16.3: Modality as quantification over possible worlds

Angelika Kratzer (1981; 1991) proposed that the English modals are not in fact polysemous. On the contrary, she suggested that English (like a number of other languages) has only one set of modal operators, which are underspecified (indeterminate) regarding the type of modality (epistemic, deontic, etc.). The strength of the modal is lexically determined, with the individual modals functioning semantically as a kind of quantifier that quantifies over situations. The specific type of modality depends on the range of situations which is permitted by the context. This section offers a brief and informal introduction to her approach.

16.3.1 A simple quantificational analysis

Kratzer’s analysis builds on a long tradition of earlier work that treats a modal auxiliary as a kind of quantifier which quantifies over possible worlds. (We can think of possible worlds as possible situations or states of affairs; in other words, “ways that things might be”.) A marker of necessity functions as a universal quantifier: it indicates that the basic proposition is true in all possible states of affairs. A marker of possibility functions as an existential quantifier: it indicates that there is at least one state of affairs in which the basic proposition is true.

In Chapter 14 we introduced two symbols from modal logic: $\Diamond = \text{‘it is possible that’}$; and $\Box = \text{‘it is necessarily the case that’}$. The use of these symbols is illustrated in the logical forms for two simple modal statements in (11).

\[
\begin{align*}
(11) \text{a. } & \text{ Arthur must be at home. logical form: } \Box \text{ AT\_HOME(a)} \\
& \text{b. Arthur may be at home. logical form: } \Diamond \text{ AT\_HOME(a)}
\end{align*}
\]

The possible worlds analysis claims that the logical forms in (11), which make use of the modal operators, express the same meaning as those in (12), which are stated in terms of the standard logical quantifiers. The $w$ in (12) is a variable which stands for a possible world or state of affairs. So under this analysis, \textit{Arthur must be}
home means that the proposition Arthur is home is true in all possible worlds, while Arthur might be home means that the proposition Arthur is home is true in at least one possible world.

(12) a. Arthur must be at home. meaning: \( \forall w [A\_HOME(a) \text{ in } w] \)
b. Arthur may be at home. meaning: \( \exists w [A\_HOME(a) \text{ in } w] \)

As we noted in §16.2, words like must and may allow both epistemic and deontic readings (among others). These different types (or “flavors”) of modality can be represented by different restrictions on the quantification, i.e., different limits on the kinds of possible worlds that the quantified variable \( w \) can refer to. Epistemic readings arise when \( w \) can range over all "epistemically accessible" worlds, i.e., situations which are consistent with what the speaker knows about the actual situation. Deontic readings arise when \( w \) can range over all "perfect obedience" worlds, i.e., situations in which the requirements of the relevant authority are obeyed. This analysis is illustrated in (13–14), using the restricted quantifier notation.

(13) Arthur must be at home

a. Epistemic: [all w: w is consistent with what I know about the actual world] AT_HOME(a) in w
b. Deontic: [all w: w is consistent with what the relevant authority requires] AT_HOME(a) in w

(14) Arthur may be at home.

a. Epistemic: [some w: w is consistent with what I know about the actual world] AT_HOME(a) in w
b. Deontic: [some w: w is consistent with what the relevant authority requires] AT_HOME(a) in w

The unrestricted quantifications in (12) express logical possibility or necessity: a claim that proposition \( p \) is true in at least one imaginable situation, or in every imaginable situation. Such statements are said to involve alethic modality. As von Fintel (2006) points out, “It is in fact hard to find convincing examples of alethic modality in natural language.” An example of logical (or alethic) possibility might be the statement, “I might never have been born.” It is possible for me to imagine states of affairs in which I would not exist (my father might have been killed in the war, my mother might have chosen to attend a different school, etc.); but none of these states of affairs are epistemically possible, because they are inconsistent with what I know about the real world. Examples of logical (alethic) necessity are probably limited to tautologies, analytically true statements, etc.; it is hard to find any other type of statement which must be true in every imaginable situation.

Analyzing modals as quantifiers accounts for a number of interesting facts. For example, the simple tautologies of modal logic stated in (15) show how either of the two modal operators can be defined in terms of the other. (15a) states that saying \( p \) is possibly true is equivalent to saying it is not necessarily the case that \( p \) is false. (15b) states that saying \( p \) is necessarily true is equivalent to saying it is not possible that \( p \) is false. It turns out that the two basic quantifiers of standard logic can be defined in terms of each other in exactly the same way, as shown.
by the tautologies in (16). This remarkable parallelism is predicted immediately if we analyze necessity in terms of universal quantification and possibility in terms of existential quantification.

(15) a. \( (◊ p) ↔ ∀(□ ¬p) \)
b. \( (□ p) ↔ ∀(◊ ¬p) \)

(16) a. \( ∃x[P(x)] ↔ ∀(∃x[¬P(x)]) \)
b. \( ∀x[P(x)] ↔ ∃(∀x[¬P(x)]) \)

We noted in Chapter 14 that combining quantifiers and modals in the same sentence often leads to scope ambiguities. The examples in (17–18) are repeated from Chapter 14. The quantificational analysis again predicts this fact: if modals are really quantifiers, then the ambiguities in (17–18) arise as expected from the interaction of two quantifiers.

(17) Every student might fail the course.\(^5\)
   a. \( ∀x[STUDENT(x) → ◊ FAIL(x)] \)
   b. \( ◊ ∀x[STUDENT(x) → FAIL(x)] \)

(18) Some sanctions must be imposed.
   a. \( ∃x[SANCTION(x) ∧ □ BE-IMPOSED(x)] \)
   b. \( □ ∃x[SANCTION(x) ∧ BE-IMPOSED(x)] \)

While this analysis works well in many respects, Kratzer points out that it makes the wrong predictions in certain cases. For example, suppose that Arthur has robbed a bank, and that robbing banks is against the law. Intuitively, we would say that sentence (19a) is true in this situation. However, the analysis shown in (19b) actually predicts the opposite, because in all possible worlds consistent with what the law requires, no one robs banks. In particular, Arthur does not rob a bank (or commit any other crime) in those worlds, and so would not go to prison. Similarly, the analysis predicts that both (20a) and (20b) should be true, because the antecedent will be false in all possible worlds consistent with what the law requires. (Recall from Chapter 4 that \( p → q \) is always considered to be true when \( p \) is false.)

(19) a. Arthur must go to prison. [Deontic]
   b. [all w: w is consistent with what the law requires]
      GO_TO_PRISON(a) in w

(20) a. If Arthur has robbed a bank, he must go to prison.
   b. If Arthur has robbed a bank, he must not go to prison.

To take another example, suppose that when a serious crime is committed, the law allows the government to confiscate the house, car, and other assets of the guilty party to compensate the victim; but that the government
is not allowed to confiscate the assets of anyone who does not commit a crime. If Arthur is convicted of a serious crime, the judge may truthfully say the sentence in (21a). But once again, the analysis in (21b) predicts that this statement should be false, since there is no possible world consistent with what the law requires in which Arthur commits a crime, so no such world in which his assets may be confiscated.

(21) a. The state may confiscate Arthur’s assets. [Deontic]
   b. [some w: w is consistent with what the law requires]
      the state confiscates Arthur’s assets in w

The problem with examples of this type is that we begin with an actual situation that is not consistent with what the law requires. The correct interpretation of the modal reflects the assumption that what happens next, in response to this non-ideal situation, should be as close to the ideal required by law as possible.

16.3.2 Kratzer’s analysis

Kratzer addresses this problem by arguing that restrictions on the sets of possible worlds available for modal quantifiers must be stated in two components. The first, which she calls the modal base, specifies the class of worlds which are eligible for consideration, i.e., worlds that are accessible. The second component, which she calls the ordering source, specifies a ranking among the accessible worlds. It identifies the “best”, or highest-ranking, world or worlds among those that are accessible. The modal’s domain of quantification contains just these optimal (highest-ranking) accessible worlds.

Let us see how this approach would apply to example (19a). Deontic modality involves a circumstantial modal base, i.e., one that picks out worlds in which certain relevant circumstances of the actual world hold true. In this case, one of the relevant circumstances of the actual world is the fact that Arthur has robbed a bank. The relevant ordering source in this example is what the law requires: the optimal worlds will be those in which the law is obeyed as completely as possible, given the circumstances. An informal rendering of the interpretation of this sentence is presented in (22b). The first clause in the restriction represents the modal base, and the second clause in the restriction represents the ordering source.

(22) a. Arthur must go to prison. [Deontic]
   b. [all w: (the relevant circumstances of the actual world are also true in w) and (the law is obeyed as completely as possible in w)]
      GO_TO_PRISON(a) in w

Epistemic modals require a different kind of modal base and ordering source. The fundamental difference between the two types of modality is summarized by Hacquard (2011: 1494) as follows:

Circumstantial [= root; PRK] modality looks at the material conditions which cause or allow an event to happen; epistemic modality looks at the knowledge state of the speaker to see if an event is compatible with various sources of information available.

The epistemic modal base, which would be relevant for epistemic modals like that in (23a), picks out worlds consistent with what is known about the actual world, i.e., consistent with the available evidence. Epistemic
modals frequently invoke a stereotypical ordering source: the optimal worlds are those in which the normal, expected course of events is followed as closely as possible, given the known facts. An informal rendering of the interpretation of (23a) is presented in (23b).

(23) a. Arthur must be at home. (=13a) [Epistemic]

b. [all w: (w is consistent with the available evidence) and (the normal course of events is followed as closely as possible in w)]
   AT_HOME(a) in w

This rendering of the meaning of epistemic must is more accurate than the analysis suggested in (13a) for the same example. That earlier analysis would lead us to predict that Arthur must be at home entails Arthur is at home, since the actual world is one of the worlds that are consistent with what the speaker knows about the actual world. But this prediction is clearly wrong; saying Arthur is at home makes a more definite claim than Arthur must be at home. By using must in this context, the speaker is implying: “I do not have direct knowledge, but based on the evidence I can’t imagine a realistic situation in which Arthur is not at home.” The use of the stereotypical ordering source in (23b) helps to account for this inferential character of epistemic must. It helps us understand why statements of epistemic necessity are usually better paraphrased with the adverb evidently than with necessarily.6

Another important part of Kratzer’s proposal is the claim that the modal auxiliaries in languages like English and French are not in fact polysemous. Kratzer suggests that the lexical entry for words like must and may specifies only the strength of modality (i.e., the choice of quantifier operator), and that they are indeterminate as to the type or “flavor” of modality (epistemic vs. deontic, etc.). The type of modality depends on the choice of modal base and ordering source, which are determined by context (linguistic or general).

Part of the evidence for this claim is the observation that type of modality can be overtly specified by adverbial phrases or other elements in the sentence, as seen in (24).7 Notice that these adverbial phrases do not feel redundant, as they probably would if the modal auxiliary specified a particular type of modality as a lexical entailment. For sentences where there is no explicit indication of type of modality, the intended type will be inferred based on the context of the utterance.

(24) a. Epistemic:
   (In view of the available evidence,) John must/may be the murderer.

b. Deontic:
   (In view of his parents’ orders,) John may watch TV, but he must go to bed at 8pm.

c. Ability:
   (In view of his physical abilities,) John can lift 200 lbs.

d. Teleological:
   (In view of his goal to get a PhD,) John must write a dissertation.

e. Bouletic:
(In view of his desire to retire at age 50,) John should work hard now.

While Kratzer’s analysis provides an elegant explanation for the unusual pattern of polysemy which we discussed in §16.2, this explanation cannot be applied to all grammatical markers of modality. In the next section we discuss examples of modals for which type of modality seems to be lexically specified.

---


6 Kratzer states that another advantage of her theory is that it provides a better way to deal with “graded modality” i.e. intermediate-strength modals of “weak necessity” like ought or should, as well as phrases such as very likely or barely possible. We will not discuss graded modality in this chapter.

7 From Hacquard (2011).