A general typology of modal categories

DIETMAR ZAEFFERER
Institut für Theoretische Linguistik, Ludwig-Maximilians-Universität
München, Deutschland

1. DEFINING ‘MODAL CATEGORY’: AN ONTOLINGUISTIC APPROACH
The first and possibly least controversial claim I am going to defend in this paper is that in order to stabilize the notion of modality in linguistics in general and in cross-linguistic research in special it is necessary to start from the outside, from the ontological and nonlinguistic domain of pure concepts. The nonlinguistic concept of modality is an entirely notional category, a dimension of variation some entities can be assigned to and others cannot, and which therefore can be defined completely independently of the ways its categories are encoded in human languages. Although this view is far from new it never seems to have been implemented in a consistent way. There are many complaints in
the literature about the vagueness of the category of modality\textsuperscript{2} and the heterogeneity of its use, and there are also several attempts around at defining it, including my own proposal,\textsuperscript{3} but there still seems to be little consensus about at least a range of acceptable definitions.

There exist in principle two strategies for defining modal categories, an expansion strategy and a restriction strategy. In general, linguists prefer the former, working from the inside out, so to speak, where the center is given by some unquestionable modal categories. A typical representative of this approach is Angelika Kratzer, who writes: "Modality has to do with necessity and possibility" (Kratzer 1991: 639). The biggest problem with this strategy is that it does not tell us where to stop: Having-to-do-with is an extremely flexible rubber band relation.\textsuperscript{4} Logicians, on the other hand, tend to prefer more precise albeit more encompassing definitions. Nicholas Rescher, e.g., writes (1968: 24): "When such a proposition is itself made subject to some further qualification of such a kind that the entire resulting complex is itself once again a proposition, then this qualification is said to represent a modality to which the original proposition is subjected."

For linguistic purposes, this is of course far too general in three respects: (a) Qualitatively, it includes a dimension of variation that ought to be kept separate, although it is clearly related, namely reference to time, traditionally the domain of tense, and temporal structure, the domain of actionalities. A logician may speak of temporal modality (as Rescher does), but linguists will not be comfortable with this terminology. (b) It also includes, although not explicitly acknowledged, reference to space, which does not belong to a linguistically fruitful concept of modality either. Modality rather refers to abstract space, to the space of propositions and their alternatives. (c) Quantitatively, even the amended Rescher proposal (after the exclusion of temporal and spatial propositional operators) leaves us with infinitely many modalities, since there are, of course, infinitely many different ways of further qualifying a proposition in a way that leads to a new proposition without being either temporal or spatial, so for linguistics this definition does not seem to be very satisfying either.
The strategy that I will adopt for this article and propose for adoption by the community of modality researchers is a combination of the restriction strategy with the expansion strategy. The restriction is an ontological one. It restricts the infinity of modalities admitted by the amended Rescher definition to those that belong to the top levels of a corresponding taxonomy (‘top’ is sufficiently vague to include the first two or three levels and to exclude the thirtieth), that is, to those levels that contain the most general and logically least powerful concepts.

This purely conceptual restriction has a double advantage. Firstly, it keeps conceptual clarifications distinct from empirical findings, which means that they don't run the risk of needing adjustment each time relevant empirical findings come in. And secondly, it allows us to formulate typological hypotheses about the ways modal categories are coded in human languages without any danger of circularity.

Since the present paper is primarily concerned with conceptual clarification and only secondarily with empirical linguistic typology, its main empirical hypotheses will be formulated only in rough outline:

(H 1) The linguistic coding means for modalities tend to reflect frequency of use in their compactness (number of syllables, morphemes and words) and generality in their degree of grammaticalization.

(H 2) The distribution of the most compact coding means for modalities over and across ontological domains reflects the central role of actions and attitudes for human cognition.

The first hypothesis is of course just a special case of a general preference law for the coding of concepts and will not be argued for in any detail in this paper, whereas the second hypothesis will be supported by confirming evidence below.

Whereas the reason for the restriction part of my strategy is obvious, the motivation for the expansion part needs a little elaboration. In order to prepare the ground for this, let us call modalities in the sense of Rescher (1968) interpropositional or internal modalities, because they are instantiated by propositional operators that map inner propositions to outer propositions which properly contain the former and thus remain internal to the realm of propositions. Alongside with these modalities and their operators there are other modalities whose operators also take propositions as arguments, but unlike the
internal operators they don’t have values of the same kind, viz. propositions, but something different like actions or attitudes. Therefore, they will be called transpropositional or external modalities. Compare the expression schemata John knows that $p$ and John’s knowledge that $p$, where $p$ is a proposition. Whereas the former codes an internal modal operator, one that maps the proposition $p$ to another, higher one, the latter codes an external operator, one which according to Rescher’s definitions does not count as modal operator, since it maps $p$ to a mental state, and mental states are not propositions.

Since internal modalities, by contrast with external modalities, are homogeneous, the operators that instantiate the former are structurally more transparent, and it is therefore appropriate to consider them first. One of their central properties is that they never come alone, they always build a group with normally four members. The reason is that both the underlying and the resulting proposition can be negated and so every modal operator $O$ has a dual counterpart $O'$, defined as the combination of $O$ with internal and external negation. The group of four operators thus defined is also called a duality group (cf. Lübner 1990). For example, if $O$ is the matrix schema it is necessary in a deontic sense, then its dual $O'$ is it is allowed (not necessary that not), its internal negation is it is disallowed (necessary that not), and its external negation is it is unnecessary (not necessary).

The question arises whether the negation operator itself should be considered as a modal operator and hence polarity as a modal category. In Zaefferer (2001a: pp. 789, 794f.) I have argued that the answer has to be split: The negation operator is by definition an internal modal operator, but this should not lead to the conclusion that polarity is a full-fledged modal category. Two reasons support this view: First and less importantly, the negation operator is selfdual, i.e., identical with its dual counterpart, so the other two elements of the duality group collapse as well, leaving a degenerated duality group consisting of negative and positive polarity operators only. Second and crucially, it does not seem appropriate to consider the latter as a modal operator at all, since it leaves its propositions unchanged. So I propose to consider polarity as a semimodal category, a category that contains both modal and non-modal operators.
We have stated above that external, by contrast with internal modalities, are not as nicely behaved, first because the operators that instantiate them do not yield propositions, but entities of other categories, and second because they are heterogeneous, i.e., the entities they yield belong to different categories. So it seems doubtful if the expansion part of my strategy makes sense at all, given the expected lack of correspondence and similarity between external and internal modalities. The core question is: Do external modalities also come in duality groups? At first glance, the answer seems to be no, because the very definition of duality groups presupposes the possibility of both internal and external negation, whereas external modal operators are compatible with internal negation, because they operate on propositions, but not with external negation, because their values are non-propositional and, worse, ontologically heterogeneous.

Let us come back to the example used above to illustrate an external modality, John’s knowledge that \( p \), where \( p \) is a proposition. What this noun phrase denotes is a mental state, not a proposition, so it cannot be canonically negated. On the other hand, it is obvious that there are ways of circumventing this categorical problem by using other forms of negation as in John’s lack of knowledge that \( p \) or John’s ignorance of the fact that \( p \) and similarly with other categories. The adjective phrase able to lift 200 pounds denotes a property. It can also be negated, but unable to lift 200 pounds denotes another property and not a proposition. The fact that there are complementary mental states and attitudes and other non-propositional entities does not help with the problem that duality groups require the possibility of not only internal but also external propositional negation.

The general solution is not too hard to come by: Every external modality has an internal counterpart that results from adding an existence predicate. A mental attitude is not a proposition, but There is this attitude in John denotes one. A property is not a proposition, but that it exists somewhere, that it is instantiated, e.g. by John, is one. Since existence predicates are not restricted by the ontological category of what they are applicable to, the solution is general: Not only internal, but also external modal operators come in duality groups, where in the latter case the members of the group are defined by the internal counterparts of the operators.
We can now define:

(D 1) A modal category is either an internal or an external modal category.

(D 2) An internal modal category is a taxonomically high-ranking duality group of interpropositional modal operators.

(D 3) An external modal category is a taxonomically high-ranking duality group of transpropositional modal operators.

(D 4) A modal operator is a propositional operator which is not exclusively time- or space-related.

2. SEMANTIC PROPERTIES OF MODAL OPERATORS

The attribute ‘taxonomically high-ranking’ in (D 2-3) above needs some further specification. In any ordering of concepts the taxonomically highest ranks are held by the weakest, i.e., most general concepts, those that are included in and entailed by most other corresponding concepts. But the members of a duality group have quite different degrees of strength, as can be seen in the following examples:

(1)  
a. The cat must be either alive or not alive
    b. The cat is either alive or not alive

(2)  
a. The cat may be dead
    b. The cat is dead

(3)  
a. The cat cannot be both alive and not alive
    b. The cat is both alive and not alive

(4)  
a. The cat needn’t be dead
    b. The cat is dead

If we compare the meanings of the modalized a.-sentences with the meanings of the plain b.-sentences we see that in the first two cases the members of the pairs stand in a relation of entailment, albeit in different directions. If we call the embedding proposition mother and the embedded one daughter, we can say that in (1) the mother entails the daughter and in (2) the daughter entails the mother (although the maxim of informativeness would prevent a cooperative speaker who believes (2) b. to be true from pronouncing (2) a.). The two other cases are different insofar as neither the mother in (3) entails the daughter nor does the inverse hold in (4), but if we negate the b.-sentences we get again two
opposite entailments: (3) a. entails the negation of (3) b. and the negation of (4) b. entails (4) a. If we use the term ‘corresponding mother’ for the result of applying a given operator to a given proposition, we can define:

(D 5) An interpropositional modal operator instantiates an **entailing or strong modality** if every corresponding mother entails her daughter.

(D 6) An interpropositional modal operator instantiates an **entailed or weak modality** if every corresponding mother is entailed by her daughter.

(D 7) An interpropositional modal operator instantiates a **neg-entailing or antistrong modality** if every corresponding mother entails the negation of her daughter.

(D 8) An interpropositional modal operator instantiates an **neg-entailed or antiweak modality** if every corresponding mother is entailed by the negation of her daughter.

Examples (1) through (4) above illustrate the four well-known modal operators of logical or alethic necessity, possibility, impossibility and unnecessity, which according to the definitions above instantiate a strong, a weak, an antistrong and an antiweak modality, respectively. In order to decide if this duality group of modal operators constitutes a modal category in the sense of (D 1), we have to find out whether it is taxonomically high-ranking, and in order to compare taxonomic ranks we have to fix the member of the group that serves as measuring stick. If we decide to appoint the weak modality, instantiated by the logical possibility operator, as the speaker of its group, then the result is obvious: The logical modal operators constitute a modal category in the sense of (D 1) since logical possibility is certainly a taxonomically very high ranking concept which dominates not only all concepts that hold of true propositions, but also all concepts that hold of propositions which are conceivable without considerable effort.

But there are many interpropositional modal operators that lack the entailment properties defined above. Let’s give the difference between the two groups a name:

(D 9) An interpropositional modal operator instantiates an **autonomous modality** if it has one of the properties described in (D 5)-(D 8), else it instantiates a **dependent modality**.

Many expressions in human languages code modalities that by themselves fail to
have one of the entailment properties that define autonomous modalities, but which acquire them as soon as some easily inferable additional condition holds. Since for allowing these entailments they depend on support from outside, they are called ‘dependent modalities’ here. Consider deontic modalities:

(5)  
a. It is required that all students attend this meeting  
b. All students attend this meeting  

(6)  
a. It is permitted to smoke at the bar  
b. People smoke at the bar  

(7)  
a. It is forbidden that people enter the lab without permission  
b. People enter the lab without permission  

(8)  
a. It is unnecessary that every letter be filed  
b. Every letter is filed  

Obviously, (5) a. does not entail (5) b., (6) b. fails to entail (6) a., (7) a. does not entail the negation of (7) b., and the negation of (8) b. does not entail (8) a. But all these entailments simultaneously become valid if a single additional premise is added:

(9)  The relevant rules are observed

If not every letter is filed and the relevant rules are observed, then it is unnecessary that every letter be filed, and analogously for the other cases. It follows that despite the difference it is possible to define notions for the dependent modalities that exactly parallel those defined in (D 5) through (D 8) and which differ only in the modifier conditionally. So we can say that the permission operator instantiates a conditionally entailed modality because every corresponding mother is entailed by her daughter on the condition that the relevant rules are observed.

If we appoint the conditionally entailed modality as the speaker of its group, analogously to the way we proceeded with the autonomous modalities, the result is again rather clear: The general deontic modal operators constitute a modal category in the sense of (D 1) since deontic possibility is certainly a taxonomically high ranking concept which dominates everything that is done.
A GENERAL TYPOLOGY OF MODAL CATEGORIES

in accordance with the relevant rules. Compare now the special deontic operator encoded in (10).

(10)  a. Mr. Miller permitted his employee Ms. Smith to use his seal
       b. Mr. Miller’s employee Ms. Smith used his seal

Assuming that the relevant rules for an employee come always from her boss, (10) b. conditionally entails (10) a. and therefore the modal operator encoded by Mr. Miller permitted his employee Ms. Smith instantiates a conditionally entailed modality, but the corresponding duality group of special deontic modal operators does not constitute a modal category in the sense of (D 1) since permission granted by Mr. Miller to his employee Ms. Smith is certainly not a taxonomically high ranking concept, but a rather special and therefore low-ranking one.

Having introduced thus two orthogonal binary distinctions, that between internal and external modalities on the one hand and that between autonomous and dependent modalities on the other, we are now in a position to discuss the central proposal of this paper, that of a general typology of modal categories.

3. FOUR MAJOR GROUPS OF MODAL CATEGORIES
The second and probably far more controversial claim I am going to argue for is that on the basis of the definition developed above the modal categories fall into the following four major groups: Action modalities, general modalities, attitudinal modalities and illocutionary modalities. The first two groups share the feature of including both autonomous and dependent modalities and owe their names to the fact that the first group is restricted to daughter propositions that are about human actions and the second has analogous subcategories with this restriction removed, so it seems plausible to assume that they originated – and should be treated – as generalizations from the former.

The relation between the second two groups is rather different. They share the property of including only dependent modalities (with one prominent exception), but they differ in being internal (first group) versus external (second group). The first group is called attitudinal modalities, because they are about propositional attitudes, the second one is called illocutionary modalities,
because they are about illocutionary act types (and therefore external). Here is the overview:

1. Action modalities
   1.1. Autonomous action modalities
      1.1.1. Action disposition modalities
      1.1.2. Action circumstance modalities
   1.2. Dependent action modalities
      1.2.1. Conditional action modalities
      1.2.2. Deontic modalities

2. General modalities
   2.1. Autonomous general modalities
      2.1.1. General disposition modalities
      2.1.2. General circumstance modalities
   2.2. Dependent general modalities
      2.2.1. General conditional modalities
      2.2.2. Alethic (ontic) modalities

3. Attitudinal modalities
   3.1. Presentation modalities
      3.1.1. Optative modalities
      3.1.2. Volitional modalities
   3.2. Representation modalities
      3.2.1. Hypothetic modalities
      3.2.2. Epistemic modalities

4. Illocutionary modalities
   4.1. Volitional illocutions
      4.1.1. General volitional illocutions
      4.1.2. Epistemic volitional illocutions
   4.2. Optative illocutions
   4.3. Hybrid illocutions
      4.3.1. Expressive assertives
      4.3.2. Exclamations
3.1. Action modalities
3.1.1. Autonomous action modalities
3.1.1.1. Action disposition modalities
The most prominent and possibly universally lexicalized action disposition modality is ability, which is an autonomous weak modality because every corresponding mother is directly entailed by her daughter, e.g., (11) a. by (11) b.

(11) a. The child is able to solve the problem
    b. The child solves the problem

This is an extremely common inference pattern: In order to find out about competence one observes performance. The whole world of tests and examinations rests on this.

Obviously the inverse inference pattern is not valid: If the child is able to solve the problem it still might not solve it, for instance because it simply does not want to, or because there is an obstacle.

Some languages offer a different modal verb for this latter case. Of the two French modal verbs savoir and pouvoir the first one codes principled and the latter circumstantial action disposition, the former being an individual level and and the latter a stage level property.\textsuperscript{11}

(12) a. Amélie sait écrire
    Amélie can write (she learned it)

    b. Amélie peut écrire
    Amélie can write (there is no obstacle)

    c. Amélie écrit
    Amélie writes

(12) c. autonomously entails both (12) a. and (12) b., and the negation of either one of (12) a. and (12) b. entails the negation of (12) c.

The strong counterpart of the weak action disposition modality ‘ability’ is not lexicalized as a single word in English, but as an idiom that requires a gerund:

(13) a. I can’t help laughing about this
    b. I’m laughing about this
Clearly, (13) a. autonomously entails (13) b., that is, without further assumptions.

3.1.1.2. Action circumstance modalities
The second category of autonomous action modalities to be mentioned here is the rich and heterogeneous group of action circumstance modalities such as in *He managed to open the door*. It is closely related with the conditional action modalities to be discussed in the next section, because it differs from them only by incorporating an assumption that makes them autonomous and which conditional modalities lack: the presence of some circumstance of the action under consideration. Most prominent are causation-related action circumstances: Reasons and counter-reasons or, more generally, causes and antagonistic forces.12

Here, an interesting observation can be made: There are cases where what looks like the dual counterpart of a strong modality, a combination of the strong modal operator with internal and external negation, turns out to be a strong modality as well, with the only difference that the direction of the force, causing or antagonistic, is inverted. How is this possible? The reason is that these operators presuppose the existence of a circumstance, so what looks like an external (weak) negation is in fact a presupposing one and what results amounts to double negation. Here are a couple of examples:

(14) a. His words made us laugh
    b. His words didn’t prevent us from laughing
       (= ... didn’t make us not laugh)
    c. We laughed

Both (14) a. and b. entail c., and both presuppose that he said something, but (14) a. entails that what he said was a reason for laughing, whereas (14) b. entails that his words were a reason for not laughing, so in the former case we laugh because of his words and in the latter case we laugh in spite of his words.

There is an action circumstance modality that is rather salient in human action, that of a counter-reason for acting, namely the circumstance that the action is dangerous. At least three languages have lexicalized this modality in a modal
verb, i.e., they have to a certain extent even grammaticalized it:\textsuperscript{13} Middle High German, Mandarin Chinese, and English. For the sake of simplicity, the example is taken from the latter.

(15) \begin{enumerate}
\item We dared to follow him
\item We didn't dare not to follow him
\item We followed him
\end{enumerate}

Here in a way the situation is the inverse of the one observed above with make. Again the c.-sentence is entailed by both of the other ones, again the a.- and the b.-sentence presuppose something, in this case that the relevant action, to follow him, is deemed risky by the agents, but here the first sentence entails that the action takes place against the presence of a counter-reason, i.e. despite the risk, whereas the second sentence, the pseudo-dual counterpart of the first one, entails that the action is in line with a reason, because of the risk, or more precisely, because the action avoids the risk.\textsuperscript{14}

3.1.2. Dependent action modalities
3.1.2.1. Conditional action modalities

By definition, for dependent action modalities to display their entailment properties the assumption of some additional premise is required. Consider the following example:

(16) \begin{enumerate}
\item If he goes to Küsnacht, he must come through this ravine
\item He comes through this ravine
\item He doesn't come through this ravine
\end{enumerate}

Obviously, there is no entailment relation from (16) a. to b. or c., nor the other way around, so these sentence pairs lack logical connectedness. But taken together with (17) a. (16) a. entails first (17) b. and then (16) b.

(17) \begin{enumerate}
\item He goes to Küsnacht
\item He must come through this ravine
\end{enumerate}

The inference from (17) b. to (16) b. is valid because this is a case of rigid physical necessity.\textsuperscript{15} So a conditional action modality is a dependent modality,
in the sample case a strong one. (Note that these modalities often depend not just on the explicitly stated salient condition, but also on some tacit background assumptions, that are taken for granted and therefore not explicitly mentioned; see below.)

Conditional action modalities show an interesting connection with volitionality, a connection that recently has been the topic of a lively debate on anankastic conditionals and related issues.\textsuperscript{16} Let’s replace the antecedent action of (16) a. by the corresponding volition, and the modalized consequent by the plain action:

\textit{(18) If he wants to go to Küssnacht, he comes through this ravine}

This does not make very much sense, since Gessler’s volition alone cannot be sufficient for his coming through here. How about (19)?

\textit{(19) If he wants to go to Küssnacht, he wants to come through this ravine}

This may be true in some cases, but often enough one has to put up with unwanted prerequisites for one’s goal. So in our case an anankastic conditional as in (20) is the most appropriate description of the situation.

\textit{(20) If he wants to go to Küssnacht, he must come through this ravine}

But here we are faced with a similar question as above: How can his volition be sufficient for this necessity? Some authors have even denied that volition plays a role here.\textsuperscript{17} This cannot be quite right, since the inference from (20) together with (21) a. to (21) b. seems to be valid:

\textit{(21) a. He wants to go to Küssnacht
          \hspace{1cm} b. He must come through this ravine
          \hspace{1cm} c. He comes through this ravine}

The problem is that by contrast with the nonvolitional case above the further inference to (21) c. does not seem to be valid. How come? My proposal for an explanation rests on two elements. The first element is practical reasoning. If
A GENERAL TYPOLOGY OF MODAL CATEGORIES

somebody rationally wants to reach a goal, then it is possible to reach this goal. And if somebody wants to reach a reachable goal, it is reasonable to reckon with the possibility that he does so. So the proposition that he realizes his plan is not entailed by, but cognitively salient in a context where a volitional attitude has been stated. The second element is covert conditionalization. I assume that the must in (20) and in (21) b. is interpreted as implicitly conditionalized by the proposition made salient by the antecedent: If Gessler realizes his plan, i.e., if he goes to Küsnacht, then he must come through the ravine. So if we add (22) to (20) and (21) a., the inference to (21) c. goes through:

(22) He does what he wants to do

This analysis of anankastic conditionals is very similar to that of von Stechow et al. (2005), but by contrast with them it takes the volitional seriously, and it explains the validity of the inference from the two premises (20) and (21) a. to (21) b. It does so by claiming that anankastic conditionals are doubly conditional: The overt condition is the wanting and the covert condition is the realization of the plan. So I submit that (20) is interpreted as shorthand for (20'):

(20') If he wants to go to Küsnacht, then if he does so he must come through this ravine

This rationality-based approach also accounts for the contrast between (23) and (24):

(23) If you want to be rich, you must work very hard
(24) If you wish you were rich, you must work very hard

Mere wish, by contrast with rational volition, does not make it reasonable to reckon with its realization, so there is no salient antecedent for a covert condition on the must, and an unconditional must doesn't make much sense in this context.

3.1.2.2. Deontic modalities

Deontic modalities have already been introduced as prototype of a dependent modality in section two above. They differ from the overt conditional action
modalities by normally hiding their type-specific condition ’All relevant rules are observed’: If people must not kill as a deontic necessity and the relevant rules are observed, then people must not kill as an analytical necessity and this entails that people don’t kill. This is another example of the covert conditionalization discussed in the last section.

The examples for deontic modalities presented in section two did not involve modal auxiliaries. The reason is that the structure of the latter is less transparent than that of the matrix clauses in the examples (5) through (8). Consider (25):

(25) a. You must drive on the left-hand side of the road
   b. All relevant rules are observed
   c. You drive on the left-hand side of the road
   d. You drive

Together with (25) b., the covert condition for deontics, (25) a. should entail (25) c., but since this in turn entails (25) d., the inference from a. and b. to d. should be valid, which of course is not the case.

The structure for the solution of this problem has been introduced in the last section: Double conditionalization. If we assume that under the intended interpretation (25) a. is a shorthand for (26), then the inference from (25) a. and b. to c. and d. is blocked unless we add d. to the premises.

(26) If you drive you must drive on the left-hand side of the road

3.2. General modalities

My claim for the general modalities is that they behave essentially like action modalities but are not restricted to operators on action propositions. Therefore, this section can be rather short.

3.2.1. Autonomous general modalities

3.2.1.1. General disposition modalities

A nice example for the distinction between action disposition and general disposition modalities is provided by a pair of English adjectives, cf. (27) versus (28):
(27)  a. The halting problem is solvable  
     b. The halting problem is solved  
(28)  a. Barium Nitrate is soluble  
     b. Barium Nitrate is dissolved  
     c. Barium Nitrate dissolves

Both are weak modalities, i.e. the daughters (b.-sentences) entail their mothers (a.-sentences), both are autonomous, but only in (27) the daughter has to be an action, whereas in (28) other events are possible, as the c.-sentence shows. Note that in its generic reading (28) c. is synonymous with (28) a., showing the phenomenon of zero marked modalization (cf. Zaefferer 2001a, section 1.4.1.).

3.2.1.2. General circumstance modalities
Example (29) shows that the phenomenon of pseudo-duality (combination of interior and presupposing exterior negation) also exists outside the realm of actions, reasons and counter-reason. As indicated above, it carries over to events, forces and counter-forces.

(29)  a. These forces made the house fall apart  
     b. These forces didn’t keep the house from falling apart  
     (= ... didn’t make it not fall apart)  
     c. The house fell apart

Both (29) a. and b. express strong modalities: Both entail c., but the a.-sentence allows furthermore the inference that the forces referred to by the subject, say, of some explosion inside, are the winning destructive forces (the house falls apart because of them) whereas b. entails that those forces, provided, e.g., by some outside props, are the losing counter-forces (the house falls apart inspite of them).

3.2.2. Dependent general modalities
3.2.2.1. General conditional modalities
Conditional modalities are ubiquitous and like the autonomous modalities they are not restricted to actions. (30) a. exemplifies a strong modality. The conditional character is reflected in the fact that the mother-daughter inference from a. to c. is valid only with the help of b.
(30)  a. If the water inside the bottle freezes, it will expand
    b. The water inside the bottle freezes
    c. The water inside the bottle will expand

3.2.2.2. Alethic (ontic) modalities
Alethic modalities have already been shortly presented to exemplify autonomous modalities in section two above. In a way they are the backbone of all the other modalities, so they should be absolutely autonomous. One could say, however, that just like the deontic modalities they depend on rules, at least the four rules mentioned in section two that define weak, strong, antistrong and antiweak operators. But this is not a problem, because unlike the deontic rules, these rules hold by definition and are thus impossible to violate. There are less trivial systems of modal logic than the bare minimum presented above (for an overview compare Garson 2003), but for the big picture to be outlined here nothing less trivial is required.

The following section takes us back to human categories, this time not human action, but its prerequisites, beliefs and preferences.

3.3. Attitudinal modalities
The capability to act is a defining criterion for the human species (or more precisely its members) and action requires propositional attitudes of two kinds. There are different labels for these kinds, the ones I will use contrast presentation attitudes with representation attitudes. Under the former, the daughter proposition functions as a blueprint, describing what could be (others speak of conative or ‘pro’ attitudes), with the latter the daughter proposition functions as a picture, describing what really or presumably is the case (others speak of cognitive attitudes). Presentation attitudes are always dependent, most of the representation attitudes too, but there is a prominent exception. This autonomous representation attitude will be discussed in section 3.3.2.2. below.

3.3.1. Presentation modalities
3.3.1.1. Optative modalities
Are optative attitudes modalities at all? According to the definitions adopted in this paper, they belong to a modal category if they constitute a taxonomically
high-ranking duality group of inter- or transpropositional modal operators. Let's look at an example:

(31)  
   a. He wishes he were at home  
   b. He doesn't wish he were not at home  
   c. He is at home

Optatives are certainly not autonomous modalities, they depend on an additional premise, but if (32) is added, the relevant entailments (a. to c., c. to b.) are valid.

(32) All of his wishes are granted

Since wishes are certainly taxonomically high-ranking, it follows that optative attitudes have to be welcomed among the modal categories.

3.3.1.2. Volitional modalities

Volitional attitudes are of course much more crucial for human action than optative attitudes and the fact that they constitute modal categories doesn't seem to require a proof. They have already been discussed above in section 3.1.2.1. in the context of anankastic conditionals, so it suffices to repeat the observation made there that although the strong modality is not autonomous, that is, wanting \( p \) does obviously not entail \( p \), there is a strong connection between wanting \( p \) and \( p \) in human cognition. A nice illustration of this fact is provided by the title of the paper quoted in footnote 20 above (Rakoczy et al., submitted), which includes the clause “that two people can have mutually incompatible desires”. This is of course not to be taken literally, what is meant is that the (different but compatible) desires have incompatible goals.

3.3.2. Representation modalities

3.3.2.1. Hypothetic modalities

Hypothetic modalities (as opposed to hypothetical ones) are attitudes like assuming and excluding, they are at the core of reasoning: If I assume that my keys can only be at one of two places and I exclude that they are at this one (I just had a look there), I will assume that they are at the other one. And there they are, if both the assumption and the exclusion were right.
3.3.2.2. Epistemic modalities

In the context of the encompassing attempt at a conceptual clarification that is outlined in this paper it is clear what the term “epistemic modalities” should mean: The propositional operators that instantiate the concepts which are based on the propositional attitude of knowing. Is this at variance with standard terminological wisdom which relates epistemic modality with indication of the degree of commitment by the speaker to what he says? Yes and no. No insofar as every speaker indicates a canonical degree of commitment to what he says as soon as he indicates that he wants his speech act to be part of an attempt to share knowledge, since knowing a proposition entails being committed to its truth and this in turn entails having convincing reasons for the assumption of its truth. However, the current approach is at variance with traditional terminology insofar as it does not restrict epistemic stance to the (outmost) speaker, but includes embedded epistemic agents.

In Zaefferer (2001a) I have proposed to distinguish three empirically related (and therefore often confused), but logically orthogonal dimensions of variation in the domain of so-called epistemic modalities: Certainty (subjective probability of information correctness), evidentiality (indication of kind of information source), and liability (social responsibility for information reliability). In the present more systematic picture it turns out that only certainty is a directly epistemic modality (knowledge implies sufficient certainty). Evidentiality is not at the same level since it is but a special case of the general circumstance modality of causation, namely the causation of epistemic attitudes and therefore a case of double modalization. Liability is still on a different level since (a) it is the only one of the three that requires language use (the other two are compatible also with silent attitudes), and (b) it is a special case of deontic modality, one that relates to acts of knowledge causation and therefore a case of triple modalization. The full consequences of this reconceptualization will have to be discussed on a different occasion, but they will already prove helpful in the following section.

There is one prominent subclass of the evidential modalities, the inferentials, whose members are sometimes treated as the prototypical epistemic modalities. Here is an example with a modal verb that has both a deontic and an epistemic interpretation (Huddleston / Pullum 2005: 55).
A GENERAL TYPOLOGY OF MODAL CATEGORIES

(33) You must be very tactful

According to Kratzer's (1991) picture, the choice between the different interpretations is determined by the kind of relevant background ('modal base'): Your tactfulness is either required in view of the current rules of conduct or assumed in view of what is known. The result is a deontic interpretation in the first case and an epistemic interpretation in the second. The big advantage of Kratzer's seminal idea is its enormous flexibility. It allows for instance a third (admittedly remote) interpretation where your tactfulness is forced upon you by your compulsive personality (a dynamic modality in Huddleston and Pullum's terms, an action disposition modality in ours). But this flexibility is also a disadvantage: It hides the fact that the conceptual difference between the epistemic and the other modalities is much deeper than the differences among the latter.

Consider first the deontic and the dispositional interpretations of (33). In both cases it is your behavior which is modalized as being required, either by the rules of good conduct or by the constitution of your personality: You play a role both in the behavior and in the rules. But the epistemic interpretation is different: Not you are required (by the rules of society or your personality) to act, but the epistemic subject, the speaker, is driven (by his premises and the rules of logic) to make an assumption about your behavior: He has to assume that you are very tactful.25

Here comes my third claim about modal categories, one that is crucial for the big picture outlined here: Epistemic interpretations arise as metonymic interpretations. Epistemically interpreted modal operators do not operate on the daughter proposition itself, but on an epistemic modalization of the latter. Let's call this the hidden operator analysis of epistemic modals:

(H 3) Epistemic interpretations of modal operators are the result of a covert epistemic modalization of the daughter proposition.

Note that the hidden operator analysis of epistemic modals is in line with the double modalization analysis of evidentiality: Evidential modalization means indication of circumstances, or more precisely, causation. A possible cause for an
another prominent cause of epistemic attitudes is of course communication. Communication in the relevant sense is intentional and therefore volitional, so epistemic interpretations should be possible under volitional modalities too, and in fact they are. In a number of Germanic languages, including Danish, Swedish, Norwegian, Dutch and German, volitional modals can be quotatively used (Brône / Feyaerts 2003). The following example shows the corresponding two modal verbs of German, autovolitional wollen and allovolitional sollen (cf. Zaefferer 2001b):

(34) a. Der Forscher will zuverlässige Resultate haben
   a.i. The researcher wants to have reliable results
   a.ii. The researcher claims to have reliable results

   b. Der Forscher soll zuverlässige Resultate haben
   b.i. The researcher is supposed to have reliable results
   b.ii. The researcher is said to have reliable results

In all cases the daughter proposition is that the researcher has reliable results. In the will sentences the subject of the volition is the researcher, in the soll sentences it is somebody else, but only in the i. interpretations the daughter proposition is the target of the volition, in the ii. it is something else. If we assume with (H 3) that this is an epistemic attitude towards the daughter proposition, the epistemic readings (which are also called quotative or reportative in the literature) are compositionally derived.

Libraries have been written about knowing, the central epistemic attitude, and the purpose of the rest of this section is not to add anything substantial to this ongoing discussion, but to point out just two things that will be needed in the following section.

First, in order to see that knowledge is a modal category it is important to distinguish, as we did with action circumstance modalities above (section 3.1.1.2.), between pseudo-duals and real duals: Both combine interior with
exterior negation, but in pseudo-duals the exterior negation is presuppositional, yielding a double negation effect, whereas in real duals the exterior negation is outmost, i.e., non-presuppositional negation. Only if the exterior negation is interpreted in that sense, i.e., without entailing that it is raining, does a sentence like (35) make sense at all:

(35)  I don't know that it is raining

In the relevant sense I don't know means not to my knowledge.26

Second, unlike belief, assumption, and most other propositional attitudes, knowledge can be described not only in a transparent way, i.e., in a way that makes the content of the attitude fully visible, but also in an opaque way, one that hides the content of the attitude. If it is raining and Jane knows that, I can truthfully say both (36) a. and b.:

(36)  a. Jane knows that it is raining
     b. Jane knows whether it is raining

In the first case, the transparent knowledge description, the content of Jane's knowledge becomes visible to my addressee, in the second case, the opaque description, it doesn't. I have proposed elsewhere (Zaefferer 2004) to call the content of the complement clause in (36) b. a near-proposition, because it is almost a proposition, but due to its lack of a truth value it cannot be a full one. However, being a strong modal operator (and, in its capacity of the only autonomous representation attitude, the exception indicated above), know requires that corresponding daughters are true in the world the knowledge is about and therefore they have to be full propositions, so in this context the near-proposition serves as a proxy for its (positive or negative) true completion. Still, being only a proxy, it also serves the purpose of hiding which full proposition it stands for, a purpose which is essential for one illocution type and which has been parasitically exploited by at least one other illocution type, as will be shown in the next section.
3.4. Illocutionary modalities
The fourth major claim of this paper is that structured illocutionary modalities\textsuperscript{27} can be defined as stacked modalities, i.e., a combination of the general autonomous modality of causation, and different attitudinal modalities.\textsuperscript{29} The common denominator is expression of an attitude in the sense of causation of its ascribability, the subtypes result from the different attitudes that are expressed or made ascribable. I propose to assume three major subtypes: volitionals, optatives, and hybrids.

3.4.1. Volitional illocutions
3.4.1.1. General volitional illocutions
General volitional illocutions present their propositional content as a blueprint for what is wanted (strong volition; directives) or tolerated (weak volition; permisives). Most languages possess a sentence mood called imperative or jussive which by default codes strong volition, but which can also be used to code its dual.\textsuperscript{29} Antistrong volition can of course be coded compositionally (strong volition with inner negation), but many languages use a different construction (often called prohibitive) for this purpose.\textsuperscript{30}

3.4.1.2. Epistemic volitional illocutions
The fifth and last central claim to be presented here is that both assertive and erotetic illocutions can be decomposed into three layers of modalization and differ basically one from the other only in the kind of propositional content they take (cf. Zaefferer 2001b, 2004). The difference between the general and the epistemic volitional illocutions is just a special case of the difference between the corresponding attitudes discussed above: the propositional content does not present the goal of the volition directly, but only the content of that goal, the goal itself being an epistemic entity, namely the existence of activated knowledge of this content. According to this analysis any unmarked use of one of the sentences (37) a. - c. has the effect that is possible to ascribe to its speaker a (strong or weak) volitional attitude, but only the a. sentence describes the goal of this attitude directly (that the addressee takes out the garbage). The goals of the other two volitionals are epistemic states, activated knowledge of either the coded proposition (with b.) or the true completion of the coded near-proposition (with c.).
(37)  a. Take out the garbage!
      b. You take out the garbage
      c. Do you take out the garbage?

3.4.2. Optative illocutions
Optative illocutions present their propositional content as a blueprint for what is wished (strong optative) or not wished away (weak optative). Their form often uses the imperative sentence mood which embeds an expression of weak (deontic or alethic) modality:

(38)  May the power be with you!

3.4.3. Hybrid illocutions
This sketch of a modal analysis of illocutions should not end without mentioning hybrid illocutions, speech acts that are a blend of one of the types mentioned so far with an additional element.

3.4.3.1. Expressive assertives
As their name indicates, expressive assertives combine assertive force with an expressive element. The subtypes are direct expressive assertions as illustrated by (39) and indirect assertions alias rhetorical questions such as the ones in (40):

(39)  I sure as heck am never bored!
(40)  a. Who loves to pay taxes?
      b. Am I a moron?

Whereas in (39) the expressiveness results from the semantics of the emphatic adverbial sure as heck, in (40) it is a result of hiding the intended proposition behind a near-proposition whose completion (toward the polar opposite) is assumed to be obvious.

3.4.3.2. Exclamations
Exclamations are a special case of expressive assertions where the expressed emotion results from the unusualness of the propositional content, or more precisely, since the proposition is always parametric, of the correct value for the parameter. This parameter can be expressed either overtly, by an interrogative
pronoun, as in (41) a., or by a demonstrative, as in b., or else covertly and only indirectly, via intonation and the initial interjection, as in c.  

(41)  
a. How clever she is!  
b. She is so clever!  
c. Boy, is she clever!  

This concludes our survey of the modal categories that constitute the major illocutions types and, as a result, the core sentence moods.

4. CONCLUSION

Based on the observation that currently available accounts of the domain of modal categories are either insufficiently structured (logico-philosophical approaches) or in want of systematicity (linguistic approaches) I have outlined an ontology of modal categories that reflects systematicity in being independent of the ways the categories are linguistically coded and that is structured by the assumption that human cognition is organized around the central categories of actions, attitudes and illocutions. Using a single entailment property for the definition of strong modality (embedding proposition entails embedded proposition) the existence of duality groups of modal operators is derived and taken to be a necessary condition for modal categories. A further requirement is generality (logical weakness). A distinction is made between autonomous (unconditionally connected) and dependent (conditionally connected) modalities, and a generalization to non-propositional modalizations (actions, attitudes, etc.) is provided. It turns out that general modalities, although logically central, are cognitively derived by generalization from action modalities. Illocution types, on the other hand, despite their affiliation to actions, are best defined on the basis of attitudinal modalities. The resulting general map of modal concepts assigns systematic coordinates to familiar notions such as deontic, epistemic and evidential operators as well as to less familiar ones such as antagonistic circumstances and covertly conditionalized modalities. It is hoped that the definitions and distinctions proposed here provide at least some orientation in the sometimes confusing jungle of modal concepts.
REFERENCES


Apostel, Leo (1972), "Negation: the tension between ontological positivity (negationless positivity) and anthropological negativity (positively described)", *Logique et Analyse*, vol. 15, no. 57-58, pp. 209-317.


Rakoczy, Hannes, Felix Warneken, Michael Tomasello (submitted), "'This way!', 'No! That way!' – 3-year-olds know that two people can have mutually incompatible desires", Cognition.

Rescher, Nicholas (1968), Topics in Philosophical Logic, Dordrecht: Reidel.


NOTES

1 I wish to thank my colleagues Daniel Hole and David Restle, who helped me reduce the number of shortcomings of this article. Neither is to be held responsible for any of my mistakes. My special thanks go to Volkmar Engerer, whose invitation to Aarhus inspired me to rethink the issue of modal categories.

2 Bybee et al. for instance express a rather pessimistic attitude when they write: "In fact, it may be impossible to come up with a succinct characterization of the notional domain of modality" (1994: 176).


4 Van der Auwera and Plungian (1998: 80), e.g., "[...] propose to use the term 'modality' for those semantic domains that involve possibility and necessity as paradigmatic variants" and go on to answer the delineation question rather arbitrarily and without further discussion: "We take this to be the case in justs four domains." These domains are participant-related and epistemic modality, with the former being subdivided into participant-internal and participant-external modality, and finally deontic modality as a subcase of the latter.

5 Kiefer’s definition (“Modality is the set of linguistically relevant accessibility relations”, Kiefer 1987: 86) deals with the restriction problem in a way that becomes more transparent if we replace his generic singular with the specific plural: "Modalities are the linguistically relevant accessibility relations". He calls this "the modified logical definition of modality". The modification consists in giving up the purely nonlinguistic way of defining modalities and relying on linguistic relevance, i.e., (cross-)linguistic constraints instead. Although I am sympathetic with this view (in Zaefferer 2001a I advocated it myself), I will show below that there is a solution to the restriction problem that does not rely on empirical linguistic data which may be difficult to obtain.

6 This is the topic of the volume Ontolinguistics. How ontological status shapes the linguistic coding of concepts (Schalley and Zaefferer, forthcoming).

7 The most famous instance of a duality group is of course the Square of Opposition, which goes back to Aristoteles’ De Interpretatione (Horn 1989).

8 The positive answer to this question given by Apostel (1972) is not relevant here, since it rests on a different notion of modality, namely expression of speaker attitude, and on a confusion of propositional negation with denial.

9 Formally, the identity function on propositions is of course the limit case of a modal operator in our sense, but including it in the definition would have the undesirable terminological consequence of formally obliterating the intuitively clear distinction between unmodalized and modalized propositions.

10 The modals in the a.-sentences are instantiations of the four vertices of the Square of Opposition, which are traditionally named A and E (for the positive members) and I and O (for the negative members). The pragmatic nature of the explanation for the well-known coding asymmetry among the three (O is the operator that is most liable for lacking lexicalization) has been reemphasized recently by Horn (2003, section 5).

11 For the distinction between individual level and stage level predicates, coding time-stable as opposed to accidental properties, compare Kratzer (1995).

12 For the close connections between modalities, causation, and force dynamics cf. Talmy (2000), chapters 7 and 8.

13 According to (H 1) this reflects high frequency and an intermediate degree of generality.


16 For the most recent available contribution cf. Huitink (2005).

17 Von Stechow et al. claim that the if-clause in anankastic conditionals is not really ‘conditional’ (von Stechow et al. 2005: 11).

18 It is well known that etymologically many of the expressions that code them have developed from deontic expressions.

19 Given standard axioms of propositional logic these four rules are equivalent. The axiom the definition of strong modality in (D 5) is based on is traditionally called M.

20 This holds already at a very early stage of child development and is based on our mutual acknowledgement as persons. "At the heart of the conceptual scheme with which we describe each other as persons – sometimes called ‘folk psychology’ – lies the ascription of two kinds of propositional attitudes to subjects: First, conative or ‘pro’ attitudes (Davidson 1963) – what someone wants, wishes, hopes to be the case, desires being the central case – and second, cognitive attitudes – what someone sees, believes or knows, belief being the central case.” (Rakoczy et al. submitted)

21 Searle (1983), borrowing from Anscombe (1957), speaks of opposite ‘directions of fit’ that either attitudes or illocutions possess: Beliefs have a mind-to-world direction of fit just as assertions have a word-to-world direction of fit.


23 “[…] the term ‘epistemic’ should apply […] to any modal system that indicates the degree of commitment by the speaker to what he says.” (Palmer 1986: 51)

24 I believe that part of the discussion about the subsumption of evidentiality under the epistemic modalities (for an opposing view cf. De Haan 2005) is due to the restriction of the concept of epistemic modality to the speaker.

25 Syntactically, this distinction is reflected in the fact that epistemically interpreted modals clearly behave as raising verbs, whereas there is some discussion whether non-epistemic modals are control or raising verbs.

26 Readers who are sceptical about this reading are encouraged to try Google on I don’t know that it. In May 2005 this yielded 89,600 hits.

27 The qualifier ‘structured’ is used to exclude holistic illocutions such as those coded by interjections, because the latter don’t have propositional content and therefore cannot, according to the definition advocated here, be conceptualized as modalities.

28 Pace opposite views like the following: “It should be kept in mind that speech acts do not constitute a separate type of modality since they cannot be used to relativize the validity of a state of affairs. The act of asserting, asking, permitting, ordering, or promising is alien to modality. There is no way to ask a meaningful question about how such an act affects modality.” (Kiefer 1997: 247)

29 This holds also for the military: Attention! and At ease! are both called ‘commands’ on the U.S. Army Infantry Homepage, but the definition of the latter (On the command AT EASE, the Soldier may move) shows its use as a permissive.

30 “Although true prohibitive markers appear to be a comparably infrequent phenomenon […] it is relatively common for languages to treat negative imperatives
differently from positive imperatives in one way or another." König and Siemund (to appear, page 28).

31 Zanuttini and Portner (2003) count only the first type as true exclamative clause and qualify the other ones as marginal. This may be due in part to a reluctance to consider intonation contours as defining characteristics of syntactic categories.