Is there a genuine modal perspective on feature structures?

Marcus Kracht

II. Mathematisches Institut
Freie Universität Berlin
Arnimallee 3
D-14195 Berlin
kracht@math.fu-berlin.de

July 24, 1996

1 Introduction

This essay is an analysis of the use of logic in the context of syntax. The specific question I want to ask is whether there is such a thing as *logic* in syntax and whether among the many proposals made in the literature there are discernible and significant differences. Moreover, I will seize the opportunity to point at some misunderstandings about using logic as a tool in syntax in general. As the title suggests I am specifically concerned with attribute-value structures. So, nothing will be said in the sequel about logic in connection with categorial grammar. There are two reasons for this. First, I lack the qualification to judge the specific developments in this area. Second, as regards the general concerns about the use of logic, they apply mutatis mutandis to that area as well (as far as I am aware of the current developments).

I became interested in attribute value logic through [Gazdar *et al.*, 1988]. [Kracht, 1989] was intended more or less as a technical note proving a theorem about the logic of category structures alias attribute value matrices (AVMs henceforth). This was taken up in [Blackburn, 1993] and developed. Others have also found uses of modal logic in that area, and the subject itself has gained some territory in the reserach on feature structures. However, as much as I am sympathetic to this development, it has remained unclear to me what significant new insights the use of modal logic would bring in this context. I have on several occasions insisted that what has been identified as the logic of feature structures may only be a weak sublogic of the the logic of feature structures once the features were given a specific interpretation. For example, if [NUM : dual] is taken to mean *the number of this item is dual*, then there is every reason to believe that [NUM : NUM : dual] is an illegimate expression (which might be equated here with being false). Yet the current research still treats this as a non-issue. Moreover, the particular properties of the concrete operators in a logic can suggest a different logical analysis than the one originally chosen. In particular, what seems to be a perfect case of an analysis in terms of attribute and value might in face of its concrete behaviour just an ordinary attribute, i. e. it might quite naturally be analyzed as a simple propositional constant. Both will be argued for in great detail with specific, non-trivial examples. This will also be a case study in formalization within linguistic theory. It has been suggested to me by Noam Chomsky that formalizing is an easy exercise, that anybody can do it, and that there is no progress made in science by doing it. Looking at the literature in linguistics one might get the impression that this is indeed so. However, I strongly disagree. On the contrary, formalization is *an art*. If not, it would not matter much if we used predicate logic or second-order logic or modal logic. Anything goes. However, comparing the various formalizations (for example, comparing the one advanced here with [Stabler, 1992]) one will very soon learn the difference. There is a case to be made for formalizations that are intellectually better digestable for humans. Any mathematics text will be rendered unreadable if written in strict first-order logic. So even *notation* matters a great deal. A very illuminating case is Leibniz' notation for the derivative. It was designed to suggest exactly what is correct, by exploiting an analogy with arithmetic. For a linguist one might, for example, compare encodings in categorial grammar with standard ones to appreciate the insistance on intellectual digestibility. From that point of view I will hold that modal logic in connection with AVMs is the right choice.

In addition to pointing at what seems to be a priori reasons to use modal logic, it still has to be seen whether things are as promised. This requires developing this approach to some extent to see whether it bears the fruit we are longing to eat, whether the formalizations yield decidable logics, or logics with other desirable properties. Feature structures, however, are barren land as it turns out. The margin between decidability and undecidability is so small that much of the effort is consumed in trying to design languages which are expressively strong but yet weak enough not to yield undecidability results. We will show here that it does not make much sense to look for such languages. The crux is the following. No matter what language we choose, we will use it to describe certain facts of natural languages. These facts have a certain inherent complexity that cannot be avoided by changing the metalanguage. We can only change the power of the metalanguage. Namely, after fixing the metalanguage we must axiomatically describe the *possible structures*. However, the *possible structures* are just the thing we are after when we study linguistics – at least if we adopt Chomsky's view here. The questions are of empirical nature and cannot be decided by some logical tricks. Chomsky's often repreated criticism of formalizations stems from the fact that before we actually know what the structures are there is little gain in strict formalizations. The life of a framework may be longer than that of a theory, but it too has a limited lifespan (approximately ten years).

This paper is formal and quite difficult for readers untrained in modal logic; I have no illusions about this and I apologize in advance if I fail to make things as clear and simple as I should. I do believe, however, that much of the complexity in this paper is unavoidable and anything that is simpler will be so at the cost of precision. Almost everything will be defined here, so that the discussion will on the whole be self-contained. But this is really not to say much when it come to mathematical topics. The reader who is seriously interested should perhaps read an introductory book on modal logic and the lucid survey article [Bull and Segerberg, 1984] to get enough background. I can also recommend [Blackburn, 1993] as an introduction into modal logic in connection with AVMs. It is impossible to go through all technical proofs in great detail; this would be tantamount to writing a book on this topic. But, I hope, the line of argumentation can be understood even without a proper understanding of the technical points. For the message is of wider importance. If I am right, then modal logic, where it fails, fails necessarily

– and no other framework I know of will not under these circumstances. Secondly, it provides enough technical apparatus to allow to prove significant results. To those who remain unimpressed I can only appeal to their sense of beauty and naturalness.

Among the persons who have quite generally helped to shape my views on syntax and logic I wish to thank explicitly those who have contributed to the present paper. These are Mark Ellison and two anonymous referees, who had the questionable pleasure of reading an earlier version of this paper. Moreover, the results on modal feature logics have been obtained in collaboration with Carsten Grefe. The errors in this paper have been obtained by myself alone.

Part I: Logical Theory

2 General Considerations

Logic may be described as the *art of reasoning*. Taking this as a definition, anything that is vaguely related to forms of reasoning may be called *logic*. This implies that there may be different uses of logic because there may be different ways of reasoning, as is now a widespread view among logicians. To wit, there is a distinction here between *logic* as a generic term and *a logic*, a technical entity. We will see that there is every reason to believe that there is a multitude of different logics each of which arises from a particular interpretation of the symbols as well as empirical facts connected with this interpretation. It is the main purpose of this section to motivate the appearance of rich classes of logics in the context of science rather than just one or a handful, as is most commonly the case in the literature. It is not intended as an introduction into logic as such, and I assume my readers to have a rudimentary understanding of logic. Nevertheless, given the complexity of the technical arguments that will follow it will be useful to motivate the necessity of going through such complicated arguments in detail. Specifically, it is helpful to understand the distinction between *logic* in general and *a particular logic*. Consider as a very simple example a propositional language with just one connective, the symbol $\wedge$, which we will read as *wegde*. Formally, $\wedge$ is just a binary connective and no interpretation has been given as yet. In transliterating this symbol as *and*, however, we are not only indicating a way of *reading* this symbol (other than calling it *wedge*) – we are also indicating what this symbol actually means. By saying that it does mean *and* we are drawing on an intuitive understanding of what *and* means (in whatever terms *that* might be expressed), and specifically, in what ways the symbol $\wedge$ may be used in reasoning, drawing here on an understanding of how *and* is used in reasoning. This example may be utterly simplistic, but even at this level we encounter typical problems. They are twofold. (a) It is not clear that we are in perfect agremeent as to what *and* means; and (b) It is not clear that we are in perfect agreement as to how to reason with *and*. Not only is *and* often claimed to be noncommutative (thus giving rise to a disagreement according to (a)), but also we are not in perfect position to say how we humans *actually* reason with *and*. Moreover, it is not always clear that we know how we should reason with *and* once we have agreed on its meaning. (This might be less obvious with *and*, but painfully clear with modal operators such as *believe*, *know* etc.)

This might be deemed more a problem with the informal meaning of *and* than with (formal) logic. But consider the theory of natural numbers as an example, this time formalized in first-order logic. By Gödels result we know that this theory is not recursively axiomatizable, there is no effective procedure that will yield an answer to any question we have about the numbers. So the theory of the natural numbers is *undecidable*. This holds even for first-order predicate logic itself. However, Gödel proved that in addition no recursively axiomatizable theory of numbers will ever be *complete*, that is, there can be no recursively axiomatizable theory such that whatever is true about numbers will also be provable in that theory. What this means, then, is that if we want to prove certain facts about natural numbers there is no way around the fact that the model of the numbers, $\mathbb{N}$ with, say, $=$, 0, the successor function, addition and multiplication etc., can be produced only by an informal sketch, relying thereby on an intuitive understanding of what these symbols mean. Ordinary mathematicians have no difficulties with this state of affairs. They simply use their intuitive model to get the theorems, relying here on the fact that this model is the one anyone else would choose. Thus, rather than (only) using the formal theory of numbers, they will resort to extraneous means to get the facts they want. In the end, however, they will produce a strict proof using induction – if they can. Some facts, such as Kruskal's Theorem, cannot be proved this way, so a proof must be found in a stronger theory, this time in set theory. Again, mathematicians find this unproblematic. All that seems necessary to them is to show that the natural numbers can be embedded (or interpreted) in set theory. Yet, it is not clear that we are able to show that. We may be able to produce a proof that the formal theory will be preserved under this embedding; but what happens to the theory of the *intended model* (the one in our heads so to speak) we do not know. Thus, there is a gap that spawns between the meaning of the symbols as fixed by the formal theory of numbers and the intended meaning. Yet this does not mean that there is nothing one can fruitfully show about numbers.

Now take as a final example the linguist in his search for *universal grammar*. Here, as in the case of natural numbers, the intended meaning of the symbols in the formalism is (reasonably) clear, even though one might in both cases have problems with the existence of these entities. To take the recurrent theme of this present essay, take the language of attributes and values. We know what an expression such as $[\textsc{case} : \mathsf{acc}]$ means. It means that the object in question has accusative case. It is less clear what the consequences are of this. For example, does it generally follow that [CASE : acc] and [CASE : nom] are exclusive, or, as it may well be the case, can there be several cases realized at a lexeme? To someone with an Indo-European mother tongue it seems almost logically true that there is one and only one case realized at a single lexeme. Yet, we cannot be so sure. Cases are realized as suffixes in Indo-European languages, and there is no reason why we cannot iterate them. A construct such as (Latin) *poetarumae* (poets-GEN.PL-NOM.PL) might be hopelelessly ungrammatical but it is not unconceivable – we have just succeeded in producing it. [1] Thus if we want to settle this question we must investigate the case marking systems of languages and see whether they allow for multiple case marking. A language where this option is realized is the Australian language Kayardild, as reported in [Blake, 1994]. We will discuss the evidence in a later section. To take a slightly less exotic language, in Georgian a possessor phrase succeeding a noun phrase will in addition to the genitive case also inherit the case of its regens. (Again this is taken from [Blake, 1994]. It is reported there that double case marking is not uncommon in Australian languages.) This being so, it is quite unclear what logical properties the formal construct [CASE : acc] has. In general we are inclined to ban nesting the attribute CASE so that expressions of the form

$$\left[\textsc{case} : \left[\begin{array}{r} \mathsf{acc} \\ \textsc{case} : \mathsf{gen} \end{array}\right]\right]$$

will be illegitimate. Apparently, for languages like Kayardild and Georgian we will need an exception to that ban. It might be worth a dispute whether we have captured the meaning of *has accusative case* properly with our formalism – but the problem itself remains. This is so because we are not in full control of the consequences of our stipulations. In contrast to mathematics, where this problem arises out of the complications of the matter itself rather than the a priori impossibility of knowing the consequences, here the questions really are of an empirical nature. Namely, with *case* defined in some or the other way, we do not know whether it has certain properties (say, that there are no nestings, that there are no more than 20 distinct cases in any language etc.). This has to be found out, by careful examination of the languages. This is one of the reasons that Chomsky rejected the use of formal theories in this enterprise. However, notice that even if not all properties of case can be fixed outright, some may be. We may, for example, comfortably settle the question whether *having nominative case* and *having accusative case*. Seen as a property of morphemes we can in any case, even in Kayardild and Georgian, postulate the exclusiveness of distinct cases. Even if there is little we know for sure, a formalization is not impossible. Notice also that even if there are no surprising theorems about case that can be

[1] In transliterations we follow [Blake, 1994] by using a hyphen when there is a corresponding segmentation in the original word and a dot when there is not. The latter case is also called cumulative exponence. In the case at hand, the suffix *-arum* indicates both genitive and plural, there is no separate plural or genitive morpheme; therefore a dot is placed.

brought to light by this formalization, the formalization may be successful in that we can write down facts about case rather succinctly. And that we may succeed in removing certain ambiguities in speaking about case, such as whether we want to regard case marking in Kayardild as *multiple case marking* rather than *nested case marking*. It seems prima facie reasonable to prefer the latter.

The position of a linguist differs somewhat from that of the number theorist. The latter has no trouble identifying correct arguments from incorrect ones – the logic itself remains fixed, so does the interpretation of the symbols. With language, however, we cannot be sure what the denotation of the symbols really is, or, to put it differently, we do not know which of the theoretical possibilities are attested in some languages and which ones are not. If all we need is to write down facts about case, we might not be bothered by that at all. For example, to say that it is a *feature* with atomic values (default) or CASE-feature value (Georgian and some others), is to allow for some minimal form of reasoning about case. We will identify the logic of this feature in the sequel as $\mathbf{K.Alt}_1$. However, it does seem likely that case has nontrivial properties beyond that. They will not be captured by this rudimentary logic, hence not be used in inferences. What can be done in this circumstance? The natural conclusion suggesting itself is that in actual fact we have *not identified the logic of case* in UG. What we have is a lower bound for it, that is, we know what axioms and rules there are *at least*. In the literature as it presents itself, this conclusion has never been drawn explicitly. Hence there has never been an inclination to study large classes of logics in order not to prejudge the case. Instead, there seems to be a hidden consensus that the base logic $\mathbf{K.Alt}_1$ is *the logic* and that the rest can be captured by adding suitable *constraints* (or whatever one may call them) later on as a kind of peripheral system around the logic of attributes and values deemed to be universal. This view has its justifications. First, there is some terminological difficulty in speaking of a logic of *case*. A fact such as the non-iterability of the case-feature does not ring of an axiom but of an accidental fact of language. Furthermore, it is usually very difficult to redo decidability proofs for logics in a general setting. Generally, the basic logics that one can settle on initially, such as the aforementioned $\mathbf{K.Alt}_1$, do turn out to be decidable, but whether or not an arbitary logic containing this logic is decidable is very difficult to prove (and mostly false). The first objection is not really serious. We settle here on two different notions of necessary properties of CASE. One that springs from the definition of CASE as a *feature* in the technical sense, whereby deriving certain rudimentary properties, and the other from interpreting [CASE : $\mathbf{x}$] by the pre-theoretic term *has case x*. The second argument is also not so serious. It might be from an architectural point of view desirable to have distinctions between core and periphery. This distinction will have its uses in diachronic linguistics or in articulate theories of language such as GB. But in the present context they camouflage parts ot the logic. Peripheral facts are facts and their consequences are facts, too. So, peripheral facts are *additional axioms.*

We will see in the course of this essay that we cannot hope for general and powerful decidability results even in the most economic setting. We just have to live with the fact that the consequences we can derive from our theories constitute a fraction of the truth. An illuminating comparison might be the notion of *space* in physics. This notion has undergone a number of revisions, from three-dimensional euclidean to four-dimensional Minkowski space-time up to higher dimensional differentiable manifolds. There is no question about the fruitfulness of this approach in physics even though the notion of *space* (such a simple notion even) is not once and for all fixed. And even if it were, the first-order theory of space is undecidable – as is the theory of numbers.

This leaves us with the following scenario. We will identify basic logical and linguistical constants auch as $\wedge$ and CASE and try to nail down their logic as much as possible. With the logical constants this will be unproblematic. However, with the linguistical constants we simply have to leave the true logic or, if you wish, their true behaviour underspecified. The reason is simply that the logic of these constants is an empirical question and cannot be solved by pure thought, unlike a mathematical problem. This is then different from the classification of finite simple groups, which is a mere consequence of the axioms of group theory and hence was provable by mere symbol crunching. Here we must confront ourselves with real data, linguistical data, such as case marking in Kayardild. In some sense, if we have settled on the constants, universal grammar might be identified with a small set of logical systems, those corresponding to particular values of the parameters, or indeed with a single logic, which itself is so strong that it allows only finitely many logical extensions, all of which are particular (core) grammars. The problem is to identify UG. But this is really, according to some researchers at least, all there is to linguistics; therefore, one should not regard the identification of UG with a logic as particularly useful or revealing. The problem is, as we will show, that we are in the uncomfortable position that there is very little that one can say about UG, if it exists at all. There are simply no significant properties of logical theories that we can prove to hold for a sufficiently rich class of logics so as to ensure that when we hit UG it is bound to have them. This point cannot be stressed often enough, because it emphasizes the unfruitfulness of the debates between proponents of GB and their adversaries. To some of the latter kind the result by Peters and Ritchie that any recursively enumerable language is the language generated by some transformational grammar is a devastating blow to transformational grammar since it shows that the power of the mechanism is by far too strong. Even if this were an argument (which I myself doubt), we will see that undecidability strikes at such a low level that it is hard to imagine one can have a formal system that is both reasonably rich and yet inherently decidable, that is, all extensions are decidable. So, if you buy my story that it is not logical systems but *classes* of extensions of these system that must be studied, then it is not enough to point at a single logic as the logic of linguistic structures – unless it is (arguably) identical to UG –

and then prove that this logic is harmless. Mostly this logic will be just the least common denominator of all conceivable interpretations. Otherwise, we might defend GB in turn by claiming that that there is a basic transformational core grammar common to all languages, which is well-behaved. To the contrary, we must be prepared to look at all possible logics that strengthen our base logic, or at least at a suitable subclass of them.

3 Modal Logic – Notation and Terminology

This section introduces basic notions from modal logic. I will make special reference to the connection with AVMs, but knowledge of the latter is presupposed. Instead I refer to [Carpenter, 1992] for AVMs and to [Blackburn, 1993] and [Blackburn and Spaan, 1993b] for the connection with modal logic. Typically, feature structures are presented as directed acyclic graphs where the arcs as well as the nodes have labels. Moreover, from each node there exist per arc label at most one arc with that label pointing to some other node. Such graphs can be seen as special kripke frames where each arc label represents a different accessibility relation. We refer to such kripke frames as *polyframes* to stress the fact that they can have any fixed number of relations, not just one. A polyframe is an object $\mathfrak{f} = \langle f, \lhd_1, \ldots, \lhd_m \rangle$ with f a set and $\lhd_j \subseteq f^2$ a binary relation for each $j \leq m$. For concreteness' sake assume that our AVMs are based on a finite set $Arc = \{\ell_1, \ldots, \ell_m\}$ of arc labels and a finite set $Node = \{\mathsf{n}_1, \ldots, \mathsf{n}_n\}$ of node labels. The node labels correspond to *types* in [Carpenter, 1992], but here the logic of types is boolean logic. We then define the following language to talk about the polyframes instantiating the AVMs. We have countably infinitely many propositional variables $p_1, p_2, \ldots$, constants $c_1, \ldots, c_n$, all boolean connectives and the modal operators $\Diamond_1, \ldots, \Diamond_m$. The constant c_i is interpreted by the node label n_i and the modality $\Diamond_j$ is the existential modality based on the relation

$$arc_j = \{\langle x, y\rangle | \text{there exists an arc with label } \ell_j \text{ from } x \text{ to } y\}$$

As usual, $\Box_j \phi := \neg\Diamond_j\neg\phi$. In order not to be overly pedantic we will write n_i instead of c_i and we will confuse the arc label ℓ_j with the set arc_j of pairs of nodes connected by an arc of label ℓ_j so that we can write $\langle \ell_j \rangle$ for $\Diamond_j$ and $[\ell_j]$ for $\Box_j$ as is common practice in propositional dynamic logic. From there we will borrow the convention to write $\langle w \rangle$ for a word over the alphabet $\{\ell_1, \ldots, \ell_m\}$ and we understand by that the composition of the elementary modalities $\langle \ell_j \rangle$ in the order and multitude specified by the word w. (And similarly for $[w]$.) So, if we have the arc labels CAT, AGR, CASE etc. then we may write $\langle$CAT; AGR; CASE$\rangle$ instead of $\langle$CAT$\rangle\langle$AGR$\rangle\langle$CASE$\rangle$. We call $\langle \ell_j \rangle$ an *arc modality*. Finally, we use the words *kripke frame* and *polyframe* interchangeably, but they are strictly different from AVMs. This will no doubt cause confusion, but necessarily so. If it helps, think of a kripke structure as a maximally specified AVM.

Let us for the moment concentrate on the case of a single modality, which we denote by $\Box$, dropping the index. The basic logic for kripke structures is called **K** and it contains all propositional tautologies, the so called *box-distribution axiom* $\Box(p \to q). \to .\Box p \to \Box q$, and is closed under substitution and modus ponens; the set of theorems is also closed under the rule $MN : P/\Box P$. A *quasi normal modal logic* is a set Λ of formulas containing **K** closed under modus ponens and substitution, but not necessarily closed under MN. If a quasi-normal logic is closed under MN, it is called *normal*. It should be stressed here that in general a logic is defined by its language and its set of *deduction rules*. Here, we have defined a logic only via its set of theorems. This is justified since it is tacitly assumed that the only rule of inference is modus ponens. This means that both substitution and MN may be applied *only to theorems*. The distinction between quasi-normal and normal is very crucial in the logic for AVMs. Typically, an AVMs satisfies a constraint only if the constraint is true if evaluated at the root the AVM, rather than at any other point. [2] If Λ is a quasi normal logic and ϕ a formula then $\Lambda + \phi$ is the smallest quasi normal logic containing Λ and ϕ. If Λ is normal then we write $\Lambda \oplus \phi$ for the smallest *normal* logic containing Λ and ϕ. This notation is borrowed from [Zakharyaschev, 1992]. To a normal logic an axiom can thus be added both quasi normally (with $+$) and normally (with $\oplus$). Obviously, a normal logic is also quasi-normal and it is standard knowledge that

$$\Lambda \oplus \phi = \Lambda + \{\Box^k \phi | k \in \omega\}$$

(As usual $\Box^k$ is the k-fold iteration of $\Box$; we have $\Box^0\phi = \phi, \Box^{k+1}\phi = \Box(\Box^k\phi)$.) We write $\mathcal{Q}\Lambda$ for the lattice of quasi normal extensions of Λ and $\mathcal{N}\Lambda$ for the lattice of normal extensions. The notation and terminology is mutatis mutandis the same in the case where we have more than one operator. We will not spell that out here. The simplest case of polymodal logics are the so-called *independent fusions* studied in [Kracht and Wolter, 1991]. For any polymodal logic Λ the mono-modal fragments Λ_j – the intersection with the language over single operator fragments – is a mono-modal logic and (quasi) normal if Λ is. Suppose that Λ is the smallest (quasi) normal logic containing all Λ_i. Then Λ is called *independently axiomatizable* and we say that it is the *independent fusion* of the Λ_i. We write $\Lambda = \bigotimes_{j \leq m} \Lambda_j$. There is no notational distinction between the quasi-normal and the normal fusion. But the fusion of normal logics is normal, and this is the only case which will arise.

Structures for normal polymodal logics are the so-called (generalized) polyframes. A *generalized polyframe* is a pair $\mathfrak{F} = \langle \mathfrak{f}, \mathbb{F} \rangle$ where $\mathfrak{f}$ is a polyframe and $\mathbb{F}$ a set of subsets of f closed under boolean operations and

$$\blacklozenge_j a = \{s \in f | (\exists t)(s \lhd_j t \mathbin{\&} t \in a)\}$$

In that case the sets of $\mathbb{F}$ form a boolean algebra with operators

$$\mathfrak{F}_+ = \langle \mathbb{F}, -, \cup, \blacklozenge_1, \ldots, \blacklozenge_m \rangle$$

[2] If this sounds cryptic, read on.

$a \subseteq f$ is called *internal* in $\mathfrak{F}$ if it is in $\mathbb{F}$ otherwise it is called *external.* $\mathfrak{F}$ is called *differentiated* if it satisfies

$$(\forall x)(\forall y)(x = y. \leftrightarrow .(\forall a \in \mathbb{F})(x \in a. \leftrightarrow .y \in a))$$

In terms of AVMs we can say that an AVM is differentiated if points are reentrant iff identical. [Johnson, 1988] calls these structures *discernible.* To use the terminology of [Carpenter, 1992], an AVM is differentiated if nodes are extensionally identical iff they are intensionally identical. Namely, reentrancy is by definition the sharing of the truth of all propositions alias structure. This merits some thought. Recall that AVMs are actually thought of as (partial) descriptions of objects, which we call polyframes. Polyframes on the other hand can be seen as fully instantiated AVMs. It is actually rather odd to interpret ordinary AVMs themselves as kripke structures, because the latter are objects, and AVMs are *sets* of objects. The fact is that AVMs are externally dynamic because they can be fused with other AVMs (a process described as unification); if two nodes of the AVM agree on all formulae prior to unification with an AVM this might not be true after unification, because some incoming information may destroy this harmony. Actually, that an individual AVMs should not be seen as a kripke structure but as sets of such structures supports the thesis which we want to advance here, namely that AVMs are axiomatic descriptions of kripke structures.

There are two other notions that have counterparts in AVMs which are essential to both. The first is that of a *generated subframe.* Consider a frame $\langle \mathfrak{f}, \mathbb{F} \rangle$, $\mathfrak{f}$ being a polyframe. Pick a set $S \in f$. Then let $\mathrm{Tr}(S)$ be the set of all points that can be reached from S in a finite number of steps following any of the elementary relations. Equivalently, define for $S \subseteq f$ the set $\mathrm{Tr}_1(S) := \{t | (\exists s \in S)(\bigvee_{j \leq m} s \lhd_j t)\}$. Further, let $\mathrm{Tr}_{n+1}(S) := \mathrm{Tr}_1(\mathrm{Tr}_n(S))$. Then

$$\mathrm{Tr}(S) = \bigcup_{n \in \omega} \mathrm{Tr}_n(S)$$

If $S = \mathrm{Tr}(S)$ we call S a *generated subset.* The structure

$$\mathfrak{s} = \langle S, \langle \lhd_j \cap S^2 | j \leq m \rangle \rangle$$

is a kripke frame. If we let $\mathbb{S} = \{a \cap S | a \in \mathbb{F}\}$, the set of subsets induced by $\mathbb{F}$ on the subset $\mathbb{S}$, then the pair $\mathfrak{S} = \langle \mathfrak{s}, \mathbb{S} \rangle$ is called a *generated subframe.* It is one-generated if $S = \mathrm{Tr}(\{x\})$ for some $x \in S$. A p-morphism between kripke frames $\mathfrak{f} = \langle f, \langle \lhd_j | j \leq m \rangle \rangle$ and $\mathfrak{g} = \langle g, \langle \blacktriangleleft_j | j \leq m \rangle \rangle$ is a function $p : f \to g$ such that (i) if $x \lhd_j y$ for some j and $x, y \in f$ then $p(x) \blacktriangleleft_j p(y)$, and (ii) if $p(x) \blacktriangleleft_j u$ for some j and $x \in f$, $u \in g$ then there exists a $y \in f$ such that $p(y) = u$ and $x \lhd_j y$.

A *model* is a triple $\langle \mathfrak{f}, \beta, x \rangle$ where β is a *valuation* and x a world. The symbol $\langle \mathfrak{F}, x, \beta \rangle \models \phi$ is standardly defined. We have $\langle \mathfrak{F}, x, \beta \rangle \models p$ if $x \in \beta(p)$. Furthermore,

$$\begin{array}{lll}
\langle \mathfrak{F}, x, \beta \rangle \models \neg\phi & \text{iff} & \langle \mathfrak{F}, x, \beta \rangle \not\models \phi \\
\langle \mathfrak{F}, x, \beta \rangle \models \phi \wedge \psi & \text{iff} & \langle \mathfrak{F}, x, \beta \rangle \models \phi \text{ and } \langle \mathfrak{F}, x, \beta \rangle \models \psi \\
\langle \mathfrak{F}, x, \beta \rangle \models \langle \ell_i \rangle \phi & \text{iff} & \text{there is } y \text{ such that } x \lhd_i y \text{ and } \langle \mathfrak{F}, y, \beta \rangle \models \phi
\end{array}$$

We write $\mathfrak{F} \models \phi$ if for all valuations β and all worlds x $\langle \mathfrak{F}, x, \beta \rangle \models \phi$. As a rule of thumb, a propositional variable in modal logic corresponds to one-generated subframes. To be more precise, a variable p can be instantiated to *true* at a world iff it reflects part of the structure generated by that world. This is immediate if we think about the definition of generated substructures. Namely, if a formula ϕ is true at a world in $\mathfrak{F}$ it is also true at the same world in the subframe generated by that world.

Any normal logic is complete with respect to differentiated generalized polyframes. (Remember that any logic is complete with respect to its *canonical* frame; the canonical frame is differentiated.) Λ is called *complete*, however, if it is complete with respect to kripke polyframes. Λ is called *df-persistent* if for all differentiated frames $\mathfrak{F} = \langle \mathfrak{f}, \mathbb{F} \rangle$ with $\mathfrak{F} \models \Lambda$ we also have $\mathfrak{f} \models \Lambda$. By a result of [Fine, 1975b] any df-persistent (normal) logic is elementary. Finally, if $\mathfrak{f} = \langle f, \lhd_1, \ldots, \lhd_m \rangle$ is a kripke frame and $a \subseteq f$ we define the subframe based on a by $\mathfrak{f} \cap a := \langle a, \lhd_1 \cap a^2, \ldots, \lhd_m \cap a^2 \rangle$. Similarly, a subframe for a generalized frame $\mathfrak{F} = \langle \mathfrak{f}, \mathbb{F} \rangle$ is the pair $\langle \mathfrak{f} \cap a, \{b \cap a | b \in \mathbb{F}\} \rangle$ where $a \in \mathbb{F}$. Following [Fine, 1985] we call a logic a *subframe* logic if its class of generalized polyframes is closed with respect to taking subframes.

Appropriate structures for quasi-normal logics are *generalized pointed polyframes*. These are pairs $\langle \mathfrak{F}, x \rangle$ where $x \in f$, f the underlying set of worlds. A model based on such a pointed frame is a triple $\langle \mathfrak{F}, x, \beta \rangle$. Thus, normal and quasi-normal logics share the notion of a *model*. But the structures that underly these models are different. One can always assume that x is a root of $\mathfrak{f}$, that is, any point in $\mathfrak{f}$ is accessible via some path from x. If this holds, we called $\langle \mathfrak{F}, x \rangle$ *rooted*. Standardly, AVMs are pointed polyframes, with the point being the root. So a constraint is evaluated at the top, so to speak. Hence, a logic for AVMs is in the deafult case a quasi-normal logic. This is worth remembering.

Modal logic can be drastically extended by two new operators, the *test* ? and the *Kleene star* *. Adding these two we get *propositional dynamic logic*. The test turns a proposition ϕ into a relation $\phi?$ between nodes defined by

$$\phi? = \{\langle x, x \rangle | \langle \mathfrak{f}, x, \beta \rangle \models \phi\}$$

Consequently, $\langle \phi? \rangle$ and $[\phi?]$ are new modalities and we have

$$\langle \phi? \rangle \psi. \leftrightarrow .\phi \wedge \psi \qquad [\phi?]\psi. \leftrightarrow .\phi \rightarrow \psi$$

The test alone is therefore a dispensable piece of notation; but in combination with the Kleene star it becomes quite forceful. Given a word w over the alphabet of arc labels, we define

$$w^* = \{w^k | k \in \omega\}$$

where w^k denotes the k-fold iteration of w, $w^0 = \epsilon$, the empty word. So w^* is the reflexive and transitive closure of w. We have

$$\langle w^* \rangle \phi. \leftrightarrow . \bigvee_k \langle w^k \rangle \phi \qquad [w^*]\phi. \leftrightarrow . \bigwedge_k [w^k]\phi$$

The interpretation of w^* relative to w can be fixed by the following axioms.

$$p. \to .\langle w^* \rangle p, \langle w^* w^* \rangle p. \to .\langle w^* \rangle p, \langle w \rangle p. \to .\langle w^* \rangle p$$

$$p \wedge [w^*](p \to [w]p). \to .[w^*]p$$

The first two force w^* to be reflexive, transitive and at least as strong as w. The last axiom, known as the recursion axiom, forces w^* to be at most as strong as the reflexive, transitive closure of w. We issue a warning here that it is possible to satisfy these axioms with structures in which w^* is different from reflexive, transitive closure of w. So this construction has to be handled with care. (See [Goldblatt, 1987].)

There is one last thing that needs explanation which will be referred to as the *irreflexivity trick*. Consider the following axiom $\mathbf{G} = \Box(\Box p \to p). \to \Box p$, known as *Gödels axiom*. To make it a bit more perspicuous, replace p by $\neg p$ and switch the arrow. This yields $\neg\Box\neg p. \to .\neg\Box(\Box\neg p. \to .\neg p)$. This is the same as $\Diamond p. \to .\Diamond\neg(p \to \neg\Box\neg p)$ or $\Diamond p. \to \Diamond(p \wedge \Box\neg p)$. It can be shown that this formula can be falsified on frames which contain a reflexive point, or on frames whose relation is not transitive. These are only sufficient conditions, however, but for finite frames they are also necessary. Now consider a basic operator $\Box$ and its iterate $\Box^*$. $\Box^*$ is based on a transitive and reflexive relation. However, $\Box^+$ defined by $\Box^+\phi := \Box(\Box^*\phi)$ is only based on a necessarily transitive relation, which may be irreflexive. If we add to any logic for $\Box$ the axiom $\mathbf{Ir} : \Box^+(\Box^+ p \to p). \to .\Box^+ p$, then we have required any structure for that logic to be such that $\Box^+$ is based on an irreflexive relation. This can only be the case if $\Box$ itself is based on an irreflexive relation. On finite structures this correspondence is exact; a finite structure satisfies $\mathbf{Ir}$ iff the relation underlying $\Box$ is irreflexive. This is also worth remembering.

4 Internal Descriptions and Axiomatic Expressivity

A logic can be seen as a description of admissible structures. If it is this description we are after via the logic the choice of the language is rather immaterial. We can characterize kripke structures using first-order logic or even second-order logic because we can view polymodal logic as a part of monadic second order logic, where modal axioms express certain special universal conditions via the so-called *standard translation* (see [van Benthem, 1983], [Blackburn, 1993]). We can be interested in an independent characterization of what properties are expressible modally. In particular, methodologically interesting is the case when the role of the second-order quantifiers over internal sets can be eliminated in favour of first-order quantifiers over worlds. It is known that neither are all axioms characterizable as first-order descriptions of the frames, neither are all conditions expressible by first-order sentences also modally expressible. Yet, it will be seen shortly that in the cases under consideration the situation drastically simplifies.

In [Kracht, 1993a] I have argued to replace the usual second-order language by a notational variant, namely two-sorted predicate logic, the so-called *external language* $\mathcal{L}^e$. The notation of $\mathcal{L}^e$ has the advantage of being closer to the intuitive understanding of the formulae. Rather than replacing every variable p by a predicate variable $p(x)$ taking a world as an argument (as in [van Benthem, 1983]), we keep variables for worlds and for propositions as separate objects and introduce the symbol ϵ for *being element of.* So, instead of $p(x)$ we write $x \epsilon p$; the intended meaning is, of course, that (the interpretation of) x should be a member of (the interpetation of) p. This conforms to the standard intuition of propositions as sets of worlds. Propositions, being one sort, are combined as before with $\neg$, $\wedge$, $\Diamond_i$ etc. So we can write $x \epsilon p \wedge \Box_j \Diamond_i \top$ to state that x belongs to the set of worlds satisfying $p \wedge \Box_j \Diamond_i \top$, that is, x satisfies $p \wedge \Box_j \Diamond_i \top$ in the standard sense. There are obvious postulates such as

$$(\forall x)(\forall p)(\forall q)(x \epsilon p \wedge q. \leftrightarrow .(x \epsilon p \,\&\, x \epsilon q))$$

which connects the conjunction of the modal language (written as $\wedge$) with the conjunction of the predicate language (written as &). We will not dwell on a precise definition of this language, as it is really not necessary; all that is required is that one is able to grasp the meaning of an occasional $\mathcal{L}^e$-formula.

The first-order predicate language $\mathcal{L}^f$ for talking about kripke structures alias directed graphs is a sublanguage of $\mathcal{L}^e$. It has the usual logical symbols (boolean connectives, quantifiers, variables, equality) plus – in our case – the constants t, t, n_i (so that $x \epsilon \mathsf{n}_i$ tells us whether n_i is true at x) and binary relation symbols $x \lhd_j y$ telling us whether there is an arc with label ℓ_j from x to y. It is better to switch from $\mathcal{L}^f$ to the language $\mathcal{R}$. It differs from predicate logic in that it has the following *restricted* quantifiers where predicate logic has ordinary quantifiers.

$$(\forall y \rhd_j x)\phi := (\forall y)(x \lhd_j y. \rightarrow .\phi)$$

$$(\exists y \rhd_j x)\phi := (\exists y)(x \lhd_j y. \,\&\, .\phi)$$

Even though the restricted quantifiers are first-order definable, the first-order quantifiers are not definable from the restricted ones, so that $\mathcal{R}$ is strictly weaker than full first-order logic. Furthermore, for the quasi-normal case we have the language $\mathcal{R}_q$ which in addition has a constant r. In pointed frames $\langle \mathfrak{F}, x \rangle$ r is interpreted by x. Only in the quasi-normal case is it possible to create sentences, i. e. formulae without free variables. In the standard case at least one variable remains free, since every quantifier binding a variable needs a variable as a restrictor. The free variables are assumed to be universally quantified (by a standard unrestricted universal quantifier). We speak of an $\mathcal{R}$-sentence ϕ whenever ϕ has *exactly* one free variable.

Given a proposition ϕ and a pointed frame $\langle \mathfrak{F}, x\rangle$ we can ask the following two questions.

- What is required for $\langle \mathfrak{F}, x\rangle$ in order to base a *model* for ϕ on it?
- What is required for $\langle \mathfrak{F}, x\rangle$ if no model *against* ϕ can be based on it?

The two questions are in nuce one and the same question, since a complete answer to one of them answers the other. But it turns out on an intuitive level that one should try to answer the first one. We say that in the first case we are interested in the property of models *described* by ϕ, and in the second case in the property *defined* by ϕ. We are of course not looking for *any* characterization – otherwise a straightforward second-order sentence will do. Namely, the effect of the modal axiom ϕ can be equally enforced by the second-order axiom $(\forall x)(\forall \overline{p})(x \in \phi(\overline{p}))$, where $\overline{p}$ collects the free propositional variables of ϕ. The intention is to develop simple criteria. A striking example of such a characterization are the *canonical formulae* of [Zakharyaschev, 1992] for extensions of **K4**. However, his formulae do not always correspond to first order properties. But here we are interested in possible reductions to first-order logic. This is the standard domain of *correspondence theory* of [van Benthem, 1983]. Traditionally, however, research has concentrated on the second question, thus at the axiomatic characterization. I have demonstrated elsewhere in [Kracht, 1991] and [Kracht, 1993a] that the theory can be much simplified if one concentrates on the first question instead. The specific advantage is that the conditions on the constitution of a model for a complex formula can in the known cases be reduced quite comfortably to that of simpler subformulae and so the conditions can in certain cases be computed algorithmically, using a calculus pairing strings of propositions with first-order conditions. Such algorithm can never cover all modal formulae. By a result of [Chagrova, 1991] there can be no general algorithm for the correspondence problem so we have to be content with a partial solution.

We will sketch here the approach of *internal descriptions* as far as it is beneficial for present purposes. Recall that a general frame is called differentiated if it satisfies

$$(\forall x)(\forall y)(x = y. \leftrightarrow .(\forall a \in \mathbb{F})(x \in a. \leftrightarrow .y \in a))$$

Alternatively, there is an $\mathcal{L}^e$-sentence characterizing this property, namely,

$$(\forall x)(\forall y)(x = y. \leftrightarrow .(\forall p)(x \epsilon p. \leftrightarrow .y \epsilon p))$$

This can be rephrased as follows. Given two worlds x and y, we know that they are different only if we can name a set $a \in \mathbb{F}$ such that $x \in a$ and $y \notin a$. Of course, if such a set exists the points must be different; the postulate of differentiation is non-trivial only in the sense that it guarantees the other direction as well. Notice that while we can always construct such sets, for example $\{x\}$, the problem is that they need not be internal, i. e. members of $\mathbb{F}$. In a final step we replace the condition on the internal set a by a condition on a valuation. Since any internal set is the value $\beta(p)$ of a proposition variable p under some β, we can say that $x \neq y$ exactly if we can find a valuation β such that

$$\langle \mathfrak{F}, x, \beta \rangle \models p \quad \text{and} \quad \langle \mathfrak{F}, y, \beta \rangle \models \neg p$$

The idea to characterize an elementary condition by the simultaneous satisfaction of some propositions at possibly different worlds is generalized to the following definition.

Definition 4.1 *A first-order condition $\alpha(x_1, \ldots, x_n)$ is* ***internally describable*** *in a class $\mathfrak{X}$ of generalized frames if there exist $P_1, \ldots, P_n$ such that for any $\mathfrak{F} \in \mathfrak{X}$ $\mathfrak{F} \models \alpha[w_1, \ldots, w_n]$ exactly if there exists a valuation β such that $\langle \mathfrak{F}, w_i, \beta \rangle \models P_i$ for all $i \leq n$. $\alpha(x_1, \ldots, x_n)$ is* ***internally definable*** *in $\mathfrak{X}$ if $\neg\alpha$ is internally describable in $\mathfrak{X}$.*

To see the simplicity in the concept of internal descriptions just try to define internal definability analogously! Internally describable conditions are closed under conjunction, prefixing with restricted $\exists$ and identification of variables. Moreover, internally describable conditions of the type $\alpha(x_1)$ are closed under disjunction as well. Constant propositions are internally describable in all generalized frames. Consequently, $\alpha(x_1)$ is internally describable in the class $\mathfrak{Df}$ of differentiated frames if it is composed from inequations and constant formulae with the help of $\wedge, \vee$ and restricted $\exists$.

The special flavour of correspondence theory is produced by the existence of two gaps. The first gap spawns between purely second-order and first-order conditions, the second between first-order and locally first-order definable conditions. How large these gaps are is only partially known. Sahlqvist's Theorem names a class of propositions first-order in descriptive frames and kripke frames; and in [Kracht, 1993a] the first-order conditions definable by them are characterized. There it is also shown that an elementary condition definable in all kripke frames must be positive. (A formula is *positive* if it is composed from atomic formulae and constant formulae using only $\wedge$, $\vee$ and the quantifiers.)

The use of internal descriptions as defined lies in the possibility to express the condition a modal axiom ϕ imposes. Notice, however, that we have distinguished two ways of adding ϕ as an axiom to a logic Λ. The first is $\Lambda + \phi$, the quasi-normal addition, and the second is $\Lambda \oplus \phi$, the normal addition. Now if $\neg\alpha(x_1)$ is internally described by $\neg\phi$ in $\mathfrak{X}$ then the logic $\Lambda + \phi$ selects from $\mathfrak{X}$ those pointed Λ-frames which satisfy $\alpha(r)$, while $\Lambda \oplus \phi$ selects all Λ-frames which satisfy $(\forall x)\alpha(x)$. Notice that 1^{st}-order definability is relative to the class $\mathfrak{X}$. For example, there are generalized frames which satisfy $\Diamond\Diamond p. \to .\Diamond p$ but are not transitive, although the two correspond for kripke frames. And, more appropriate in this context, there exist generalized frames which satisfy a certain reentrancy postulate without reentrant points being identical. So they satisfy an axiom of the form $\langle x\rangle p. \leftrightarrow .\langle y\rangle p$ for some words x, y, but the x-successor of the root r and the y-successor are not identical. This can happen if the frame is not differentiated.

We give a brief exposé of the calculus of internal descriptions. A frame is *tight* if it satisfies

$$(\forall x)(\forall y)(x \lhd_j y. \leftrightarrow .(\forall a \in \mathbb{F})(y \in a. \rightarrow .x \in \blacklozenge_j a))$$

or, equivalently, if it satisfies the $\mathcal{L}^e$-sentence

$$(\forall x)(\forall y)(x \lhd_j y. \leftrightarrow .(\forall p)(y \epsilon p. \rightarrow .x \epsilon \Diamond_j p))$$

Tightness guarantees that $x \ntriangleleft_j y$ is internally describable. A frame is *compact* if $\bigcap U \neq \emptyset$ for every ultrafilter $U \subset \mathbb{F}$. A frame is *descriptive* if it is differentiated, compact and tight. On kripke frames as well as descriptive frames a limited version of a rule for $\forall$-introduction can be derived. The following theorem can now be proved.

Theorem 4.2 *An $\mathcal{R}$-formula which is positive and in which every non-constant atomic subformula contains at least one inherently universal variable is internally definable in the class of kripke frames and descriptive frames.* ($\dashv$)

Here, a formula is called *constant* if it is composed from the constant atomic formulae t and f. A variable is *inherently universal* if it is bound by a universal quantifier which itself is not in the scope of an existential quantifier. Notice that the occurring constant subformulae can have any shape. The modal axioms to which such properties correspond are the so-called *Sahlqvist formulae*. Call a formula **strongly positive** if is composed from constant formulae and variables with only $\wedge$ and the boxes $\Box_i$, $i \leq m$.

Theorem 4.3 (Sahlqvist) *A modal axiom corresponds to an elementary property of both kripke and descriptive frames if it is of the form $A \rightarrow B$ where A is composed from constant formulae and strongly positive formulae using $\wedge, \vee, \Diamond$, while B is composed from constant formulae and variables using $\wedge, \vee, \Diamond, \Box$.* ($\dashv$)

An important subclass of Sahlqvist formulae are **primitive** formulae. A formula is **simple** if it is composed from constant formulae and variables with the help of $\wedge$, $\vee$ and $\Diamond$. A formula is **primitive** if it is of the form $A \rightarrow B$ where both A and B are simple. Notice that simple formulae may contain instances of $\Box\psi$ but only when ψ is constant. It can be shown that a primitive axiom can be written in such a way that each variable in B occurs exactly once in A. Primitive formulae are all Sahlqvist and define frame properties of the form $\forall\exists$, ignoring the constant subformulae (though not all of the latter are definable by primitive formulae). They are thus of the complexity Seq_1 in the Sahlqvist hierarchy defined in [Kracht, 1995a].

5 The wonderful world of limited choice logics

Of particular significance are the logics of limited alternatives. These logics are characterized by the following postulates.

$$\mathbf{alt}_n \qquad \bigwedge_{i \leq n+1} \Diamond p_i . \to . \bigvee_{i<j} \Diamond(p_i \wedge p_j)$$

These axioms are Sahlqvist and express the condition that any point has at most n different successors. In $\mathcal{R}$ this is expressed by

$$\mathrm{alt}_n \qquad (\forall y_1 \rhd x)(\forall y_2 \rhd x) \ldots (\forall y_{n+1} \rhd x)(\bigvee_{i<j} y_i = y_j)$$

A polymodal logic where each relation satisfies a principle $\mathbf{alt}_n$ for some n is called *limited choice logic*. Moreover, polymodal logics where each relation satisfies $\mathbf{alt}_1$ will be called *no choice logics*. Quite a lot is known about these logics ([Bellissima, 1988], [Segerberg, 1986] and [Grefe, 1994]).

Remembering that we are interested in logics for feature structures the connection should be clear. Feature structures admit per node and arc label ℓ_j at most one related node, and so each single arc fragment is a structure for $\mathbf{K.Alt}_1$. Hence, the logic of feature structures for m features is – constants aside – exactly the m-fold independent fusion $\bigotimes_{i \leq m} \mathbf{K.Alt}_1$. This logic and its extensions are the main object of our investigation. As [Bellissima, 1988] first observed, every limited choice logic is canonical. The argument is extremely simple and consists in an application of what I will call the *principle of finite effect*. It will be used here to show a much stronger property, *df-persistence*. It consists in the observation that in any polyframe of this sort only a finite number of points can be reached from any given point with paths of given length. Thus, if $\langle \mathfrak{F}, x \rangle \models \phi$ for a differentiated $\mathfrak{F}$, and the modal depth of ϕ is δ then the set algebra induced on the subframe consisting of all points reached from x in at most δ steps, is the powerset algebra over this set. Hence we can pass with impunity to the underlying kripke frame, and so $\langle \mathfrak{f}, x \rangle \models \phi$. This shows that any socialist logic is df-persistent and thus Δ-elementary. Another consequence of the principle of finite effect is worth remembering.

Theorem 5.1 *Let Λ be a finitely axiomatizable socialist subframe logic and let Φ be a finite set of formulae. Then $\Lambda + \Phi$ has the finite model property and is decidable.*

Proof. Let δ the maximum of the modal depths of all $\phi \in \Phi$. To know whether a rooted frame $\langle \mathfrak{f}, x \rangle$ satisfies Φ it suffices to study the set of points reachable in at most δ steps. This set is finite and hence this task can be carried out for any frame. Now let a formula ψ be given and assume it has a model $\langle \mathfrak{f}, x, \beta \rangle$. Let $\widehat{\delta}$ be the maximum of δ and the modal depth of ψ. Let $\mathfrak{g}$ be the result of dropping all points not reachable from x in $\widehat{\delta}$ steps. Simple induction shows that $\langle \mathfrak{g}, x, \beta\rangle \models \psi$. $\mathfrak{g}$ is finite and satisfies the postulates of Λ by the fact that the latter is a subframe logic. But $\langle \mathfrak{g}, x\rangle \models \Phi$ as well since the subframe of δ-reachable points of $\langle \mathfrak{g}, x\rangle$ is isomorphic to the subframe of δ-reachable points of $\langle \mathfrak{f}, x\rangle$. $\dashv$

That subframe logics of limited choice have the finite model property was first remarked in [Wolter, 1993], where it was also shown that subframe logics do not necessarily enjoy the finite model property, contrary to what one might expect. These results can be improved drastically in the no choice case. Notice, namely, the following basic equivalences.

$$\Box(p \vee q). \leftrightarrow .\Box p \vee \Box q \qquad \Diamond(p \wedge q). \leftrightarrow .\Diamond p \wedge \Diamond q$$

$$\Box p. \leftrightarrow .\Box\bot \vee \Diamond p \qquad \Diamond p. \leftrightarrow .\Diamond\top \wedge \Box p$$

We will show all of these equivalences. First, in the top row, in both formulas one direction is valid in **K**. Namely, $\Box p. \rightarrow .\Box(p \vee q)$ and $\Box q. \rightarrow .\Box(p \vee q)$ hold in **K** (instances of box-distribution), so that $\Box p \vee \Box q. \rightarrow .\Box(p \vee q)$, by propositional calculus. Likewise, $\Diamond(p \wedge q). \rightarrow .\Diamond p \wedge \Diamond q$ is a theorem of **K**. The converse implication, $\Diamond p \wedge \Diamond q. \rightarrow .\Diamond(p \wedge q)$, is an instance of $\mathbf{alt}_1$. Now replace p by $\neg p$ and q by $\neg q$. This gives $\Diamond(\neg p) \wedge \Diamond(\neg q). \rightarrow .\Diamond(\neg p \wedge \neg q)$. By propositional calculus this is equivalent to $\neg\Diamond\neg(p \vee q). \rightarrow .\neg\Diamond\neg p \vee \neg\Diamond\neg q$, that is, $\Box(p \vee q). \rightarrow .\Box p \vee \Box q$. This shows the upper two equivalences. By replacing q with $\neg p$ in the first equivalence we get that $\Box(p \vee \neg p). \leftrightarrow .\Box p \vee \Box\neg p$. Now, the left hand side is a theorem, $\Box\top$, and so we have as a theorem $\Box p \vee \Box\neg p$, or $\Diamond p. \rightarrow .\Box p$. The latter is often used as a charactistic axiom of $\mathbf{K.Alt}_1$. Since $\Box p \wedge \Diamond\top. \rightarrow .\Diamond p$ is a theorem of **K**, we have shown the lower right equivalence. The lower left equivalence is similar. These equivalences allow to create very simple normal forms (see [Fine, 1975a] for the general case). Recall from standard propositional logic that we can transform any formula into disjunctive normal form. For each formula ϕ there exists an equivalent formula which is a disjunction of formulae ϕ_i such that each ϕ_i is a conjunction of either a variable or its negation. The proof is by showing that negation 'commutes' in some sense with disjunction and conjunction (via the de Morgan laws) and that disjunction and conjunction commute via the distribution laws. The same method can be applied here as well. First, we can move negation inside before variables as usual. We just have to observe that in addition to de Morgan's laws we have $\neg\Diamond p. \leftrightarrow .\Box\neg p$ and $\neg\Box p. \leftrightarrow .\Diamond\neg p$. Thus we can move negation inside. Double negation can as usual be removed. Next we move $\wedge, \vee$ outside the scope of any modal operator. That this is possible is due to two facts. In **K**, we have the equivalences $\Box(p \wedge q). \leftrightarrow .\Box p \wedge \Box q$ and $\Diamond(p \vee q). \leftrightarrow .\Diamond p \vee \Diamond q$, allowing to drive conjunction out of $\Box$ and disjunction out of $\Diamond$. Moreover, by the additional equivalences proved above, conjunction can be moved out of $\Diamond$ and disjunction out of $\Box$. Finally, any subformula $\Box_j\phi$ can be replaced by the conjunction of $\Box_j\bot$ and $\Diamond_j\phi$. Thus, at the cost of introducing some subformulae of the kind $\Box_j\bot$, $\Box_j\phi$ can be replaced by $\Diamond_j\phi$. After having done this we can drive the conjunction out of the scope of the modal operators.

Hence any formula ϕ can be written as a conjunction of formulae of the following kind

$$(\ddagger) \qquad \bigvee_i [u_i]\bot \vee \bigvee_j \langle v_j\rangle\top \vee \bigvee_k \langle w_k\rangle p_k \vee \bigvee_\ell \langle x_\ell\rangle\neg q_\ell$$

Here, u_i, v_j, w_k, x_ℓ are words over the alphabet of arcs. The p_k, q_ℓ need of course not be distinct.

Proposition 5.2 *Every extension of $\bigotimes_j$ **K.Alt**$_1$ can be axiomatized by primitive formulae.*

Proof. Recall that a primitive formula is a formula of the form $A \to B$ where both A and B are made from constant formulae and variables with the help of $\wedge$, $\vee$ and $\Diamond_j$. Now, any formula can be brought into a conjunction of formulae of the form ($\ddagger$). However, adding a conjunction of formulae as an axiom is equivalent to adding the formulae individually, so we can assume our formula to have the form ($\ddagger$). In ($\ddagger$), replace disjunction over negative formulae by an arrow. This gives

$$(\P) \qquad \bigwedge_j [v_j]\bot \wedge \bigwedge_\ell [x_\ell]q_\ell . \to . \bigvee_i [u_i]\bot \vee \bigvee_k \langle w_k\rangle p_k$$

The consequent is a simple formula. Now reduce the subformula $\bigvee_\ell [x_\ell]q_\ell$ into normal form. This procedure will not introduce any negation. Thus after doing this reduction, the antecedent of the conditional is a simple formula as well. $\dashv$

Often enough a modal operator also satisfies the axiom $\Diamond\top$, konwn as **D**. If it does, then we can derive $\Box p. \leftrightarrow .\Diamond p$. With this axiom it is more or less obvious that every formula can be brought into primitive form. Moreover, every constant subformula is equivalent to either $\bot$ or $\top$, so that we derive extremely simple normal forms.

Similar considerations lead to reduced first-order descriptions. We have the equivalences

$$(\forall y \rhd x)(\phi \vee \psi). \leftrightarrow .(\forall y \rhd x)\phi \vee (\forall y \rhd x)\psi$$

$$(\exists y \rhd x)(\phi \,\&\, \psi). \leftrightarrow .(\exists y \rhd x)\phi \,\&\, (\exists y \rhd x)\psi$$

$$(\forall y \rhd x)\phi. \leftrightarrow .(\forall y \rhd x)\mathsf{f} \vee (\exists y \rhd x)\phi$$

$$(\exists y \rhd x)\phi. \leftrightarrow .(\exists y \rhd x)\mathsf{t} \,\&\, (\forall y \rhd x)\phi$$

Remember that the (local) elementary condition expressed by an axiom ϕ above a limited choice logic is positive. Using the reductions above we can eliminate all existential quantifiers in favour of a universal quantifier plus a constant formula. It follows that the first-order conditions are such that all occurring variables are universal and a fortiori inherently universal. Moreover, above polymodal **K.Alt**$_1$.**D** every axiom corresponds to a $\forall$-sentence. The conditions are therefore equivalent to Sahlqvist formula.

Theorem 5.3 *A positive elementary $\mathcal{R}$ ($\mathcal{R}_q$) condition is internally definable in differentiated generalized polyframes for $\bigotimes_{i\leq m}$ $\mathbf{K.Alt}_1$.* $\dashv$

Corollary 5.4 *A condition on (pointed) frames for $\bigotimes_{i\leq m}$ $\mathbf{K.Alt_1}$ is internally definable in differentiated generalized polyframes iff it satisfies a positive $\mathcal{R}$ ($\mathcal{R}_q$) sentence.*

We have stated above that a quasi-normal extension of a subframe logic is decidable if finitely axiomatizable. We will see that this is not so for normal extensions. The tool to prove this is to code Thue-processes. (See [Kracht, 1995a] for the connection with modal logic as described here and [Baader *et al.*, 1993] for similar undecidability result for feature logics.) A **Thue-process** $\mathfrak{T}$ is a finite set of equations $v_i \approx w_i$, $i \leq k$, over a finite alphabet $L = \{\ell_1, \ldots, \ell_m\}$. One can think of the Thue-process as a specification of a semigroup, with generators ℓ_i and certain equations. In mathematical terms we call $\mathfrak{T}$ a **presentation** of that semigroup. It is obtained from that specification by forming the semigroup of all L-words with a binary operation $\cdot$ of concatenation (this is quite an easy object); and then lumping together in one equivalence class all those words that can be obtained from each other by successively replacing a substring identical to a v_i or w_i by w_i (and v_i, respectively). We write $\mathfrak{T} \vdash x \approx y$ for some words x, y if $x = y$ in the semigroup $\mathfrak{F}_{SG}(L)/\mathfrak{T}$, where $\mathfrak{F}_{SG}(L)$ is the free semigroup over L. The free semigroup over L can be interpreted as a kripke frame; the elements are words over the alphabet, and $v \lhd_i w$ iff $w = v \cdot \ell_i$. This frame satisfies exactly the logic $\bigotimes_{i\in m} \mathbf{K.Alt}_1.\mathbf{D}$. Now, the semigroup presented by a Thue-process is a homomorphic image of the free semigroup. It is proved by direct calculation that the homomorphism corresponds to a p-morphism onto a frame whose nodes are the equivalence classes of $\mathfrak{T}$ under the lumping. For two such equivalence classes, $[v]$ and $[w]$ we put $[v] \lhd_i [w]$ iff there are $\widehat{v}$ and $\widehat{w}$ such that $\mathfrak{T} \vdash v \approx \widehat{v}$, $w \approx \widehat{w}$ and $\widehat{w} = \widehat{v} \cdot \ell_i$.

It is known that there exist $\mathfrak{T}$ such that '$\mathfrak{T} \vdash x \approx y$' is undecidable. Take such a $\mathfrak{T}$ and let

$$\Lambda_{\mathfrak{T}} = \bigotimes_{i\leq m} \mathbf{K.Alt}_1.\mathbf{D} \oplus \{\langle v_i\rangle p. \leftrightarrow .\langle w_i\rangle p | i \leq k\}$$

Then $\mathfrak{T} \vdash x \approx y$ iff $\Lambda_{\mathfrak{T}} \vdash \langle x\rangle p. \leftrightarrow .\langle y\rangle p$. Thus $\Lambda_{\mathfrak{T}}$ is undecidable. Now let T_1 be the first-order theory of bi- or polymodal $\mathbf{K.Alt}_1$-frames. Then $\Lambda_{\mathfrak{T}}$ is an extension of polymodal $\mathbf{K.Alt}_1.\mathbf{D}$. The postulate $\langle x_i\rangle p \leftrightarrow \langle y_i\rangle p$ is not only first-order but universal. It says that any x_i-successor (i. e. the one and only such successor) is equal to any y_i-successor. Now take a Thue-process $\mathfrak{T}$. It induces a strengthening T_2 of the theory T_1 by the axioms corresponding to the equations. Hence $T_2 = T_1 \cup U$, where U is a finite set of universal sentences. By construction, $T_2^{\forall}$ is undecidable. But $T_2^{\forall} = (T_1 \cup U)^{\forall}$. Suppose that T_1 is decidable. Then since $T_2^{\forall}$ is the set of all ϕ such that $T_1 \vdash \bigwedge U \rightarrow \phi$, $T_2^{\forall}$ is decidable, contrary to fact. So T_1

is undecidable. Now, modulo T_1, $\bigwedge U \rightarrow \phi$ always reduces to a universal sentence (the reduction is effective). Hence $T_1^\forall$ is undecidable. Thus we have derived that the universal theory of two binary, (quasi-)functional relations is undecidable; we have also shown that the universal theory of two unary function symbols is undecidable.

It is known to be undecidable whether or not a modal axiom expresses an elementary condition (proved in [Chagrova, 1991]). The following converse in bimodal logics due to [Grefe, 1994] holds.

Theorem 5.5 (Grefe) *($m \geq 2$) It is undecidable whether a first-order condition is modally definable in classes of frames for* $\bigotimes_{i \leq m} \mathbf{K.Alt}_1$.

Proof. We work with two modal operators. Recall that $T_1^\forall$ is undecidable. The formula

$$\alpha_0 = (\forall x)[(\forall y \rhd_1 x)(y \neq x) \wedge (\forall y \rhd_2 x)(y \neq x)]$$

expresses the irreflexivity and is not modally definable, as is standardly known in modal logic. The culprit is the fact that the formula has a negative matrix (see [Kracht, 1993a]). Now consider the formula $\beta = \alpha_0 \vee \gamma$, where γ is arbitrary. Then if for a kripke frame $\mathfrak{f}$ we have $\mathfrak{f} \models \alpha_0$, then also $\mathfrak{f} \models \beta$. Now suppose that β is modally definable. Then it must hold on all kripke frames for $\mathbf{K.Alt}_1 \otimes \mathbf{K.Alt}_1$ by the fact that every frame is the p-morphic image of an irreflexive frame. Thus $T_1 \vdash \alpha_0 \vee \gamma$, whence $T_1; \neg\alpha_0 \vdash \gamma$. Suppose now that β is not modally definable. Then $T_1 \nvdash \beta$, that is, $T_1; \neg\alpha_0 \nvdash \gamma$, for otherwise β holds in all frames and is therefore modally definable (for example by the true constant). Hence if we are able to show that $T_2 = T_1 \cup \{\neg\alpha_0\}$ is undecidable, we have succeeded in showing that modal definability (of β) is undecidable.

Now consider the theory $T_3 = T_1 \cup \{\alpha_1\}$ with

$$\begin{array}{rl} \alpha_1 = (\exists x) & [(\forall y \rhd_1 x)(y = x) \wedge (\forall y \rhd_2 x)(y = x) \\ \wedge & (\forall z)\{(\forall y \rhd_1 z)(z \neq x) \wedge (\forall y \rhd_2 z)(y \neq x)\}] \end{array}$$

Since $\vdash \alpha_1 \rightarrow \neg\alpha_0$ we have that if T_2 is decidable, so is $\{\zeta | T_2 \vdash \alpha_1 \rightarrow \zeta\}$ $= \{\zeta | T_1; \neg\alpha_0; \alpha_1 \vdash \zeta\} = \{\zeta | T_1; \alpha_1 \vdash \zeta\} = \{\zeta | T_3 \vdash \zeta\}$. So we are done if we have shown that T_3 is undecidable. Now, consider a frame $\mathfrak{f}$ for T_1. If we add an inaccessible, reflexive point, that is, if we form the disjoint union $\mathfrak{f} \oplus \mathfrak{r}$, where $\mathfrak{r}$ is the one-point, reflexive frame, then we have a T_3-frame. And a T_3-frame is a T_1-frame. Thus, by standard model theory, $T_3^\forall = T_1^\forall$. Hence, since $T_1^\forall$ is undecidable, so is $T_3^\forall$ and a fortiori T_3. $\dashv$

In addition to certain Thue-processes being undecidable it is also undecidable whether a given process $\mathfrak{T}$ is decidable. This result can be sharpened. Instead of the above axiomatization we can consider an axiomatization where we use a subframe axiom to code a Thue-equation, namely

$$(\langle x_i \rangle p. \rightarrow .[y_i]) \wedge (\langle y_i \rangle p. \rightarrow .[x_i]p)$$

This states that if we start at a given point and an x_i-path exists, then it is equal to all y_i-paths starting from that point, and if a y_i path exists it is equal to all x_i-paths. Actually, only one half of this axiom is sufficient.

$$\Sigma_{\mathfrak{T}} = \bigotimes_{i \le m} \mathbf{K.Alt}_1 \oplus \{\langle v_i \rangle p. \leftrightarrow .[w_i]p | i \le k\}$$

The logic $\Sigma_{\mathfrak{T}}$ is a subframe logic and therefore has the finite model property and is decidable. However, if we add the postulates $\langle \ell_j \rangle \top$ for all $j \le m$ we get the logic $\Lambda_{\mathfrak{T}}$.

Theorem 5.6 *$(m \ge 2)$ There exist undecidable, finitely axiomatizable normal extensions of* $\bigotimes_{i \le m} \mathbf{K.Alt_1}$. $\dashv$

Call an extension **trivial** if it has only one model, with a single point x reached from itself by all features. Equivalently, if corresponds to the logic $\Lambda_{\mathfrak{T}}$ where $\mathfrak{T} = \{\ell_i \approx \epsilon | i \le m\}$. Using a result by Rabin the following can be proved.

Theorem 5.7 (Grefe) *$(m \ge 2)$ Whether a finitely axiomatizable extension of* $\bigotimes_{i \le m} \mathbf{K.Alt}_1$ *is decidable or trivial is undecidable.* $\dashv$

Finally, consider the question of finite model property. It, too, is undecidable. For suppose it is decidable. Then we show that it is decidable whether or not a Thue-process is trivial, a contradiction. Thus let $\mathfrak{T}$ be a Thue-process. First decide whether or not $\Lambda_{\mathfrak{T}}$ has the finite model property. If not, it is not trivial. If it does have the finite model property, then it is decidable (!) and so $\Lambda_{\mathfrak{T}} \vdash \langle \ell_i \rangle p. \leftrightarrow .\langle \epsilon \rangle p$ is decidable, showing the decidability of triviality.

Theorem 5.8 *$(m \ge 2)$ It is undecidable whether an extension of* $\bigotimes_{i \le m} \mathbf{K.Alt}_1$ *has the finite model property.* $\dashv$

In [Segerberg, 1986] it is shown that for $m = 1$ this theorem is false. All normal extensions are finitely axiomatizable and have the finite model property and are thus decidable. Although there are uncountably many quasi-normal extensions, the finitely axiomatizable extensions have the finite model property as well, as follows from Theorem 5.1. However, as the previous theorem shows, even *recursively* axiomatizable quasi-normal extensions need not be decidable.

The results can be summarized as follows. The lattice $\mathcal{N}(\bigotimes_j \mathbf{K.Alt}_1)$ is a sublattice of $\mathcal{Q}(\bigotimes_j \mathbf{K.Alt}_1)$. The first contains undecidable logics, so does the latter. However, the semilattice $\mathcal{N}_f(\bigotimes_j \mathbf{K.Alt}_1)$ of finitely axiomatizable normal extensions is *not* a subsemilattice of the lattice $\mathcal{Q}_f(\bigotimes_j \mathbf{K.Alt}_1)$

of finitely axiomatizable quasi-normal extensions. [3] The latter consists exclusively of decidable logics, while the former does not. Nothing changes if we replace $\bigotimes_j \mathbf{K.Alt}_1$ by an extension – provided this is a subframe logic.

These results show the following.

Theorem 5.9 *Let Λ be a finitely axiomatized subframe logic extending polymodal $\mathbf{K.Alt}_1$. Then the lattice $\mathcal{Q}_f(\Lambda)$ is decidable. That is, for two logics $\Theta_1, \Theta_2 \in \mathcal{Q}_f(\Lambda)$ the problems '$\Theta_1 \subseteq \Theta_2$' and '$\Theta_1 = \Theta_2$' are decidable.*

Proof. Let $\Theta_1 = \Lambda + \Phi_1$ and $\Theta_2 = \Lambda + \Phi_2$ with finite Φ_1 and Φ_2. Then for each $\phi_2 \in \Phi_2$ it is decidable whether $\phi_2 \in \Lambda + \Phi_1$, by Theorem 5.1. Hence $\Phi_2 \subseteq \Lambda + \Phi_1$ is decidable and so '$\Lambda + \Phi_2 \subseteq \Lambda + \Phi_1$' is decidable. Consequently, '$\Theta_1 = \Theta_2$' is also decidable. $\dashv$

6 Extensions and fragments of the language

Various enrichments of the standard modal apparatus have been considered ([Blackburn, 1993], [Blackburn and Spaan, 1993a], [Blackburn and Spaan, 1993b], [Gazdar *et al.*, 1988]). In addition there is the 'classical' language of Kasper and Rounds (see [Kasper and Rounds, 1990]) with it's various extensions, for example [Baader *et al.*, 1993]. We will discuss them in turn concentrating on the aspects of definitional strength and their decidability.

Kasper and Rounds in their work advance a language L^{KR} that has conjunction, disjunction, arc modalities $\langle \ell_j \rangle$ and a device that is equivalent to having path equations $v \approx w$. Indeed, here (and only here) only the question of descriptive power as opposed to the defining power makes sense. But this is only a consequence of the fact that there is no negation. Consider now a formula ϕ in L^{KR} and ask what it tells us about a model for it. Suppose there are no path equations. Then by the fact that there is no negation only constant conditions are describable. So, only path equations introduce non-trivial, non-constant conditions. The elementary condition described by ϕ is thus a conjunction of disjunctions of constant formulae or path equations. Compare this now with $\bigotimes_{i \leq m} \mathbf{K.Alt}_1$. A formula ϕ defines (!) a conjunction of disjunctions of constant formulae or path equations. However, in contrast to the logic of Kasper and Rounds we can also define the non-existence of paths. By the reduction to normal form (‡) one can show that this is the exact limit. Call a modal formula *impartial* if it does not contain positively a subformula asserting the inexistence of paths.

Theorem 6.1 *Any satisfiability condition of L^{KR} is modally definable (in differentiated polyframes) by an impartial modal formula.* $\dashv$

[3] Notice that in contrast to quasinormal extensions, the intersection of two finitely axiomatizable normal extensions need not be again finitely axiomatizable.

Given a formula ϕ of L^{KR}, the satisfiability problem for can be transformed into the problem

$$\bigotimes_{i \leq m} \mathbf{K.Alt}_1 + \widehat{\phi} \vdash \bot$$

for a suitable translation $\widehat{\phi}$. This translation is as follows. Translate booleans by booleans and arc modals by arc modals; translate $v \approx w$ by

$$\langle v \rangle \top \wedge \langle w \rangle \top \wedge (\langle v \rangle p. \leftrightarrow .\langle w \rangle p)$$

(See [Blackburn, 1993].) As an immediate consequence we get the decidability of the logic of Kasper and Rounds by reference to Theorem 5.1.

We can consider also axiomatic extensions of Kasper and Rounds' logic, for example by path existence statements and by path equations. To let this be nontrivial, we will assume these statements to hold globally, that is, if $v \approx w$ is added as an axiom then $xv \approx xw$ as well as $vx \approx wx$ should be derivable as well. It is possible to axiomatize Kasper and Rounds' logic by axioms and rules in such a way that this is possible. The interesting question is whether all extensions are decidable. Here the answer depends on the interpretation of $\approx$. There are three plausible alternatives. (1) $v \approx w$ holds at a node if the two paths are equal if both exist; (2) $v \approx w$ is true at a node if whenever one path exists, the other exists as well, and if both exist they are equal; (3) $v \approx w$ is true at a node if both paths exist and are equal. (1) is a condition closed under taking subframes, so by reference to Theorem 5.1 we know that all extensions by finitely many axioms are decidable. (2) and (3) are stronger and both allow to recreate undecidability via the word problem alias Thue-process. Incidentally, Kasper and Rounds' have chosen (3).

Rather than using the method of axiomatic classification [Blackburn, 1993] uses a more powerful language to ensure that path equations can be described. He adds *nominals*; these are a special sort of propositional variables (constants) which can and must be made true at singleton sets. If i is a nominal and β a valuation, $\beta(i) = \{x\}$ for some x. This poses some problems for generalized frames here, but the logics we consider here are all complete, so for the correspondence problem we can settle comfortably on Kripke frames, which removes the worry about non-internal singleton sets. It is clear that by using nominals we can describe equality of worlds in kripke frames (that is, we can define inequality). All that is required for x and y to be equal is that some nominal i is simultaneously true at both x and y under some valuation. So, modal logics with nominals are more powerful than modal logics without nominals. Interestingly, an inspection of typical axioms for attribute-value structures show that they can be written with exclusive use of nominals. Ordinary propositional variables can be eliminated. Call a formula **purely nominal** if it contains no ordinary propositional variable.

Theorem 6.2 *Every extension of* $\bigotimes_j \mathbf{K.Alt}_1$ *can be axiomatized by means of purely nominal axioms.*

Proof. (*Fast proof.*) Observe that for primitive formulae the singleton sets are decisive in the sense of [Kracht, 1993a], whence 1^{st}-order correspondence is determined by singleton valuations. So nothing is changed if proposition variables are exchanged by nominals. $\dashv$

Proof. (*Slow proof.*) Any extension is complete with respect to kripke frames, so the axiomatic equivalence can be checked on kripke frames. We know that any formula can be reduced to primitive form. So we have an axiom of the form $A \to B$ where A and B contain $\Box_j \psi$ only if ψ is constant. Moreover, we can assume that a variable occurs in A at most once. It is enough if we show that $A \to B$ is axiomatically equivalent to $(A \to B)^n$ where $(A \to B)^n$ results from $A \to B$ by replacing a propositional variable p by a nominal i_p not already occuring in $A \to B$. So assume $\mathfrak{f} \nvDash A \to B$. Then there exists a valuation β and a x such that

$$\langle \mathfrak{f}, x, \beta \rangle \models A, \neg B$$

A is a disjunction of constant formulae or formulae of the form $\langle w \rangle q$ for some word w and variable q. If $q = p$ then we have in particular

$$\langle \mathfrak{f}, x, \beta \rangle \models \langle w \rangle p$$

Since $\mathfrak{f}$ is functional with respect to the relations, there is exactly one point y such that y is w-related to x. Thus

$$\langle \mathfrak{f}, y, \beta \rangle \models p$$

Define now β^+ by $\beta^+(q) = \beta(q)$ if $q \neq p$ and $\beta^+(p) = \{y\}$. Then we have

$$\langle \mathfrak{f}, x, \beta^+ \rangle \models A, \neg B$$

Namely, we have seen that A is satisfied with new valuation, simply by construction. For $\neg B$, however, we need not worry, because the set of points on which $\neg B$ is true grows when the sets on which the variables are true shrinks, because $\neg B$ contains every variable negatively. With β^n like β^+ but $\beta^n(i_p) = \{y\}$ we thus have

$$\langle \mathfrak{f}, x, \beta^n \rangle \models A^n, \neg B^n$$

And this had to be shown. Assume conversely that there is a valuation γ and a point x

$$\langle \mathfrak{f}, x, \gamma \rangle \models A^n, \neg B^n$$

Then let γ_n be like γ except for p where we put $\gamma_n(p) = \gamma(i_p)$. Then

$$\langle \mathfrak{f}, x, \gamma_n \rangle \models A, \neg B \quad \dashv$$

Notice also that the formulae $\mathbf{alt}_1$ can be replaced by purely nominal formulae because they are primitive. The internal descriptive power of the language with both nominals and standard propositions is therefore the same as the descriptive power of the sublanguage with just nominals. Moreover, the novum is the describability of identity, which shows by means of the calculus of internal descriptions that with nominals any $\mathcal{R}$ ($\mathcal{R}_q$) condition which is existential modulo constant formulae is internally describable. Every condition reduces to such a condition in no choice logics.

Theorem 6.3 *Every $\mathcal{R}_q$-sentence is internally both describable and definable in frames for $\bigotimes_j \mathbf{K.Alt}_1$ in logics with nominals; every $\mathcal{R}$ sentence is internally definable.* $\dashv$

By the finite effect principle it can be shown that the logic $\bigotimes_{i\leq m} \mathbf{K.Alt}_1$ enriched with nominals is decidable, and an analogue of Theorem 5.1 can be proved. The technique of selecting points is the one used in [Blackburn and Spaan, 1993b] as well as [Blackburn, 1993]. The addition of nominals has the advantage to allow for defining inequality – which is not possible in ordinary modal logic. [Carpenter, 1992], while rejecting the use of negation (which would result in a boolean logic of types), argues for inequations, but the only real example is that of disjoint reference for pronouns. This, however, is not such a strong case; it has been argued that the disjoint reference is not syntactically required but follows from pragmatic considerations on top of the fact that the pronouns allows the option of same or different referent, while the reflexive does not (see [Reinhart, 1983]). For an excellent exposition arguing that disjoint reference is a non-issue see [Fiengo and May, 1994]. This being the only case where non-identity has been claimed to be necessary, we can therefore not discern much necessity in the introduction of nominals.

Another concept which enjoys only marginal status is that of the relational converse. Suppose that R is a binary relation. Then $R^{\smile} = \{\langle y, x\rangle | xRy\}$ is called the *(relational) converse* of R. R being $\lhd_j$, one of the accessibility relations in a polyframe, we write $\rhd_j$ instead of $\lhd_j^{\smile}$. The converse relation of an accessibility relation plays a fundamental role in tense logic, which is essentially the logic of bi-frames in which the relations are converses of each other. The converse relation of an attribute appears for example in [Johnson and Moss, 1994], and we will make use of this construction further down. Here, let us only show how one can axiomatically characterize that two operators, call them $\overleftarrow{\Diamond}$ and $\overrightarrow{\Diamond}$ are inverses of each other. These are

$$p \rightarrow \overleftarrow{\Box}\overrightarrow{\Diamond}p \qquad p \rightarrow \overrightarrow{\Box}\overleftarrow{\Diamond}p$$

Via Corresponence Theory it can be shown that these two axioms are sufficient. The reader may try to prove that for himself, or consult a textbook. I should also add that the addition of the converse is not just a matter of definitional strength. One has to be careful in that the *meaning of a variable changes*. This is due to the fact that there are now less generated substructures. If a frame is connected via $\lhd$, then it has no nontrivial generated subframes. We will return to this point below.

In [Gazdar *et al.*, 1988] it was proposed to add a *master modality* on top of the arc modalities. This master modality should allow to look arbitrarily deep into the structure. I proposed an axiomatization in [Kracht, 1989] and proved that the logic has the finite model property. However, the axiomatization only guarantees the master to be at least as strong as the intended one; the reason was that a priori there was no bound on the number of arcs. The result is therefore of little practical value. The case where we have only a finite number of arc modalities is actually covered by a result of [Ben-Ari *et al.*, 1982]. Call a program π **deterministic** if the corresponding modality $\langle\pi\rangle$ satisfies $\mathbf{alt}_1$, that is,

$$\langle\pi\rangle p \wedge \langle\pi\rangle q. \rightarrow .\langle\pi\rangle(p \wedge q)$$

Theorem 6.4 (Ben-Ari, Halpern and Pnueli) *The dynamic logic of deterministic programs is decidable.* $\dashv$

Corollary 6.5 *The logic of category definitions is decidable.* $\dashv$

The master modality is the reflexive, transitive closure of the union of all basic modalities. So, with $\blacklozenge$ denoting this union, $\blacklozenge^*$ will be the sought after master modality. The definability and describability questions are not so straightforward. Since $\blacklozenge^*$ is not limited choice, there is no hope that the second-order definition can be reduced to a first-order one; moreover, there exist no exchange laws for existential and universal quantifiers.

The principal motivation of a master modality is actually to overcome the distinction between quasi-normal extensions and normal ones. The problem is that, as explained, traditionally only the satisfaction problem with respect to the basic logic has been studied, or, equivalently, the decidability of quasi-normal extensions. However, some conditions on AVMs amount to adding an axiom normally rather than quasi-normally, which means if seen as a satisfaction problem, we are asking not for the *local* but for the *global* satisfaction of a formula. Namely, in presence of a master modality all finitely axiomatizable normal extensions turn out to be finitely axiomatizable as quasi-normal extension. For it is the case that $\Lambda \oplus \phi = \Lambda + \blacksquare^*\phi$. This has for consequence that there are finitely axiomatizable quasi-normal extensions of the logic of category definitions which are undecidable, and that it is not decidable which of them is decidable.

Instead of a master modality, [Blackburn and Spaan, 1993b] study the addition of a *universal modality* $\boxplus$ (as from [Goranko and Passy, 1992]). With this modality we get back our standard unrestricted quantifiers. The universal modality is the reflexive and transitive closure of $\blacklozenge$ and it's converse $\blacklozenge^{\smile}$ or, alternatively, as the fusion with **S5** plus the axiom $\boxplus p. \rightarrow .\blacksquare p$. The primary advantage of the universal modality, which it shares with the master modality, is that we can enforce global constraints by a local condition. However, from an intuitive point of view, the master modality, and not the universal modality, is the right kind of object to look for. The reason is that we want to think of the initial relations as fundamental and when moving around in the structure we must follow these relations rather than jump anywhere we want. This has consequences also for the notion of a *constituent*, as we will see below. In [Blackburn and Spaan, 1993b] it is proved that the universal modality is from a complexity point of view better behaved than the master modality. This is to be expected, if we consider the fact that the universal modality is doing not much more than allowing to collapse the global and the local derivability relation (see [Goranko and Passy, 1992], while the master modality has greater expressive power. I remain unimpressed, however, by such an argument. First of all, it is unclear whether the complexity bonus of the universal modality will really show up in practice, for the simple reason that linguistic statements may be simpler (and in fact are simpler) than those that are responsible for the worst case in the complexity result. Second, even though low complexity is an advantage, it should not be a criterion when choosing the language. Much more than that, I claim, we must strive for *natural* translations, those which are by themselves fitted to the problem. In addition, adding the universal modality is tantamount to appreciating the use of first-order predicate logic in this context. Recall that if a first-order formula ϕ is modally definable if it can be rewritten with restricted quantifiers such that the matrix is positive and the variables meet a certain condition. In this context, a restricted quantifier corresponding to the universal modality is really an unrestricted quantifier. So, with the other qualifications remaining in place, we are a good step closer to using predicate logic. I do think that the hallmark of modal logic as opposed to predicate logic is in this context to use of restricted quantification, so that adding the universal modality removes that distinctive feature.

These extensions can be mixed. We can add both nominals and a master modality or a universal modality. In both cases the defining power increases drastically, due to the fact that identity can be described. Without any master modality this was harmless by the finite effect principle but now that there is no such principle we are caught. Consider a Thue-process $\{v_1 \approx w_1, \ldots, v_k \approx w_k\}$ and an equation $x \approx y$. Write the following formula

$$\blacksquare^*(\bigwedge_j \langle v_j \rangle i_j \leftrightarrow \langle w_j \rangle i_j). \rightarrow .\langle x \rangle i_{k+1} \leftrightarrow \langle y \rangle i_{k+1}$$

This formula defines the decision problem '$\mathfrak{T} \vdash x \approx y$' for the Thue-process $\mathfrak{T}$. Hence there are finitely axiomatizable quasi-normal extensions of the logic with *and* a master modality (or universal modality) for which the decision problem is undecidable. [Blackburn and Spaan, 1993b] and [Blackburn and Spaan, 1993a] explore variations on that theme, but the outcome is always the same. Replacing the nominals with path equations exacerbates the problem in that it is now the *base logic itself* which is undecidable. Similarly if we add the connective $\Rightarrow$ defined by

$$\phi \Rightarrow \psi := \boxplus(\phi \rightarrow \psi)$$

Notice namely that $\boxplus\phi$ is definable via $\top \Rightarrow \phi$, so that little is changed passing to this new language. Notice that the connective $\Rightarrow$ is exactly the one used in constraint programming, cf. [Carpenter, 1992].

Finally, [Baader *et al.*, 1993] define a language that extends Kasper and Rounds' language by negation and what is called *functional uncertainty*. The latter means nothing but that for any regular set L of words over the alphabet of arcs we have the expression

$$\langle L\rangle\phi := \bigvee_{w\in L} \langle w\rangle\phi$$

The addition of negation brings us into plain modal logic. Functional uncertainty is equivalent to the introduction of the *Kleene Star*

$$\langle w^*\rangle\phi := \bigvee_i \langle w^i\rangle\phi$$

Notice that $x \approx y$ becomes internally definable via $\blacksquare^*(\langle x\rangle p. \leftrightarrow .\langle y\rangle p)$. So, unsurprisingly, even the base logic is undecidable. If we want to express modally in test free dynamic logic we need in fact to add nominals. Or, alternatively, we might replace this by dynamic predicate logic, where the only predicate is equality.

In subsequent sections I will argue that the most suitable language for describing linguistic entities is dynamic logic, in particular deterministic dynamic logic. It will *include* the master modality, indeed *any* master modality. In will *exclude* the use of nominals, the use of converse relations and the universal modality. I have argued two of these cases. The reasons for banning te converse relation(s) will be more subtle. There will be a certain split in the feature language in that the *star* will only be needed when talking about syntactic structures in the traditional sense, while it will be dispensable when talking about categories (again in the traditional sense).

Part II: Feature Logic in Linguistic Practice

7 Coding Linguistic Structures

While in GPSG the feature structures where finite objects, giving specifications for the type of a constituent, HPSG and other frameworks took this one step further and used feature structures throughout to code linguistic entities (see [Pollard and Sag, 1987]). We will see below how this is done and show that there is a reduction of trees to feature structures. However, two questions arise after this observation has been made. First, should we then replace trees by feature structures? And second, what significance does this observation have in linguistics? To answer the first question, let us just observe that if indeed there is such a translation then it does not really matter in which of these formalisms we express ourselves – this is just a question of personal taste and whether one is at ease with one or the other. This will also answer the second question: there is no linguistic significance in this observation.

Nevertheless, there is a point being made by this in HPSG namely that one should regard the distinction between categories and tree structure as artificial. Looking at a code of a traditional syntactic tree we see that there is no distinction made between what is inside or outside a category. It has lost its own status. Thus, in spite of the previous criticism, one has to adduce good arguments in favour of the traditional split. I guess that the real reason for distinguishing between categories and structure is that the number of categories is bounded, while the number of distinct structures is not. So, we take side here with GPSG as concerns the analysis. However, this does not mean we are claiming that languages are context-free. Even if there are only finitely many categories, the conditions on the structures (or trees) can be more complex. We might, for example, consider a TAG-grammar instead of an ordinary CFG. The former can generate non-context free languages using finitely many symbols. A typical transformational grammar can even be more powerful. Be this as it may, the question remains as to why there are only finitely many categories. I cannot really produce a knock-down argument. What I want to say is that there are only finitely many words. However, this argument meets two objections. (1) There obviuosly are infinitely many words, (2) The are more symbols in the grammar than there are lexical categories. To answer the second objection, even though there are more non-lexical categories, each of them must be lexically motivated. If we believe in X-bar-syntax or something of that sort, then each non-lexical category is a projection of a lexical one, so if the latter is finite, so is the former. Objection (1) is somewhat harder. We must acknowledge the fact that the lexicon is productive. Yet, it is arguably clear that simple lexotactic processes do not give rise to an increase in categories, so in languages like German, where huge words can be built by compounding, these words nevertheless behave syntactically like corresponding non-compounded words. Similarly, in most languages a word displays only a finite amount of morphological variation, due to inflection and derivation. The increase is only by a finite factor. So a challenge derives only from processes which change the nature of the item and are iterable at the same time. Such morphological processes are subsumed under the header *incorporation* in [Baker, 1988]. The distinctive nature of incorporation processes is that they change the syntactic properties of the incorporating element. For example, case assignment possibilities may change. A verb may suddenly assign case to quite different NPs after incorporation. Notice, however, that Baker assumes the following principle.

> CASE FRAME PRESERVATION PRINCIPLE.
> A complex X^0 of category A in a given language can have at most the maximal Case assigning properties allowed to a mor-

phologically simple item of category A in that language.

What this says is that from the point of view of case relations, an incorporation process is licit only if it transform the item into an item of an already realized syntactic class. So, iterable morphological processes do not expand the class of pre-terminal symbols. We conclude that it is justified to assume a finite amount of syntactic symbols out of which complex structures are built. The potential infinity of syntactic entities then is entirely within syntax, that is, the structure (but see below for a closer look).

We will argue in detail in the subsequent sections the various approaches to formalization. In the remainder of this section we will introduce a logic for formalizing tree structures. This logic will be something of a reference point for my own views on the subject. The basic idea is that an ordered tree is a quadruple $\langle T, r, <, \sqsubset\rangle$, where T is the set of nodes, r the root, $<$ the proper dominance relation, and $\sqsubset$ the linear precedence relation for nodes. We assume that the reader is familiar at least with the notion of a tree; suffice it say that for T to be a tree it is sufficient that $<$ is an irreflexive, transitive relation such that for all x, if $x < y, z$ then either $y < z$ or $y = z$ or $y > z$. r is the root if there is no x such that $r < x$. A syntactic tree specifies a structured event, namely that of an utterance. The nodes correspond to parts of that utterance, subevents, that is. The dominance relation describes whether a subevent of the utterance is a proper subevent of the other. The relation $\sqsubset$ specifies which event precedes another. $\sqsubset$ is irreflexive and transitive. Moreover, two events x, y *overlap* if $x < y$ or $x = y$ or $x > y$. Two no-overlapping events must be $\sqsubset$-comparable. It is possible to write down explixit first-order axioms, but we will refrain from doing that. We will turn an ordered tree into a pointed kripke frame $\mathfrak{f} = \langle T, r, \mathit{down}, \mathit{right}, \mathit{left}\rangle$. We define x *down* y iff $x > y$ and there is no z such that $x > z > y$. Thus, *down* is immediate dominance. We define *right* to be immediate precedence and *left* immediate succedence between daughters of a single node only. So x *right* y iff $x \sqsubset y$ and there is a z such that $z \prec_1 x, y$ and no u exists such that $x \sqsubset u \sqsubset y$. Finally, x *left* y iff y *right* x. From the primitive relations *down*, *right* and *left* any complex relation using the resources of dynamic logic may be formed. So we have defined in effect a dynamic logic over three basic relations with the following interaction postulates.

(**coh.r**)	$\langle \mathit{down};\ \mathit{right}\rangle p. \to .\langle \mathit{down}\rangle p$
(**coh.l**)	$\langle \mathit{down};\ \mathit{left}\rangle p. \to .\langle \mathit{down}\rangle p$
(**con**)	$\langle \mathit{down}\rangle p \wedge \langle \mathit{down}\rangle q. \to .\langle \mathit{down}\rangle(p \wedge \langle \mathit{right}^* \cup \mathit{left}^*\rangle q)$
(**lr**.+)	$p \to [\mathit{right}]\langle \mathit{left}\rangle p$
(**lr**.−)	$p \to [\mathit{left}]\langle \mathit{right}\rangle p$
(**alt**$_1$.**r**)	$\langle \mathit{right}\rangle p \wedge \langle \mathit{right}\rangle q. \to .\langle \mathit{right}\rangle(p \wedge q)$
(**alt**$_1$.**l**)	$\langle \mathit{left}\rangle p \wedge \langle \mathit{left}\rangle q. \to .\langle \mathit{left}\rangle(p \wedge q)$
(**Ir.d**)	$\langle \mathit{down}^+\rangle p. \to .\langle \mathit{down}^+\rangle(p \wedge [\mathit{down}^+]\neg p)$
(**Ir.l**)	$\langle \mathit{left}^+\rangle p. \to .\langle \mathit{left}^+\rangle(p \wedge [\mathit{left}^+]\neg p)$
(**Ir.r**)	$\langle \mathit{right}^+\rangle p. \to .\langle \mathit{right}^+\rangle(p \wedge [\mathit{right}^+]\neg p)$

The first two ensure a coherence between moving down and left/right. They mirror the fact the left and right are defined only among daughters. The third postulates ensures that the daughters are connected by precedence. Then follow two axioms stating that *left* and *right* are inverses of each other. Then there are the standard no choice postulates for *left* and *right*. Finally, for each of the three basic operators we have added a postulate that ensures the irreflexivity of the operators. This language and logic is similar to that studied in [Kracht, 1995b], but the *up* relation, the converse of *down* is left out here. We will see why this has to be so. Notice, finally, that models must always be trees with the distinguished world being the root. Technically, without the relation *up* this cannot be achieved. The best approximation is to add *quasi-normally* these axioms

$$\neg\langle right\rangle\top \qquad \neg\langle left\rangle\top$$

They will make sure that we are standing at a sisterless node. We will in sequel not insist on these postulates.

Feature structures have been designed also to exhibit the fine strucure of linguistic categories, so it is perhaps not wise to treat the nodes of a tree as atomic entities. Hence, we will need a few more modal operators to cater for the internal structure of categories. This is not hard to do. A syntactic object now looks pretty much like a HPSG-type feature structure, so we have by this move also abandoned the distinction between being internal and being external of a linguistic category. But we are at the moment not arguing that this is the best possible formalization of matters, so we leave that aside here, pursuing only the formal consequences. We now have modal operators that take us down, left and right in a tree and those which take us into a node. We must ensure now that the tree structure and the category structure do not mingle too much. In [Blackburn *et al.*, 1993] this is achieved by *layering* two languages, one for talking about the structure and another for talking about the internal shape of a linguistic category. *Layering* them consist in banning all formulae which stack structure formulae inside category formulae. For example, while we might happily speak of the case feature of a daughter of a noun phrase, speaking likewise of the daughter (in the tree sense) of the case feature of a noun phrase, is talking nonsense. Inasmuch as layering achieves the right kind of restrictions, it is quite successful. However, it is unnecessary to impose layering since by using the right kind of axioms we can directly force the models of the logic to be of the right shape. Notice namely that layering puts a ban only on the way we may express things, not on the structures themselves. So, rather than layering the languages we use axioms. Suppose then that in addition to the tree-relations *left*, *right* and *down* we have category relations FEAT_i, $i \leq m$, of unspecified nature. Then in addition to the postulates above and the no choice axioms for the features, we add the following axioms.

$$\begin{array}{l}[\mathrm{FEAT}_i]\,\neg\langle down\rangle\top\\ [\mathrm{FEAT}_i]\,\neg\langle left\rangle\top\\ [\mathrm{FEAT}_i]\,\neg\langle right\rangle\top\end{array}$$

These axioms are Sahlqvist, hence elementary, and express the fact that when we follow a feature relation, that is, go inside a category, then no daughter and no sister is defined. This has the desired effect of layering the two structure and leaving the tree, as it where, *outside* the categorial structure. We call the so defined logic the *constituent logic*, **CL**.

It will be this logic – or better, the family of logics of the above kind – that we will be considering. Variation may consist in adding propositional constants. Moreover, if we believe with [Kayne, 1983] that syntactic trees are binary branching, then the relation *down* actually satisfies $\mathbf{alt}_2$. In general, if an upper limit δ on the number of daughters can be found, then the *down*-relation is of limited choice. In that case we can interpret the language in the language of deterministic dynamic logic as follows. Instead of the relation *down* we propose to have a number of relations $down_i$, $i \leq \delta$. The relations *left* and *right* will be dropped. Instead, it is implicitly assumed that if $x\, down_i\, y$ and $x\, down_j\, z$ then $i < j$ iff $y\, right\, z$. It is a somewhat longwinded procedure to show that with these stipulations talk of sisters can in the linguistically relevant cases be eliminated in this new language. It is an extremely simple language. It has the relations $down_i$ and FEAT_j, each satisfying $\mathbf{alt}_1$. There is no interdependency between these relations except for layering. Thus, if there is an a priori bound on the number of daughters, we can design an attribute-value language to talk about the trees in this way. We call this *vertical coding* of trees into AVMs. However, in [Pollard and Sag, 1987] another method is described, which we refer to as *horizontal coding*. It is related to the treatment of lists in LISP. Introduce an operator *first* and an operator *rest*. Both will be $\mathbf{alt}_1$-operators. We will require that *rest* and *first* be irreflexive, and that if there is a *first*-successor, there must also be a *rest*-successor. (We might also introduce a constant nil to denote the empty list.) The daughters of a node in a tree are then thought of as a list, the first entry being the first daughter, the second entry the second daughter etc. The elements of the list are accessed as follows. We have $down_i = (rest^{i-1}); first$. This way of coding allows to define *down* and *right*, but not *left*, unless we grant ourselves the converse modality *up*. For various reasons we will not do that; first, decidability of the constituent logic of trees is a simple matter without *up*, and secondly, it seems that for linguistic applications we can do without. We will announce here without formal proof the following result.

Theorem 7.1 *CL has the finite model property and is hence decidable.*

The proof is analogous to the proof of the finite model property of **DPDL** by [Ben-Ari *et al.*, 1982], using computation traces. We have opted for excluding the proof from this paper since it is rather long and unrevealing for the present purposes. Notice the following immediate condequences.

Corollary 7.2 *Any extension of **CL** by a finite set of constant formulae has the finite model property and is decidable. In particular, the logic of* *at most n-branching trees and finite constant extensions thereof have these properties.*

Proof. Let X be a finite set of constant formulae. Put $\chi = \bigwedge X$. Now consider $\mathbf{CL} \oplus \chi$. We want to show that this logic has the finite model property. To that end assume $\phi \notin \mathbf{CL} \oplus \chi$. Then a finite countermodel must be found. Now, from the assumption follows $\boxdot^* \chi. \rightarrow .\phi \notin \mathbf{CL} \oplus \chi$, since $\boxdot^* \chi$ is a theorem of this logic. A fortiori, $\boxdot^* \chi. \rightarrow .\phi \notin \mathbf{CL}$, since the latter logic has less theorems. By the previous theorem there is a finite model $\langle \mathfrak{F}, \beta, x \rangle \models \boxdot^* \chi \wedge \neg \phi$ and we can assume that $\mathfrak{F}$ is generated by x. Hence, $\mathfrak{F} \models \mathbf{CL}$, by the fact that $\mathfrak{F}$ is a frame for $\mathbf{CL}$, and $\mathfrak{F} \models \chi$ by the fact that $\langle \mathfrak{F}, \beta, x \rangle \models \boxdot^* \chi$ and (1) the frame is generated by x, so every point is reachable from x by going down. (Remember we have stipulated via non-normal postulates that x has no sisters; but that can be avoided in the proof if necessary.) Thus $\langle \mathfrak{F}, \beta, x \rangle \models \chi$, whence $\langle \mathfrak{F}, \beta \rangle \models \chi$. (2) χ is constant, so it does not depend on the valuation so that $\mathfrak{F} \models \chi$. This proves the first claim. For the second observe that *to be at most n-branching* is expressible by a constant formula namely $\neg \diamondsuit^n \top$, to be added as a normal axiom. $\dashv$

8 Grammatical Theories as extensions of CL

Grammatical theories of any sort can be viewed as axiomatic extensions of the constituent logic **CL** for some fixed set of features. It is impossible to describe in detail all possibilities that this approach offers, but we will give the reader an impression of how large the spectrum is. The first example is the variety of grammars of the GPSG-type. Recall from [Gazdar *et al.*, 1985] that GPSG has two principal components. One component for defining the admissible categories and the other for defining the grammatical rules. The architecture of GPSG is quite close to that of **CL**. In order to translate a specific GPSG-grammar into an extension of **CL** we first introduce for each GPSG-feature a corresponding modality. Thus for the the features CASE, COMP, AGR, ADV, SUBCAT etc. we introduce arc labels with the same name, and each in turn gives rise to a modal operator satisfying $\mathbf{alt}_1$. Furthermore, boolean constants will be introduced for the atomic values of the features, which are, for example, nom, ACC, $+$, $-$, for, that, $\mathsf{whether}$, if, nil. [4] Now we start with the axioms. First, we will have to state that the constants are all *mutually exclusive*. So, if n_i and n_j are constants with $i \neq j$ then an axiom $\mathsf{n}_i \rightarrow \neg \mathsf{n}_j$ is added. Next, the value range of an operator must be defined. If an operator has atomic values only, like CASE, we must add axioms of the form $[\text{CASE}]\neg\langle\text{FEAT}\rangle\top$, for each feature FEAT that is to be excluded. Furthermore, we must state for each atom n_i that is outside the range of CASE the axiom $[\text{CASE}]\neg\mathsf{n}_i$. Alternatively, if the permitted range is $\{\mathsf{acc}, \mathsf{nom}\}$ then the axiom $[\text{CASE}]\mathsf{nom} \vee \mathsf{acc}$ will do.

[4] In order to appreciate this list, it is advisable to look at the appendix of [Gazdar *et al.*, 1985].

In GPSG there are also feature co-occurrence restrictions and feature specification defaults. The former are straightforwardly translatable, since they are axioms. The latter are tricky, because they do not have the status of an axiom. Thus, a priori there is the possibility that defaults cannot be translated. It seems, however, quite plausible that defaults are just a shorthand for an otherwise clumsy formula. A typical case is the so-called *elsewhere*-condition the principle that a more specific rule overrides a less specific one. This is in some sense a meta-axioms because it does not reduce to rules simpiciter, but given a rule system that employs it, this rule system can be rewritten into a rule system that does not. Naturally, the latter will be much more complex in its formulation. I will not say more the reducibility of defaults; in the case at hand I trust that the reader finds it as obvious as I do that they are a shorthand notation.

Next comes the rule component. It is split into *rules* and *metarules*. Again, it is possible to remove the metarules at the price of inflating the rules. After having done this we can translate the rule system as follows. Let $R = \{\rho_1, \ldots, \rho_m\}$ be the totality of rules. Then

$$\gamma = \bigvee_{i=1}^{m} \rho_i^t$$

where each individual rule $\rho = C_0 \to C_1 \ldots C_m$ is translated as

$$\rho^t = C_0^t \to \bigwedge_{i=1}^{m} \langle down_i \rangle C_i^t \wedge \neg \langle down_{m+1} \rangle \top$$

The translation of the categories is straightforward. The given translation is not the most efficient, but it serves the purpose of illustrating our point. There is actually more to GPSG, such as the head-feature convention, the factorization of rules into ID/LP-format etc. By now, however, it should be clear that they all can be rendered more or less as axiomatic extensions. A fast argument is a proof that GPSG allows only a finite number of rules. However, this proof is not so illuminating because unravelling the rule system completely destroys the structure of the system completely. Ideally, however, one wants the internal architecture of GPSG to be preserved as much as possible, so a closer analysis is to be preferred. Let us note that all axioms of GPSG are constant formulae, so GPSG corresponds to a constant extension of GPSG. This is a consequence of a theorem of [Kracht, 1995b] which states that an extension of **CL** for m-branching trees (n some number) is context-free iff it is axiomatizable by constant axioms. This is may not be exciting, but it shows directly that the logic of these grammars is decidable. Moreover, with some effort one can show that for two such grammars it is decidable whether or not they generate the same constituent structures. (See [Kracht, 1993b] for an extensive discussion. There the same result is proved by reducing the context-free grammars to a special normal from preserving the constituent structures.)

At the other end of the spectrum lies GB with its very rich internal structure. GB, however, does fit the logical approach rather well. This is partly due to the fact that the logic **CL** has arisen from studying formalizations of GB. The whole process of destilling a logic from a particular GB-grammar is a rather long and difficult procedure, so will have to be content with a sketch, focussing on some details only. Curious readers might consult [Kracht, 1995b] for an example of formalizing *Relativized Minimality* and [Kracht, 1993b] for an extensive study. First of all, GB is a multistratal theory, unlike GPSG. This leaves two options. The first is to encode in one structure the total derivation by introducing a new modal operator NEXT which mimicks the flow of time. Move-α can be understood as specifying how a tree at point t_i is related to a tree at point t_{i+1}. This can be written down axiomatically. This approach is rather unsatisfactory because it does not reveal much about the empirical consequences of the theory. Much better is the second approach of *stratification.* It consists in replacing BG by an empirically identical monostratal version. [5] It is not difficult to get rid of D-structure, because the derivation can be read off S-structure. The real question, therefore, is the reducibility of LF and PF. Largely, PF is still an unarticulated component, so we may ignore it here. LF, however, cannot be ignored, although [Koster, 1986] has argued that monostratal variants of GB can be produced. He mainly argues that the arguments in favour of LF are not so compelling. I have been convinced especially be [Fiengo and May, 1994] that this is not so. What can be done? We use the following trick. Complete a full derivation from D-structure to LF. Do not delete traces, just mark them by a special feature to make them transparent. (Thus we fudge LF a bit.) In addition, leave the physical element in its original place at S-structure and move a sort of anti-trace, which we call a *shadow.* Thus, we obtain a mixture of S-structure and LF. It is S-structure extended by additional empty categories that mimick the movement of certain items to LF. Shadows will be identifiable by a feature, just as traces and other elements.

If it is clear how GPSG works, the phrase-structure component of GB can be understood with ease. Of course, we need to introduce enough features and values to create the categories. So the basic categories of GB such as cp, $\bar{\mathsf{v}}$, infl^0 and $\mathsf{agr\text{-}o}$ can all be produced from a list of features and values.

[5] It needs to be clarified what counts as *empirically identical.* I propose the following definition. Two grammars are empirically identical if they generate the same bracketed strings alias constituent structures. This notion is somewhat in between weak and strong equivalence. Alternatively, we might also take certain empty categories as if they were part of a string, so that theories are counted different if they distribute PRO differently, everything else being equal. However, Noam Chomsky (p. c.) insists on the reality of the generative process itself, so that two theories must be counted as different when they differ only in the way the structures are being produced. Two remarks on this are in order. First, this kind of definition might be useful in identifying the right theory, but is unhelpful in the present circumstance because it trivializes the notion of equivalence. Notice, that for independent reasons we must be interested in the formal consequences of a theory, thus testing equivalence with respect to other theories in whatever sense is a reasonable thing to do. If the term *empirical* hurts, replace it by another. Second, I know of very few linguistic articles that use data other than alignment facts in languages. Even if most linguists have higher aspirations, they use strings paired with grammatical judgements as (raw) data. Rejecting the word *empirical* in this context is thus simply unfair, since I use the same data as a criterion.

Much of what remains implicit in GB-notation will have to made explicit, for example the θ-grid, case assignments features etc. There will be no indices, however. The θ-grid is syntactically relevant, so it clearly must be reflected in the categorial system. X-bar-syntax is straightforward.

9 Cross Serial Dependencies

A particularly challenging construction in natural language is that of cross-serial dependencies. Languages in which they occur are difficult to handle with many syntactic formalisms. Two rather interesting solutions will be discussed here and we will see how the modal constituent language **CL** can be put to use. The syntactic problem with them is that they lead to what are known as *copy languages*. These are languages of the form ww, w a word over a finite alphabet. These languages are known not to be generable by context-free grammars. In the present context, the argumentation must be careful. If we are interested only in syntactic acceptability, it is not immediately clear that cross serial dependencies lead to copy languages. In essence, they may on a very simple account only lead to languages of the form $\{a^n b^n | n \in \omega\}$, which are indeed context-free. The grammar generating this language will not reflect the underlying associations between the elements; however, since the associations are not part of the representation, the difference will not show up. If, on the other hand, the associations do exert an influence (e. g. when we have a selectional restriction holding between associated items), then we can indeed detect it. In the latter case there is non-context-freeness. In the former case we note that a context-free grammar even when it generates the right set of string will never generate the right kind of constituent structures. For the type of construction discussed below, the syntactic structure is in fact rather complex. We will not question it here and follow [Johnson, 1988]. Recall that nested sentences in English are syntactically unproblematic; the subordinate clause begins after the sequence of nominal complements is complete, e. g. in *I told Mary not to let Bill help Peter.*. In German, the subordinate clauses can also be center embedded.

…, [*daß ich*$_1$ [*Maria*$_2$ [*den Kindern*$_3$ *aufzuräumen*$_3$] *helfen*$_2$] *verboten habe*$_1$].
…, that I forbade Mary to help the kids clear up.

In Dutch, finally, the dependencies cross. We do not have the sequence 1-2-3-3-2-1 but 1-2-3-1-2-3. A particularly striking sentence is the following one.

…, *dat Jan*$_1$ *Piet*$_2$ *Marie*$_3$ *de kinderen*$_4$ *zag*$_1$ *helpen*$_2$ *laten*$_3$ *zwemmen*$_4$
… that Jan saw Piet let Mary help the kids swim.

A particularly simple analysis is provided in GB. We will explain that solution here in an extremely simplified fashion, in order to be able to concentrate on the case at hand. We assume for Dutch and German the same

D-structure. However, for independent reasons we assume that while for German we may let the verb remain in place until S-structure the verb has to be raised in Dutch and adjoined to the next higher verb; when it adjoins it will swap places with that verb. Thus we have the following rudimentary grammar with two rules for deriving the D-structure [6]

$$\mathsf{s} \to \mathsf{np}\ \mathsf{v} \qquad\qquad \mathsf{s} \to \mathsf{np}\ \mathsf{s}\ \mathsf{v}$$

In addition the movement process generates the following S-structure

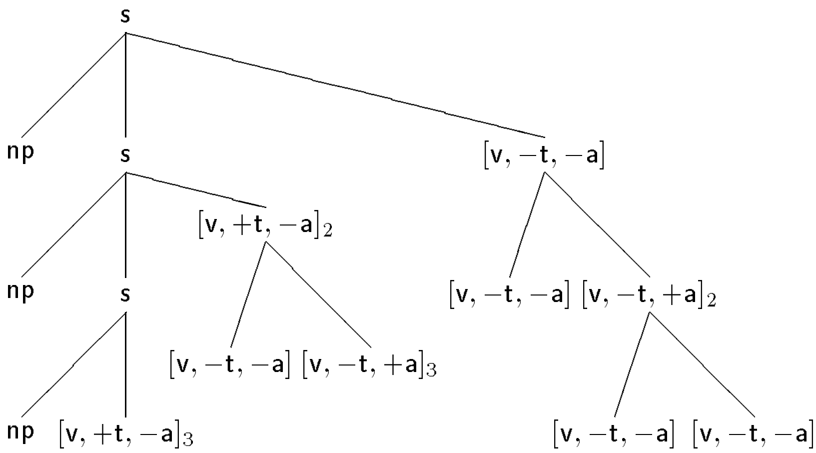
$$\begin{array}{lll} (i) & [\mathsf{np}_1\ [\mathsf{np}_2\ [\mathsf{np}_3\ \mathsf{v}_3]\ \mathsf{v}_2]\ \mathsf{v}_1] & \leadsto \\ (ii) & [\mathsf{np}_1\ [\mathsf{np}_2\ [\mathsf{np}_3\ t_3]\ [\mathsf{v}_2\ \mathsf{v}_3]\ \mathsf{v}_1] & \leadsto \\ (iii) & [\mathsf{np}_1\ [\mathsf{np}_2\ [\mathsf{np}_3\ t_3]\ t_2]\ [\mathsf{v}_1\ [\mathsf{v}_2\ \mathsf{v}_3]]] & \end{array}$$

Each individual process follows a local pattern. The verbal constituent dominated by **s** is adjoined on the right of the verb sister to **s** and leaves behind a (coindexed) trace. Let us see whether we can approach this analysis with an axiomatic constituent theory by reducing GB to a single stratum, S-structure. It seems at first look impossible to redo this analysis using the modal constituent language. The simple reason is that each time the verbal complex is raised we have to check for the admissibility of the movement, that is, we check whether the trace left behind is bound. However, in the above derivation we can see that in the structure (iii) t_3 is not bound. It need not be, because it has been created by moving from (i) and (ii). Its legitimacy has been checked already. This is the advantage of the derivational component of GB. If we want to write a representational account of the same construction, we need to remember the justification for the traces. We do this by using a simple trick. Rather than using unstructured traces we will let the trace have the shape of its antecedent. In particular, it will be *completely isomorphic* with an important exception. We need two boolean features, **ant** and **trc**, roughly equivalent to *being an antecedent* and *being a trace*. Now the root node will be labelled **trc** in the (feature structure of the) trace while the root node of the antecedent will be labelled **ant**. To let no misunderstandings arise, **trc** does not exactly identify the traces. Any node which is itself **trc** or directly or indirectly dominated by a **trc**-node will be phonetically empty and be thought of as part of one and the same trace in the original sense of the word. In effect, then, the modification suggested here will make no difference in the output string nor in its physical (acoustic/grapic) constituent structure. We have just articulated the internal structure of a trace somewhat more. Let us see how the derivation from

[6] This grammar just serves the purpose of illustration. No further claims are being made here. In particular, this grammar does not conform to X-bar syntax. Moreover, it is not implied here that a similar verb raising for German is excluded. We can perform the same analysis as with Dutch with the difference that the verb adjoins to the left, thus recreating the same order. This is on the one hand banned in GB on the grounds that such movement is string vacuous, i. e. reproduces the original aligment. On the other hand, there might be syntactic facts pointing to a constituent structure similar to Dutch. As an additional note observe that the mere notion of *string vacuousness* as opposed to *constituent structure vacuousness* seems dubious if we are interested in structural description. (See also the footnote on *empirically identical theories*.

D- to S-structure looks like.

$$
\begin{array}{lll}
(i) & [\mathsf{np}_1\ [\mathsf{np}_2\ [\mathsf{np}_3\ \mathsf{v}_3]\ \mathsf{v}_2]\mathsf{v}_1] & \rightsquigarrow \\
(ii) & [\mathsf{np}_1\ [\mathsf{np}_2\ [\mathsf{np}_3\ \mathsf{v}_3]\ [\mathsf{v}_3\ \mathsf{v}_2]\ \mathsf{v}_1] & \rightsquigarrow \\
(iii) & [\mathsf{np}_1\ [\mathsf{np}_2\ [\mathsf{np}_3\ \mathsf{v}_3]\ [\mathsf{v}_2\ \mathsf{v}_3]\ [\mathsf{v}_1\ [\mathsf{v}_2\ \mathsf{v}_3]]] &
\end{array}
$$



The picture shows the fully articulated S-structure. $\pm\mathsf{t}$ means that the element is/is not trc and $\pm\mathsf{a}$ means that the element is/is not ant. All nonverbal nodes are $-\mathsf{t}, -\mathsf{a}$. Notice that all four combinations are legitimate. In a three element chain $\langle t_1, t_2, \mathsf{xp}\rangle$ the first element is $+\mathsf{t}, -\mathsf{a}$, the middle element $+\mathsf{t}, +\mathsf{a}$ and the head of the chain is $-\mathsf{t}, +\mathsf{a}$. The combination $-\mathsf{t}, -\mathsf{a}$ represents unmoved elements, or one-element chains. Head-movement chains are two-membered, so the combination $+\mathsf{t}, +\mathsf{a}$ – characteristic of intermediate traces – does not appear here. The indices are not part of the representation, they are just helpful devices for referring to constituents. Now the locality restriction on trace-antecedent pairs can be monitored at any subsequent level. We will show how. Let us agree to call a *trace* a constituent being labelled trc such that it is not properly dominated by any constituent which is also trc. We introduce the constant trace denoting traces. Thus, traces in the old sense are constituents whose root satisfies

$$\mathsf{trace} := \mathsf{trc} \wedge \boxdot^{+} - \mathsf{trc}$$

We are now in a position to write down an *axiom* defining that a tree has proper trace-antecedent relations for head-movement of the illustrated kind. This axiom needs to be satisfied for a S-structure to be correctly derivable from a D-structure by head-movement.

$$\text{V-TRACE} : \mathsf{s} \wedge \langle down; \mathsf{v}?\rangle(\mathsf{trace} \wedge \langle down\rangle p. \to .\langle right; down\rangle \mathsf{v} \wedge \mathsf{ant} \wedge \langle down\rangle p$$

This axiom is not constant. It contains the variable p, standing here for the open constituent dominated by the v-node. The content of the axiom is that if there is a s node with a v-daughter which is a trace, then s is sister to a v dominating a constituent which is isomorphic to the trace with the exception of the trace/antecedent labelling. Given the fact that adjunction can be iterated any number of steps we do not know how large the constituent is that is being moved. Consequently, we must use a variable to stand for the (open) constituent dominated by the trace.

To get the full picture, we also need to revise the grammar, since we want to generate S-structures directly. Thus write the following grammar.

$$\mathsf{s} \rightarrow \mathsf{np}\ \mathsf{v} \qquad \mathsf{s} \rightarrow \mathsf{np}\ \mathsf{s}\ \mathsf{v} \qquad \mathsf{v} \rightarrow \mathsf{v}\ \mathsf{v}$$

We have seen how to convert this such a grammar into an axiom for **CL**. This grammar will overgenerate. Hence we need to restrict it in some ways. Among some minor adjustments (to do with the fact that we cannot afford base generated adjunction) we need to add the axiom V-TRACE. Thus, with cross-serial dependencies we have an example of a syntactic analysis leading to a rather nontrivial modal logic of its trees. This logic is not a constant extension, since it would then yield a strongly context-free language, which we have shown to be false.

It is interesting to compare this with a different approach, namely LFG. The fact that LFG can come to grips with cross serial dependencies shows that its internal make-up must be more complex than that of GPSG. The idea is that in addition to an ordinary syntactic tree (called C-structure) another structure is created hand in hand with the rules, a structure that reflects the argument structure of the sentence. This structure is called F-structure. Each node x in the syntactic tree will get a unique associated node fx in the F-structure. It may be that two syntactic nodes will be associated with one and the same node. Well-formedness conditions therefore operate not only on syntactic trees but also on the accompanying functional analysis. We will explain this using the formalism of attribute-value grammars in [Johnson, 1988]. Recall that LFG has rules of the following kind.

$$\begin{array}{cccc} \mathsf{s} & \rightarrow & \mathsf{np} & \mathsf{vp} \\ & & (\uparrow \textsc{sub}) = \downarrow & \uparrow = \downarrow \end{array}$$

This means the following. Suppose we are at node x in the C-structure alias syntactic tree; x is associated with fx is the F-structure. Suppose that x is expanded according to the above rule. This means that x dominates y and z, y a np and z a vp, that the subject of fx is fy and that $fy = fz$. LFG can handle cross-serial dependencies because it generates in addition to the syntactic tree the comparatively speaking shallow F-structure, which allows to encode the cross-serial dependencies in a simple way. We will not present the specific solution. Rather, I will seize the opportunity to show how the framework itself can be translated into **CL**. First of all, we will drop the independent structure and treat the relations SUB, OBJ etc. as relations over the syntactic tree. To make this possible we must recreate the shallow analysis. This we do by introducing an equivalence relation $\approx$ on the nodes. The intention is that we want to put $x \approx y$ in a syntactic tree whenever $fx = fy$ in the F-structure. So, in the case of the rule above we will say that the second row expresses the following. x dominates a np y and a vp z; moreover, y is the subject of x and $z \approx x$. To use an equivalence relation other than identity only has the disadvantage that the relation *object of* is *not* identical with OBJ. [7] Rather, a node u is object of v if u is equivalent (via $\approx$) to a node w which is related via OBJ to z and $z \approx v$. In other words, *object of* is the relational composition $\approx \circ \textsc{obj} \circ \approx$.

To translate this into **CL** assume a modal operator for each basic functional relation, such as SUB and OBJ, and assume a separate operator $\boxtimes$ based on $\approx$. In dynamic logic notation, $\boxtimes$ is the same as $[\approx]$.

$$p \wedge \mathsf{s}. \wedge .\langle down_1\rangle(\mathsf{np} \wedge q) \wedge \langle down_2\rangle(\mathsf{vp} \wedge r) \wedge \neg\langle down_3\rangle\top. \rightarrow .\langle\approx \circ \textsc{obj} \circ \approx\rangle q \wedge \langle\approx\rangle r$$

By the correspondence between modal formulae and first-order statements it can be shown that the above formula defines that the admissible local trees are those defined by the (unique) LFG rule we have given above. Of course, a grammar has several such rules, but we know that a disjunction of first-order axioms is again first-order, and the formula corresponding to this disjunction can be computed. In this way any LFG grammar can be effectively reduced to a modal logic. Moreover, it will be an extension of polymodal $\mathbf{K.Alt}_1$ unless we admit functional uncertainty.

10 On alternatives to features

We have seen that there is a division between tree structure and feature structure. Roughly (though not always agreed on) the typical feature structures have bounded size, so that there is a bounded number of them. This suggests that rather than viewing them as attribute-value structures we might actually use different apporaches, for example a straight boolean approach, without any modal operators. In this section we will show that this is a justifiable option, even though the arguments for and against will be balanced. We discuss this using concrete examples.

INDO-EUROPEAN CASE MARKING. Consider the arc modality $\langle\textsc{case}\rangle$ in Indo-European languages. We write $\langle\textsc{case}\rangle\mathsf{acc}$ to state that the object under consideration has *accusative case*. The syntax of the modal language allows to repeat the operator and create the following statements.

$$\langle\textsc{case}\rangle\langle\textsc{case}\rangle\mathsf{acc}$$
$$\langle\textsc{case}\rangle\mathsf{dat}. \wedge .\langle\textsc{case}\rangle\langle\textsc{case}\rangle\mathsf{acc}$$

In Indo-European languages, case is a simple property of nouns and adjectives, there is no sense in iterating the operator $\langle\textsc{case}\rangle$. Thus, for such languages (constituting the majority of languages in fact) this has to be ruled out. Two alternatives are plausible. One fix is to forbid the reapplication of

[7] It wouldn't be in LFG either since we are literally speaking not talking about the same node when we talk about fx instead of x.

the operator. So we postulate the axiom $[\mathrm{CASE}][\mathrm{CASE}]\bot$. The other would be to postulate that the reapplication must receive the same interpretation, i. e. we add the postulate $\langle\mathrm{CASE}\rangle\langle\mathrm{CASE}\rangle p. \leftrightarrow .\langle\mathrm{CASE}\rangle p$. (Or, equivalently, the path equation $\mathrm{CASE} : \mathrm{CASE} \approx \mathrm{CASE}$.) Notice that the latter postulate is weaker, because in addition to the irreflexive singleton (a) it allows the reflexive singleton (b) as a model structure.

Of course, this condition has to hold everywhere, so the postulate is added normally. Both approaches, however, do have their problems. Up to now it is still conceivable that the atom acc is assigned at the root and the $\langle\mathrm{CASE}\rangle$ successor. So, the formula $\mathbf{nom} \wedge \langle\mathrm{CASE}\rangle\mathsf{acc}$ still needs ruling out. We can do this by requesting that acc or any other atom for cases be false at the root of feature structures for constituents, and can only be true if prefixed by $\langle\mathrm{CASE}\rangle$. This sounds like a difficult thing to do, but is possible. But let us let us reflect instead on what constitutes a reason for using feature structures at all.

Namely, I think even though it is optically pleasing to regard CASE as intrinsically relational, the initial attraction results from a superficial linguistic analogy of $\langle\mathrm{CASE}\rangle\mathsf{acc}$ with *the case of this item is accusative*, thus suggesting that there is an underlying relation. Yet, the expression *my car is red* does not indicate that the expression *car* is relational, neither does *this item's case is accusative* lead to the conclusion that *case* denotes a relation. There are only cars and in particular red ones, and there are cases, in particular accusative. So it seems that *case* is a group or *sort* with certain subsorts, one of which is *accusative*. This can be implemented as follows. Simply add a boolean constant case in addition to the constants for all cases rather than a feature, and write down a disjunction of the following kind.

$$\mathsf{case}. \leftrightarrow .\mathsf{nom} \vee \mathsf{gen} \vee \mathsf{dat} \vee \mathsf{acc} \vee \mathsf{abl} \vee \mathsf{inst} \vee \mathsf{loc}$$

(This reflects the original Indo-European case system. For arguments that vocative is not a case see [Blake, 1994].) Hence it seems that the formalization of *case* should involve a plain boolean constant, not a relation. This is in line with our intuition that case is a property of $[+\mathsf{N}]$-phrases.

There is actually an argument in favour of a feature analysis stemming from feature sharing. Suppose, we need to say that two different nodes share a property, say their case. Let us for simplicity assume that the nodes are mother and daughter, though any other structural relation between them will do. The case at hand is a specific instance of the head-feature convention, which requires head-features of the mother be passed on to the head daughter. Whatever the general principle is, suppose we want to state the fact that mother and head daughter agree in case. With attribute-value formalisms this is easy.

$$\langle \textsc{case} \rangle p. \rightarrow .[down; \mathsf{head}?; \textsc{case}]p$$

In conjunction with the system of axioms for GPSG, this will ensure exactly what we want. Recall, namely, that we have axioms stating that although p can stand for any partial description of a generated substructure (in the modal sense), the structure that is being generated is the single world at which we stand, and there only case-atoms can be true. So p stands for information about case, nothing more. If we instantiate p to the case of the mother node, then whichever way we go down to the head (which is a unique daughter) and look into the case feature, it too must be p, and this enforces the identity. The same effect can be achieved by using typed variables. However, in with booleans we need to state seven different axioms, each instantiating a distinct case. For example, sharing nominative must be expressed by

$$\mathsf{nom}. \rightarrow .[down; \mathsf{head}?]\mathsf{nom}$$

There is no way to simplify this set of axioms other than using typed variables or a CASE-feature. Thus, it is not possible to write down certain general laws.

CASE MARKING IN KAYARDILD AND GEORGIAN. [Blake, 1994] reports the case of a language in which several cases are marked on a single lexical item. The example sentence is the following

Maku-ntha	*yulawu-jarra-ntha*	*yakuri-naa-ntha*
woman-OBL	catch-PAST-OBL	fish-MABL-OBL
dngka-karra-nguni-naa-ntha		*mijil-nguni-naa-ntha*
man-GEN-INST-MABL-OBL		net-INST-ABL-OBL

'The woman must have caught fish with the man's net.'

Here, OBL stands for *oblique* and MABL for *modal ablative*. It seems legitimate to consider Kayardild as a language in which case can be arbitrarily nested. However, this has to be analysed with great care. [Blake, 1994] speaks of several layers of case marking. There is adnominal case, adverbal case, and two outer layers of what are etymologically case markers. It is not clear that any of the layers allows for iteration. In fact, it is stated that there are something like two adverbal cases. In Georgian, and some other languages, there is double case marking in that there is independent adnominal case marking (genitive case) and adverbal case. Consider the following example from Old Georgian.

sarel-ita	man-isa-jta
name-INST	father-sc gen-INST

with father's name

The instrumental case is realized on the dependent noun despite the fact that it bears genitive case already. [8] In many languages that do exhibit double case marking we find only this constellation, that adverbal case may be added in addition to adnominal case. One might speak here of *independent possessor marking*. Georgian has the additional twist that the noun phrase is possessor marked only when it follows the governing noun phrase. It is not clear whether possessor marking can be iterated.

Agreement in Indo-European languages. A similar story can be told about agreement, in this case between verb and noun phrase. Indo-European languages have nominative-accusative-agreement systems, and there is agreement between the subject (in nominative case) and the verb. They agree in person and number. In [Gazdar *et al.*, 1985], agreement is implemented by using a category valued feature AGR, which is mediated between subject and verb. Again, a closer analysis (see [Kracht, 1993b]) shows that, technically speaking, no such feature is needed and we can use the percolation mechanisms of GPSG to do the job. The principal line of argumentation is that it is not necessary to use a property *agrees in number x and person y* in addition to *has number x and person y*, since there will be no cases in which an item has different number or person than the one which it will be required to agree (if it has to agree with anything at all). There are intermediate categories of which it may make little sense to say that they are singular or plural, and that is why the argument is technical in nature. Nevertheless it points at an information overload in the system making the category inventory unnecessarily large.

Agreement in Basque. There are languages in which the verb agrees with two, sometimes three argument. Basque is such a language. Following [Laka, 1991], Basque displays agreement with subject, object and indirect object, corresponding to three cases, ergative, dative and absolutive. (Basque is an ergative language.) Unlike the Indo-European system, we cannot simply pass on the agreement features of the arguments as properties of the intermediate categories. They may be inconsistent. In a sentence like *I give the boys two books.* the subject is singular, while the object is plural. So if the subject passes on its number feature, sing, and the object passes on its number feature, plu, they will both arrive at the verb and clash. Moreover, the verb must be able to identify each argument as to where it comes from. So, GPSG has in fact given a better analysis – instead of introducing an single AGR, we just need to correct GPSG by introducing three agreement features. As with case, it is not clear what the upper limit is, at least there is no a priori plausible number of arguments with which a verb can agree in a natural language. It seems that Basque represents an extreme, but this has to bet left open. If that is so, however, we might ask whether it is better not to introduce distinct agreement features but simply allow AGR to be iterated. So, effectively we keep a list of agreement facts, and let additional axioms

[8] Thus, a typical interventionist approach to case à la Barriers as proposed in [Chomsky, 1986] and elaborated in [Fanselow, 1991] will not work for languages like Georgian (and Kayardild).

regulate the length of such a list. (For example, add $[\text{AGR}^4]\perp$ for Basque.)

So we see that there is a non-trivial question of what is the right language within which we formalize, and which is the correct logic of let us say *sensible* attribute value structures under a particular interpretation. Let me close by pointing at an alternative approach which saves the finiteness of the feature component of syntactic categories. We may analyze the complex words of Kayardild as syntactically complex objects, treating each case ending as a different terminal node in the structure. This is similar to the GB analysis of morphology, e. g. that of [Baker, 1988], in which the morphological complexity is due to head-movement. However, while in Baker's own analysis the adjoining head and the adjoined head occupy different terminal nodes, [Chomsky, 1993] assumes that lexical items are not spread over several terminal nodes. Instead, the items are considered to be tagged onto the tree in the process of derivation as fully derived and/or inflected and the syntactic derivation that ensues will only serve to identify (or check, as it were) the features of the lexeme. From an explanatory point of view, Baker's theory has the advantage. [9] However, it depends also on the exact demarcation line between syntax and morphology whether one wants to take side with one or the other. In the present context, Baker's analysis has the advantage to leave the number of syntactic categories finite and relegate the apparent infinity of them to the syntactic structure. Yet, Baker can obviously explain with his approach only part of the morphological variety, and he does not tell us what happens in the cases discussed above. They do not, in his view, constitute cases of incorporation. Nevertheless, should it turn out that case is iterable without bound, it seems plausible to give each case ending a separate terminal node, albeit not derived via head-movement. The necessity of such a strategy depends (in my view) on questions of iterability of case marking. Nevertheless, an analogical treatment of agreement will not be so straightforward. Considering the multiple agreement system of Potawatomi or the Georgian proclitics (see [Halle and Marantz, 1993]) we see that the agreement information surrounds the verb in rather unusual fashion suggesting that the approach in terms of different nodes will be tantamount to reproducing the morphological analysis in the syntax. In addition, this irregular distribution of agreement information makes it plausible that languages prefer the verb to agree with only a small number of arguments. To analyse in what ways it does this, is a fascinating study, but feature logic will remain completely neutral in this affair.

11 Genuine contributions of modal logic

From a formal point of view, anything goes. We can use Kasper and Rounds' language, predicate logic, partial semigroups – and modal logic. From an

[9] The differences between these views are discussed in an appendix to [Halle and Marantz, 1993].

intuitive point of view these approaches have their own strengths and weaknesses. Yet, if we have to choose between all of these ways of talking about AVMs and linguistic strictures such as trees, which one is to be preferred? I wish to argue contra [Johnson, 1991] and others – as did P. Blackburn – that modal logic is the best candidate. But my reasons are quite different. Initially, modal logic was used there (as generally in linguistics) as a mere language of analysis, for the simple reason that modal logic – by the fact that it is a well developed logical discipline – allowed to state the problem in simple terms and proceed to its solution without worrying too much about the surrounding technical appratus. There is no sufficiently general result on the decidability of polymodal logics in general and logics for AV structures in particular. It goes without saying that this has nothing to do with the fact that we used modal logic here; the same holds for all sufficiently expressive logics for AVMs. On the other hand, due to advances in correspondence theory it was possible to characterize the strength of modal axioms and show that all existing languages for AVMs can be interpreted quite comfortably as modal languages. Here lies what I consider a good argument in favour of modal logic. For modal logic has developed all necessary classificatory tools just as it has for intermediate logics. First, there is a rough classification according to the power of the language. We have polymodal logic, test free propositional dynamic logic etc. There is no need for extravagant devices and even less need for constantly changing notation. Secondly, there are well-known theories about lattices of logics allowing a global view on the spectrum of possible logics. Moreover, the specific advantage is that modal uses the relations in an essential way, more intrinsically at least as predicate logic.

It has been thought, however, that modal logic is nevertheless too weak and has to be boosted up by extra operators to make it useful. I think, in that respect the introduction of *nominals* has been a mistake that damaged the image of modal logic more than it helped to promote its case. What, the attentive observer asks, is the purpose of modal logic if we need to add what effectively reduces to first-order variables? And what sort of theorems can I apply from nominal modal logics? There are none of significance. Nominals have been a desparate attempt to guarantee the *token identity* of those nodes marked for reentrancy. HPSG makes a point in requiring that the marked nodes be really identical, not just that they agree in the structure that is tagged onto them (= that they generate in the modal sense). We have seen that the postulate

$$\langle x\rangle p. \leftrightarrow .\langle y\rangle p$$

added as a quasi-normal postulate enforces identity on differentiated frames, but not necessarily on all generalized frames. But so what? There is no need to assume that there is something that is actually moved in our heads by **move**-α in order to understand GB. [Koster, 1986] has shown that we can do without. And there is likewise no need to assume real identity in reentrancy. Of course, HPSG needs that in order to explain *why* structure sharing happens, but this is really only a superficial explanation in view of the fact that the nodes need not correspond to anything in the world; they are – as explained above – quite often artefacts of the visual representations. Tagging AVMs is a nice metaphor but is under threat by Ockham's razor. Above all, however, HPSG prides itself with an information based approach. And the modal postulate above expresses the second-order condition

$$(\forall p)(r \in \langle x\rangle p. \leftrightarrow .\langle y\rangle p)$$

which satisfies that every p true at the x-successor of the root is also true at the y-successor of r. In other words, these two successors *share all information.* What more than that do we need?

Well, there actually *is* more. Such apparatus is namely only needed if the amount of information that is shared across or between nodes is a priori unlimited. By the fact that HPSG allows in principle unbounded nestings, this case may theoretically arise and so the notion of structure sharing rather than information sharing becomes vital. It is up now unclear whether we actually need an unbounded amount of information. In [Kracht, 1993b] I investigated this question rather closely. The argumentation is complex but can be summarized as follows. Standard locality restrictions between moved phrases and traces, subjacency, etc. are all expressible in **CL**, as we have seen. Then if the number of intersecting dependency paths is bounded a priori (which is commonly the case except for marginal cases leading to non-context free languages such as Swiss German), the amount of information that is to be stored at a single node is finite, and there is no need for infinite AVMs, provided we are not using AVMs to code the structure of the entire tree.

The difference between the approach I have sketched above and the typical approach in constraint based grammars cannot be overestimated. If, namely, we use feature structures on the basis of the tree then the relational interpretation of features makes real sense; sisters in a tree are different objects, a mother node is really different from its daughter. [10] Different in a sense that the value of the attribute CASE cannot be, because the value of a feature is a property of the same object to which we apply the feature. The graphical representation of knowledge about structure – which is no doubt successful – has created the impression that there is a traceable difference between the nodes of that representation and that it makes no difference whether we talk about a feature EMPLOYER or about a feature COLOUR. But reality gets us all in the end. [Carpenter, 1992] quotes an example of Carl Pollard based on employers and employees where we would rather not make the AVM differentiated – that is collapse it as much as possible – for reasons that are highly unsurprising.

[10] This can be debated if empty categories are admitted in the grammar, because it is not straightforward to argue for a *physical* difference between different nodes in an analysis tree. For example, [V *t*] is phonetically identical to V. But if we use the constituent structure as the underlying kripke frame, there is really no doubt about this.

12 Conclusion

One can adduce more arguments to show that for logics of AVMs modal logic is the most appropriate candidate. I am thinking here in terms of the specific results that are folklore in modal logic and have been reinvented in the context of AVMs, such as p-morphisms, generated substructures, unravellings, contractions, differentiated frames etc. Obviously, the use of modal logic would have allowed to concentrate on real issues instead. One caveat, however, should be made. [Carpenter, 1992] does not use a boolean logic of types, so his notions are somewhat different, but the spirit is the same.

The outsider to modal logic must, however, get the impression that there is no such thing as modal logic, only a bewildering multitude of languages, with or without nominals, with or without masters, etc. Indeed, there is no real consensus on what counts as a language of modal logic. Such a consensus need also not exist – there is a variety of languages for AVMs, some notational variants of the other, some slightly different, others markedly stronger or weaker. In this essay I have tried to show that to some extent it makes no difference whether we use 1^{st}-order logic or modal logic, and many equivalences with other calculi follow from there. However, as we should strive for simplicity, it is good to consider the simplest language that allows to express all that we need to express. In that respect the modal logics extending $\bigotimes \mathbf{K.Alt}_1$ (or in fact the constituent language) are the best candidates. First, they are stronger as Kasper Rounds' type logics because of the possibility to forbid path existence. Second, they are weaker as logics with nominals and related systems because the non-identity of paths cannot be defined. I do not know of any compelling argument that it should be definable.

The use of ordinary modal logic without masters and without nominals and other devices has a price. Now we must be interested not in a single logic for which we can study the satisfiability problem but in an infinite number of such logics. This seems to be very dangerous situation. However, notice first of all that there is no real difference in studying a spectrum of logics rather than a single one. We have seen, for example, that reentrancy postulates can be modelled either by (non)normal axioms in our polymodal logic or by formulae containing nominals. So in the latter case we study the satisfiability problem in the base logic while in the former we study the derivability in an axiomatic extension of the base logic. The original problem remains the same – so whatever solution is produced using one method will give a solution for the other. To go to the extreme, everything can be understood as a problem to satisfy a certain $\mathcal{L}^e$-formula. The twist is that $\mathcal{L}^e$ is so vastly more complex that we see immediately the hopelessness of this translation. So, by the very fact that our approach via axiomatic extensions of polymodal logic allows to state less problems than the one with nominals shows that the former has an advantage over the latter.

In addition to the explicit commitment to studying axiomatic extensions of $\bigotimes \mathbf{K.Alt}_1$ or the constituent logic, the other approaches have an implicit commitment to studying extensions of their base logic. We have shown, namely, that the base logic does not filter out those structures that are meaningless in the intended interpretation, or simply inadmissible for empirical reasons. Such reasons can be universal facts of language. To remedy this situation we need to pass to an axiomatic extension that has exactly those structures that are admissible; admissibility of course is an intentional notion here. The special problem that arises is that while typical descriptions of frames (or AVMs) specify local properties corresponding to quasinormal extensions, the admissibility conditions arising from the base logic are typically global, hence correspond to normal extensions. For example, the appropriateness and well-typing conditions of [Carpenter, 1992] are global. So even if the literature suggests the contrary, the study of the satisfiability problem for the base logic is of little theoretical significance. Rather, we have to be interested in the whole lattice of finitely axiomatizable extensions. However, we have seen that even the lattice of normal extensions of $\mathbf{K.Alt}_1 \otimes \mathbf{K.Alt}_1$ is so vastly complex that no reasonable results of sufficient generality can be expected. This is the situation as it presents itself throughout the literature and if this line of theoretical research is pursued, we just have to get used to this fact.

References

[Baader *et al.*, 1993] Franz Baader, Hans-Jürgen Bürckert, Bernhard Nebel, Werner Nutt, and Gert Smolka. On the expressivity of feature logics with negation, functional uncertainty and sort equations. *Journal of Language, Logic and Information*, 2:1 – 18, 1993.

[Baker, 1988] Mark C. Baker. *Incorporation. A theory of grammatical function changing.* Chicago University Press, 1988.

[Bellissima, 1988] Fabio Bellissima. On the lattice of extensions of the modal logics $K.Alt_n$. *Arch. Math. Logic*, 27:107 – 114, 1988.

[Ben-Ari *et al.*, 1982] M. Ben-Ari, J. I. Halpern, and A. Pnueli. Deterministic propositional dynamic logic; finite models, complexity and completeness. *Journal of Computer and System Sciences*, 25:402 – 417, 1982.

[Blackburn and Spaan, 1993a] Patrick Blackburn and Edith Spaan. Decidability and undecidability in stand-alone feature logics. In *Proceedings of the 6^{th} Conference of the EACL*, 1993.

[Blackburn and Spaan, 1993b] Patrick Blackburn and Edith Spaan. A modal perspective on the computational complexity of attribute value grammar. *Journal of Language, Logic and Information*, 2:129 – 169, 1993.

[Blackburn *et al.*, 1993] Patrick Blackburn, Claire Gardent, and Wilfried Meyer-Viol. Talking about trees. In *Proceedings of the* 6^{th} *Conference of the EACL*, 1993.

[Blackburn, 1993] Patrick Blackburn. Modal logic and attribute value structures. In Maarten de Rijke, editor, *Diamonds and Defaults*, Synthese No. 229, pages 19 –65. Kluwer, 1993.

[Blake, 1994] Barry J. Blake. *Case*. Cambridge Textbooks in Linguistics. Cambridge University Press, 1994.

[Bull and Segerberg, 1984] Robert A. Bull and Krister Segerberg. Basic modal logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 1 – 88. Reidel, 1984.

[Carpenter, 1992] Bob Carpenter. *The Logic of Typed Feature Structures*. Cambridge Tracts in Theoretical Computer Science 32. Cambridge University Press, 1992.

[Chagrova, 1991] L. A. Chagrova. Undecidable problems in correspondency theory. *Journal of Symbolic Logic*, 56:1261 – 1272, 1991.

[Chomsky, 1986] Noam Chomsky. *Barriers*. MIT Press, Cambrigde (Mass.), 1986.

[Chomsky, 1993] Noam Chomsky. A minimalist program for linguistic theory. In K. Hale and Keyser S. J., editors, *The View from Building 20: Essays in Honour of Sylvain Bromberger*, pages 1 – 52. MIT Press, 1993.

[Fanselow, 1991] Gisbert Fanselow. *Minimale Syntax*. Number 32 in Groninger Arbeiten zur germanistischen Linguistik. Rijksuniversiteit Groningen, 1991.

[Fiengo and May, 1994] Robert Fiengo and Robert May. *Indices and Identity*. Number 24 in Linguistics Inquiry Monographs. MIT Press, 1994.

[Fine, 1975a] Kit Fine. Normal forms in modal logic. *Notre Dame Journal of Formal Logic*, 16:31 – 42, 1975.

[Fine, 1975b] Kit Fine. Some connections between elementary and modal logic. In Stig Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 15 – 31. North-Holland, 1975.

[Fine, 1985] Kit Fine. Logics containing K4, Part II. *Journal of Symbolic Logic*, 50:619 – 651, 1985.

[Gazdar *et al.*, 1985] Gerald Gazdar, Ewan Klein, Geoffrey Pullum, and Ivan Sag. *Generalized Phrase Structure Grammar*. Blackwell, Oxford, 1985.

[Gazdar *et al.*, 1988] G. Gazdar, G. Pullum, R. Carpenter, T. Hukari, and R. Levine. Category structures. *Computational Linguistics*, 14:1 – 19, 1988.

[Goldblatt, 1987] Robert I. Goldblatt. *Logics of Time and Computation.* Number 7 in CSLI Lecture Notes. CSLI, 1987.

[Goranko and Passy, 1992] Valentin Goranko and Solomon Passy. Using the universal modality: Gains and Questions. *Journal of Logic and Computation*, 2:5 – 30, 1992.

[Grefe, 1994] Carsten Grefe. Modale Logiken funktionaler Frames. Master's thesis, Dept. of Mathematics, FU Berlin, 1994.

[Halle and Marantz, 1993] Morris Halle and Alec Marantz. Distributed morphology and the pieces of inflection. In K. Hale and Keyser S. J., editors, *The View from Building 20: Essays in Honour of Sylvain Bromberger*, pages 111 –176. MIT Press, 1993.

[Johnson and Moss, 1994] David E. Johnson and Lawrence S. Moss. An overview of stratified feature grammar. In Carlos Martín-Vide, editor, *Current Issues in Mathematical Linguistics*, number 56 in North-Holland Linguistic Series, pages 103 – 120. North-Holland, 1994.

[Johnson, 1988] Mark Johnson. *Attribute-Value Logic and the Theory of Grammar*, volume 16 of *CSLI Lecture Notes.* CSLI, 1988.

[Johnson, 1991] Mark Johnson. Logic and feature structures. In *Proceedings of the* 12^{th} *IJCAI*, pages 992 – 996, 1991.

[Kasper and Rounds, 1990] Robert T. Kasper and William C. Rounds. The logic of unification in grammar. *Linguistics and Philosophy*, 13:35 – 58, 1990.

[Kayne, 1983] Richard Kayne. *Connectedness and Binary Branching.* Foris, Dordrecht, 1983.

[Koster, 1986] Jan Koster. *Domains and Dynasties: the Radical Autonomy of Syntax.* Foris, Dordrecht, 1986.

[Kracht and Wolter, 1991] Marcus Kracht and Frank Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56:1469 – 1485, 1991.

[Kracht, 1989] Marcus Kracht. On the logic of category definitions. *Computational Linguistics*, 15:111 – 113, 1989.

[Kracht, 1991] Marcus Kracht. *Internal Definability and Completeness in Modal Logic.* PhD thesis, FU Berlin, 1991.

[Kracht, 1993a] Marcus Kracht. How completeness and correspondence theory got married. In Maarten de Rijke, editor, *Diamonds and Defaults*, Synthese, pages 175 – 214. Kluwer, 1993.

[Kracht, 1993b] Marcus Kracht. Nearness and syntactic influence spheres. Manuscript, 1993.

[Kracht, 1995a] Marcus Kracht. Highway to the danger zone. *Journal of Logic and Computation*, 5:93 – 109, 1995.

[Kracht, 1995b] Marcus Kracht. Syntactic Codes and Grammar Refinement. *Journal of Logic, Language and Information*, 1995.

[Laka, 1991] Itziar Laka. Negation in syntax: on the nature of functional categories and projections. *International Journal of Basque Linguistics and Philology*, 25:65 – 138, 1991.

[Pollard and Sag, 1987] Carl Pollard and Ivan A. Sag. *Information-Based Syntax and Semantics, Part I.* Number 13 in CSLI Lecture Notes. CSLI, 1987.

[Reinhart, 1983] Tanya Reinhart. Coreference and bound anaphora: A restatement of the anaphora question. *Linguistics and Philosophy*, 6:47 – 88, 1983.

[Segerberg, 1986] Krister Segerberg. Modal logics with functional alternative relations. *Notre Dame Journal of Formal Logic*, 27:504 – 522, 1986.

[Stabler, 1992] Edward P. Stabler. *The Logical Approach to Syntax. Foundation, Specification and Implementation of Theories of Government and Binding.* ACL-MIT Press Series in Natural Language Processing. MIT Press, Cambridge (Mass.), 1992.

[van Benthem, 1983] Johan van Benthem. *Modal and Classical Logic.* Bibliopolis, 1983.

[Wolter, 1993] Frank Wolter. *Lattices of Modal Logics.* PhD thesis, FU Berlin, 1993.

[Zakharyaschev, 1992] Michael Zakharyaschev. Canonical formulas for K4, Part I: Basic results. *Journal of Symbolic Logic*, 57:377 – 402, 1992.