

I Modes of Modality

1 *Pragmatics: Modality and Speaker Orientation*

1.1 *The Human as an Animal Catoptricum*

When grown up, we are the way we are just because we are (also) catoptric animals and have developed a double ability to look at ourselves (insofar as it is possible) and the others in both our and their perceptive reality and catoptric¹ virtuality. (Eco 1986: 207)

In the history of language philosophy, the metaphor of language as a mirror of reality has a long tradition. Accordingly, language, thought, and world are linked in an iconic (imaging) relationship. As a system of signs standing for something else, language, on its different levels of representation, is the mirror image of the real world. Beyond this relationship of language–thought–world, the concept of the mirror is present in a very concrete sense also in the process of signification to the extent that the realization of one’s own mirror image is presupposed for identifying one’s own subject as well as the constitution of semiotic consciousness. In this sense, the human is not only an *animal semioticum*, but also an *animal catoptricum* (Eco 1986: 207) excelling in the capacity of perceiving him- or herself as well as others in their perceptive reality and, beyond that, in their reflected virtuality. Cognition-psychological studies tell us that this process of knowledge acquisition embraces several stages.

Mirror Principle: Deictic Origo Binding and Category Splits

Deixis is pointing at something. It is the function of grammatical deixis to locate the position of the viewer. Consider (1). The source (origo: speaker’s *Here-and-Now*) of deixis has to be identified. For each grammatical category, the place has to be identified from which reference is signaled. The place has to be reconstructed from which reference is taken up. We are not enabled to

¹ *Catoptrics* (from Greek κάτοπτρον/kátoptron ‘mirror’) is the study of the reflection of light through the mirror.

14 *Pragmatics: Modality and Speaker Orientation*

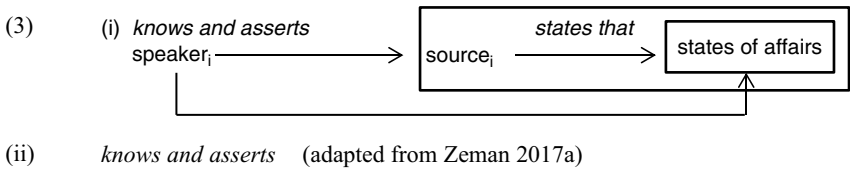
identify and refer to the extralinguistic context independent of the *Here-and-Now* until the morphological markers of the grammatical categories are encoded (Leiss 1992: 7; Zeman 2017b: 98). Consider the following example illustrating the split of the *Ego/I*.

- (1) Ich sollte den Artikel fertigschreiben.
 I should the paper finish

There are two readings of (1). See (2a, b).

- (2) a. 'I am obliged to finish the paper.' (→ realization of the event lies in the future and is therefore uncertain)
 b. 'It happened in due course that I finished the paper.' (→ realization of the event also lies in the future but it is certain that it happened)

While the referential distinction between the two individual *Ego/I* dimensions is maintained, the split is based on the resolution of 'awareness' about the knowledge of the speakers. At the same time, example (2b) shows that the perspectives linked to the different *Ego/I* dimensions must be compatible with each other. Following Evans (2005: 104), one would have to speak of a metaperspective (Evans' "X with respect to perspective 2, which is considered from perspective 1") in which two different awareness contents are integrated (also see Zeman 2017b modeling the binary relation as a process of triangulation). See (3).



The dependency hierarchy in (1) reflected in (3) arises not only from the speaker knowledge qua thought content, but also from the relative structural embedding of the discourse roles in (1). This is demonstrated by the exemplary use of the preterit *sollte* 'should' importing what is termed 'fate future' (German *Schicksalsfutur*; see Abraham 2014b). Example (1) is ambiguous allowing for the two different readings in (2a, b). The interpretation rests on how the perspective starting point ('source') is then anchored. Given the discourse configuration, in which the current speaker is the illocutionary subject, (2a) signifies the obligation that the paper be finished. The realization of the event is thus uncertain, i.e. it remains open as to whether the paper will be completed. In contrast, given a narrative context where the narrator has

1.2 Modality, Deixis, and Orientation in Displaced Worlds 15

discourse sovereignty over the respective progress of the story, a high degree of certainty is attributed to the realization of the event. The storyteller *Ego/I* ‘knows’ more than the propositional subject *I*. This shows that the hierarchical viewpoint (‘source’) relationship is not necessarily linked to contents of mental awareness but is constituted by the functional discourse role.

1.2 Modality, Deixis, and Orientation in Displaced Worlds

Pragmatic competence consists in anchoring our mental representations of reality in context. As mental representations pertain to form, whereas reality is made of real substance, our mental representations are reduced models of the world. Mental representations are simplified categorizations that are not able to refer per se. Pragmatic competence consists in relating mental representations to the context. It is important to keep in mind that mental representations are categorizations devoid of any time-space-person coordinates. In other words, mental representations, such as phonemes or lexemes, do not refer to existing entities in the world. They categorize reality, which means that they reduce the complexity of perceived items, i.e. the possible infinite sensory input, to finite classes or categories and their build in terms of features. We know that instinct does the same. It reduces potential infiniteness of sensory input by providing innate models of categorization. Is pragmatic competence then instinctive in nature in the first place?

In order to define pragmatic competence, we have to understand the difference between instinct-driven categorization by animals and language-driven human-specific categorization. Instinct is a quite reliable and robust means of orientation in the world and of categorization of the world. In contrast to instinct, human-specific categorizations are highly adaptive in nature. Instinct is directly innate, whereas processes of human categorizations are not directly innate, but constructions of the human mind. Nevertheless, it is commonly held that the tool that creates adaptive categorization systems in the *Homo sapiens* mind is innate. In other words, the pragmatic sections of this book are based on the axiomatics of Universal Grammar, albeit of Universal Grammar in a non-Cartesian format (see Leiss 2009a/2012b; Hinzen 2014a,b; Hinzen and Sheehan 2014). Universal Grammar (UG) cannot be reduced to the Chomskyan and to the Cartesian or Port-Royal approach to UG. Following Leiss (2009/2012a), the present basic UG approach is that of Sebastian Shaumyan (1987, 2006), Roman Jakobson (1957/1971), the Modist Universal Grammar of late Mediaeval Europe, and the universal dependency syntax of Lucien Tesnière (1959/1965). As to the universal pragmatic component, the present approach is largely indebted to Charles S. Peirce

16 *Pragmatics: Modality and Speaker Orientation*

(1982) as well as to Karl Bühler (1934/1982; 1934/2011). See, above all, Leiss (2009a, and second edition thereof in 2012a).

As soon as we understand that instinct is replaced in the *sapiens* mind by a more adaptive system of orientation in the world, we are in a position to explain why systems of lexical semantics as well as phoneme systems differ from language to language, while nevertheless displaying architectures built by the same universal principles. In sum, language provides the human species with a universal tool that filters the input in an adaptive way. As a consequence, we hold, in sharp contrast to Steven Pinker, that language is not an instinct. Quite in contrast and what is more, language is defined as a universal tool that is able to overwrite instinct.

Let us come back to the definition of pragmatic competence. Pragmatics has to bridge the functional abyss between human-specific mental representations and the instinct-driven categorization of animals. Animals remain rooted in the world when communicating, which is because they remain in the prison of HERE, NOW, and ME. This natural origo (in the sense of Bühler 1934/1982; 1934/2011) adds space-time-person indexes per se, which, however, remain reduced to the present context and which can be shared with other members of the species. In contrast to this, the linguistic animal, which we are, is able to overcome the natural viewpoint (origo) and to anchor it in distant places (by aspect), distant times (by tense), distant worlds (by mood), and in distant modes of security as to truth veridicality (by modality). Mood arises when we recategorize the world in variable ways, with the advantage that we are able to construe possible worlds (moods such as irrealis and optative), but with the disadvantage that we can never be sure that our constructions of distant worlds are completely reliable. The tool of modality helps us out of this dilemma. Modality is a tool kit that is specialized for the evaluation of certainty, or uncertainty, of information provided by the propositional part of a sentence. Overall, modality helps us as a compass of orientation in displaced worlds.

As human categorizations are neither mainly instinct-driven nor essentially reduced to the HERE/NOW/ME, the linguistic animal that we are needs bridges between the factual origo and displaced possible worlds in order to be able to refer to, and to anchor, mental representations in the world. This anchor is reference, especially higher order reference created by linguistic tools, as pointed out before: aspect (reference to distant spaces), tense (reference to distant times), mood (reference to possible worlds), and modality (reference to secure, or less secure, mental representations of the world in one's own mind as well as in others' minds).

The universal linguistic tool that liberates the human species from the immediate context (origo) involves complex displacement techniques. These

1.2 Modality, Deixis, and Orientation in Displaced Worlds 17

techniques are linguistic in nature and not of some general cognitive quality. Leiss (2012a) goes with Bickerton (1990, 2012) in assuming that the so-called general cognitive faculties are reducible to those faculties that we share with the animal brain. In contrast to the animal brain, however, the mind of the human species can be defined as consisting of an animal brain plus additional analytic techniques. These analytic and feature-based techniques optimize the animal brain and transform it into the human mind. These techniques are grammatical techniques involving syntax and grammatical semantics. The latter cannot be reduced to formal semantics alone but involves the classical inventory of grammatical categories. With respect to modality, the involved grammatical categories are aspect, tense, mood, as well as all modes of modalities that elaborate further these basic building blocks.

Reference to the displaced world enables the human mind to establish reference to displaced minds, better known under the label of others' minds or the faculty of Theory of Mind. In order to understand this process, we have to delve deeply into the nature of the aspect-tense-mood (-epistemic modality) (ATM(E)) architecture of the proposition.² A thorough understanding of the ATM-architecture is a prerequisite to understanding the illocutionary function of the sentence, which consists to a large degree of different modes of modalities.

At the very core of ATM, as well as of human-specific pragmatic competence, we find deixis. Deixis is involved in, and operates under, the condition that two members of a species share the same context. Deixis begins already with eye contact and, consequently, with shared attention to some part of the surrounding context. This kind of 'pointing with the eye' is present in a large range of animals, such as in birds and in some domesticated mammals. What is special in humans and protohumans is that they begin to point with their hands, whereby one hand is specialized in pointing. The non-deictic hand, by contrast, remains animal-like, to put it somewhat provocatively. The intricate relationship between handedness and the evolution of language has been thoroughly studied by the epochal work of Leroi-Gourhan (1964–1965, 1987). The specialized deictic hand creates shared focus markers in shared contexts. This is the foundation of communication and of mind sharing. Animals without the tool of language do not share their minds when communicating offline, i.e. beyond the immediate context. They instead share instinct.

² ATME = aspect, tense, mood, epistemic modality. These categories are ordered with hierarchically growing complexity. This is the reason that ATM rather than the more familiar abbreviation, TMA, has been chosen.

18 *Pragmatics: Modality and Speaker Orientation*

In human communication, speaker deixis comes in various forms of complexity. To cope with this complexity, we have to distinguish between techniques of simple speaker deixis and double speaker deixis. Emphasis will be put on the claim that these different techniques of simple and double deixis enable us to encode different modes of displacement in modality, the first being simple displacement, the second double displacement. With respect to Hockett's design feature of language, the feature of displacement is defined as the ability of a sign to refer to objects remote in space and time (Hockett and Altmann 1968; Nöth 2000: 155f.). It is still a matter of controversy as to whether the faculty of displacement can really be attributed to humans only. For instance, in the pertinent literature we find reports of spatial displacement in the honeybee and spatial as well as temporal displacement in trained apes. The decision whether displacement is a faculty unique to the human species depends on our ability to distinguish between two qualities of displacement techniques: simple displacement and double displacement.

1.3 **Simple and Double Displacement as Basic Building Blocks of Modality**

Displacement (Leiss 1992; see German *Versetzung*, as used in Zeman 2017) refers to concepts of shifting (Jakobson 1957/1971; Diewald 1999), which can be captured under the heading of perspectivization. The latter concept has been developed by cognitive psychologists (among them notably Perner 1991) and linguists interested in narrative structures (Eckardt 2012, 2014; Zeman 2017 referring to Dancygier 2012a,b,c). It seems appropriate to insert general remarks on perspectivization before we return to linguistic displacement in the framework of modality.

1.3.1 *Cognitive and Linguistic Perspectivization: The Viewpoint Constellation*

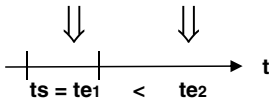
In a wide sense of the term, grammar is genuinely perspectival in that every grammatical paradigm offers a choice of alternatives. All grammatical phenomena are linked to perspectivization insofar as they are based on the speaker's deictic origo that determines the 'view' on the verbal event situation. In cognitive linguistics, this is captured in the premise that "inherent in every usage event is a presupposed viewing arrangement, pertaining to the relationship between the conceptualizers and the situation being viewed" (Langacker 2002: 16). It follows from this that grammatical

1.3 Simple and Double Displacement 19

(unlike lexical) markers do not indicate a simple viewpoint, but rather a viewpoint constellation (see Verhagen 2005; Dancygier and Sweetser 2012; see Zeman 2016, 2017 for illustrations).

Let us illustrate this by taking a look at the category of TENSE. Consider (4) below. The temporal perspectivization thus refers to the relation between primary and secondary reference points and, hence, a multiperspectival constellation between viewpoints of different qualities. As such, it requires a three-point description, as reflected in the ternary system of Reichenbach (1948) that models tenses as relations between the time of event and the time of speech as perspectivized by a third reference point (see in detail Zeman 2013, 2016; see also Evans 2005 with respect to the multiperspectivity of complex tenses in typology).

- (4) Biphasic structure of a (present) modal verb in its standard root meaning: notice the three deictic anchor points, t_s , t_{e1} and t_{e2} (corresponding roughly to Reichenbach's (1948) $\langle s, e, r \rangle$). [t =time line, t_s =speech act time, t_e =event time, t_r =reference time] Consider a modal expression like *He must come* with the obligation lying on the subject at speech act time, t_s , from which the event proceeds onwards ($t_{e1} <$), and the event happening at event time t_{e2} ; the vertical arrows signaling the two foci of the sentence meaning. Root modals thus have clear future reference just as perfective aspect and telic verbs have (Abraham 1989 et sequ.).



Temporal adverbials are simple shifters. By contrast, grammatical tense markers are prototypical double shifters insofar as they shift the focus on the displaced viewpoint while the primary origo is only maintained in the background. The same holds, on a different level, for markers of evidential and epistemic modality that display a more complex perspectival structure. As laid out in the introductory section, there seems to be a strong link between the development of ToM capabilities and the comprehension of epistemic and evidential modality. These categories are late developments in language acquisition (Perner 1991: 150; Papafragou 1997: 16, 399; Papafragou 2001; Leiss 2012b). Epistemic and evidential meanings share with propositional attitudes the feature that they operate on propositions and as such have acquired the status of overt markers of speakers' attitudes. Both evidentiality and epistemicity thus display a similar perspectival structure by embedding veridical statements that may be true to fact or false. This is seen in the fact

20 *Pragmatics: Modality and Speaker Orientation*

that only (5b), but not (5a), allows for translation into the structure of an epistemic modal verb.³

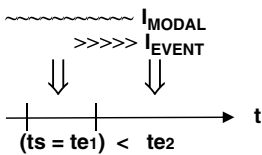
- (5) a. Elisabeth **must** walk her dog Schmeckl through the English Garden.
 → ‘She is obliged to do so.’ = DMV/modal root-reading
- b. Elisabeth **must** right now be walking her dog Schmeckl through the English Garden.
 → ‘I (i.e. the speaker) guess that she is doing so right now’ = EMV-reading (but reality could be different)

In contrast to (5a), (5b) leads to the potential for viewpoint contrasts between the level of the sentence subject and the level of the speaker, and as such, the potential for an internal1 and an external1 view in the structural sense (i.e. inside vs. outside of the proposition). This is also a crucial difference of (5b) when compared with the perspectival structure of root modals as in (5a) (see also Zeman 2017).

Root modals like *must* in (5a) are characterized by their biphasicness, since they refer to two different time intervals, i.e. the temporal interval of the modal in the real world (i.e. the time for which the obligation holds) and the time interval of the event denoted by the infinitive complement in a possible world. As such, root modals display the potential for a focus shift from the present viewpoint to the future event, as seen in the fact that modals constitute a grammaticalization source for future tenses. The main difference with respect to the epistemic example (5b) is that the latter introduces the speaker’s viewpoint, and, hence, an additional potential contrast between two belief systems (external2 vs. internal). While the obligation in (5a) holds for the subject, i.e. Elisabeth, the epistemic meaning in (5b) scopes over the whole proposition p [Elisabeth is walking through the English Garden] and thus requires an external viewpoint of evaluation. See (6) (Abraham 1989, here adapted from Zeman 2017).

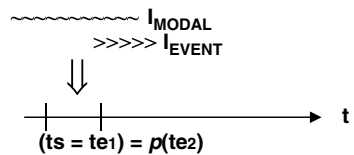
(6) Biphasic structure of (present) modal verbs

a root meaning



(I = interval; ts = speech time; te = event time; p = proposition)

b epistemic meaning



³ The obvious contingency of the MV-bias, DMV vs. EMV, and the aspect of the verb modified by the MV will be discussed at length in Chapters 4 and 5.

1.3 Simple and Double Displacement 21

The perspectival complexity of (6b) is thus based on the structural embedding of viewpoints (i.e. inside vs. outside the proposition), the distinction between two belief contents, and a contrast of veridicality on the propositional level, i.e. the fact that the proposition could be false. Thus, once again, the construction integrates external and internal viewpoints at the same time. As such, they display a pattern of viewpoint integration. The difference in degree of complexity between (6a) and (6b) is also reflected in language acquisition. Complementation structures like *think that p* are a later acquisition development than constructions like *think of p* that can be analyzed in terms of perspective-taking. In sum, it is not the potential viewpoint switch – rather than perspective-taking – that is crucial with respect to the difference between both constructions, root vs. epistemic. As has been pointed out already, the viewpoint constellation is based on the contrast between two different hierarchical levels (internal₁ vs. external or, with reference to aspect, inner vs. outer₁), two different belief systems (internal vs. external).

1.3.2 Simple and Double Displacement

Simple displacement is present in temporal adverbials such as *yesterday*. Double displacement is involved in grammatical categories such as TENSE. We will demonstrate that there is a fundamental difference between simple and double displacement, the first being less complex in its architecture. This statement implies that languages that don't have the grammatical category of tense cannot compensate for this by more frequent use of temporal adverbials. Temporal symbols for *yesterday* can be mastered by trained higher primates (Savage-Rumbaugh and Fields 2011). However, the technique of double displacement, which is present in Tense, cannot be trained. Temporal adverbials such as *yesterday* are shifters (in the sense created by Jakobson 1957/1971). In order to understand a shifter such as *yesterday*, the hearer has to shift to the viewpoint of the speaker, otherwise he would not be able to calculate the reference of the time adverbial. This is not the case with adverbials coding absolute time, such as *in 1936*, where even a simple shifting process is missing.

A successful shifting process involves the mental displacement of the speaker to the viewpoint of the hearer. Otherwise, the hearer would not be successful in calculating temporal reference. In contrast, to master a double shifter (double displacement) such as PAST TENSE or FUTURE TENSE, the hearer has to shift to the viewpoint of the speaker. With grammatical categories such as TENSE, the remit of displacement therefore increases in complexity. The speaker as well as the hearer has to split additionally into a speaker (I) and a viewer (ME). The speaker remains placed in the present, whereas the viewer is