structures themselves), which I will present in section 6.4.

The first measure is Gower's (1971) coefficient,[24] discussed already in section 6.3.3, and it was selected because of its widespread use and easy and transparent interpretation. The second measurement is Goodall's (1966) similarity

[22]Please note that this is one of the main reasons I use the terms *measure(ment)* or *index* instead of *metric* here, as the zero assignment fails to adhere to the strict definitions of metrics (see Schweizer & Sklar, 1960, p. 315). The measures are, however, suitable and well-tested for many datasets, as we will see below.

[23]These measurements were calculated using the *nomclust*-package for R (Šulc & Řezanková, 2015).

[24]Gower's coefficient is also called 'Simple Matching'.

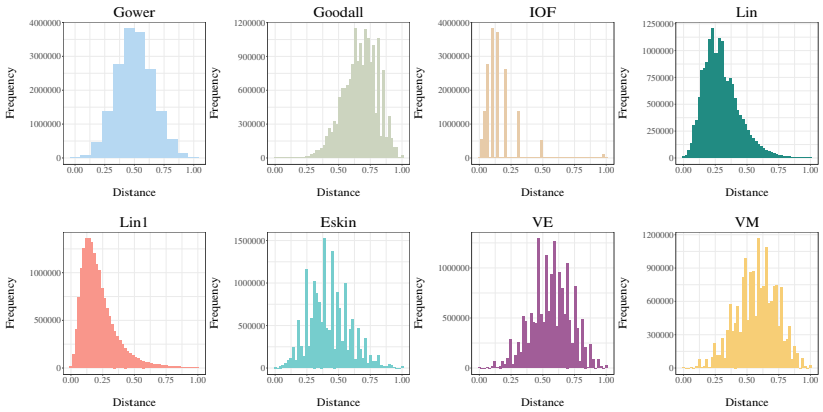
index, discussed in section 6.3.4. The third measure is Spärck Jones's (1972) *Inverse Occurrence Frequency (IOF)*, which adds more weight to non-matching pairs on less frequent values and less weight to non-matching pairs on more frequent values (see also Boriah, Chandola & Kumar, 2008; Šulc, Cibulkova & Řezanková, 2020). The fourth measure is *Lin's similarity measure* (Lin, 1998), which adds more weight to matches on frequent categories and less weight to non-matching pairs on less frequent categories. The fifth measure is the *Lin1 measure*, which is a modification of the previous measure. While it is based on the same definition and assumptions (see Boriah, Chandola & Kumar, 2008, p. 249), it has a more complex weighting system (cf. Šulc & Řezanková, 2015), in which lower weight is given to mismatches if the mismatching values are frequent, or if the values have a frequency in between the frequency of the mismatching values, whereas higher weight is given to mismatches on infrequent values when there are only a small number of other infrequent values. In case of matching pairs, lower weight given to matches on frequent values and to matches that have other values with corresponding frequencies, whereas higher weight is given to matches on infrequent values. The sixth measurement is the *Eskin* measure (Eskin et al., 2002), which adds more weight to non-matching pairs on variables with more categories. The seventh measure is the Variable Entropy (VE) measure, which was recently introduced by Šulc and Řezanková (2019, pp. 63–64). A match of two conditionals on a certain feature is weighted by the variability in the feature, resulting in more weight in case of for matches on rare (i.e., infrequent) values. Variability in this measure is defined in terms of entropy (a measure of the randomness of the data) using the relative frequencies of all categories. The eight and last measure, the Variable Mutability (VM) measure, was also introduced by Šulc and Řezanková (2019, pp. 63–64) and differs from Variable Entropy in its operationalisation of variability, for which not entropy, but mutability is used, which is the nominal variance or 'Gini coefficient' of the data (see e.g., Gastwirth, 1972; Han et al., 2016).

In the remainder of this section, I will evaluate the results from each of the measures briefly discussed above. First, the distributions of distances will be used to check for multimodality, and second, I will use dimension-reduction techniques to see whether the variance in the dataset can be described by a limited number of components.[25] Both methods indicate to what extent the dataset has an underlying structure (see Adolfsson, Ackerman & Brownstein, 2019), in order to decide whether subsequent clustering of the data is eligible. Testing a distance matrix for multimodality assumes that the data come from a unimodal distribution. This assumption constitutes the null hypothesis. If the actual distribution differs strongly enough from the assumed unimodal distribution, this indicates that there is evidence for multiple modes in the data,

[25]Note that 'multimodality' is used here as a statistical term referring to probability distributions with more than one mode, which is the most frequent value.

which could reflect multiple clusters (see Adolfsson, Ackerman & Brownstein, 2019, pp. 6–7). The distributions of the distance matrices are presented below in Figure 6.2.

Figure 6.2:
Distribution of distances per measure



Note. Measures are presented in consistent colours throughout this chapter.

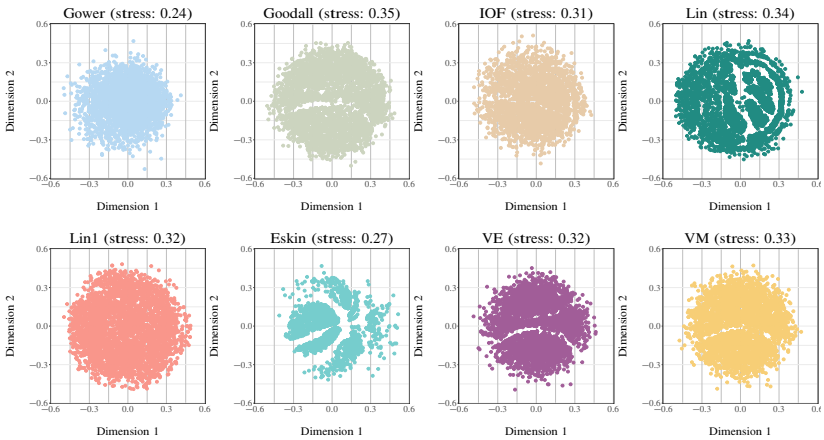
In case of multimodality, the (multiple) modes suggest multiple clusters, as cluster will have conditionals with similar distances. In other words, we would like to see distributions with clearly identifiable 'peaks' in the distribution. The histograms in Figure 6.2 do not show clear multimodal distributions, however. This becomes especially clear when comparing the distance distributions in Figure 6.2 to truly multimodal distance distributions, such as the examples by Ackerman, Adolfsson and Brownstein (2016, p. 5). Unfortunately, the multimodality tests do not provide conclusive results, which is likely due to the categorical nature of the data.[26] Therefore, a second type of evaluation was performed on the distance matrices to select the most promising one for clustering.

Because the evaluation of multimodality does not provide clear indications of clusterability, as they may be ambiguous between showing an effect the categorical data used, or indeed an indication of low clusterability, the evaluation of distance matrices was supplemented with an evaluation based on dimension reduction. Reducing the number of dimensions in the dataset aids finding out whether there are inherent structures present in the data. Dimension reduction was performed using *non-metric multidimensional scaling (NMDS)*, which is

[26] See Appendix C.5.2 for technical details, references, tests, results and a discussion on the relation of this finding to clustering categorical data.

comparable to the more familiar *Principle Component Analysis (PCA)* (see e.g., Hay & Baayen, 2003; Levshina, 2015, chapter 18 for examples). I chose NMDS because it works with any distance measure, not just Euclidean distances. NMDS is a so-called 'ordination technique', as it orders observations, placing similar observations close and dissimilar object further apart (for an introduction and explanation, see Cox & Cox, 2001, chapter 3; see Kruskal & Wish, 1978, for origins; see also Borg & Groenen, 2005, chapter 1). Furthermore, an index of *stress* is calculated, which indicates the level of distortion introduced by reducing the set of variables into a low number of dimensions (cf. Kruskal, 1964). In general terms, dimension reduction techniques try to group observations using a smaller number of variables by combining those variables. As this introduces a decrease in information, the groups of data will be less detailed, but more efficiently described. This stress level is thus a ratio between the fit of a model, which should be as high as possible, and the information needed to produce the model, which should as low as possible. The lower the stress, then, the less distortion the dimension reduction introduces, consequently indicating that there are groups to be discovered in the data. The results of the NDMS calculations are plotted below in Figure 6.3.[27]

Figure 6.3:
NMDS configurations and stress levels for distance matrices (full feature set)



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

There are at least two important observations to be made regarding Figure 6.3. First, Gower's measure, which is the only measure that is not probability-based, results in the lowest stress score, i.e., it results in the least distortion

[27] For technical details and discussion, see section C.5.3 of Appendix C.

of the data. This is remarkable, given that a number of features with skewed distributions were ranked high based on theory (see section 6.3.2 above), which would suggest better results for measures that incorporate this skewedness. Furthermore, all stress levels are above the threshold of 0.20, which means the results are 'dangerous to interpret' (cf. Clarke, 1993, p. 126; see also Appendix C.5.3 for an important and more detailed discussion of this interpretation). I suggest three possible causes. First, there may be a possible influence of mode and register. As we saw in chapter 5, a number of features showed associations with mode or register, or both. The different distributions of these features on those dimensions may introduce unwanted variance that prohibits clustering the complete data as a whole (i.e., without distinguishing between modes and registers). This influence was critically assessed for each mode-register combination, and multimodality tests and dimension reduction produced roughly the same results as reported above for each individual combination (for details, see section C.5.3 of Appendix C). Second, the dataset used here is much larger than usual in experiments with stress levels of dimension reduction techniques on real data, which only relatively recently started to gain access to samples with sizes above 100 observations (see e.g., Bollens et al., 2014; Hassett et al., 2017). Dexter, Rollwagen-Bollens and Bollens (2018, p. 437) show how 'stress increases with increasing sample size and decreases with increasing ordination dimensionality [...] essentially irrespective of the underlying data' (i.e., by increasing the number of dimensions resulting from the reduction), which is in line with early studies on the subject (see Kruskal & Wish, 1978; McCune, Grace & Urban, 2002, p. 132). A third possible cause, which was already discussed in section 6.3, is 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. In other words, variables that do not contribute to forming clusters, or variables that point towards different clusters may have a negative impact on the results. Therefore, we will return to variable selection next in order to investigate whether clusterability can be improved by removing uninformative or distorting variables from the dataset. We will then evaluate the distances matrices again and compare results to choose the optimal set of variables for finding clusters of grammatical features in conditionals.

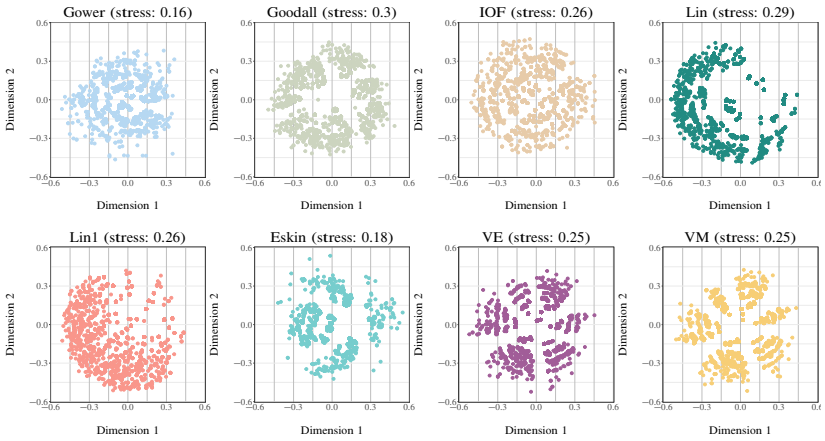
6.3.6 Final variable selection

As we saw earlier in this section, feature selection for unsupervised machine-learning tasks, especially with non-numerical variables, is non-trivial, because the observations do not have labels that can be used for evaluation. The NMDS configurations and stress levels, however, provide arguments to improve the dataset before clustering. The insights from the initial variable selection discussed in section 6.3.2 can now be used to pursue the aim of 'selecting the minimum number of variables while preserving as much information for the interest variable of the system to be modelised' (cf. Bouhamed, Lecroq & Rebaï,

2012, p. 10). For the initial variable selection I only used univariate insights, and I will complement these analyses with dimension reduction as a multivariate method to assess and improve the clusterability of the dataset.

Removing features that were indicated as problematic by either a high frequency ratio (focus particles) or low theoretical relevance (aspect, person and number) improved the NMDS configurations (for details, see Appendix C.6). Therefore, only the remaining features (clause order and syntactic integration, verb tense, modality, sentence type, and negation) will be considered in the analyses in the remainder of this chapter. The results of dimension reduction on these remaining features in the dataset are presented in the NMDS-configurations in Figure 6.4 below.

Figure 6.4:
NMDS configurations and stress levels for distance matrices (reduced feature set)



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

In Figure 6.4 the 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 on page 501 in section C.5.3 of Appendix C). As was the case with the results in Figure 6.3, Gower's measure continues to perform well despite its lack of any probability information processing (see previous section). At this point, however, it seems warranted to conclude that the degree in which the distance matrices indicate possible underlying structure in the dataset has been enhanced, with the important remark in order that stress levels are used here as indices of possible underlying structures only, and that these measures, as mentioned above, are not flawless. For example, Dexter, Rollwagen-Bollens and Bollens (2018, p. 440) show how simulation data that

were designed to be 'moderately structured' failed to reach stress levels below 0.20. Therefore, evaluation procedures in subsequent steps of the exploration will be used to minimise the aforementioned risk of misinterpretation, and to monitor the performance of Gower's measure in subsequent steps.[28] Before doing so, however, we will inspect the distances in a more qualitative manner to find out which conditionals in the corpus are most representative of all conditionals, and which are least representative.

6.3.7 Identification of representative conditionals

In most studies (see references above) distance matrices are only used as input for further analyses. Levshina (2011, pp. 71–72), however, shows how a distance matrix can be insightful in itself (see also Levshina, 2015, chapter 15). More specifically, a distance matrix can be used to calculate the representativity of what will be clustered, i.e., conditionals in this study. Please note that this is overall representativity, as the data has not yet been clustered. As this step is not needed for clustering, neither directly addresses any of the research questions in this dissertation (see section 2.7), I will present the most and least representative conditionals in the complete dataset as an intermezzo. The reason for doing so is that having an overview of how the features are collectively distributed in the corpus does provide us with a picture of the grammatical form of conditional constructions in Dutch, which is relevant to the study as a whole. Furthermore, as the previous evaluations were all primarily quantitative, this step provides an opportunity to return to the a more qualitative look at the object of this study before moving on to the actual clustering.

In line with Levshina (2011), who reports on the minimum, maximum and mean distances in her dataset, we will inspect the distribution of distances for the current dataset. Although in subsequent steps all the distance matrices discussed will be used for evaluative purposes, we need to answer the question which distance matrix to use for the identification of the most and least representative conditionals. The results of Gower's measure introduced the least amount of stress (see discussion above), but its NMDS configurations did not show clear and separable groups of data, and the histogram was, even given the categorical data used, most reflective of a unimodal distribution. For these reasons, the Eskin measure was used for identifying the most and least representative conditionals, as it did show some signs of a multimodal distribution, resulted in the lowest stress level after Gower's measure, and it showed clearly separated groups in the NMDS configurations, all indicating possibilities of detecting structures underlying the data.

Calculating the average distance from every conditional to every other conditional provides an index of representativity (or 'prototypicality'; cf. Levshina, 2011, p. 72), which can then be used to inspect the most and least representative conditionals in the corpus, which correspond to conditionals with the lowest

[28] This will, of course, also be done for the other measures.

and highest average distance respectively. Using the Eskin distance matrix, the minimum and maximum mean distances are 0.21 and 0.88 (on a scale of 0-1), and the average mean distance was 0.33 with a standard deviation of 0.12.[29] All conditionals were ranked based on their average distance to all other conditionals. As a number of conditionals have equal average distance, which is not surprising given the number of conditionals in the corpus, the five representative conditionals in (402)-(406) below are a random selection from the conditionals with the lowest average distance (i.e., highest representativity).

(402) Nederland is voor werkgevers goedkoop als het gaat om loonkosten. (fn006272)
The Netherlands is cost-effective for employers when it comes to labour costs.

(403) De overlevingskans van een baby wordt kleiner als de moeder rookt. (fn006645)
A baby's survival rate decreases if the mother smokes.

(404) Voor de nieuw toegetreden verzekeraar ligt dat anders als de gemelde omstandigheid tegelijkertijd dateert van voor het oversluiten naar hem. (WR-X-A-A-journals-nthr-008)
This is different for the newly joined insurer if the reported circumstance simultaneously dates from before the transfer.

(405) Nederland staat nog steeds bij de top drie in de wereld als [het] gaat ook om kwaliteit van bulkproducten. (fn000221)
The Netherlands is still among the top three in the world {if/when} [it] also comes to the quality of bulk products.

(406) De Noren komen terug als de regering en de president hun ruzie bijleggen. (WR-P-P-G-0000116371)
The Norwegians return {if/when} the government and the president settle their quarrel.

Looking at these examples, the question may rise why these examples are considered representative conditionals. Here, 'representativity' is a purely quantitative notion, based on the difference between conditionals in terms of their features. What we see in the examples in (402) to (406) is that these conditionals all agree on the features selected in the previous section: they have simple present tense in both clauses, no modal marking and no negation in either clause, declarative consequents, and, more surprisingly, sentence-final antecedents. This is surprising because the sentence-initial order is presented as the default clause order throughout the literature (see section 5.2). What this

[29] Please note that this does not mean that no conditionals had a distance lower than 0.31 or higher than 0.84, as these summary data represent mean distances.

shows, is that conditionals with sentence-final antecedents show stronger agreement on the other features than conditionals with sentence-initial or sentence-medial antecedents.

Having carried out the same procedure on the distances calculated using all variables, we can see that a small amount of variation indeed only occurs in person and number of subjects in antecedents (23% 2ps, 77% 3ps; variation occurring in the bottom of the top-100), and no variation in person and number of subjects in consequents (100% 3ps). None of the clauses contain any form of negation, and all have simple present tense in both clauses. Modal marking is absent too from the top-100 in both clauses. Aspect does show some variation in the antecedent (15% achievement, 85% state), but none in the consequent (100% state). There was considerable variation in syntactic patterns (comprised of clause order and syntactic integration, see section C.2 in Appendix C), with 58% sentence-final antecedents, 33% integrative conditionals and 9% resumptive patterns.[30] What this suggests, is that the selected variables do not vary for the most similar conditionals and provide the most robust basis for similarity measuring in the dataset, and that the variables showing variation have less impact on the similarity between conditionals. This does not hold for syntactic integration, which, as we have seen before, may introduce unused variation in the dataset. From these results, we can already observe that only (403) and (406) are of the 'prototypical' predictive (i.e., causal) type. We should not draw strong conclusions from this, however, although it might indicate that the features used are not able to differentiate between the types discussed in the literature. This will point will be taken up later on in this chapter.

Next, we will look at the least representative examples, i.e., those conditionals which, on average, share the least features with all other conditionals, presented in (407)-(411).

(407) Als we dit niet hadden gewild dan hadden we er maar niet moeten gaan wonen. (WR-U-E-A-0000000078)
If we hadn't wanted this we shouldn't have decided to live there.

(408) Ik zou van [mevrouw] Van Gent willen vragen of zij dan niet juist zou vinden dat als je verlof zou betalen tegen een bepaald percentage van het wettelijke minimumloon juist dat niet meer uh nivellerend zou werken dan wanneer je het zou betalen tegen een percentage van het salaris. (fn000167)
I would like to ask Ms Van Gent whether she would agree that if you paid leave at a certain percentage of the statutory minimum wage, it would work more levelling than if you paid leave at a percentage of the salary.

[30]Please note that the distribution of these 100 conditionals is evenly spread over mode (47% spoken, 53% written) and somewhat less evenly over register (61% formal, 39% informal).

(409) En als we ergens mee kunnen helpen zou al uh ongepast zijn als je dan niet zou doen. (fn008220)
And if we can help with anything, uh, it would be inappropriate if you didn't.

(410) De Franse president Chirac en de Duitse kanselier Schröder *zouden* Prodi bij wijze van spreken om de nek *zijn gevlogen* als hij Solbes – de man die met ijzeren hand regeert over de begrotingstekorten in de lidstaten – in deze economisch zware tijd onschadelijk *zou hebben gemaakt*.[31] (WR-P-P-G-0000105269)
The French President Chirac and the German Chancellor Schröder would have hugged *Prodi, so to speak, if he* would have defused *Solbes – the man who rules the budget deficits in the Member States with iron – in this economically difficult time.*

(411) Met de afspraak van als niet zou lopen dat we dan meteen vrij snel zouden gaan zakken. (fn008261)
With the agreement that if it were not to work out, we would immediately go down rather quickly.

What we see here is variation across features in these examples. Although all features show variation, I focus here on the most notable difference with respect to the representative examples in (402) to (406). As we can see in (407) to (411), all examples license an implicature of unassertiveness, which we discussed in terms of 'epistemic distancing' in section 2.5.4, and which corresponds mostly to a counterfactual interpretation. This implicature seems to be licensed by past verb tenses in combination with modal marking, as all examples except the conditional in (407) feature the past tense of the modal auxiliary *zullen* 'will' (*zou* 'would'). More specifically, there is a higher percentage for simple past in both clauses (49% and 48%), and lower frequencies for simple present (22% and 35%), past perfect (20% and 16%), and present perfect (9% and 1%). With respect to modality, we see that antecedents are most frequently unmarked for modality (40%), followed by marking of epistemic, dynamic, deontic and evidential modality (39%, 12%, 5% and 4% respectively), whereas consequents are most frequently marked for epistemic modality (49%), followed by no modal marking, marking of dynamic, deontic and evidential modality (27%, 12%, 9% and 3% respectively). Another indication of the counterfactuality of these examples can be observed in the high frequency of negation, which was already hypothesised by Wierzbicka (1997) as discussed in section 3.2.10, as most of the clauses feature syntactic negation (72% for antecedents, 63% for consequents), with lower frequencies for non-negated clauses (15% and 21% respectively) and morphologically negated clauses (13% and 16%). These distanced or counterfactual conditionals thus differ from other conditionals on the features mentioned, with the strongest deviations in tense, modality and negation patterns.

[31] This example is repeated from page 269.

In other words, as most conditionals have simple present tense in both clauses, and no modal marking or negation in either clause, diverging tense, modality and negation patterns quickly add up to the average distance, and it is therefore expected that these least representative conditionals will form a group in consequent clustering results in section 6.4. To complete the description of the least representative set of conditionals, I note here that there is more variation in clause order and syntactic integration too. Whereas the representative examples showed an exclusive preference for sentence-final antecedents, non-representative examples also have a preference for this clause order, but less strongly so (33%), followed by the integrative pattern (26%), resumption (25%), sentence-medial antecedents (9%) and, finally, non-integration (7%).

The inspection of the most and least representative examples, and especially the latter, give but an indication of how conditionals are ranked by means of their average dissimilarity to the other conditionals in the corpus. Nevertheless, the fact that the least representative examples all appear to be clearly marked for epistemic distance or counterfactuality is an interesting result in itself, because it suggests that this specific implicature of unassertiveness is indeed marked by grammatical means, and it is expected that this result will be reflected in the clustering results, which we will discuss next.

6.3.8 Conclusion

Based on literature discussed in chapter 3, I selected features of conditionals that were suggested to be related to implicatures of unassertiveness and connectedness. As discussed in terms in the previous section, it is expected that the repeated use of certain patterns of these features has conventionalised into grammatical constructions, i.e., pairings of form and meaning. While the literature does indeed suggest incorporating the grammatical form of the conditional as a complete complex sentence, it is of course possible that the implicatures central in this dissertation remain particular instead of generalised or conventionalised, and cannot in fact be analysed as grammatical constructions. In that case, the dataset is not expected to show signs of underlying structures. To address this very issue, I evaluated the clusterability of the dataset in this section. Both the theoretical and statistical assessment of the distance matrices suggested removing aspect, person and number, and focus particles from the dataset to improve clusterability. The evaluations of the final feature set indicate Gower's measure and the Eskin measure to be the most promising for the current dataset, as they produce stress levels indicating reasonable levels of clusterability. The distance matrices also allowed for the identification of the most and least representative conditionals in the corpus, and the especially the latter group showed clear signs of patterns of features related to implicatures of unassertiveness. None of the tests provided definitive grounds for definitive conclusions on clusterability, however. A likely reason for this is that the available tests and evaluations used originate from the field of machine learning and are tested mainly on numerical data, whereas the current linguistic dataset

consists of categorical data. The results were therefore treated as preliminary evaluations, rather than definitive assessments clusterability. Consequently, all distance measures will be included in the final clustering analyses, which we will turn to in the next section.

6.4 Clustering and evaluation

6.4.1 Introduction

Now we have performed all preparatory steps, we can finally turn to the clustering of conditionals to explore and assess relations between the form and meaning of conditionals. In more specific terms, we will explore the extent to which the grammatical features of Dutch conditionals selected systematically license implicatures of unassertiveness and connectedness, and thus can be seen as grammatical constructions.

In this section, the distance matrices presented in the previous section function as input for two types of cluster analysis. In section 6.4.2, I will discuss the main clustering algorithms used, and in section 6.4.3, the evaluation methods of clustering solutions will be introduced. In section 6.4.4 I will present, evaluate and discuss the results of the hierarchical clustering approach used, and in section 6.4.5, I will do the same for the partitional approach. By means of a choice for the optimal clustering solution, a conclusion is presented in section 6.4.6, before moving on to the analysis of the clusters in section 6.5, in which I will focus on the relation between the clusters and the implicatures of unassertiveness and connectedness central in this dissertation.

6.4.2 Clustering algorithms

While there exist many clustering algorithms, the most notable types are *partitional*, *hierarchical* and *model-based* or *density-based* clustering algorithms (for overviews, see Aggarwal, 2014, chapter 1; Ester, 2014). I discuss each of these algorithms briefly, before selecting the appropriate algorithms to use in the remainder of this section.[32]

Partitional algorithms such as *K-means clustering* (MacQueen, 1967; Kaufman & Rousseeuw, 1990, pp. 113–114) and *K-medoids clustering* or *Partitioning Around Medoids (PAM)* (Kaufman & Rousseeuw, 1990, chapter 2) simultaneously attempt to divide a dataset into k groups, with k set to a constant value by the researcher. While setting k to a constant value may seem

[32]Model-based clustering assumes the data come from a number (k) of Gaussian distributions (see Fraley & Raftery, 2002, p. 612), as is discussed in section C.4 of Appendix C. Important limitations exist, however, including the aforementioned assumption of Gaussian-distributed data, and problems with high-dimensional and large datasets (cf. Fraley & Raftery, 2002, pp. 625–628). This type of clustering was therefore discarded for use in this study, and will not be discussed further.

problematic, as the number of groups is not always known beforehand, an often used solution is to generate clustering solutions for a range of k-values and then selecting the best solution from the results. Partitional algorithms search for partitions in which the within-cluster variance is minimised, while the between-cluster variance is maximised. Although this might be said to be the aim of all types of clustering algorithms, the difference with hierarchical clustering (see below), is that partitional algorithms do not work by step-wise grouping or dividing pairs of (clusters of) observations, and it therefore does not suffer from 'the defect that it can never repair what was done in previous steps' (Kaufman & Rousseeuw, 1990, p. 44; see also Oyelade et al., 2016), which is the case for hierarchical algorithms. On the other hand, it does not provide a hierarchy of clusters, which is sometimes wanted, for example when uncovering evolutionary trees in biological studies (Rohlf, 1970; Sneath & Sokal, 1973; Murtagh & Contreras, 2017, for a recent overview). A partitional clustering algorithm 'simply' divides the dataset into non-overlapping subsets. The requirements are that there cannot be empty clusters, and each observation must belong to exactly one cluster (cf. the concept of 'mutual exclusivity' discussed in section 6.2.3). This latter requirement is loosened somewhat in a special type of partitional algorithms, namely *fuzzy partitioning*, implemented in algorithms such as *Fuzzy c-means clustering* or *FCM* (cf. Bezdek, Ehrlich & Full, 1984), and *Fuzzy Analysis* or *FANNY* (see Kaufman & Rousseeuw, 1990, chapter 4). These algorithms do not provide so-called 'hard clustering' solutions, in which each observation must belong to one cluster only, but they provide 'soft clustering' solutions, in which each observation is scored on 'degree of belonging' for each cluster, quantified as membership coefficients between 0 and 1 (see Kaufman & Rousseeuw, 1990, p. 164). The advantage of this 'fuzzy approach' is that it takes into account that real data do not always contain clear-cut cases only. The disadvantage is that computation times are considerably longer for large datasets and that the results are harder to interpret, because many observations may belong to multiple clusters.

In contrast to partitional algorithms, hierarchical clustering algorithms work incrementally, either top-down or bottom-up (for an accessible introduction, see Andritsos & Tsaparas, 2010). Top-down or *divisive* algorithms divide the data into two clusters recursively until all data are clustered, implemented in divisive algorithms such as *Divisive Analysis* or *Diana* (see Kaufman & Rousseeuw, 1990, chapter 6). This algorithm starts off from one cluster in which all data are gathered, and then splits it until every observations forms its own cluster. At each step, the observation with the largest distance to the other observations is treated as a cluster, and all other observations closer to this cluster than to the rest of the observations are added to the cluster. Bottom-up or *agglomerative* algorithms such as *Agglomerative Nesting* or *Agnes* (see Kaufman & Rousseeuw, 1990, chapter 5), begin with a cluster for each observation and combine nearest clusters until the top of the hierarchy is reached, i.e., until the number of clusters k is 1. Divisive and agglomerative clustering algorithms produce a hierarchical clustering solution, which can be represented in a tree-structure, known as a

dendrogram. As mentioned above, depending on theoretical foundations, such tree-structures can be interpreted as 'a sufficiently accurate model of underlying evolutionary progression' (Murtagh & Contreras, 2017, p. 3). In contrast to partitional algorithms, the number of clusters k is thus not pre-defined and rather refers to the moment (or 'height' in the tree-structure) at which groups are defined.

Choosing an algorithm depends on the nature of the data and theoretical considerations.[33] Theoretically, and with respect to cognitive theories, the choice for both a partitional or a hierarchical approach can be defended. As discussed by Divjak (2010, pp. 9–10), on the one hand, Langacker (1987, pp. 369–371) argues complex categories are best viewed as '(hierarchical) schematic networks of interrelated senses', in which a schema is an abstraction compatible with all (more concrete) members it defines, while, on the other hand, Lakoff (1987, pp. 83–84) views categorisation as a radial structure, in which no hierarchy exists, but categories are built around a 'central case' and conventionalised (i.e., non-rule generated) variations on this central case. Langacker's view would promote a hierarchical approach, whereas Lakoff's view would promote a partitional approach. With respect to conditionals, given the arguments by Dancygier (1998, pp. 184–185) and Athanasiadou and Dirven (1997a, p. 89), already briefly discussed in section 6.2 and in sections 3.3.7 and 3.3.9 before, hierarchical clustering is favoured because less prototypical conditionals can be seen not as part of distinct subsets, but as specifications of the properties of conditionals in general. In both clustering approaches, prototypes can be identified by measuring how well they fit or define their cluster, for instance by calculating the 'silhouette widths' of clusters, as we will see below. Another, more methodological consideration is that hierarchical methods have been tested more extensively on categorical data, especially in in corpus linguistic studies (Faure & Nédellec, 1998; Gries & Stefanowitsch, 2010; Divjak & Gries, 2006; Divjak & Gries, 2008; Berez & Gries, 2008; Hilpert & Gries, 2009; Gries, 2010; Gries & Otani, 2010; Gabrielatos, 2010, pp. 52–53; Levshina, 2011, chapters 4, 5; see also Tang, 2017). A downside is that this is a form of 'hard clustering', and it may be said that conditionals may exhibit less clear-cut boundaries and membership should be expressed in probabilistic instead of deterministic terms.

Both hierarchical and partitional approaches have their theoretical merits. While the hierarchical approach is more frequently applied in linguistic studies, this is not a reason to discard the partitional approach to clustering. As this study can be seen as both an exploration of Dutch conditionals as grammatical constructions, and as an exploration of applying advanced data analysis methods to linguistic data, I will subject the dataset to both types of clustering, and I will use several indices to evaluate their results. We will discuss these measures of cluster evaluation in the next section.

[33]The choice for an algorithm also depends on available implementations, and although this should, of course, not be the primary reason for algorithm choice, in a practical sense, it is an important factor.

6.4.3 Measures of cluster evaluation

As is suggested in most of the literature on unsupervised machine-learning techniques (see section C.4 of Appendix C), the most common strategy for choosing a type of clustering, a specific algorithm and determining its parameters, such as number of clusters k, is to try different solutions and evaluate the results, both internally and comparatively. As we discussed in section 6.2, direct validation using external labels is not possible for unsupervised machine-learning, and Jain and Dubes (1988) provocatively describe the nature of cluster validation as follows.[34]

> The validation of clustering structures is the most difficult and frustrating part of cluster analysis. Without a strong effort in this direction, cluster analysis will remain a black art accessible only to those true believers who have experience and great courage. (Jain & Dubes, 1988, p. 222)

As no class labels are available, the question remains how to evaluate outcomes of clustering algorithms. I will validate the results of both hierarchical clustering (section 6.4.4) and partitional clustering (section 6.4.5) using a number of indices of cluster validity. The reason for multiple indices is that each index measures another aspect (or combination of aspects) of the quality of a particular clustering solution, for instance a measure of how how well clusters are separated, or how consistently conditionals have been assigned to clusters. Unfortunately, this means that no single index can answer the question of what grouping of data is best, somewhat like the example of cutlery discussed at the start of this chapter. I will describe the indices used in general terms here, but they are discussed in (technical) detail, including references, in section D.2 of Appendix D.

For cluster homogeneity, Within-Cluster Entropy (WCE) (Šulc & Řezanková, 2019) was used. A low WCE value indicates low within-cluster variability and high homogeneity. For cluster separation, the Pseudo F Coefficient based on Entropy (PsFE; Sevcik, Rezankova & Husek, 2011; Šulc, 2016) was used. A high PsFE value indicates strong cluster separation. For cluster consistency, the Silhouette Coefficient (Kaufman & Rousseeuw, 1990) was used. A high Silhouette Coefficient indicates strong cluster structures. For dispersion over clusters, Deviation from the Mode (DM; Wilcox, 1973) was used. Low values indicates extreme differences in cluster size, i.e., the risk of ending up with clusters of only one or two conditionals. Finally, for cluster stability, the Jaccard Coefficient (Hennig, 2007) was used. A high coefficient indicates stable clustering.

[34] There exist alternative approaches, such as 'semi-supervised' algorithms which evaluate classification based on a small number of labels. See e.g., Chapelle, Schölkopf and Zien (2006).

As mentioned, there is no one value that will indicate 'the best' clustering solution, and therefore I will combine these indices and compare them between solutions. This will be the topic of the following two sections, starting with the evaluation of hierarchical clustering solutions in 6.4.4, followed by the evaluation of partitional clustering solutions in 6.4.5.

6.4.4 Hierarchical clustering

As discussed in section 6.4.2 above, hierarchical cluster algorithms can be either of the *agglomerative* or *divisive* kind, although most studies only mention 'hierarchical cluster analysis' in referring to agglomerative clustering. Because computation for agglomerative clustering is both more efficient than divisive clustering (Kaufman & Rousseeuw, 1990, p. 253) and used more widely in various fields, including corpus linguistics (Gries & Stefanowitsch, 2010, see e.g., Divjak, 2010; Levshina, 2011), I used agglomerative cluster analysis as well. The most important parameters of hierarchical algorithms are the number of clusters k, as discussed above, and *linkage*. Linkage determines how an algorithm calculates the distance between two clusters, i.e., how the 'closeness' of two clusters is defined (see Kaufman & Rousseeuw, 1990, pp. 45–48). Using single linkage, the similarity between two clusters is defined as the distance between their two most similar members, and consequently, this local approach merges the two clusters with the smallest distance between their most similar members. Using complete linkage, 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. Average linkage is a compromise between single and complete linkage, and 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), which calculates the distance from each observation to the centroid (the mean distance) of the cluster it is assigned to. All combinations of k and linkage were systematically evaluated using the the evaluation criteria discussed. A detailed account of the results can be found in section D.3 of Appendix D.

After evaluating clustering solutions in terms of homogeneity, separation, consistency, dispersion, and stability, the next step was to select the optimal solution. This was, as discussed, not a trivial task, as the evaluation measures mentioned all reflect different qualities of the solutions generated, and no one combination of linkage methods and distance measures proves uniformly superior. As a first step, I have discarded solutions with low dispersion values. Dispersion allowed for discarding solutions that score high on the other measures, but in reality propose one very big cluster and a small number of clusters with only a small number of conditionals. Second, unstable solutions were discarded, as low stability indicates that the results are highly dependent on sampling. This excluded solutions with single and complete linkage, and thus left solu-

tions generated using average and Ward's linkage (see section D.3 in Appendix D for details). Third, using Silhouette Coefficients, we see relatively high and stable consistency values for 4- to 6-cluster solutions for the Lin measure. Most solutions, however, should be interpreted with caution, as the structures found have consistency values mostly between 0.4 and 0.5 on a scale from 0 (no structure) to 1 (perfect structure). As the Lin measure produced solutions with Silhouette Coefficients just below 0.5, this suggests that somewhat reasonable structures were found. It must be noted, however, that these clustering results are not in line with the evaluation of clusterability by dimension reduction in discussed in section 6.3.5.[35] Fourth, evaluation of within-cluster variation shows lowest values for solutions using the Lin measure, average linkage and k 4-6, which means that these solutions hold the most homogeneous clusters. After discarding remaining solutions based on low Silhouette Coefficients (VE, VM) or low dispersion values (IOF), cluster separation was measured in terms of PsFE. This, again, was highest for 2- to 6-cluster solutions using the Lin measure with average linkage, which means that these solutions not only hold homogeneous clusters, but also that these clusters are more clearly separated than in other solutions. These combined evaluations suggest the 4- to 6-cluster solutions based on the Lin measure with average linkage to be the optimal hierarchical solutions for the current dataset.

In Table 6.6 below, the membership distributions of the selected hierarchical clustering solutions are presented. These figures show the sizes of the clusters for the 4, 5 and 6 k solutions. We can, for instance, see that the third cluster is the biggest, followed by cluster 1 and 2, whereas clusters 4, 5 and 6 are relatively small.

Table 6.6:
Membership distributions of Lin average-linkage solutions (4-6 clusters)

	Cl. 1	%	Cl. 2	%	Cl. 3	%	Cl. 4	%	Cl. 5	%	Cl. 6	%
4 cl.	597	*14.53*	546	*13.29*	2774	*67.51*	192	*4.67*				
5 cl.	597	*14.53*	546	*13.29*	2401	*58.43*	373	*9.08*	192	*4.67*		
5 cl.	597	*14.53*	546	*13.29*	2021	*49.18*	373	*9.08*	192	*4.67*	380	*9.25*

The solutions are stable in their membership distributions in the first two clusters, and in the fourth cluster of the 4-cluster solution, which is the fifth cluster in the 5- and 6-cluster solutions. This is due to the hierarchical nature

[35]Note that this should not have large repercussions for the discussion of representative conditionals in section 6.3.7, as the the average Eskin distances used there and the Lin measure selected here are positively correlated, as indicated by a Spearman's correlation test ($R_s = 0.90$, $p<0.001$). This test was chosen over Pearson's Correlation Coefficient because it does not require normally distributed variables (see section 6.3.5).

of the clustering. The added clusters in these latter solutions all come from the biggest cluster (cluster 3), which was split into a new cluster in the 5-cluster solution and into two new clusters in the 6-cluster solution.

While a possible next step is to review the actual contents of these clusters, we will first review the evaluations of clustering solutions using partitional algorithms, which then can be compared to the results of hierarchical clustering, before analysing the results with respect to implicatures of unassertiveness and connectedness, and their possible relations to the feature distributions in each cluster.

6.4.5 Partitional clustering

As discussed in section 6.4.2 above, partitional algorithms do not incrementally build a structure in either top-down (*divisive*) or bottom-up (*agglomerative*) fashion, but they consider all distances at once. In general, partitional algorithms first select the k most representative observations from the dataset (*centrotypes*, or *medoids*), and then k clusters are formed around these representative observations by choosing the closest representative object for each of the other observations. The two main parameters are the specific algorithm used, and the number of clusters k. Two algorithms were used in this study, of which the first was 'Partitioning Around Medoids' (PAM), described in section 6.4.2. This algorithm 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). The algorithm works in two steps. First, in the so-called 'build phase', the algorithm selects k 'medoids' (i.e., most representative points) and it allocates 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 dissimilarity no longer decreases. As an observation can only be member of one cluster, this is a form of hard-clustering. The second algorithm used was 'Fuzzy Analysis' (FANNY), which is a form of soft-clustering, as it assigns to each object a membership coefficient indicating how well that particular object fits within each cluster. In contrast to PAM, this approach does not choose representative observations as medoids, but it minimises the dispersion over all clusters for each observation. The algorithm is also capable of hard-clustering by simply selecting the cluster with the highest membership coefficient for each object. The second parameter is the number of clusters k, which should be defined on a theoretical basis, and/or, as was done in this study, evaluated for a range values for k. Note that linkage is not relevant for these algorithms, because each membership assignment is determined by comparison of two objects only: the cluster-representative and the observation to assign membership to. The results of the combinations of algorithm and number of clusters were systematically evaluated using the same evaluation criteria as discussed in the previous section. A detailed account of the results can be found in section D.4 of Appendix D.

As was the case with the evaluation of hierarchical clustering solutions, no single combination of algorithm and distance measures proved uniformly superior to other solutions, and selecting the optimal solution remains a non-trivial, to some degree interpretive task of the researcher. First, I discarded solutions with low dispersion values, and second, those with low stability values. These evaluations allowed discarding all solutions using the Goodall and VE measures, and in case of the FANNY algorithm, solutions generated using the VM measure. Third, using Silhouette Coefficients, I identified relatively high and stable values for 2- to 4-cluster solutions for the Lin and Lin1 measures. These solutions have Silhouette Coefficients around 0.5, which is around the lower bound of what Kaufman and Rousseeuw (1990, p. 88) call 'reasonable structure'. Fourth, within-cluster variation was similar for most measures in the PAM solutions, but, again, for FANNY the lowest values, reflecting the most homogeneous clusters, were found for 2- to 4-cluster solutions for the Lin and Lin1 measures. Cluster-separation, measured in terms of PsFE, was high and most stable for these solutions too. While PAM seems to produce better on average, we can see here that the FANNY algorithm using the Lin1 measure produced the highest Silhouette Coefficients for 2- to 4-cluster solutions, while having low within-cluster variability, highest between cluster separation values, average to high dispersion values, and high stability values. The combined evaluations suggest the 2- to 4-cluster FANNY solutions based on the Lin1 measure to be the optimal partitional solutions for the current dataset.

In Table 6.7 below, the membership distributions of the selected partitional solutions are presented. These figures show that partitional solutions involve more evenly distributed cluster memberships in contrast to the hierarchical solutions selected in the previous section.

Table 6.7:
Membership distributions of Lin1 FANNY solutions (2-4 clusters)

	Cl. 1	*%*	Cl. 2	*%*	Cl. 3	*%*	Cl. 4	*%*
2 cl.	1638	*39.86*	2471	*60.14*				
3 cl.	1337	*32.54*	1633	*39.74*	1139	*27.72*		
4 cl.	1058	*25.75*	658	*16.01*	1394	*33.93*	999	*24.31*

In all three solutions, there is one larger cluster and one or a number of smaller cluster, which, however, are still sizeable. Please note that, for reasons of comparison to the results of hierarchical clustering, Table 6.7 presents the hard-clustering results. See section D.4 of Appendix D for membership coefficients for each of the solutions above.

6.4.6 Conclusion

In the previous two sections, we discussed the evaluations of a large number of systematically generated clustering solutions, using both hierarchical and partitional algorithms. For both approaches, one set of solutions (comprised of a small range of cluster numbers k) was selected. For the hierarchical approach to clustering, this was the Lin measure using average linking and k 4-6. For the partitional approach to clustering, the FANNY algorithm using the Lin1 measure and k 2-4 was selected. As the Silhouette Coefficients for these solutions were around 0.5 (the minimum for 'reasonable structure' cf. Kaufman and Rousseeuw, 1990, p. 88), and only quantitative measures were used to arrive at these two sets of solutions, the real test is, of course, to interpret the results in qualitative terms, i.e., can the data-driven clusters be motivated theoretically with respect to grammatical features and implicatures? This will be the main question in the next section.

6.5 Analysis of hierarchical clusters

6.5.1 Introduction

In this section, I analyse the clusters present in the hierarchical solutions in terms of their distributions of grammatical features and possible implicatures of unassertiveness and connectedness. I will discuss these implicatures, introduced at the start of this dissertation in chapter 2, in relation to the types distinguished in the accounts discussed in chapter 3 and the features distilled from these accounts and inventoried in chapter 5. For each cluster, I will first discuss its internal feature distribution, and I will present conditionals representative of that cluster. Next, the conditionals in the cluster are analysed in term of the implicatures discussed in chapter 2, and these are compared to possible matches on types of conditionals from the accounts discussed in chapter 3.

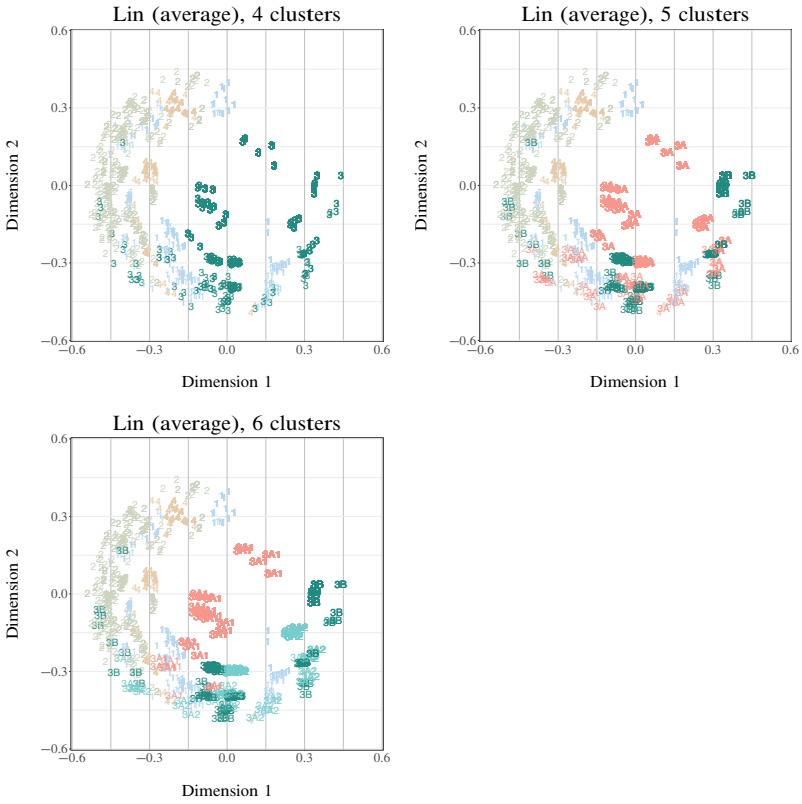
In section 6.5.2, I will present an overview of the hierarchical clustering solution and the feature distributions for each cluster. Then, in section 6.5.3, I will offer a preliminary remark with respect to the comparison of clusters and types of conditionals, which is needed before we can move on to sections 6.5.4 to 6.5.7, in which I will discuss each cluster in the fashion outlined above. As the evaluations in section 6.4.4 did not provide definitive arguments for a 4-, 5- or 6-cluster solution, the additional clusters will be discussed in section 6.5.8. In section 6.5.9 I will provide a brief conclusion on the results of hierarchical clustering, before moving on to the analysis of the partitional clusters in section 6.6.

6.5.2 Clusters and feature distributions

In section 6.4, I selected the 4- to-6 cluster solutions that were generated using the Lin measure and average linkage. In this section, I inspect the characteristics of each cluster. In order to visualise the clusters, the same dimension reduction technique as in section 6.3.5 was used, i.e., non-metric dimensional scaling (NMDS). As the clustering has been performed at this point, however, it is possible to add cluster memberships to the existing configuration to see whether memberships are systematically placed on the ordination axes, as can be seen in Figure 6.5.[36]

[36]Due to the large number of observations, the traditional visualisation of hierarchical clustering solutions, i.e., a dendrogram, provides a less insightful, and harder to read overall picture. As it is the standard, however, a dendrogram is included in section D.5 of Appendix D.

Figure 6.5:
NMDS configurations with memberships from hierarchical clustering



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

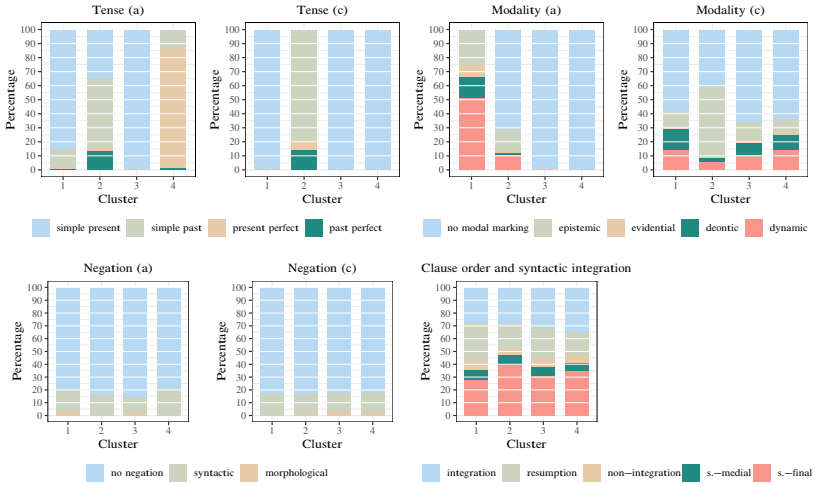
As a reminder of what was already observed in sections 6.3 and 6.4, in terms of stress levels and separation in the NMDS configurations (see Figure 6.4 in section 6.3.6), Lin did not perform as well as other measures. The measure did, however, perform best with respect to the cluster evaluations presented in section 6.4. As these evaluations are more specific to the clustering aim in this study, they are well-tested on categorical data, and the provided converging evidence, Lin was chosen over the other measures.

What we see in the left panel of Figure 6.5, which presents the results of the 4-cluster solution, is that the largest cluster, cluster 3 (67.51%), is positioned towards the lower-right corner of the configuration. The second largest cluster, cluster 1 (14.53%), is not well-separated from the other clusters, whereas cluster

2 (13.29%) is, being positioned at the upper-left hand of the configuration. The smallest cluster, cluster 4 (4.67%), finally, is positioned at the left, with most of the observations on the upper-half of the configuration. In the middle panel of Figure 6.5, we can see how cluster 3 is sub-divided into clusters 3A (middle) and 3B (lower-right), and in the right panel we see that, when a 6-cluster solution is selected, cluster 3A from the middle panel is sub-divided into clusters 3A1 and 3A2, which show clear separation occupying the groups of conditionals in the middle of the panel, and, predominantly, the lower-right respectively. In summary, we can see that clusters 2 and 4 overlap strongly, but combined they are clearly separated from the largest cluster, cluster 3. Cluster 1 shows overlap with virtually all other clusters.[37]

Before discussing each cluster in more detail, their feature distributions are presented in Figure 6.6 below.

Figure 6.6:
Feature distributions for the hierarchical 4-cluster solution



Note. These distributions represent clusters generated using the Lin measure and average linkage. Numbers on the horizontal axes correspond to cluster numbers reported in this section.

In sections 6.5.4 to 6.5.7, we will review the clusters in the 4-cluster solution.

[37] Any hierarchical information is lost in these two-dimensional NMDS configurations. Therefore, see also the dendrogram in section D.5 of Appendix D.

6.5.3 A note on comparing clusters and types

Before discussing the clusters in the next sections, a remark on the comparison between these clusters and types of conditionals proposed in the literature is in order. As discussed in chapter 4, there were several reasons not to annotate each conditional in the corpus for the type of conditional (see section 4.3.3). In short, I provided the following three arguments. First, annotating types of conditionals based on language-specific features of English conditionals may not be applicable to Dutch conditionals, as it would assume universal or non-language specific types to exist. Second, choosing a number of classifications to apply would introduce theoretical bias, possibly discarding useful classifications. Third, and most pressing for the remainder of this chapter, applying theoretical classifications to actual corpus data revealed low reliability. Therefore, comparing the conditionals in the clusters found to types from classifications, as I will do below, is not without problems, and I will briefly discuss this point here.

The comparison between types and clusters of conditionals must be seen as an attempt to interpret the results in light of the theory, not as applying a 'gold standard' and thereby reintroducing the problems addressed above (see also section 6.2.5). Although, of course, not all conditionals will equally likely resist clear-cut classification (for an elaborate discussion, see section 4.2), and there will undoubtedly be conditionals that do constitute clear types from the literature, the comparisons in what follows must be seen as what could be called a 'silver standard' approach,[38] akin to Beekhuizen, Watson and Stevenson's (2017) approach in comparing clusters of indefinite pronouns to their semantic function in terms of Haspelmath's (1997) account. To be clear on terms, note that multiple uses of the concept of 'silver standard' annotations can be found in the literature on evaluating clustering results. Kang, van Mulligen and Kors (2012) for instance, use the term to refer to part-of-speech tags that are 'automatically generated by combining the outputs of multiple chunking systems' in order to circumvent the expensive and time-consuming creation of a gold standard. Estiri, Klann and Murphy (2019) on the other hand define their silver standard in terms of expert judgement, literature search and data distributions. Ménard and Mougeot (2019) use the term, in line with Rebholz-Schuhman et al. (2010), to refer to annotations of lower quality than gold standards, as they are not produced by 'expert annotators', but 'manually by human agents' or 'automatically by tools or trained prediction models'. In what follows, I use the term to refer to judgements based on a thorough discussion of the theory, as reflected in chapters 3 and 5, with the notable difference that the 'standard' here involves inspection of representative examples of each cluster, instead of a complete label set, for which the aforementioned arguments against attempting to constructur a gold standard would apply. This approach takes into account

[38] Suggested by B.F. Beekhuizen (personal communication, July 8, 2020).

those arguments, while utilising the insights the extensive literature provides. With this remark in place, we can continue by discussing the first cluster of conditionals in the next section.

6.5.4 Unmarked or default conditionals (cluster 3)

The largest cluster, cluster 3 in Figure 6.5 and Figure 6.6, holds 2774 or 67.51% of all conditionals in the corpus. It is therefore expected that this cluster can be described as the unmarked type of conditional. Indeed, as we can see in Figure 6.6 above, this cluster is the one least marked in terms of verb tense, as it adheres mostly to the prevalence of the simple present tense in both clauses as discussed in section 5.4. With respect to modality, antecedents rarely contain any modal marking, as expected, whereas consequents are marked for modality in 34.35%, mostly for epistemic modality (12.83%), as in (412), followed by dynamic modality (10.56%) and deontic modality (8.69%), as in (414). In a minority of cases, antecedents and consequents in this cluster contain negation, and as Figure 6.6 suggests, this in line with the other clusters formed. Representative examples of this cluster are presented in (412) to (414) below.[39]

(412) En ja als ik de alinea goed lees dan slaat dat op uh parallelimporten na vierennegentig en de mogelijke betrokkenheid van invuele [internal] overheidsfunctionarissen daarbij. (fn000142)
And yes, if I read the paragraph correctly, it refers to uh parallel imports after ninety-four and the possible involvement of internal government officials.

(413) Door deze ziekte kan hij maar drie vingers gebruiken en dat is lastig als je piano speelt. (WR-P-P-G-0000098919)
Because of this disease, he can only use three fingers and that is difficult if you play the piano.

(414) Als je dat gelooft zal het zeker zo lopen. (WR-P-E-A-0005330763)
If you believe that, it will certainly work out that way.

Cluster 3 thus looks like a cluster of default conditionals, which corresponds to its dominance in size. In terms of implicatures, these default conditionals do not share specific implicatures of unassertiveness or connectedness, although, as was argued for in chapter 2, they remain unassertive and they do implicate a connection between antecedent and consequent.

[39]These examples were selected not based on highest silhouette widths per se, but based on a combination of high-ranking silhouette widths and variation in features. The reason for this is to show some of the distributional differences occurring within the cluster. In example (414), for instance, the consequent is marked for epistemic modality, although conditionals with unmarked consequents have higher silhouette widths, as they resemble the rest of the members in the cluster more closely. These examples are thus relatively representative of the cluster and its variance.

With respect to implicatures of unassertiveness, the cluster largely holds what could be called neutral conditionals, i.e., those conditionals without an implicature of, for instance, certainty or 'actuality', epistemic distance or counterfactuality. In terms of the accounts discussed in chapter 3 (see section 3.2), the conditionals in this cluster would be classified as present conditionals (Goodwin, 1879; section 3.2.2), undetermined conditionals (Gildersleeve, 1882; section 3.2.3), present non-implicative conditionals (cf. Sonnenschein, 1892; section 3.2.4), real and future conditionals (Kaegi, 1905; section 3.2.5), open, non-past conditionals (cf. Funk, 1985; section 3.2.6), and open conditionals (Huddleston & Pullum, 2002; section 3.2.9).

The cluster of unmarked conditionals includes those conditionals in which the uncertainty often ascribed to conditionals occurs (see section 2.5), but, with respect to implicatures of connectedness, this cluster does not differentiate between these (uncertain) predictive conditionals on the one hand, as in (414), and, speech-act conditionals as in (412), or evaluative (epistemic) conditionals, as in (413). In terms of the accounts discussed in chapter 3, the conditionals in this cluster would be classified mostly as performance conditionals (Davies, 1979; see section 3.3.3), direct-open conditionals (Quirk et al., 1985; see section 3.3.4), partially determined conditionals (Johnson-Laird, 1986; see section 3.3.5), and now conditionals (Nieuwint, 1992; see section 3.3.6). The distinction between the actualising and inferential sub-types of case-specifying conditionals from Declerck and Reed's (2001) classification is not found in this cluster, and the distinction between predictive (content) and epistemic conditionals (Dancygier & Sweetser, 2005; see section 3.3.7) is not found in this cluster either, nor is the distinction between event and premise conditionals from Haegeman's (2003) account (see section 3.3.10). This can be explained by the fact that the former type is defined by *will*-deletion in antecedents of English conditionals, whereas in Dutch conditionals, the simple present without *zullen* 'will' is used most frequently for future reference, also outside the domain of conditionals (see sections 5.4 and 5.5). With respect to future reference in consequents, the example in (414) is part of a minority of conditionals in this cluster, as the non-modalised type of consequent (i.e., a consequent without *zullen* 'will') found in the example in (415) below is much more frequent.

(415) Als die niet tevreden is over de afhandelingen wordt de betrokken politieman daarop aangesproken. (fn005684)
If he is not satisfied with how the case is dealt with, the police officer involved will be approached.

This can also be seen in the distributions of epistemic modality in Figure 6.6, which shows that the majority of conditionals in this cluster features consequents that are not marked for modality.

Although the characterisation of this largest cluster is only general, the main use of this largely unmarked cluster of what could be called default conditionals lies in its contrast with the other, smaller, and as we will see, more specialised clusters. Furthermore, this cluster does already show that a number of main

types of conditionals found in the literature were not detected by the hierarchical clustering algorithm. In the discussion in chapter 7 we will come back to the implications of this observation for the relation between grammatical features and implicatures of unassertiveness and connectedness.

6.5.5 Conditionals with antecedents marked for modality (cluster 1)

The second largest cluster holds 597 or 14.53% of all conditionals and is represented as cluster 1 in Figures 6.5 and 6.6. It has a strong preference for the simple present in the antecedent (85%), which reflects the overall distribution of that feature (see section 5.4). Consequents in this cluster have simple present verb tense exclusively, which also reflects most other clusters. Negation in both clauses too reflects the general trends reported in section 5.9. These reflections may explain why this cluster is not well separated in Figure 6.5, as was observed earlier. What differentiates this cluster from the other clusters is mostly that all of the conditionals have antecedents marked for modality, with the highest frequency for dynamic modality, as in (416) below, followed by epistemic modality and accompanied by simple past tense, as in (417), consequently followed by deontic modality, as in (418).

(416) Als ik een proefrit wil maken dan regelen ze dat. (fn007730)
If I want to take a test drive, they will arrange that.

(417) Als je zou versnellen dan moet het CLB in principe eerst schoolrijpheidstesten afnemen. (WR-P-E-A-0004834951)
If you were to speed up, the CLB should in principle first conduct school readiness tests.

(418) Als wij, gynaecologen, jonge zwangere vrouwen niet mogen aanbieden te testen of hun ongeboren kind een verhoogde kans op een afwijking heeft, dan verzinnen we daar wel wat op. (WR-P-P-G-0000076619)
If we, gynaecologists, are not allowed to offer young pregnant women a test to determine whether their unborn child has an increased risk of an abnormality, we will come up with a solution.

Consequents have a higher frequency of modalisation than the first cluster (42.21%), which is largely due to higher frequencies of deontic modality (15.41%) and dynamic modality (14.07%). The relative frequency of epistemic modality (12.40%) is comparable to that in the first cluster (12.83%). With respect to syntactic integration, we see a higher frequency of sentence-final antecedents (39.93%), almost solely at cost of the resumptive pattern (19.41%).

When we try to connect the features of the conditionals in this cluster to implicatures of unassertiveness and connectedness, there is no clear unified meaning aspect to be found, apart from the fact that their antecedents are marked for *event modality* (Palmer, 2001, cf.) and express ability or willingness

mostly (i.e., dynamic modality). In a minority of cases, epistemic modality (i.e., propositional modality) is expressed by means of modal verbs in past tense, used by the speaker to distance herself from p in the antecedent. In terms of the literature discussed in chapter 3, we can interpret this cluster as double decision conditionals from Davies (1979) account, which, according to her, contain a 'decision modal' and are mostly used for making polite requests, purely case-specifying conditionals from Declerck and Reed's (2001) account, which 'just specify[...] the case(s)' in which the consequent actualises, potential conditionals from Kaegi's (1905) account, in which both the antecedent and consequent are presented as 'purely imaginable', conceivable situations, and the condition sub-type of hypothetical conditionals from Athanasiadou and Dirven's (1996) account, which expresses desirable outcomes in the antecedent. Again, it is clear that there is no perfect overlap between this cluster and the types and sub-types of conditionals discussed.

6.5.6 Past tense conditionals with modalised consequents (cluster 2)

The third largest cluster holds 546 or 13.29% of all conditionals and is represented as cluster 2 in Figures 6.5 and 6.6. It is characterised, as can be seen in the examples below, by simple past (49.08%) and past perfect tense (13.92%) in the antecedent, and a strong preference for simple past tense in the consequent (78.39%), followed by past perfect (14.29%). This readily shows how strong this cluster is focused on past tense. In comparison to the unmarked cluster 3, we see a higher frequency of modalised antecedents (29.49%), with epistemic modality being most frequent (15.93%), followed by dynamic modality (10.07%). Consequents, however, have an even higher frequency of modal marked clauses (59.34%), largely marked for epistemic modality (50.55%), which can be seen in the representative examples of this cluster in the examples in (419) to (421) below.

(419) Ik zou toch wel vaker fietsen als ik op Vossenveld woonde. (fn000573)
I would cycle more often if I lived on Vossenveld.

(420) Als de rijkswachters eind 95 beter zijn gruwelhuis in Marcinelle hadden doorzocht, hadden zij de meisjes nog levend uit zijn kelder kunnen halen. (WR-P-P-G-0000045321)
If the gendarmes had searched his horror house in Marcinelle in late 95 better, they would have been able to get the girls out of his basement alive.

(421) Een val in het ziekenhuis werd vastgesteld als deze in het dossier vermeld stond of bij de valincidentenregistratie was gemeld. (WR-X-A-A-journals-001)
A fall in the hospital was registered if it was mentioned in the file or was reported in the fall incident registry.

With respect to negation, the cluster does not differ from the other clusters, with a minority of 15% to 20% of antecedents and consequents containing negation. In terms of clause order, this cluster has a slightly higher percentage of sentence-final antecedents than the other clusters (39.93%), mainly at the cost of resumptive conditionals (19.41%), but otherwise, syntactic integration is comparable to the other clusters.

In contrast to the previous cluster, this cluster can be connected to a specific implicature, namely that of epistemic distancing. As can be seen in the examples in (419) and (420), the conditionals in this cluster are used to express distance, disbelief, or, depending on theoretical predisposition, counterfactuality (see section 2.5). For this, tense is instrumental, as is, to a lesser degree, modal marking. This would make a case for a what some would call a counterfactual conditional construction, but, as can also be seen in the example in (421) above, past-tense conditionals, especially of the recurrent or 'course-of-event' type (cf. Athanasiadou & Dirven, 1996; see section 3.3.9), are also present in this cluster. This means that the clustering algorithm did not readily differentiate between the past tense being used for epistemic distance or temporal distance. Although the algorithm has, of course, no internal knowledge of concepts like time and belief, it was expected that this distinction could have been identified based on the combination of tense and modality distributions, as epistemic distancing would have been more frequently marked by modals. I expected to see higher frequencies for negation in this cluster, as implicatures of counterfactuality are often supported by negation (see section 2.5 and especially section 3.2.10), and the least representative conditionals discussed in section 6.3.7 showed both past tense and negation. However, antecedents are negated in only 16.22% of all cases in this cluster, and consequents in 17.03%. As in the previously discussed clusters, these numbers seem to reflect the general distribution of negation mostly, which means that negation has probably not played a large role in the clustering. As already discussed in section 3.2, the ambiguity between remoteness and past time as expressed by the past tense is an issue in many accounts, for instance those by Funk (1985; see section 3.2.6), and Huddleston and Pullum (2002; see section 3.2.9). Whereas, for instance, past conditionals and 'future conditionals with less vivid form' as distinguished by Goodwin (1879) are not differentiated in this cluster, the cluster does reflect Funk's (1985) category of closed conditionals, which involve both neutral and hypothetical or marked conditionals.

In terms of the accounts discussed in section 3.2, the conditionals in this cluster would be classified as conditionals implying non-fulfilment (cf. Goodwin, 1879; section 3.2.2), unreal conditionals (cf. Gildersleeve, 1882; section 3.2.3), implicative non-fulfilment conditionals (cf. Sonnenschein, 1892; section 3.2.4), unreal conditionals (cf. Kaegi, 1905; section 3.2.5), closed hypothetical conditionals (cf. Funk, 1985; section 3.2.6), imaginative conditionals, both hypothetical and counterfactuals (cf. Celce-Murcia & Larsen-Freeman, 1999; Wierzbicka,

1997; sections 3.2.7 and 3.2.10), theoretical conditionals (cf. Declerck & Reed, 2001; section 3.2.8), and remote conditionals (cf. Huddleston & Pullum, 2002; section 3.2.9).

In terms of implicatures of connectedness, the representative conditionals in this cluster all implicate a causal connection between the antecedent and consequent. This is likely to be related to the implicature of epistemic distance discussed above, as, for instance, pragmatic conditionals are not frequently distanced (see section 3.3.4). However, content or predictive conditionals (cf. Dancygier & Sweetser, 2005; section 3.3.7) or hypothetical conditionals (cf. Athanasiadou & Dirven, 1997a; section 3.3.9) have been argued to be the most frequent and prototypical type of conditionals. In terms of the accounts based on connections between antecedents and consequents, discussed in section 3.3, the conditionals in this cluster are most comparable to what direct-hypothetical conditionals (cf. Quirk et al., 1985; section 3.3.4), not-now conditionals (cf. Nieuwint, 1992; section 3.3.6), unreal conditionals (cf. Gildersleeve, 1882; section 3.2.3, cf. Kaegi, 1905; section 3.2.5), implicative conditionals (cf. Sonnenschein, 1892; section 3.2.4), and imaginative conditionals (cf. Celce-Murcia & Larsen-Freeman, 1999; section 3.2.7).

In conclusion, this cluster is marked by means of the combination of past tense in antecedents and especially consequents, and epistemic modality in consequents. These features support an implicature of unassertiveness, and more specifically, epistemic distance towards the situations expressed.

6.5.7 Conditionals with present perfect antecedents (cluster 4)

The fourth and smallest cluster holds 192 or 4.67% and is represented as cluster 4 in Figures 6.5 and 6.6. It can be characterised by the high frequency of present perfect tense in the antecedent (84.90%), followed by the simple past (13.54%) in the antecedent. Consequents in this cluster feature simple present tense exclusively, which largely reflects tense in consequents of clusters 1 and 3. Negation does not deviate from the general trend reported in section 5.9 and seen in the other clusters. Antecedents are not modalised in this cluster, and consequents are marked for modality in 36.46% of the cases, which is largely in line with clusters 1 and 3, although this cluster is marked for dynamic modality more often (14.06%) than cluster 3, followed by deontic modality (10.94%), epistemic modality (9.90%) and, in a minority of cases, evidential modality (1.56%). In terms of syntactic integration patterns, the cluster is comparable to the other clusters, with slightly more sentence-final antecedents (34.90%), slightly more integrative conditionals (34.90%), and less resumptive conditionals (17.71%). Examples of representative conditionals in this cluster are presented in (422) to (424) below.

(422) Als je nog nooit op de HCC geweest bent, is het erg moeilijk om overzicht te bewaren. (WR-X-B-A-discussion-lists-tweakers-63565)
If you have never been to the HCC, it is very difficult to keep an overview.

(423) Als 'ie de aanklacht goed heeft begrepen moet 'ie zeggen of 'ie zich schuldig vindt. (fn004379)
If he has understood the charges correctly, he must say whether he is guilty.

(424) Als er toen fouten in zaten die niet door klanten gemeld zijn, zullen die fouten er ook nu nog zijn. (WR-P-P-D-0000000006)
If there were errors back then that were not reported by customers, those errors will still be there today.

As we see in the feature distributions and these examples, this cluster is based mostly around verb tense in the antecedent.

As with the conditionals in cluster 1 (see section 6.5.5), there does not appear to be a clear specific implicature of either unassertiveness or connectedness licensed by the conditionals in this cluster, or, to be more specific, by the divergent verb tense in the antecedent. Using the accounts discussed in chapter 3 as a guide, the cluster could be said to reflect Davies's (1979) knowledge conditionals (section 3.3.3), Johnson-Laird's (1986) completely determinate conditionals (section 3.3.5), and Dancygier and Sweetser's (2005) epistemic conditionals (section 3.3.7), although the degree to which the exemplars in this cluster are truly of these types is debatable. As explicitly discussed with respect to Gildersleeve's (1882) logical conditionals (see section 3.2.3), whether or not the antecedents here are 'accepted as true' is largely a matter of context. Furthermore, although the examples in (423) and (424) could be interpreted as such, their causally-reversed counterparts (cf. Sweetser, 1990, p. 123; section 3.3.7) show the actual inference-chain that would be present in epistemic conditionals.

(425) Als moet 'ie zeggen of 'ie zich schuldig vindt, heeft 'ie de aanklacht goed begrepen.
If he must say whether he is guilty, he has understood the charges correctly.

(426) Als ze er nu ook nog zijn, zaten ze die fouten die niet door klanten gemeld zijn er toen ook in.
If they are still in there, those errors that were not not reported by customers were in there back then.

A large number of conditionals in this cluster involve dynamic and deontic modality in consequents, expressing a 'true condition' (cf. Athanasiadou & Dirven, 1997a; see section 3.3.9) in the antecedent, and a resulting necessary action to be undertaken, as in (422) and (423). Again, these are informal comparisons, and, given the small size of the cluster, they should be interpreted with caution.

6.5.8 Additional clusters

As the cluster evaluations did not clearly indicate a preference for a 4-, 5- or 6-cluster solution, I will also inspect the additional clusters in the latter two solutions. As a feature of hierarchical clustering, this does not mean a completely different clustering solution, but a sub-clustering of, in this case, the largest cluster, which held the most general, unmarked kind of conditionals. In the 5-cluster solution, cluster 3 discussed above is divided into two clusters, cluster 3A and 3B. The largest sub-cluster, cluster 3A, holds 86.6% of cluster 3, whereas cluster 3B holds 13.4%. Cluster 3A roughly adheres to the characterisation of cluster 3 in section 6.5.4, except for negation in the antecedent, which is used by the algorithm to create cluster 3B. Representative examples for the latter cluster are provided in (427) to (429) below.

(427) De NOS krijgt overigens geen korting als Oranje zich niet voor het WK plaatst. (fn002418)
By the way, the NOS will not receive a discount if the Dutch soccer team does not qualify for the World Cup.

(428) Het is ons probleem niet als je het niet haalt. (WR-X-A-A-journals-003)
It's not our problem if you don't make it.

(429) Geen idee, ik ga eerst lekker F1 kijken:-) Duurt nog een uur als er geen doden vallen. (WR-U-E-D-000000030)
No idea, I'm going to watch F1 first:-) It will take another hour if there are no casualties.

The only difference between the conditionals in the two sub-clusters is the presence of negation, which rose to 86.33% of antecedents being syntactically negated, and 13.67% of antecedents being morphologically negated (all antecedents thus contain negation), and 24.66% of consequents being syntactically negated and 1.61% of consequents being morphologically negated, compared to 14.74% and 2.67% in cluster 3. Tense, modality, and syntactic integration remained stable mostly. This cluster reflects what we discussed in terms of 'negative polarity' in section 3.3.8, i.e., it presents a relation between the non-fulfilment of the situation in the antecedent and the situation expressed in the consequent, which may be, but does not have to be negated itself.

When we look at the 6-cluster solution, cluster 3B discussed above remains the same, and cluster 3A is split into two sub-clusters, clusters 3A1 and 3A2, which hold 84.17% and 15.83% percent of the conditionals in cluster 3A respectively. As cluster 3A1 resembles cluster 3A closely, we will focus on cluster 3A2. Representative examples are presented in (430) to (432) below.

(430) Als gmail een POP 3 of IMAP server heeft is het niet zo moeilijk. (WR-U-E-A-0000000301)
If Gmail has a POP 3 or IMAP server, it is not that difficult.

(431) Uh als je toch doodgaat maakt ook niet uit als je verslaafd bent. (fn000559)
Uh if you will die anyway it doesn't matter whether you are addicted.

(432) Als het aan de regeringspartijen ligt komen er geen verschillende tarieven. (fn003811)
If it is up to the government parties, there will be no different rates.

This new cluster is formed mainly on basis of negation in the consequent instead of the antecedent, which was the case for cluster 3B discussed above. The other features show distributions comparable to the main cluster, namely simple present in both clauses, non-modalised antecedents and consequents modalised in 36% of the cases. Syntactic integration also showed a distribution comparable with the first cluster. What we see in (430) to (432) is the denial of the situation expressed in the consequent, as 'caused', either in content or epistemic terms (cf. Dancygier & Sweetser, 2005) by the situation in the antecedent.

As may be expected by the discussion and examples above, there appear to be no clear, specific implicatures of unassertiveness or connectedness licensed by these sub-clusters, beyond the addition of negation to either the antecedent (cluster 3B) or the consequent (cluster 3A2). The meaning aspect contributed by negation of the antecedent (cluster 3B) can be explained in terms of negative conditions in the Cognitive approach to Coherence Relations (cf. Sanders, Spooren & Noordman, 1992; see section 3.3.8), i.e., the non-fulfilment of the condition in the antecedent, such as there being no casualties in (429) causes taking the race (just) another hour. As discussed in section 3.3.8, polarity is independent of 'source of coherence', and consequently, no more specific implicatures of connectedness, such clear preference for causal or inferential implicatures, were found in this cluster. Conditionals in cluster 3A2 appear to implicate that the situation in the consequent can be prevented by the situation in the antecedent. In (430), for example, the antecedent (Gmail has a POP 3 or IMAP server) 'causes', in an epistemic sense, (cf. Dancygier & Sweetser, 2005) the denial of it being difficult expressed in the consequent, which is comparable to the preclusive conditionals ('P prevents Q') discussed by Declerck and Reed (see section 3.3.11).

6.5.9 Conclusion

In the analysis of hierarchical clusters presented in this section, I aimed to provide insights into groups of conditionals that can be formed based on their grammatical features, and I attempted to interpret the resulting clusters with respect to implicatures of unassertiveness and connectedness. From the results and analyses, a number of conclusions can be drawn.

First, it became clear that the clusters, with the exception of cluster 2 (past tense conditionals with modalised consequents), did not license clear implicatures of unassertiveness or connectedness discussed in chapter 3. In short,

there never appeared to be a clear agreement between (theoretical) types distinguished in the literature and (data-driven) clusters of conditionals. The influential distinction between content, epistemic and speech-act conditionals (cf. Sweetser, 1990; Dancygier & Sweetser, 2005), for instance, was not reflected at all by any of the clusters discussed in this section. It is important to remark that the expectation of an agreement between meaning and form is warranted, given the suggestions of links between types of conditionals and their grammatical features in the literature (see chapters 3 and 5). Furthermore, as Gabrielatos (2010, 2021) shows, a clustering approach is viable to uncover types of conditionals of theoretical distinction, as his results show how modality can be used to differentiate between direct and indirect conditionals (Quirk et al., 1985). The current results, however, show why this is not the case for Dutch conditionals, as consequents of Dutch direct and indirect conditionals are not marked by the presence or absence of the modal verb *zullen* 'will' respectively, whereas English conditionals are. We will discuss these points in more detail in the next chapter.

Second, the clustering solution produced a large, unmarked cluster, cluster 3, which consists mostly of conditionals in the present tense, with a minority of consequents marked for modality, mostly of the epistemic kind. This cluster, then, can be seen as the default type of conditional in Dutch. In terms of prototype theory, this 'type' has the highest frequency (as shown by the cluster size), the highest number of shared attributes and an internal prototypicality range consisting of a limited number of deviations from the default verb tense and modality, although clear characterisations of these deviations in functional terms, as Athanasiadou and Dirven (1997a) do in terms of *cause*, *condition*, and *supposition* sub-types of hypothetical conditionals (see section 3.3.9), cannot be given. Whereas Athanasiadou and Dirven's prototypical type of conditional expresses the strongest (i.e., causal) dependency between antecedent and consequent, alike Dancygier's prototypical predictive conditional (see section 3.3.7), types of dependency such as causality, epistemic inference, pragmatic or speech-act relations, analysed as implicatures of connectedness in this study, were not identified by the clustering algorithm, i.e., the features included in this study do not seem to differentiate clearly between such types of degrees of dependency.[40] The three remaining main clusters are less prototypical, reflected in their lower overall frequency, and smaller number of cases sharing features such as tense and modality. These clusters thus have a less stable, but more specific set of defining features. Whereas the second main cluster, cluster 1, can be described as expressing either willingness, ability or epistemic distance in the antecedent, cluster 2 is perhaps most identifiable, as

[40]To be clear, these types of conditionals do occur in the corpus. Speech-act conditionals, for instance, although not found among the representative examples of any cluster, can be found in multiple clusters, such as the example in (a) from the unmarked cluster, cluster 3.

(a) Ik zie dat toch echt anders hoor als ik het wetsvoorstel lees. (fn000152)
I really see that differently if I read the bill.

it consists mostly of conditionals that express epistemic distance by means of modal auxiliaries and past tense in antecedents and especially consequents. The fourth and smallest cluster holds mostly conditionals in which the antecedent presents a proper condition and the consequent an action to be undertaken, or in a smaller number of cases, a conclusion to be drawn.

Third, the feature distributions of the clusters show that one of the most promising features for Dutch conditionals in relation to implicatures of connectedness, namely syntactic integration, does not contribute clearly to the formation of clusters, whereas the literature on Dutch conditionals suggests otherwise (see section 5.3). The degree of syntactic integration was hypothesised to be reflective of the degree of semantic integration in terms of Dancygier and Sweetser's (2005) distinction between *content*, *epistemic* and *speech-act conditionals*. This, however, does not mean that the literature is wrong on this point, as it might be the case that the contribution of tense and modality, as reflected in Figure 6.6, is stronger, or points towards different dimensions on which clusters are formed. The assumption of clustering is that certain features go together ('cluster') to form groups that have theoretical, empirical or practical importance. As applied to linguistics, this means that certain linguistic features cluster together to support a certain (range of) interpretation(s), analysed here as implicatures. This does, at least for the results in this section, not seem to be the case, which does not exclude the possibility that in Dutch, syntactic integration is the only feature of importance, or a feature that operates in relative isolation of the other features.[41] Furthermore, there is, as discussed in chapter 4, the question of language specificity (see section 4.5 especially). These points will be taken up further in the discussion in the next chapter. Before doing so, however, we will look at the results of the partitional clustering next.

6.6 Analysis of partitional clusters

6.6.1 Introduction

In this section, I analyse the clusters present in the partitional solutions in terms of their distributions of grammatical features, and possible implicatures of unassertiveness and connectedness. As the aims of the partitional approach are equal to those of the hierarchical approach discussed in the previous section, I will use the same steps in the analysis by discussing the internal feature distribution of each cluster, representative examples, their implicatures, and a comparison to possible matches on types of conditionals from the classifications discussed in chapter 3.

[41] Note that this explanation is not in conflict with the results by Gabrielatos (2010, 2021) mentioned earlier, as he distinguished between modal load and modal spread as separate features.

In section 6.6.2, I will present an overview of the clustering solutions. As the evaluations in section 6.4.5 did not provide definitive arguments for a 2-, 3- or 4-cluster solution, I will first select the most promising solution, after which I will present its feature distributions per cluster. Then, in sections 6.6.3 to 6.6.5, I will discuss each cluster in the fashion outlined above. In section 6.6.6 I will provide a brief conclusion on the results of partitional clustering, before moving on to conclusion to this chapter in section 6.7.

6.6.2 Clusters and feature distributions

In section 6.4, I selected the 2- to-4 cluster solutions that were generated using the Lin1 measure and the FANNY algorithm. In this section, I will inspect the characteristics of each cluster. As in the previous section, NMDS was used to visualise the clusters, and I added the cluster memberships to the configurations to see how they are distributed on the ordination axes, as can be seen in Figure 6.7.

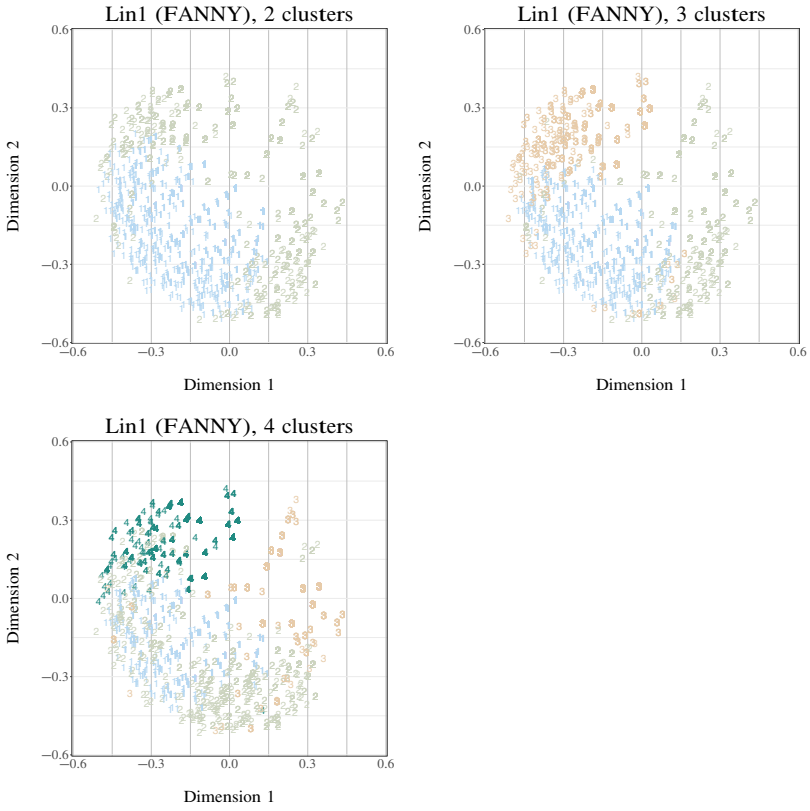


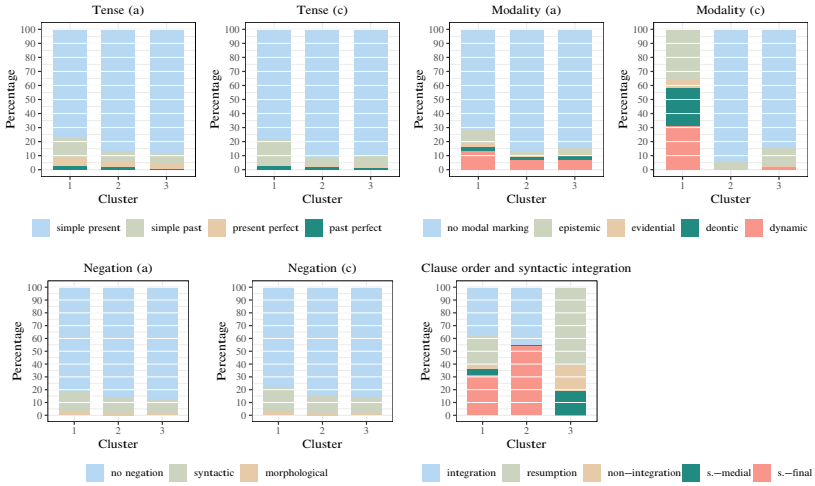
Figure 6.7: NMDS configurations with memberships from partitional clustering

As partitional clustering does not embed clusters (see section 6.4.5), the clustering itself can change depending on the given number of clusters. The consistency of the clusters, in terms of their average silhouette widths, is relatively stable: 0.52 and 0.50 for the 2-cluster solution, 0.40, 0.51, and 0.55 for the 3-cluster solution, and 0.42, 0.36, 0.60, 0.56 for the 4-cluster solution. Increasing the number of clusters introduces less consistent clusters. For instance, a fifth cluster with an average silhouette width of 0.21 and in the 6-cluster solution a cluster with an average silhouette width of 0.09 is introduced. The configurations in Figure 6.7 resemble those resulting from the Lin measure in the previous section, but show slightly less separation. In the 2-cluster solution, we see all conditionals from cluster 1 are in the bottom-left corner. Cluster 2 is scattered around the first cluster on the left and right. In the 3-cluster solution in the middle panel we see dimension reduction is able to preserve the difference between clusters 2 and 3, which are on the bottom-right and the top-left

respectively. The 4-cluster solution does not differentiate groups well in the right panel. We see more overlap between the first two clusters, and while the third cluster is somewhat concentrated in the top-right of the configuration, we see that the fourth cluster largely overlaps with the second cluster. Based on the evaluations in section 6.4, the average silhouette widths and the configurations in Figure 6.7, I will discuss the 3-cluster solution in the remainder of this section.

In terms of cluster membership, the 3-cluster solution shows a relatively even distribution. Cluster 1 holds 32.54% of all conditionals, the second and largest cluster holds 39.74%, and the third cluster holds 27.72% of all conditionals in the corpus. As we can see, the memberships are more evenly distributed compared to the hierarchical clusters. The feature distributions of each cluster are presented in Figure 6.8 below.

Figure 6.8:
Feature distributions for the partitional 3-cluster solution



Note. These distributions represent clusters generated using the Lin1 measure. Numbers on the horizontal axes correspond to cluster numbers reported in this section.

In sections 6.6.3 to 6.6.5, we will review the clusters in the 3-cluster solution.

6.6.3 Unmarked conditionals (cluster 2)

As with the discussion of the selected hierarchical clustering solution, we will start by inspecting the largest cluster, cluster 2, which holds 1633 conditionals (39.74%).[42] As we can see in Figure 6.8, this cluster holds conditionals with present tense in antecedents and in consequents mostly (86.47%, 91.43% respectively), followed by simple past (6.92%, 5.57%), present perfect (4.72%, 1.22%), and past perfect (1.90%, 1.78%). Negation reflects the overall frequencies with 85.85% non-negated antecedents and 84.63% non-negated consequents. Antecedents contain modal marking in 13.29% of all cases, mostly of the dynamic kind (6.61%), as in (434), followed by epistemic, deontic, and evidential modality (3.06%, 2.33% and 1.29% respectively). Modal marking is mostly absent in consequents (94.49%), and consequents that are marked for modality are marked exclusively for epistemic modality (5.51%). Finally, we see sentence-final antecedents in the majority of cases (54.44%), followed by the integrative pattern (45.56%). The other patterns of syntactic integration are not found in this cluster, which is reflected below in the representative examples in (433) to (435).[43]

(433) Tijddwang treedt op als er klanten wachten. (WR-P-P-F-0000000012)
Time constraints occur {if/when} customers are waiting.

(434) Als je toch nog wilt komen lever ik graag een bijdrage. (WR-U-E-D-0000000307)
If you still want to come I [would] like to contribute.

(435) Mogelijk zijn de verbanden tussen privacyschending en conflict heel anders als het een kwestie betreft die jongeren privé vinden [...]. (WR-X-A-A-journals-003)
Possibly the links between privacy violation and conflict are very different if it concerns an issue that young people consider private [...].

It appears that this largest cluster mostly holds the unmarked conditionals discussed in the previous section. Accordingly, in terms of implicatures, these default conditionals do not seem to share specific implicatures of unassertiveness or connectedness. Although the examples above may suggest that the conditionals in this cluster license implicatures of causal connections between antecedent and consequent, this is an effect of the frequency of such implicatures, as, for instance, inferential (i.e., epistemic) and pragmatic (i.e., speech-act) implicatures of connectedness can also be found, as in the examples in (436) and (437) from this cluster.

[42] The reason for this order is that the largest cluster can be used in comparison to smaller, more specialised clusters, although, in this particular solution, cluster sizes are more similar than in the hierarchical solution discussed in the previous section.

[43] Note that silhouette widths were used for consistency in selection of representative examples.

(436) Oke als ik het goed begrijp heeft Arsenicem een systeem voor mij wat er voor gemaakt is, dat zoek ik ook maar het hoeft niet zo profesioneel te zijn. (WR-X-B-A-discussion-lists-tweakers-980460)
Okay if I understand correctly Arsenicem has a system for me that is made for it, I am looking for that too, but it does not have to be that professional.

(437) Europarlementarier Max van den Berg weet duidelijk niet waar hij het over heeft als hij zegt dat een koe in Nederland omgerekend drie euro subsidie per dag krijgt. (WR-P-P-G-0000024358)
Member of the European Parliament Max van den Berg clearly does not know what he is talking about {if/when} he says that a cow in the Netherlands receives a three euro subsidy per day.

With respect to implicatures of unassertiveness, this cluster largely holds conditionals without any marking of certainty, uncertainty or counterfactuality. Alike the conditionals in the unmarked cluster in the hierarchical solutions presented in the previous section, the conditionals in this cluster would be classified as present conditionals (cf. Goodwin, 1879; section 3.2.2), undetermined conditionals (cf. Gildersleeve, 1882; section 3.2.3), present non-implicative conditionals (cf. Sonnenschein, 1892; section 3.2.4), real and future conditionals (cf. Kaegi, 1905; section 3.2.5), open, non-past conditionals (cf. Funk, 1985; section 3.2.6), and open conditionals (cf. Huddleston & Pullum, 2002; section 3.2.9). These conditionals thus present the antecedent and consequent mostly in a neutral fashion, without implication of fulfilment.

As with the cluster of unmarked conditionals in the hierarchical solution, this cluster does not include conditionals with implicatures of epistemic distancing. The characterisation of this cluster is comparable to that of the unmarked conditionals discussed in the previous section, which is not surprising, given their distributions of features, and their overlap, as most of the conditionals in this unmarked partitional cluster were members of the unmarked hierarchical cluster in the previous section.[44] This comparison extends not only to the lack of shared implicatures of unassertiveness, but also to implicatures of connectedness, as this cluster also does not differentiate between 'uncertainty' implicatures of unassertiveness on the one hand, as in (435), and recurrent or iterative implicatures on the other, as in (433). The same goes for implicatures of connectedness, as no distinction is made between, for instance, direct and indirect conditionals (cf. Quirk et al., 1985; see section 3.3.4), actualising and inferential conditionals (cf. Declerck & Reed, 2001; see section 3.3.11), or predictive, epistemic and speech-act conditionals (cf. Dancygier & Sweetser, 2005; see section 3.3.7). We can conclude that the main types of conditionals found in the literature based on connections between antecedents and consequents

[44] 1211 of 1633 members (74.16%) of this partitional cluster are part of the unmarked hierarchical cluster. Note that this percentage is lower for the reverse perspective (43.66%), as the unmarked hierarchical cluster is larger.

are not detected by the partitional clustering algorithm. We will continue by inspecting the two remaining clusters to find out whether their feature distributions do give rise to more specific implicatures.

6.6.4 Conditionals with modalised consequents (cluster 1)

The second largest cluster, cluster 1, holds 1337 conditionals (32.54%). This cluster is similar to the first cluster in most respects. A difference can be observed in clause order and syntactic integration, as this cluster holds less sentence-final antecedents, and, in contrast to unmarked cluster, a relatively large number of resumptive conditionals. The clearest difference, however, is that all of the consequents in this cluster are marked for modality, as can be seen in the representative examples in (438) to (439) below.

(438) En als de VUT in klap wordt afgeschaft zou zelfs de spanning op de arbeidsmarkt in keer zijn opgelost. (fn000242)
And if the early retirement fund is repealed, even the tension on the labor market would be resolved in one go.

(439) Ik vind als ik uh ga kijken naar een stripper dan wil ik ook alles zien en uh toen zei Catherine maar mevrouw dat u dat durft te zeggen. (fn000578)
I think if I uh look at a stripper then I also want to see everything and uh then Catherine said but madam how dare you say that.

(440) Als de aanvraag op tijd is ingediend en de panelen tijdig zijn geïnstalleerd moet het geld worden uitgekeerd aan de energiebedrijven. (WR-P-P-G-0000160102)
If the application is submitted on time and the panels are installed in time, the money must be paid to the energy companies.

Simple present is frequent in both clauses (77.64% and 77.71% respectively). In most cases, thus, these conditionals are not the counterfactual types found in cluster 3 in the previous section, but rather conditionals in which the consequent is marked for epistemic modality, dynamic and deontic modality (36.13%, 30.67%, and 27.75% respectively), as in (438) to (440). In the case of epistemic modality, as in (438), in a minority of cases epistemic distance is expressed, but in most cases, the modal marking expresses future reference and promise (see section 5.4.5), as in (441) and (442) below.

(441) Als dat het geval is zullen al om deze reden de effecten van plaatsgebonden maatregelen op verschillende plaatsen verschillend uitpakken. (WR-X-A-A-journals-001)
If this is the case, the effects of site-specific measures will have different effects for different reasons.

(442) Als dat niet lukt, zullen wij ons niet aan onze taak onttrekken. (WR-P-P-G-0000134919)
If that does not work, we will not evade our task.

Conditionals in this cluster show integrative (38.52%), sentence-final (31.94%) and resumptive patterns mostly (22.44%), followed by a minority of sentence-medial antecedents (4.41%) and non-integrated conditionals (2.69%).

Inspecting the conditionals in this cluster, and the feature distributions, there does not appear to be a clear relation to any of the types of conditionals discussed in chapter 3, neither with respect to implicatures of unassertiveness, nor with implicatures of connectedness. Whereas in the hierarchical clustering solution, modal marking and verb tense clearly clustered together, especially in case of epistemic modality and past tense, to license implicatures of epistemic distance (see hierarchical cluster 2 in section 6.5.6), such an identifiable feature combination was not found in this cluster.

6.6.5 Resumptive, non-integrated and sentence-medial conditionals (cluster 3)

The third and last cluster holds 1139 conditionals (27.72%). Whereas the previous cluster was clearly formed by modal marking of the consequent, the third cluster only deviates from the other clusters in terms of syntactic integration. The cluster shows a prevalence of simple present tense in both clauses (88.67% and 89.73% respectively), as in the examples in (443) to (445) below, followed by simple past (7.11%, 7.55%), present perfect (3.25%, 1.23%) and past perfect (0.97%, 1.49%). Frequencies of negation in antecedents and consequents are comparable with the previous clusters too (12.99%, 14.57%). Antecedents are marked for modality in a minority of cases (16.33%), mostly for dynamic modality (7.11%), followed by epistemic, deontic and evidential modality (5.09%, 2.81%, 1.32%). Consequents show a comparable frequency of modal marking (15.36%), but with a much more pronounced preference for epistemic modality (13.35%), as in (444), followed by dynamic modality in only 2.02% of cases. Below, representative examples of this cluster are presented.

(443) Als het iets later op de middag wordt, dan melken we vanavond ook maar iets later:-((WR-P-E-A-0004240623)
If it gets a little later in the afternoon, then we will milk a little later tonight too:-(

(444) Zal ik, als de wegen droog zijn, het zonnetje schijnt en er geen regen voorspeld wordt naar jou toe komen? (WR-U-E-D-0000000007)
Shall I, if the roads are dry, the sun shines and no rain is predicted, come over to you?

(445) En als je toch een kwalitatief uh goed besluit wilt nemen en draagvlak wil dan heb je daar veel maatschappelijke organisaties bij nodig. (fn000162)
And if you still want to make a good quality decision and want support, then you need a considerable number of social organisations.

As mentioned above, the largest difference between this cluster and the other clusters can be found in syntactic integration patterns. Whereas the previous clusters, clusters 2 and 1 discussed in sections 6.6.3 and 6.6.4 respectively, both had high frequencies of integrated (sentence-initial) antecedents (45.56%, 38.52%) and sentence-final antecedents (54.44%, 31.94%), this cluster has a higher frequency of resumptive conditionals (60.32%), followed by non-integration (21.07%), and sentence-medial antecedents (18.53%). The latter two patterns are largely absent from the other clusters. The high frequency of non-integrative conditionals is to a large extent a consequence of including interrogative and imperative consequents in this category (see sections 5.3, and 5.8, and section C.2 of Appendix C), as can be seen in the examples below.

(446) Zou Geert Wilders 7 of 18 zetels halen als er nu verkiezingen waren? (WR-P-P-G-0000049699)
Would Geert Wilders get 7 or 18 seats if there were elections now?

(447) En Johan als jij toevallig een pijp krijgt wat doe je dan? (fn007858)
And Johan if you happen to get a pipe what do you do then?

(448) Als er iets is, bel me. (WR-U-E-D-0000000050)
If anything is wrong, call me.

Whereas syntactic integration seems to be ignored largely in the hierarchical clustering, we see its influence in these partitional results. It does not seem to interact with the other features, however.

In terms of implicatures, the conditionals in this cluster do not seem to share implicatures of unassertiveness. We must be careful, however, and refrain from concluding that the conditionals in any of the three clusters do not license implicatures of unassertiveness. The reason for this is that, on the whole, conditionals licensing implicatures of epistemic distance by means of past tense and modal marking, which were identified by the hierarchical clustering algorithm (see section 6.5.6), are distributed over different clusters in the partitional solutions. The consequence of this is that it is not possible to identify types of conditionals in terms of the accounts discussed in section 3.2, which were largely based on implicatures of epistemic distancing.

With respect to implicatures of connectedness, as we discussed in section 3.3, this cluster seems to hold a large number of conditionals licensing an indirectness implicature, as can be seen in examples (447) and (448) above. This is not surprising, as almost all non-integrated conditionals (86.96%) are in this cluster, and low degrees of syntactic integration were linked to low degrees of semantic integration in section 5.3. Add to this the fact that many conditionals in this cluster have non-declarative consequents, which partly explains

the number of non-integrated conditionals, and it becomes clear why the examples above are ranked high on representativity for this cluster. We do see, however, that no distinction is made between speech acts about conditionals, as in (446), and conditional speech acts (i.e., questions), as in (447) (cf. van der Auwera, 1986; for details and discussion, see section 3.3.7). So while this cluster holds, in terms of the classifications discussed in section 3.3, the largest number of telling conditionals (cf. Davies, 1979; section 3.3.3), indirect conditionals (cf. Quirk et al., 1985; section 3.3.4), non-determinate conditionals (cf. Johnson-Laird, 1986; section 3.3.5), speech-act conditionals (cf. Dancygier & Sweetser, 2005; section 3.3.7), pragmatic conditionals (cf. Athanasiadou & Dirven, 1997a; section 3.3.9) and rhetorical conditionals (cf. Declerck & Reed, 2001; section 3.3.11), the algorithm does not distinguish between conditionals licensing an indirectness implicature and those that do not license such an implicature. This can be seen in the examples in (443) and (445) above, which do not license any implicature of indirectness, but of rather of directness or causality, and inferential reasoning.

Next to causal implicatures of connectedness, a considerable number of resumptive conditionals in this cluster appear to license an implicature of inferential connection, as in (449) and (450).

(449) Als de muren om ons heen instorten – 'en onze oude maatschappij morgen vervangen kan zijn door een nieuwe maatschappij', zoals Smalbrugge stelt – dan is dat omdat we in weerwil van alle lessen van de geschiedenis, opnieuw in zwart-wit tegenstellingen zijn gaan geloven, en op basis daarvan een tweedeling in de maatschappij in de hand werken. (WR-P-P-G-000012571)
If the walls around us collapse – 'and our old society may be replaced by a new society tomorrow,' as Smalbrugge states – it is because, despite all the lessons of history, we have again started to believe in black-and-white contradictions on the basis of which we promote a dichotomy in society.

(450) En als die vakantie echt tegenvalt, dan zal dat bij jullie allebei zo zijn, waarschijnlijk of niet [...]. (WR-U-E-A-0000000171)
And if that vacation is really disappointing, it will be for both of you, probably or not [...].

These conditionals are comparable to logical conditionals (cf. Gildersleeve, 1882; section 3.2.3), knowledge conditionals (cf. Davies, 1979; section 3.3.3), epistemic conditionals (cf. Dancygier & Sweetser, 2005; section 3.3.7), subjective conditionals (cf. Sanders, Spooren & Noordman, 1992; section 3.3.8), and premise conditionals (cf. Haegeman, 2003; section 3.3.10) Although this cluster includes different types of connections, the inclusion of inferential conditionals in the representative examples may reflect the relation between resumptive *dan* 'then' and inferential conditionals in Dutch. Recall from section 5.3 that Renmans and van Belle (2003, p. 148) observed a considerably higher frequency

of inferential relations between antecedent and consequent in their set of resumptive conditionals as compared to non-resumptive conditionals (see also Verbrugge & Smessaert, 2011; Reuneker, 2020).

In contrast to the remark made at the end of the previous section regarding the lack of influence of syntactic integration on the cluster formation, the partitional algorithm has singled out the feature of syntactic integration to form a cluster, but in doing so, it did not include other feature distributions. This suggests the importance of syntactic integration for clustering, but the cluster does not clearly reflect a construction formed by a relation between this single feature and implicatures of connectedness. The number of indirect conditionals, as far as they can be reliably identified (see chapter 4), is high, but next to these uses, the cluster includes a considerable number of conditionals that do not license any implicature indirectness, but rather of causality and inferential reasoning. Another indication that syntactic integration is, in this solution, not a sufficient predictor for implicatures of connectedness is the size of the cluster. While it is the smallest cluster (27.72%), it is much larger than the relative frequencies or mentioned low frequencies of indirect (pragmatic, speech-act) conditionals reported in other studies. Reuneker (2017b, p. 142) reports that only 6.1% of conditionals in his corpus license speech-act implicatures of connectedness, while 90% of all conditionals license causal implicatures of connectedness. Even sentence-medial conditionals show such a strong preference for causal implicatures (83.90%). Renmans and van Belle's (2003, pp. 152, 154) figures show that even though conditionals with resumptive patterns license an inferential implicature of connectedness in 41% of their 155 cases, the remaining 59% licenses other implicatures of connectedness, most notably causal implicatures (22%).[45] This corroborates the observation that the resumptive, non-integrated and sentence-medial patterns in this cluster are used frequently to license implicatures of connectedness beyond those of indirectness. Even though syntactic integration was singled out by the algorithm, it does not seem a strong predictor for implicatures of connectedness.

6.6.6 Conclusion

As with the the analysis of hierarchical clusters presented in the previous section, I aimed to provide insights into groups of conditionals formed by the partitional algorithm in this section. I attempted to interpret the resulting clusters with respect to both feature distributions, and implicatures of unassertiveness and connectedness. From the results and analyses, a number of conclusions can be drawn.

The partitional clusters did not, or only very weakly reflect types based on relations between antecedents and consequents as discussed in chapter 3. The 3-cluster partitional solution reflects one main category of unmarked conditionals (cluster 2), but in contrast to the hierarchical results, in which conditionals

[45] Renmans and van Belle's (2003) corpus did not contain any non-integrated conditionals.

seem to be clustered based on the interplay between a number of features, most notably tense and modality, this seems not to be reflected in the results of the partitional clustering. With respect to prototypicality, we see unmarked conditionals here too, but prototypicality is not clearly reflected in frequency, as the largest cluster is far less dominant in terms of membership frequency than in the hierarchical results. Also, the degree in which the conditionals in this cluster share attributes is lower, making their attribute spaces less clearly identifiable. Looking at the second largest cluster (i.e., cluster 1), we can see it is formed almost exclusively on the basis of modal marking in the consequent, and the last cluster (i.e., cluster 3) is based on syntactic integration and non-declarative consequents. Although these clusters thus have clear defining and identifiable characteristics, they do not seem to be connected to the characteristics of the other clusters, which may be due to the non-hierarchical nature of the clustering algorithm. Contrary to expectation for this approach to clustering, which was linked in section 6.4.2 to a more radial type of categorisation, inspecting the conditionals in the clusters did not reveal clear links to implicatures of either unassertiveness or connectedness. A noticeable exception was cluster 3, which holds many conditionals licensing indirect (i.e., pragmatic, speech-act) implicatures of connectedness in cluster 3. However, this cluster also holds conditionals licensing implicatures of, for instance, causality and inferential connections too, and it is likely that the high number of indirectness implicatures is a direct reflection of the high number of non-integrated conditionals in the cluster.

6.7 Conclusion

The primary aim of this chapter was to test the extent to which the feature distributions of Dutch conditionals presented in the previous chapter can be used to identify grammatical contexts licensing (generalised) implicatures of unassertiveness and connectedness. To do so, a number of data-driven, unsupervised machine learning techniques were used, and the results were analysed and evaluated.

In the first part of this chapter (sections 6.2 to 6.4), I provided arguments for analysing conditionals as form-meaning pairings, i.e., constructions, in order to investigate relations between grammatical features and implicatures of conditionals. As the features are expected to 'work together' in licensing implicatures of unassertiveness and connectedness, a clustering approach to the data was chosen to form groups that exhibit the smallest amount of within-group variance and the largest amount of between-group variance. Based on the literature, it was expected that the repeated use of certain patterns of grammatical features would have conventionalised to some extent into grammatical constructions, and together with a number of quantitative indices of feature distributions, I selected those features which maximised the chance of finding such structures underlying the data. As became apparent in this chapter, applying

standard procedures to categorical features, especially those with skewed distributions, proved problematic. While this is not uncommon in the literature, as a number of references in this chapter attest to, it did show that clustering is not a simple and objective 'go-to approach' for all datasets, especially in fields such as linguistics in which most features are of categorical nature. Because of this, a wide array of measures, algorithms and evaluations was used to select the most promising basis for further clustering, and to maximise the chance of finding structures. Both a combination of proven and state-of-the-art machine learning techniques, and theoretical evaluation were used to solve these problems, and to assess the clusterability of the dataset, enabling the selection of the most promising features for clustering. Evaluations of the distance matrices suggested removing aspect, person and number, and focus particles from the dataset to improve clusterability. None of the tests provided definitive grounds for conclusions on clusterability, however, which was linked to the focus in the clustering literature on numerical data, whereas the current dataset involves categorical data only. Two main approaches of clustering, hierarchical and partitional clustering, were selected based on their applicability to the data, and their theoretical relation to prototype theory. I evaluated the clusterability of their various implementations and parameters in detail, to arrive at the most promising clustering solutions. The selected solutions indicated 'reasonable structure' (cf. Kaufman & Rousseeuw, 1990, p. 88), which, although not uncommon in clustering applied to real data as opposed to controlled, generated data, already suggests not a very strong basis was found for grouping the conditionals in this study.

In the second part of this chapter (sections 6.5 and 6.6), I analysed the results of the cluster analyses. It became clear that most clusters did not license clear implicatures of unassertiveness or connectedness discussed in chapters 2 and 3. In short, there never appeared to be a clear agreement between (theoretical) types distinguished in the literature and (data-driven) clusters of conditionals. Types found in influential accounts of conditionals, such as direct and indirect conditionals (cf. Quirk et al., 1985), or content (predictive), epistemic and speech-act conditionals (cf. Dancygier & Sweetser, 2005), were not identified in the data, although the former was found earlier by Gabrielatos (2010, 2021) using only modal marking as input for clustering. The features of Dutch conditionals included in this study thus do not seem to differentiate clearly between types of conditionals based on unassertiveness and connectedness distinguished in the literature. We can compare this observation to Verhagen's (2021) analysis of translating Latin into English, in which the English language 'forces' one to make a choice between different modal verbs to present a dilemma as a moral one or as of various options (*should* or *must*, and *shall* or *can* respectively), whereas in Latin, the subjunctive does not require such a choice, leaving 'the interpretive possibilities open, including the option of complete irrelevance of a choice'. In the same vein, consequents of direct and indirect conditionals in Dutch are not marked by the presence or absence of the modal verb *zullen* 'will' respectively, as they are in English, which points towards the importance of lan-

guage specificity in this study. Whereas the hierarchical solutions did provide interpretable groups, most prominently a large unmarked group of conditionals, which was analysed as the prototypical type of conditional in Dutch, and a group of distanced conditionals, the partitional solution did not offer much basis for interpretation of the groups, with the exception of a cluster of conditionals licensing indirect implicatures of connectedness. This cluster, however, was formed almost exclusively on the basis of syntactic integration, a feature deemed of theoretical importance for the current purposes, but neglected mostly by the hierarchical algorithm, and the large number of conditionals licensing indirectness implicatures was explained by the large number of non-integrated conditionals, including those with non-declarative consequents. Although this suggests the importance of a single feature for clustering (i.e., syntactic integration), the cluster does not strongly indicate construction status, i.e., a pairing of this specific form to a clear meaning, because the conditionals in this cluster license various implicatures of connectedness beyond indirectness, such as causality and inferential reasoning, without a strong preference for one specific implicature.

As discussed in this chapter, reasonable structures were found in terms of quantitative evaluations. Closer inspection, analysis, and comparison of clusters to the literature on conditionals, however, indicated that none of the solutions directly or strongly reflected any of the implicatures discussed in chapter 2 and the types discussed in chapter 3. The question now is what implications these results have. Not finding a systematic relation between grammatical features and implicatures after all does not prove there is no such relation. Or, put differently, 'absence of evidence is not evidence of absence' (cf. Wright, 1888, p. 59; Sagan, 1977, p. 6). In the next and final chapter, I will take the liberty to discuss this issue and related issues raised in this study in more detail.