In: Proceedings of the Sixth Meeting on Mathematics of Language, Orlando, FL, July 23–25, 1999, pp. 89–104

On Conditional Information in Feature-Based Theories

Rainer Osswald
Applied Computer Science VII
University of Hagen
Germany
rainer.osswald@fernuni-hagen.de

Abstract

Two approaches to conditional information used in feature-based linguistic theories, especially head-driven phrase structure grammar, are compared and their interrelation is explicated on a formal and a conceptual level. For this purpose concepts of locale theory are introduced that allow to define feature descriptions and structures in a unified manner and, in particular, to make the difference between both approaches transparent. In addition, a foundation for attribute-value logic as a predicate-functor logic based on regimented and formalized descriptions is proposed.

It turns out that the relative pseudo-complement version of conditional constraints as put forward by Pollard and Sag (1987) is the wrong choice. From a formal perspective, the mistake is due to the confusion of the category of Heyting algebras with that of frames. Only the former are equipped with an operation corresponding to the conditional, that is, with means to represent conditional information as an element within the algebra itself. With respect to cognitive processing the central point is that conditional information is (in general) not actual information acquired during cognitive action but background information applied in processing.

An appropriate formalism to keep conditional and actual/observational information apart is given by geometric or observational logic (Vickers, 1989). Since this logic is on the other hand the logic of frames and locales it provides a promising framework for grammatical theory.

1 Introduction

Most modern grammatical theories are formulated as a system of constraints. Pushing this viewpoint, head-driven phrase structure grammar (HPSG) regards the entire grammar of the language in question as a single constraint on linguistic signs in form of a conjunction of universal and language-specific principles, including disjunctive constraints on lexical entries and phrase-structure templates (also known as immediate dominance schemata).

An explication of this proposal is given by Pollard and Moshier (1990) and Carpenter (1992, chapter 15). For them, a grammar or constraint system is roughly a *relation* between attribute-value descriptions. Carpenter prefers to write '$x \Rightarrow y$' iff the ordered pair $\langle x, y\rangle$ is a constraint, that is, an element of the constraint system.

This notation, on the other hand, reminds the attentive scholar of the extensive footnotes in Pollard and Sag's pioneering work (1987, chapter 2). The symbol '$\Rightarrow$' denotes the *relative pseudo-complement operation*, by means of which Pollard and Sag formulate *conditional information*. One of the more famous examples is the *Head Feature Principle*:

$$\textit{phrase} \Rightarrow \text{SYN}|\text{LOC}|\text{HEAD} \doteq \text{DTRS}|\text{HEAD-DTR}|\text{SYN}|\text{LOC}|\text{HEAD}\,.$$ [1]

What is the relation between these two uses of '$\Rightarrow$'? Obviously, it denotes a two-place relation in the first case, and a two-place function in the second. Perceiving relations as functions makes the comparison more direct: a function with possible values true and false vs. one with attribute-value descriptions as values.

In what follows we show, after reviewing some well-known insights, that in general the relative pseudo-complement operation is the *wrong* choice in regard to Pollard and Sag's intention, pin down the reason for that mistake, and propose a formal framework to overcome this obstacle. To prevent misunderstandings: we are *not* going to furnish the trend to classical logic, which Pollard (1999, p. 283f) regards as a side effect of a sharp distinction

[1] It would be more precise to replace '*phrase*' by '*headed-phrase*' and to abandon the anachronistic 1987 feature architecture, but such details are of no concern here.

between feature structures and feature descriptions. On the contrary, we propose a logic (*geometric* or *observational*) that has *no* internalized conditional.

Subsequent to these primarily formal considerations we focus on the question what conditional information is, and is about, in grammatical and cognitive theory.

(Readers primarily interested in an argument from within Pollard and Sag's (1987) approach contra the choice of relative pseudo-complements as constraints are encouraged to skip to section 3.1.)

2 Recapitulation

The following synopsis belongs more or less to folklore (cf. Pereira and Shieber (1984), Pollard and Moshier (1990), Carpenter (1992), Shieber (1992), Moshier (1993), Rounds (1997), Pollard (1999)).[2] Style of presentation differs insofar as it draws on concepts from locale theory. See appendix for background.

2.1 Attribute-Value Logic

Elementary (or *subbasic*) *attribute-value* (or *feature*) *terms* over finite sets L and A are of the form $p \doteq q$ and $p{:}a$ with p and q being elements of L^* and a element of A. To be slightly more pedantic than usual these terms are defined through *formal specification*. Ingredients are an operation | realizing the concatenation of strings over L, the so-called empty string ϵ (a 0-ary operation), and two "ordered pair" operations $\doteq$ and :, of which the first takes elements of L^* as arguments and the second elements of L^* and A respectively. In other words, the set E of elementary attribute-value terms is specified as

(1) $$L^* \times L^* \ \uplus \ L^* \times A\ .$$

If one likes to admit elements of A as attribute-value terms then it is reasonable to require that $\epsilon{:}a = a$.[†]

Emphasis on such formalistic aspects is programmatic to the viewpoint of universal algebra taken up in the following. The apparently harmless starting point of elementary attribute-value terms presupposes equational (or functorial) specifications. Initial models can be constructed as quotients of free term algebras modulo the congruence generated by all instances of the given equations.[3] An equivalent representation is given in (1) as a direct sum of direct products.[4] Abstract specification at the beginning allows to switch representations when appropriate.

Let E be the set of elementary attribute-value terms. They are subject to the constraint system, henceforth *Avl* (attribute-value logic), given by the following schemata:[5]

$$\begin{array}{ll}
p \doteq q \ \preccurlyeq\ q \doteq p & (\textit{Symmetry})\\
p \doteq q\,,\ q \doteq r \ \preccurlyeq\ p \doteq r & (\textit{Transitivity})\\
p \doteq q \ \preccurlyeq\ p \doteq p & (\textit{Reflexivity})\\
p{:}a \ \preccurlyeq\ p \doteq p & \\
p \doteq q\,,\ p|r \doteq p|r \ \preccurlyeq\ p|r \doteq q|r & (\textit{Substitutivity})\\
p \doteq q\,,\ p{:}a \ \preccurlyeq\ q{:}a & \\
p|q \doteq p|q \ \preccurlyeq\ p \doteq p & (\textit{Prefix Closure})
\end{array}$$

Our object of interest is the frame $\mathcal{F}(\mathit{Avl})$ presented by the constraint system *Avl* or, more precisely, the corresponding locale $\mathcal{L}(\mathit{Avl})$, henceforth called the *feature* (or *attribute-value*) *locale*. Its *opens*, i.e. the elements of $\mathcal{F}(\mathit{Avl})$, are the *attribute-value descriptions* formed by finite conjunction and infinite disjunction of elementary terms modulo equivalence with respect to *Avl*. The *points* of $\mathcal{L}(\mathit{Avl})$ are the (abstract, possibly infinite, conjunctive) *feature structures*.

Readers acquainted with the somewhat extensive construction given by Pollard and Moshier (1990) and Carpenter (1992, chapter 12) may wonder how fast our definition comes to the point(s). Two answers: First, the category of locales (or frames, respectively) is the perfect choice for the intended purpose in that it supplies the desiderata by universal constructions. Second, specification is one thing, representation is another. The Pollard/Moshier/Carpenter definition is lengthy because it involves a specific construction whereas our definition makes recourse to a general fact about frame existence.

Points of $\mathcal{L}(\mathit{Avl})$, i.e. feature structures, can be described in several equivalent ways: as elements of $\mathcal{F}(\mathit{Avl})^\star$ (frame morphisms from $\mathcal{F}(\mathit{Avl})$ to **2**),

[2] We stick to "ordinary" attribute-value structures. For set valued variants see Moshier and Pollard (1994).

[†] This trifling matter is missed, typically enough, in Carpenter's (1992, p. 63) axiomatization (but see ibd., p. 66).

[3] see e.g. Wechler (1992) for background.

[4] For a functorial specification of the list structure L^* see e.g. Manes and Arbib (1986).

[5] Equivalence to Shieber's (1992) system

$$\begin{array}{ll}
p \doteq q \ \preccurlyeq\ q \doteq p & (\textit{Symmetry})\\
p|q \doteq r \ \preccurlyeq\ p \doteq p & (\textit{Reflexivity})\\
p{:}a \ \preccurlyeq\ p \doteq p & \\
p \doteq q\,,\ p|r \doteq s \ \preccurlyeq\ q|r \doteq s & (\textit{Substitutivity})\\
p \doteq q\,,\ p{:}a \ \preccurlyeq\ q{:}a &
\end{array}$$

is easily checked via identity, weakening, and cut.

as completely prime filters of $\mathcal{F}(Avl)$, or more concretely, as subsets of *E* closed under *Avl*.[6] (The representation as rooted feature systems given in section 2.2 hinges in contrast on the specific structure of *Avl*.)

Note that the notions of *satisfaction* and *subsumption* are already determined by $\mathcal{L}(Avl)$: a description, represented as the disjunction of a set S of finite conjunctions of elementary descriptions is satisfied by a feature structure x if and only if there is an element $\phi_1 \wedge \ldots \wedge \phi_n$ of S such that x (as *Avl*-closed subset of *E*) is a superset of $\{\phi_1, \ldots, \phi_n\}$. A feature structure x subsumes y if and only if y satisfies every description that is satisfied by x. Furthermore, the set of feature structures ordered by subsumption is directed complete.[7] The join (*unification*) $\bigsqcup X$ of a set of feature structures, existence assumed, satisfies ϕ iff at least one element of X satisfies ϕ.

So far, no specific properties of *Avl* came into play. Since *Avl* is coherent, $\mathcal{L}(Avl)$ is a *spectral* locale, that is, $\mathcal{F}(Avl)$ is the ideal completion of the distributive lattice freely generated by *Avl*. (Elements of this lattice are composed of elementary descriptions by *finite* disjunction and conjunction.) In addition, $\mathcal{L}(Avl)$ is spatial, which means that two descriptions are identical iff they are satisfied by the same set of feature structures. Because *Avl* is definite, the partially ordered set of feature structures is bounded complete, and even complete since there are no exclusions in *Avl*.[8]

Algebraicity of $\mathcal{L}(Avl)$ is another immediate consequence of the definiteness of *Avl*. $\mathcal{L}(Avl)$ is even ω-algebraic since *E* is countable. The poset of $\mathcal{L}(Avl)$ points therefore is a Scott-domain. Spectral algebraic locales can be characterized by the condition that every compact open has a unique representation as a finite irredundant join of compact, coprime opens. (The latter correspond to compact points.) This serves as a blueprint for the construction carried out by Pollard and Moshier (1990) and Carpenter (1992, chapter 12) starting with compact feature structures represented as finite rooted feature systems (see below).[9]

[6] Appendix A.3, lemma 2.

[7] see again lemma 2.

[8] Since $\{\epsilon \doteq \epsilon\}$ is the least element with respect to subsumption, the partial order of feature structures is a complete lattice. Another implication of definiteness is that $\mathcal{F}(Avl)$ is the universal frame over the semilattice freely generated by *Avl*.

[9] Lack of space does not allow an explication of the notions used in the last paragraph; see e.g. Vickers (1989, chapter 9)

2.2 The Structure of a Point

In light of representation (1) each subset of *E* *is* the (disjoint) union of a subset of $L^* \times L^*$ and one of $L^* \times A$. From this perspective it is common to call *Avl*-closed subsets *Nerode* representations of feature structures. Being closed with respect to *Avl* is then rephrased as being the union of an equivalence relation on L^* which is conditionally closed with respect to concatenation from the right (i.e., contains $p|r \doteq q|r$ if it contains $p \doteq q$ and $p|r \doteq p|r$) and a subset R of $L^* \times A$ such that all sets of the form $\{p \mid Rpa\}$ are congruence classes.

An immediate consequence of this representation is that $\mathcal{L}(Avl)$ points can be represented as *rooted feature systems*.

Definition 1: A *feature system* over L and A is a pair $\langle N, I\rangle$ consisting of a set N, the *carrier*, and a function I which takes each element l of L to a functional relation[10] on N and each element a of A to a subset of N.

It is convenient to extend I by converse relational composition inductively to L^*, that is, $I(p|l) = I(l) \circ I(p)$.

Definition 2: A *rooted feature system* is a triple $\langle *, N, I\rangle$ such that $\langle N, I\rangle$ is a feature system, $*$ is an element of N, and each element n of N is "reachable" from $*$ in the sense that there is an element p of L^* with $\langle n, *\rangle \in I(p)$.

If x is a feature structure (thought as its Nerode representation) then the set N_x of congruence classes $\{[p] \mid p \doteq p \in x\}$ of x serves as the carrier of the rooted feature system $\langle[\epsilon], N_x, I_x\rangle$ where I_x is defined such that

$$
\begin{aligned}
I_x(l) &= \{\langle [p|l], [p]\rangle \mid p|l \doteq p|l \in x\}\,,\\
I_x(a) &= \{[p] \mid p{:}a \in x\}\,.
\end{aligned}
$$

$I_x(l)$ is *functional* because $[\ ]$ is a conditional right congruence.

This correspondence works the other way around as well: if $\langle *, N, I\rangle$ is a rooted feature system then

$$
\begin{aligned}
&\{p \doteq q \mid \exists n(\langle n, *\rangle \in I(p)\ \&\ \langle n, *\rangle \in I(q))\}\\
&\cup\ \{p{:}a \mid \exists n(\langle n, *\rangle \in I(p)\ \&\ n \in I(a))\}
\end{aligned}
$$

[10] Though afflicted with the danger of confusing the reader a relation R is understood to be functional iff $\forall xyz(Rxz \,\&\, Ryz \rightarrow x = y)$. This Quine/Gödel/Peano convention draws its justification from the natural way it allows to formalize natural language expressions like 'is father of'; see section 4.1 for further examples.

is closed with respect to *Avl*. Applied to I_x this construction returns x. And, with morphisms defined in the obvious way, it takes isomorphic rooted feature systems to identical feature structures. For those who enjoy the language of category theory:

Proposition 1: The poset of $\mathcal{L}(Avl)$ points is (as category) equivalent to the category of rooted feature systems ∎

Translation

Each element p of L^* determines a function t_p from E to $\mathcal{F}(Avl)$ taking $q \doteq r$ to $p|q \doteq p|r$ and $q{:}a$ to $p|q{:}a$. We call it the *translation by* p and allow to write '$p{:}\phi$' for '$t_p(\phi)$'. Since *Avl* is invariant under translation, there is an induced endomorphism on $\mathcal{F}(Avl)$.[11]

The switch to $\mathcal{L}(Avl)$ deserves attention: when translation is taken as the frame part of a locale morphism then point translation goes in the opposite direction taking each point x to $x \circ t_p$, i.e. (as *Avl*-closed set) to $\{\epsilon \doteq \epsilon\} \cup \{p{:}\phi \mid \phi \in x\}$.

Converse translation is more interesting. It does not give a function but at least a functional relation because translation is injective. Converse translation by p applies to every point x which satisfies $p \doteq p$ and takes it to $\{\phi \mid p{:}\phi \in x\}$. By defining $U(l)$ as the converse translation by l and $U(a)$ as the extent $\{x \mid x \vDash a\}$ of a one gets a feature system carried by the set of $\mathcal{L}(Avl)$ points.

3 Conditional Constraints

3.1 Two Positions

According to HPSG and related paradigms, well-formed linguistic entities are modelled by attribute-value structures which satisfy all constraints of the grammar (provided its adequacy). Pollard and Sag's original approach (1987) to constraints uses the relative pseudo-complement operation, which is characterized by:

$$\phi \leqslant \psi \Rightarrow \chi \;\leftrightarrow\; \phi \wedge \psi \leqslant \chi\,,$$

that is, $\psi \Rightarrow \chi = \bigvee\{\phi \mid \phi \wedge \psi \leqslant \chi\}$. Constraints are thus attribute-value descriptions.[12]

Pollard and Moshier (1990) and Carpenter (1992, chapter 15) present a somewhat different picture of constraints. They define a *grammar* or *constraint system* respectively as a function C which takes elements of A to (finite) attribute-value descriptions. The concept of satisfaction needs explication in this context. According to Carpenter, a feature structure x satisfies the constraint system iff x is *resolved*, which means that "every one of its substructures satisfies the inherited constraint on its type" (ibd., p. 229). One possible formulation of this is:

$$x \vDash \bigwedge\{p{:}C(a) \mid x \vDash p{:}a\}\,,^{\dagger}$$

or equivalently,

(2) $$\forall a \forall p (x \vDash p{:}a \rightarrow x \vDash p{:}C(a))\,.$$

In other words, x is closed with respect to all constraints of the form $\langle p{:}a, p{:}C(a)\rangle$.

Implicit closure of constraints under translation is also appropriate for the relative pseudo-complement version. For gain of elegant formulation we slightly abuse the so-called *master modality* $[*]$ introduced by Gazdar et. al. (1988):[13]

$$[*]\phi \;=\; \bigwedge\{p{:}\phi \mid p \in L^*\}\,.$$

The relative pseudo-complement version of the constraint corresponding to a constraint system C then leads to:

$$x \vDash [*]\bigwedge\{a \Rightarrow C(a) \mid a \in A\}\,,$$

which is equivalent to:

(3) $$\forall a \forall p (x \vDash \; p{:}a \Rightarrow p{:}C(a))\,.^{\ddagger}$$

Diagnosis

The relation between (3) and (2) is straightforward. (3) implies (2) whereas the reverse is generally not the case:

$$x \vDash a \Rightarrow b \;\rightarrow\; \forall y (x \sqsubseteq y \rightarrow (y \vDash a \rightarrow y \vDash b))\,.$$

Proof: Suppose that $x \vDash a \Rightarrow b$, $x \sqsubseteq y$, and $y \vDash a$. By definition of $a \Rightarrow b$ there is a c such that $c \wedge a \leqslant b$ and $x \vDash c$. Therefore, by definition of subsumption, $y \vDash c$. So, $y \vDash a \wedge c \leqslant b$, which implies that $y \vDash b$ ∎

To give a simple illustration, suppose there are three different elementary descriptions a, b, and c, and one (nontrivial) constraint $\langle a, b\rangle$. According to (2), x is resolved iff $x \vDash a \rightarrow x \vDash b$. Resolved, i.e. closed sets are $\{b\}$, $\{c\}$, $\{a, b\}$, $\{b, c\}$, and $\{a, b, c\}$. (3) on the other hand calls for x such that $x \vDash a \Rightarrow b$. But $a \Rightarrow b$ equals $\bigvee\{c \mid c \wedge a \leqslant b\}$, that is, b. Therefore, $a \Rightarrow b$ is satisfied by the closed sets $\{b\}$, $\{a, b\}$, $\{b, c\}$, and $\{a, b, c\}$, obviously lacking $\{c\}$.

[11] by the universal property of definition 10, appendix A.2.

[12] Each element is permissible since $\chi = \top \Rightarrow \chi$.

[†] Note that in general this incorporates inherited constraints; see section 3.2.

[13] see also Kracht (1995) and Rounds (1997).

[‡] Translation is a Heyting algebra endomorphism on $\mathcal{F}(Avl)$. A proof is omitted since for our main argument we could equally well choose (3) in the first place.

A closer look to Pollard and Sag's reasoning reveals that what they regard as "most important consequence" (1987, p. 43) of introducing pseudo-complements does *not* make them prerequisite. It is the demand that if $x \vDash c$, $c \leqslant a$, and $x \vDash a \Rightarrow b$ then $x \vDash b \wedge c$. But for this conclusion, the premise '$x \vDash a \to x \vDash b$' in place of the stronger '$x \vDash a \Rightarrow b$' is sufficient. Contrary to what Pollard and Sag seem to have in mind, when x satisfies $a \Rightarrow b$ this does not only mean that if x satisfies a then it must also satisfy b. It rather means that everything subsumed by x that satisfies a must also satisfy b.

Possible complaints about non-monotonic effects of the condition '$x \vDash a \;\to\; x \vDash b$' are out of place since in this case there is no analogue to $a \Rightarrow b$ which may turn wrong when more information is available. In the above example $\{c\}$ belongs to $\{x \mid x \vDash a \to x \vDash b\}$ because $\{c\} \nvDash a$. In contrast, $\{a, c\}$ does not, since $\{a, c\} \vDash a$ but $\{a, c\} \nvDash b$. Its "resolution" therefore is $\{a, b, c\}$. This effect is best described as "growth" and not as "revision" of information.

To reiterate, if x does not satisfy a then it trivially satisfies the constraint $\langle a, b\rangle$ – the constraint "does not apply" to a, as intended. But if a specialization of x satisfies a but not b then x does *not* satisfy $a \Rightarrow b$. So, a possible explanation for Pollard and Sag's mistake may be their inattentive constriction to the case that specialization "into" the premise of constraints is excluded. This means that each feature structure x either satisfies the premise a of a constraint or is incompatible with it in the sense that none of its specializations satisfies a. In short, either $x \vDash a$ or $x \vDash a \Rightarrow \bot$. Abbreviating '$a \Rightarrow \bot$' as usual by '$\neg a$' we get that $x \vDash a \vee \neg a$ for every x, that is, $a \vee \neg a = \top$. Thus, finally, we are led back to Boole.

3.2 A Framework for Grammatical Theory

Geometric/Observational Logic

Using relative pseudo-complements as constraints does not serve the intended purpose. From a categorical perspective, the choice of complete Heyting algebras is a category mistake (in the real sense of the word). The appropriate candidate is the category of frames (or locales respectively). One source of confusion is the fact that the objects of both categories are indistinguishable as ordered sets.

The preferable alternative is to keep descriptions and constraints apart. This is precisely the setting of (*propositional*) *geometric* or, as Vickers calls it, *observational logic*.[14] With 'description' and 'constraint' in place of 'formula' and 'axiom' the following passage from Vickers (1999, section 2) describes our viewpoint perfectly: "*The observational intuition is that formulae represent observations, while axioms – how observations relate to each other – represent scientific hypotheses or background assumptions.*" Observational logic is thus well suited for both a scientific theory about "external" language (types of utterances, inscriptions, etc.) and a theory about the grammatical knowledge of speakers/hearers (see section 4.3).

As Vickers (1989) explains, it is reasonable to take *finite observations* (or *affirmative assertions*) as closed under *finite conjunctions* and *arbitrary disjunctions* (only one assertion has to be affirmed in case of the latter). Negation and conditional, however, do *not* preserve affirmability because negation transforms affirmative assertions into refutive ones. Therefore, the logic of affirmative assertions/finite observations is algebraized by frames. Given a set of elementary observations (or descriptions) the domain of observations is the frame freely generated by it. Identification of descriptions which are equivalent with respect to background knowledge in form of a constraint system (or geometric theory) L leads to the frame presented by L.

Though there is no negation operator, there are several ways to include "negative" information. Negative observations can be handled, for example, by *term negation*:[15] $-\phi$ is regarded as affirmative and not as a refutation of ϕ. That ϕ and $-\phi$ are incompatible can be enforced by an axiom scheme: $\phi, -\phi \;\preccurlyeq\; \varnothing$. Even the following scheme can be adopted: $\varnothing \preccurlyeq \phi, -\phi$; nothing is observed, but either ϕ or $-\phi$. Exclusions, i.e. constraints of the form $\langle \Phi, \varnothing\rangle$ serve as a surrogate for conditionals with negative consequent, for in classical propositional logic, $\phi \to \neg\psi \;\leftrightarrow\; \neg(\phi \wedge \psi)$.

Grammars as Constraint Systems

Pollard and Sag (1994, section 1.10) formulate HPSG explicitly as a system of constraints. A characteristic example is the specification of the lexicon as a constraint on words:

$$\textit{word} \;\preccurlyeq\; \omega_1 \;\vee\; \omega_2 \;\vee \ldots\,,$$

where the ω_i are feature descriptions, i.e. elements of the frame $\mathcal{F}(\textit{Avl})$. Though not a constraint in the narrow sense of appendix A.1 it does not prevent

[14] The name 'geometric' goes back to algebraic geometry.

[15] see Horn (1989) who traces the idea back to Aristotle.

our general theory from applying (see the remarks at the end of section A.2).

As in section 3.1 let us first close the constraints of HPSG with respect to translation and call the resulting system *Hpsg*. Then the locale $\mathcal{L}(Avl \cup Hpsg)$ is all we have to look for because its point set is the *denotation* of HPSG. It constitutes the space of models of the grammar. Given its coherence (no infinite disjunctions) then soundness and completeness are by-products of this construction.[16]

Let us proceed at a lower pace. The well-trodden path is to start with (compact) feature structures, i.e. with points of $\mathcal{L}(Avl)$, i.e. with elements of $\mathcal{C}(Avl)$, and then to pick out those which are closed under *Hpsg*. But that means nothing else than restricting $\mathcal{L}(Avl)$ to its *sublocale* $\mathcal{L}(Avl \cup Hpsg)$. Compare diagram (7) of appendix A.3.

The special status of *Avl* derives from its roots in *logical* truth; see section 4.1. This is of course not the case for certain subsystems of *Hpsg* which have been granted special status too, particularly the *sort hierarchy* and *feature declarations*. Pollard and Sag (1994) restrict their attention explicitly to so-called *well-typed* and *sort-resolved* feature structures.

King's (1994) formalization of these assumptions has led him to an approach within the realm of classical logic that incorporates certain appropriateness constraints into the *signature* and thereby into the *interpretation* of his formal language. This kind of attempt has been criticized by Moshier (1993, p. 195) for whom "the 'signature' approach to appropriateness conditions is unsatisfying simply because appropriateness conditions are *constraints*, and constraints are exactly what logic is good at." We fully agree.

Happily there is a way out of this dilemma: simply put all constraints at question into a constraint system *App* and move to the sublocale $\mathcal{L}(Avl \cup App)$ of $\mathcal{L}(Avl)$. Its points are exactly those feature structures that satisfy all constraints of *App*.

Full Distinguishability

Another motivation for imposing appropriateness conditions seems to be the wish for *full distinguishability*. By that we mean that any two feature structures x and y can be separated by descriptions ϕ and ψ in the sense that x satisfies ϕ but not ψ and y satisfies ψ but not ϕ. In other words, there is no subsumption of feature structures in $\mathcal{L}(Avl \cup App)$ besides identity, that is, $Avl \cup App$ is *Post complete*. A further paraphrase of this condition is that the frame $\mathcal{F}(Avl \cup App)$ is a *complete Boolean algebra*.[17]

There are two principal alternatives to gain full distinguishability: disconnect the points, or erase all non-maximal points. Topologically speaking, the first method means to switch to a topology fine enough to separate every pair of points. This corresponds to an enrichment of the descriptive vocabulary in the obvious way: for each elementary description ϕ add $-\phi$ and two constraints: $\phi, -\phi \preccurlyeq \varnothing$ and $\varnothing \preccurlyeq \phi, -\phi$.[18] The other, more involved variant is to add constraints by keeping the descriptive vocabulary constant such that each closed set is maximal consistent.

Prospects

The picture given so far is conceptual in nature. Its main intent is to present a framework that allows an almost trivial explication of such seemingly involved tasks like defining the denotation of a grammar given as a constraint system. It does not address any "real-life" questions like how to compute a deductive closure let alone what kind of representation and algorithm to choose best.

What is offers, however, is a more flexible interface to domain theory – where such questions are addressed – than the usual one which is restricted to the Scott domain of feature structures. To give an example, it appears that grammatical constraints, especially those of HPSG, are *disjunctive* in the sense that for each constraint $\langle \Phi, \Psi \rangle$ every two-element subset of Ψ is an exclusion. The point spaces represented by this type of constraint system are known (Zhang, 1992) to coincide with the so-called *L-domains*.[19] A natural question to ask concerns the calculation of fix-points of recursive equations.[20]

4 Foundations

This section addresses questions of ontology and empirical content of feature-based grammatical theory. It rests to a considerable degree on Quinean positions.[21]

[16]Note that, though we do not make use of it, infinite disjunction would allow for example the formulation of *regular path equations*.

[17]Vickers (1989, section 9.2).

[18]Vickers (ibd.) calls it a *Booleanization*.

[19]An L-domain is a cpo such that every principal ideal is a complete lattice – which is why Jung (1990) calls them "lattice-like". They form a Cartesian closed category.

[20]Zhang (1992) makes some encouraging remarks about this.

[21]see Quine (1995) for a crisp introduction.

4.1 Attribute-Value Ascription

Referential Roots

The general picture Pollard and Sag (1994, p. 6) give of linguistic theory is that "[t]he theory itself does not talk directly about the empirical phenomena; instead, it talks about, or is *interpreted by*, the modelling structures." Empirical adequacy then "arises from the conventional correspondence between the model and the empirical domain."

In section 4.2 we plead for rejecting these "methodological assumptions" because spelling out the "correspondence" inevitably involves natural language to refer to entities of the empirical domain. So why not start here and derive from these descriptive means by way of appropriate regimentation and formalization expressions of predicate logic or an equivalent formalism, which then can serve as formulae of a theory. This program is pursued in the following for attribute-value ascriptions.

The basic assumption of feature-based theories is that the objects of interest bear certain relations ("attributes", "features") to other objects ("values") of a certain type. These properties are naturally ascribed to objects by use of ordinary language expressions of a certain syntactical form. People and their cars give a simple non-linguistic example. A typical attribute-value description is 'someone whose car is a convertible'. Regimentation and formalization[22] leads to:

(4) $$\{x \mid \exists z(z = \iota y Fyx \;\&\; Pz)\}\,,$$

with 'F' for '$\{yx \mid y$ is car of $x\}$' and 'P' for 'convertible'. For the moment, '$\{x \mid \ldots x \ldots\}$' is nothing more than a suggestive surrogate for 'x such that $\ldots x \ldots$'. Elimination of the definite description along Russellian lines leads to

$$\{x \mid \exists z(Fzx \,\&\, Pz) \,\&\, \forall yz(Fyx \,\&\, Fzx \rightarrow y = z)\}.$$

Presupposing 'F' as functional, (4) becomes equivalent to

(5) $$\{x \mid \exists y(Fyx \,\&\, Py)\}\,,$$

which coincides with the so-called *Peirce product* or *inverse image* of 'P' by 'F', usually written as '$F\!:\!P$'.[23] (We still adhere to syntactical manipulations although the notation may evoke set-theoretical associations.)

[22] e.g. in the style of Quine (1960), (1982).

[23] Note that a widespread rival convention calls '$\{x \mid \exists y(Fyx \,\&\, Py)\}$' the *image* of '$P$' by '$F$'. Compare footnote 10.

The second type of expression used in feature-based theories describes an entity as having two attributes with identical value. 'Someone whose father is her employer' will do as example. Regimentation and formalization in similar vein as above transforms this type of predicate into

$$\{x \mid \exists y(Fyx \,\&\, Gyx)\}\,,$$

which we define as the result of a predicate functor '$\doteq$' applied to two dyadic predicates 'F' and 'G'.

Finally, attribute composition as in 'car of the father of the employer' has to be taken care of. Not surprisingly, '$F|G$' stands for '$G \circ F$', that is, for

$$\{xy \mid \exists z(Fzy \,\&\, Gxz)\}\,.$$

In this way, attribute-value logic is developed as a special kind of *predicate-functor* or *operator logic*.[24] Logical implication of attribute-value terms reduces to implication in predicate logic.[25] In order to make our operator logic autonomous it has to be equipped with a *sound* and *complete* set of inference rules in the sense that they allow to deduce an attribute-value term from a set of such terms if and only if it is deducible by means of predicate logic. That is exactly what *Avl* does.

Completeness

Completeness of *Avl* in the above sense is first-order derivability of attribute-value descriptions as expressions of predicate logic granted the functionality of attribute predicates.

Let L and A be finite sets of dyadic and monadic predicates respectively. Let *Fun* be the set of axioms given by functionality, that is, the first order theory determined by means of the axiom schema:[26]

$$\forall xyz(lyx \,\&\, lzx \rightarrow y = z)\,.$$

The rooted feature system representation I_x of a feature structure x is by definition a first-order model of *Fun*. It has the following key property:

Lemma 1: $[\epsilon] \in I_x(\phi)$ iff $\phi \in x\,.$

Proof: According to (5), $I_x(l\!:\!a)$ is

$$\{m \mid \exists n(\langle n, m\rangle \in I_x(l) \;\&\; n \in I_x(a))\}\,.$$

[24] see e.g. Quine (1976).

[25] The attentive reader presumably has detected the gap of *reflexive* attributes. Predicates like 'someone who is her own lawyer' are not expressible within the operator logic introduced so far. The many ways to overcome this obstacle are left to the reader.

[26] *Fun* is a natural starting point for approaches to feature logic within the realm of first-order logic; see e.g. Smolka (1992), Aït-Kaci et al. (1994).

Since $\langle [l], [\epsilon] \rangle \in I_x(l)$, functionality of $I_x(l)$ implies that $[\epsilon] \in I_x(l\!:\!a)$ iff $l\!:\!a \in x$. The other cases of elementary feature terms can be handled by analogous reasoning and induction on elements of L^* ∎

(This type of observation is commonly called a *truth lemma*. Our line of reasoning is similar to that of the usual ad hoc semantics for attribute-value logic.)

Proposition 2: *Avl* is sound and complete with respect to first-order derivability in *Fun*, that is,

$$\mathit{Fun} \vdash \bigwedge \Phi \subseteq \psi \quad \text{iff} \quad \Phi \vdash_{\mathit{Avl}} \psi\,.^{\dagger}$$

Proof: Soundness is proofed by straightforward application of predicate logic. To show completeness, suppose $\mathit{Fun} \vdash \bigwedge \Phi \subseteq \psi$. Let x be any *Avl*-closed superset of Φ. The feature system corresponding to x is a first-order model of *Fun* (section 2.2). By lemma 1, it follows that $[\epsilon] \in I_x(\phi)$ for every element ϕ of Φ. Therefore, by assumption, $[\epsilon] \in I_x(\psi)$. Hence, $\psi \in x$ by lemma 1. Now apply proposition 7, appendix A.1 ∎

Note that we have to look "into points" of $\mathcal{L}(\mathit{Avl})$ to proof completeness. The feature system of $\mathcal{L}(\mathit{Avl})$ points (section 2.2) will not do. If, for example, *A* is empty and *L* consists of two elements F and G then the point $\{\mathrm{F} \doteq \mathrm{F}, \mathrm{G} \doteq \mathrm{G}\}$ is taken to $\{\epsilon \doteq \epsilon\}$ by both $U(\mathrm{F})$ and $U(\mathrm{G})$ and is therefore an element of $U(\mathrm{F} \doteq \mathrm{G})$. Contrary to what Rounds (1997, p. 521) seems to assume, it is thus *not* possible to recover attribute-value logic from the feature system of $\mathcal{L}(\mathit{Avl})$ points if one disregards subsumption.

There are two possible views of attribute-value logic as geometric/observational logic. In section 3.2 we introduced the "propositional" version. There is also a predicate variant of geometric logic, whose formulae are restricted to conjunction, disjunction, equality, and existential quantification, whereas axioms are universally quantified conditionals with antecedent and consequent restricted to these formulae.[27] Obviously, the above formalization of attribute-value descriptions and axioms belongs to this class.

It is tempting to take this "double nature" of attribute-value terms as "propositions" and predicates for an explanation for the vexation that the point set of $\mathcal{L}(\mathit{Avl})$ carries a feature system and at the same time each point *is* a (rooted) feature system. The latter interpretation corresponds to the view of $\mathit{Fun} \cup \mathit{Avl}$ as a predicate geometric logic for which the space of models, i.e. the system of feature structures represents the *classifying topos*. Completeness presumably will then follow from topos theory.[28]

Reflections

An advantage of treating attribute-value descriptions as regimented pieces of natural language is that its limitations are brought to the surface. Once adopted it puts its user to the procrustean bed of the attribute-value idiom. For example, all relations in the basic vocabulary are assumed to be functional. It might be worth to explore *extensions within predicate geometric logic*.

On the other hand, it can be argued that attribute-value predicates have more expressive power than needed in linguistic theory. For the main motivation for introducing the operator '$\doteq$' is the description of *agreement phenomena* in natural language. The prototypical example is 'something such that the person of its subject coincides with the person of its verb', which denotes sentences.

'Person' is thus understood as a *dyadic* predicate. It denotes roughly certain ordered pairs composed of natural language expression and entities denoted by 'first person', 'second person', or 'third person'. Therefore, we have to call in criteria of identity for these new inhabitants of our universe of discourse (see section 4.2). After a short reflection one recognizes that 'first person' denotes *exactly one* entity which in addition takes part in *exactly one* relation, namely the relation denoted by 'person'. This type of argument has led some researchers (typically logicians working with modal variants of feature logic) to the conclusion that attribute-value identity might be abandoned altogether.[29]

4.2 The Nature of Linguistic Theory

Scientific Theories

It is common in the philosophy of science to distinguish *syntactic* from *semantic* approaches to scientific theories.[30] To put it in a nutshell, theories are axiomatic systems for the first and collections

† The '$\subseteq$' on the left side is of course not set-theoretic inclusion but an operator which takes two monadic predicates 'P' and 'Q' to '$\forall x(Px \rightarrow Qx)$'.

27 see Vickers (1993) for an introduction.

28 see Makkai and Reyes (1977) and Vickers (1993), (1999) for background. Spelling out these remarks is part of future work.

29 e.g. Kracht (1995).

30 see e.g. van Fraassen (1980, chapter 3), Lambert and Brittan (1987, chapter IV).

of models for the latter. This coarse distinction has to be relativized insofar as axioms have their place in semantic theories as well as models are of use for syntactic approaches. But their status then is thought to be only second class.

For a semantic approach, the role of language therefore is neither basic nor unique. The choice of a class of structures as its models comes first whereas a (preferably complete) description of them is a secondary task. On the other hand, to equip syntactic approaches with a model-theoretic semantics and, thereby, with models, involves also a certain degree of freedom. Though the models of model theory differ at first sight to a considerable extent from those of scientific theories, it seems reasonable, as e.g. van Fraassen (1980, pp. 44, 199) points out, that both usages of 'model' can be reconciled.

We have noticed at the beginning of section 4.1 that the scientific picture of linguistic theory given by Pollard and Sag (1994) is semantically oriented. For them, modelling linguistic entities by attribute-value structures is what comes first. Furthermore, their usage of 'model' carries both of the meanings mentioned above.

Modelling and Empirical Adequacy

The checkpoint for scientific theories is their *empirical adequacy*. For semantic approaches, this means that the models have to "fit" the empirical phenomena, i.e., that they "represent" them correctly. For this purpose the intended relationship between the model and certain aspects of the real world has to be made precise. Here the notion of *isomorphism* comes into play. Following van Fraassen (1980, p. 64), a theory is *empirically adequate* if the structures described in measurement reports are *isomorphic* to (certain substructures of) some model of the theory.[31]

Moshier and Pollard (1994, p. 613) are quite explicit on the modelling relation, that is, on the question how feature structures relate to the entities they are supposed to represent. They propose a *modelling convention* along the following lines: a rooted feature system $\langle *, N, I\rangle$ models an entity o iff there is an (injective) function f (the isomorphism) from N to the empirical domain which takes $*$ to o such that if $n \in I(a)$ then $f(n)$ "is of a sort conventionally named a" and if $\langle n, m\rangle \in I(l)$ then $f(m)$ "has an attribute conventionally named l, the value of which is" $f(n)$, or, better, $f(n)$ bears to $f(m)$ a functional relation conventionally named l. The last two conditions reformulated: if $n \in I(a)$ then $f(n)$ satisfies a and if $\langle n, m\rangle \in I(l)$ then $\langle f(n), f(m)\rangle$ satisfies l. (a and l are monadic and dyadic predicates respectively.)

Leaving it to the reader to fill out some minor details (using e.g. proposition 1, section 2.2), we get as an immediate consequence the following equivalent formulation of Moshier and Pollard's modelling convention, now abstracting away from the specific representation as rooted feature systems: a feature structure x models an entity o iff

(6) $x \models \phi \quad \text{iff} \quad o \text{ satisfies } \phi$

for every elementary attribute-value description ϕ. Therefore, when it comes to empirical content the semantic approach – at least in the case of HPSG – has to rely on the language-based approach of section 4.1.

Grammatical theories like HPSG, as we understand it, are sets of universally quantified conditionals with antecedent and consequent restricted to a certain type of monadic predicate, namely attribute-value descriptions. Such a theory may be said to have empirical content if it allows to deduce from certain observations falsifiable predictions. Since science strives for intersubjectivity, there has to be agreement what assertions to count as true and what as false. As for the ascription of a predicate 'P' to an entity x, this means that any two researchers can *identify* x to make sure they refer to the same thing and that they agree whether x satisfies 'P' or not.

Utterances and inscriptions are good candidates to start with because they are identifiable by their extension in time and space whereas phonemes, morphemes and word forms give a collection of predicates for which it seems reasonable to expect agreement.

Ontological Indifference

Tying down identity is according to Quine on a par with establishing ontology: "We cannot know what something is without knowing how it is marked off from other things. Identity is thus of a piece with ontology" (1969, p. 55).

Remarkably, identity and thus ontology can be defined anew relative to a given theory. The idea is to *identify indiscernibles*.[32] If, for example, our linguistic theory does not provide descriptive means to specify the spatio-temporal extension of inscriptions but only e.g. word form predicates then it is

[31] cf. Giere (1995) for a more modest view.

[32] Quine (1960, §47).

more than natural to identify all inscriptions of the same word form since there is no way to distinguish them within the theory. We thereby *reduce* our ontology of inscriptions to one of word forms.[33]

Spatio-temporal identity and ontology are still part of the *background theory*, i.e. physics. As long as we are working within our theory we do not have to bother about utterances and inscriptions. But we have to be always prepared to *invert* the reduction and move to the background theory when empirical content is the issue. One can only hope that Pollard and Sag (1994) have this in mind when they propose that theorizing with empirical content is possible without decisions whether the entities are in the mind or in the (external) world and how they relate to tokens.

If the basic vocabulary of a theory consists of attribute-value predicates as introduced in section 4.1 then its ontology does not distinguish objects which satisfy the same set of predicates. But we should go further: since *Avl* captures logical equivalence it makes no sense to say that different sets of predicates which are equivalent with respect to *Avl* are of use to differentiate objects. Consequently, the ontology of the theory is reduced to that of feature structures – the points of the locale $\mathcal{L}(Avl)$. (6) then defines ontological reduction.

By the same line of reasoning one could regard the extension of an (observational) theory while keeping the basic vocabulary as an example of ontological reduction. Points of the corresponding locale that cannot be distinguished anymore by non-equivalent descriptions are identified. Taking up the discussion in section 3.2 again, the move from $\mathcal{L}(Avl)$ to the sublocale $\mathcal{L}(Avl \cup App)$ is ontological reduction.[34]

Partiality

The notion of *partiality* is of some virulence in the community interested in constraint-based approaches. First, it is more than obscure to speak of *partial objects*. One can of course refer to *parts* of objects, given a part-whole relation for the objects in question. A primary candidate in our context is the relation $\sqsubseteq$ on the point set of the locale presented by an observational theory – using identification of indiscernibles and ontological reduction again. However, it is by no means clear what to count as part and what as whole since the natural language expressions ‘part’ and ‘whole’ do not apply anymore. Indeed, both choices can be defended.

As for the *partial description* view,[35] again it makes only sense *relative* to a given frame of descriptions (in the informal or formal sense of the word). It has to be clear what it means that one description is less partial then the other. The locale of a theory continues to serve as a generic example.

4.3 Cognitive Theory

Knowledge of Grammar

According to the Chomskian paradigm, linguistics is concerned with the study of the knowledge of language internally represented in the human mind.[36] Pollard (1999, p. 281) subscribes to this position as he writes that “grammars exist in the real world, more specifically the part of the world inside language knower’s minds.” A formal grammar then “is a mathematical idealization of the mental grammar.” Such an attitude cannot be called scientific if it does not take serious its ontological commitments, that is, gives identity criteria for minds, knowledge, and mental grammars.

Following for example Davidson (1989) and Quine (1995) one can abandon such things as minds altogether. The crucial idiom ‘to know that’ which serves to express a *propositional attitude* is best analyzed as *relational*: states of human beings (and other animals) are related to sentences.[37] States are here understood to be *physical*, that is, ontologically determined by their spatio-temporal extension. Objects of knowledge, on the other hand, are abstract objects like sequences of signs or related structures which gain their identity by way of formal specification.

An analogy might help: ascribing knowledge to people is like ascribing weight to objects. Reflection on the predicational structure of the latter ascription reveals it at relational: *numbers* are related to objects by, for example, the dyadic predicate ‘is weight in kilogram of’. Hence, “the entities we mention to help specify a state of mind do not have to play any *psychological* or epistemological role at all, just as numbers play no physical role” (Davidson, 1989, p. 11). Some readers may feel more at

[33] This is not to say that word forms are enough. Linguistics needs to talk about *sequences* of word forms. And sequences must not be identified with classes of tokens; see Quine (1969, p. 42). Their identity is defined by way of formal specification.

[34] The injection on the point side is Quine’s *proxy function*; e.g. (1969, p. 56).

[35] see e.g. Johnson (1988).

[36] see e.g. Chomsky (1986), (1995).

[37] The idea goes at least back to Carnap.

ease with this view when the nature of grammatical knowledge is recalled: it is *tacit* knowledge which the language user is *not* aware of. Exactly for this reason Chomsky suggested to use the neologism ‘to cognize’ instead of ‘to know’.

Ontology is one thing, empirical content is another. A theory of grammatical knowledge in itself lacks empirical content. What is needed in addition is a theory of language use, that is, a theory of how a speaker/hearer puts her grammatical knowledge to use. The paradigmatic behavioral task (which is tacitly also adopted by Chomsky) is that of grammatical judgment. Its underlying assumption is that native speakers/hearers use their grammatical knowledge in “unconscious acts of inference” to deduce whether an observed utterance is grammatical or not.

Cognitive Dynamics

A theory of grammatical knowledge is, again in Chomskian terms, about the state a person’s language faculty has attained after language acquisition, or, better, about the “objects of knowledge” this state bears the “cognizing” relation to. Within our framework we can assume this relation to be borne to a grammatical theory in form of a constraint system.

Grammatical knowledge remains, under reasonable idealizations, constant during language processing. Conditional constraints belong to the background knowledge of the speaker/hearer. Language processing during speaking and hearing is thus not concerned with the acquisition of grammatical constraints but with their application to momentary knowledge about certain parts of speech. Observational logic nicely reflects this distinction between permanent and transient information.

A fully developed theory of language processing has to be explicit about the time course of the inferences which the speaker/hearer draws (unconsciously) from its background grammatical knowledge. This asks for a setting similar to epistemic logic, not in the Hintikka style with its strongly idealized assumption of logical omniscience, but for a formal framework rich enough to express the succession of cognitive states and the dynamics of processing in detail.[38]

One of the merits of our framework for theories of cognitive processing is that it is uncommited with respect to which “pieces of information” to count as *cognitively equivalent*, that is, as characterizing identical processing states. For example, assuming a feature-based constraint system, even cognitive equivalence with respect to *Avl* could be denied, in which case *Avl* is used for cognitive inference. The other extreme is to count *every* equivalence of the grammar as cognitive. In that case, perceiving word forms coincides with perceiving all grammatical information about the speech segment – an assumption which in a way caricatures J. J. Gibson’s (1979) idea of *direct perception*. An obvious problem for this Gedankenexperiment is the presence of alternatives, that is, ambiguity. Decisions and therefore cognitive acts are then unavoidable.

An even more demanding task is to develop a *theory of language acquisition*. It would be a theory of how conditional constraints are acquired, that is, a theory of theory dynamics. And, maybe, the intuitionistic conditional is of some use in this context.

A Constraints, Frames, and Locales

This appendix is to a large part distilled from Vickers (1989), Johnstone (1982, chapter II) and Mac Lane and Moerdijk (1992, chapter IX) (on frames and locales), Droste and Göbel (1990) and Barwise (1992) (on constraint systems), as well as Davey and Priestley (1990) (on lattices and order).

A.1 Constraint Systems

Let G be a set of elementary descriptive elements (affirmative assertions, observations, predicates, infons, or whatever).

Definition 3: A *constraint system*[39] is a pair $\langle G, \preccurlyeq\rangle$ consisting of a set G and a binary relation $\preccurlyeq$ borne by finite subsets of G to arbitrary subsets of G. The constraint system is called *coherent* if it is restricted to finite consequents, and *definite* if the latter contain at most one element.

The intended reading of ‘$U \preccurlyeq V$’ is that if all elements of U are satisfied then at least one of V. Convention: ‘$U \preccurlyeq x, y$’ stands for ‘$U \preccurlyeq \{x, y\}$’, etc.

Definition 4: A subset X of G is *closed* with respect to $\langle G, \preccurlyeq\rangle$ iff it is non-empty and

$$U \preccurlyeq V \;\rightarrow\; (U \subseteq X \;\rightarrow\; X \cap V \neq \varnothing)\,.$$

[38] See e.g. van Benthem (1996) and Muskens et. al. (1997) for promising frameworks.

[39] also known as *sequent structure* or *non-deterministic information system* or *theory*.

Let $\mathcal{C}(L)$ be the partially ordered set (poset) of closed sets of a constraint system L ordered by inclusion.

In other words, a non-empty set is closed if it contains with each antecedent of a constraint at least one element of its consequent. Constraints of the form $\langle U, \varnothing \rangle$ are referred to as *exclusive*.

Proposition 3: $\mathcal{C}(L)$ is upwards directed-complete. It is also downwards directed-complete if L is coherent.

Proof: Easy; e.g. Droste and Göbel (1990, p. 292) ∎

Note that since we do not count $\varnothing$ as closed, $\mathcal{C}(L)$ does not necessarily have a least element.

Definition 5: Non-empty subsets of closed sets are called *consistent*. A *deductive closure* of a consistent set X is a minimal element of the set of closed supersets of X.

Definition 6: A constraint system L is *Post complete* if $\mathcal{C}(L)$ is an antichain, that is, if each closed set is maximal consistent.

Proposition 4: If a constraint system is coherent then every consistent set has a deductive closure.

Proof: Proposition 3 and Zorn's lemma ∎

If L is definite then $\mathcal{C}(L)$ is obviously closed under intersection. Therefore, each consistent set X has a *unique* deductive closure, namely

$$\bigcap \{Y \in \mathcal{C}(L) \mid X \subseteq Y\}.$$

Proposition 5: If L is definite, $\mathcal{C}(L)$ is bounded complete ∎[40]

The following definition presents the standard method to extend a constraint set $\preccurlyeq_L$ without changing $\mathcal{C}(L)$.

Definition 7: A coherent constraint system $\langle G, \preccurlyeq \rangle$ is *normal* iff $\preccurlyeq$ is closed under *identity*, *weakening*, and *cut*, that is, iff

$$a \preccurlyeq a\,,$$
$$\text{if } U \preccurlyeq V, U \subseteq S, \text{and } V \subseteq T \text{ then } S \preccurlyeq T\,,$$
$$\text{if } U, a \preccurlyeq V \text{ and } U \preccurlyeq a, V \text{ then } U \preccurlyeq V$$

for every element a and all finite subsets $S, T, U,$ and V of G.

[40] 'Bounded complete' means for us that every *non-empty*, bounded subset has a least upper bound.

Normal, definite constraint systems are more or less the same as *Scott information systems*.[41]

Proposition 6: For every coherent constraint system L their is a smallest normal set of constraints $\vdash_L$ containing $\preccurlyeq_L$, called the *entailment relation* determined by L ∎

Proposition 7: Identity, weakening, and cut are sound and complete in the following sense: Given a coherent system L, then $U \vdash_L V$ iff

$$U \subseteq X \;\rightarrow\; X \cap V \neq \varnothing$$

for every element X of $\mathcal{C}(L)$.

Proof: Standard; e.g. Barwise (1992, p. 176) ∎

A.2 Frames

As for combinations of descriptions we allow finite conjunctions and arbitrary disjunctions subject to the usual logical identities such as commutativity, idempotency, etc. and the *infinite distributive law*:

$$x \wedge \bigvee Y \;=\; \bigvee \{x \wedge y \mid y \in Y\}\,.$$

More formally, descriptions are elements of the *frame* $Fr\langle G\rangle$ *freely generated by* G.

Definition 8: A *frame* is a partially ordered set where subsets have joins, finite subsets have meets, and binary meets distribute over arbitrary joins. *Frame morphisms* are functions preserving finite meets and arbitrary joins.

A simple example is the power set $\wp(X)$ of a set X with union and intersection as join and meet. More interesting are *subframes* thereof, that is, subsets of $\wp(X)$ closed under infinite union and finite intersection, better known as a *topology* on X.

Note that in a frame *every* subset has a meet, namely the join of its lower bounds. Frames are therefore the same as *complete lattices* which satisfy infinite distributivity. Consequently, defining $x \Rightarrow y$ as $\bigvee\{z \mid z \wedge x \leqslant y\}$ makes frames, as ordered sets, indistinguishable from *complete Heyting algebras* (which is a common source of confusion).[42]

Definition 9: A frame A is *freely generated* by a set G iff there is a function η from G to A such that for every frame B and every function θ from G to B there is a unique frame morphism f from A to B with $\theta = f \circ \eta$.

[41] see e.g. Zhang (1994). Constraints of the form $\langle \{a\}, \varnothing \rangle$ have to be excluded.

[42] Recall that $x \leqslant y \;\leftrightarrow\; x \wedge y = x$.

It follows immediately from this *universal property* that a frame freely generated by a given set is unique up to isomorphism.[43] *Existence* has to be proved separately:

Proposition 8: Every set freely generates a frame.

Proof (sketch):[44] Let G be a set and S the set of finite subsets of G ordered by inclusion. (S is the semilattice freely generated by G.) Take A as the set of decreasing subsets of S ordered by inclusion and η as the function taking each element x of G to $\{\{x\}\}$. Show that A is a frame and check the universal property of definition 9 ∎

It is important to note that there can be *no* complete Heyting algebra freely generated by a countably infinite set.[45] (Replace 'frame' in definition 9 by 'Heyting algebra'. The crucial difference is that morphisms of Heyting algebras preserve the Heyting conditional.)

With respect to a constraint system we regard a description more "informative" then another if and only if the first entails the latter. In particular, $\bigwedge U \leqslant \bigvee V$ whenever $U \preccurlyeq V$.[46] This motivates the following

Definition 10: A frame A is *freely generated* or *presented* by a constraint system $\langle G, \preccurlyeq\rangle$ iff there is a function η from G to A such that

$$U \preccurlyeq V \ \rightarrow\ \bigwedge \eta(U) \leqslant \bigvee \eta(V)$$

and every function from G to a frame B satisfying this condition factors uniquely through η by a frame morphism f from A to B.

Existence can be checked by using proposition 8 and standard methods of universal algebra:

Proposition 9: Every constraint system freely generates a frame.[47]

Proof: Take the *quotient frame* of $Fr\langle G\rangle$ modulo the *congruence* given by the least *congruence preorder*[48] on $Fr\langle G\rangle$ containing $\langle \bigwedge U, \bigvee V\rangle$ for every constraint $\langle U, V\rangle$ of $\langle G, \preccurlyeq\rangle$ ∎

[43] Take any introduction to universal algebra.

[44] e.g. Johnstone (1982, section II, 1.2)

[45] see e.g. Vickers (1989, p. 50) and Johnstone (1982, p. 33f). Another reflection of this categorical difference is that the frame freely generated by a unit set consists of three elements whereas the Heyting algebra is infinite.

[46] Recall that $\bigwedge \varnothing = \top$ and $\bigvee \varnothing = \bot$.

[47] A natural task to pursue, which we have to put aside, is to make constraint systems into a category by defining appropriate morphisms such that the free generation of frames gives rise to an adjunction of categories.

[48] see e.g. Vickers (1989, p. 72) for background.

Constraint systems are essentially the same as propositional *geometric theories*. *Geometric formulae* (over G) are elements of $Fr\langle G\rangle$. *Axioms* are pairs of geometric formulae, intended as implications. A straightforward adjustment of definition 10 gives rise to the notion of a frame presented by a theory. Existence can be proved in the same way as above.

Constraints correspond to axioms of a certain normal form consisting of conjunctions of subbasics in the first and disjunctions of subbasics in the second component. On the other hand, every axiom corresponds to a set of constraints: represent its antecedent as a disjunction of finite conjunctions and its consequent as a finite conjunction of disjunctions and split the axiom in the obvious way. (An immediate consequence of this equivalence between constraint systems and geometric theories is that every frame is freely generated by a constraint system: simply take all inequalities that hold in the frame as axioms.)

A.3 Locales

Let **2** be the frame $\{\bot, \top\}$ freely generated by $\varnothing$. To study *duality* means to investigate how A relates to the set $A^\star$ of frame morphisms from A to **2**. Note that any set of functions into **2** carries a partial order: $f \leqslant g$ iff $f(a) \leqslant g(a)$ for every a.

Lemma 2: If the frame A is freely generated by a constraint system L then $A^\star$ and $\mathcal{C}(L)$ are isomorphic posets.

Proof: Let L be $\langle G, \preccurlyeq\rangle$. By definition, there is a one-to-one correspondence between frame morphisms from A to **2** and those functions g from G to **2** which satisfy

$$U \preccurlyeq V \ \rightarrow\ (\bigwedge g(U) = \top \ \rightarrow\ \bigvee g(V) = \top)\,.$$

The latter coincide with the characteristic functions of the closed subsets of G ∎

It is extremely useful for technical and conceptual clarity to study duality within the more general setting of topological systems introduced by Vickers.

Definition 11: A *topological system* is a triple $\langle X, A, \models\rangle$ consisting of a set X, whose elements are called *points*, a frame A, whose elements are called *opens*, and a relation $\models$ being borne by points to opens such that

$$x \models \bigwedge F \quad \text{iff} \quad \forall a (a \in F \rightarrow x \models a)$$
$$x \models \bigvee S \quad \text{iff} \quad \exists a (a \in S \ \&\ x \models a)$$

for finite subsets F and arbitrary subsets S of A. If $x \models a$ then x is said to *satisfy* a and a is said to *denote* x.

For example, a *classification* $\langle X, G, \models \rangle$ in the sense of Barwise and Seligman (1997) gives rise to the topological system $\langle X, Fr\langle G \rangle, \models \rangle$.

Definition 12: A *morphism* of topological systems from $\langle X, A, \models \rangle$ to $\langle Y, B, \models \rangle$ is a pair $\langle \gamma, f \rangle$ consisting of a function γ from X to Y and a frame morphism f from B to A such that

$$x \models f(b) \quad \text{iff} \quad \gamma(x) \models b\,.$$

Definition 13: A *locale* is a topological system of the form $\langle A^\star, A, \models \rangle$ with $x \models a$ iff $x(a) = \top$.

A morphisms of locales from $\langle A^\star, A, \models \rangle$ to $\langle B^\star, B, \models \rangle$ is by definition 12 of the form $\langle f^\star, f \rangle$ where f is a frame morphism B to A (note the reverse direction) and $f^\star$ takes each element x of $A^\star$ to $x \circ f$.

Definition 14: x *subsumes* y (or y *specializes* x or x *approximates* y) iff for every open a, if $x \models a$ then $y \models a$. Notation: $x \sqsubseteq y$.

A point x of a locale subsumes a point y iff $x \leqslant y$. Therefore, via lemma 2, if the frame is freely generated by a constraint system L, subsumption means inclusion in $\mathcal{C}(L)$.

In the following, the frame presented by a constraint system L is denoted by '$\mathcal{F}(L)$' and the corresponding locale by '$\mathcal{L}(L)$'.

Let L and L' be constraint systems with the same set G of subbasics and suppose $\preccurlyeq_L$ is a subset of $\preccurlyeq_{L'}$. Then each L'-closed subset of G is also L-closed, i.e. $\mathcal{C}(L')$ is a subset of $\mathcal{C}(L)$. Definition 10, on the other hand, gives an induced frame (epi)morphism from $\mathcal{F}(L)$ to $\mathcal{F}(L')$.

$$(7)\qquad \begin{array}{ccccc} \preccurlyeq_L & : & \mathcal{C}(L) & \models & \mathcal{F}(L) \\ \cap\!\mid & & \cup\!\mid & & \downarrow \\ \preccurlyeq_{L'} & : & \mathcal{C}(L') & \models & \mathcal{F}(L') \end{array}$$

One verifies easily that this defines a locale morphism from $\mathcal{L}(L')$ to $\mathcal{L}(L)$. $\mathcal{L}(L')$ is then called a *sublocale*[49] of $\mathcal{L}(L)$.

We conclude with the observation that coherent constraint systems give rise to "good" locales in the sense that they have enough points.

Definition 15: The *extent* of an open a of a locale D is the set of points of D that satisfy a.

[49]see Johnstone (1982) or Vickers (1989) for the general definition.

Definition 16: A locale D is *spatial* (or said to *have enough points*) if opens with identical extent are identical.

Definition 17: A locale is *spectral* if its frame is coherent.[50]

Proposition 10: Spectral locales are spatial.

Proof: Vickers (1989, p. 120) or use proposition 7 ∎

[50]Vickers' terminology; Johnstone calls them coherent.

References

Hassan Aït-Kaci, Andreas Podelski, and Gert Smolka. 1994. A feature logic constraint system for logic programming with entailment. *Theoretical Computer Science*, 122:263–283.

Jon Barwise and Jerry Seligman. 1997. *Information Flow. The Logic of Distributed Systems*. Cambridge Tracts in Theoretical Computer Science 44. Cambridge University Press, Cambridge.

Jon Barwise. 1992. Information links in domain theory. In S. Brookes, M. Main, A. Melton, M. Mislove, and D. Schmidt, editors, *Proceedings of the 7th International Conference on the Mathematical Foundations of Programming Semantics*, LNCS 598, pages 168–192, Berlin. Springer.

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

Noam Chomsky. 1986. *Knowledge of Language*. Praeger, New York.

Noam Chomsky. 1995. Language and nature. *Mind*, 104:1–61.

B. A. Davey and H. A. Priestley. 1990. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge.

Donald Davidson. 1989. What is present to the mind? In Johannes Brandl and Wolfgang L. Gombocz, editors, *The Mind of Donald Davidson*, Grazer Philosophische Schriften, pages 3–18. Rodopi, Amsterdam.

Manfred Droste and Rüdiger Göbel. 1990. Non-deterministic information systems and their domains. *Theoretical Computer Science*, 75:289–309.

Gerald Gazdar, Geoffrey K. Pullum, Robert Carpenter, Ewan Klein, Thomas E. Hukari, and

Robert D. Levine. 1988. Category structures. *Computational Linguistics*, 14(1):1–19.

J. J. Gibson. 1979. *An Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA.

Ronald N. Giere. 1995. The skeptical perspective: Science without laws of nature. In Friedel Weinert, editor, *Laws of Nature*, pages 120–138. de Gruyter, Berlin.

Laurence R. Horn. 1989. *A Natural History of Negation*. University of Chicago Press, Chicago.

Mark Johnson. 1988. *Attribute-Value Logic and the Theory of Grammar*. CSLI Lecture Notes, No. 16. CSLI Publications, Stanford, CA.

Peter T. Johnstone. 1982. *Stone Spaces*. Cambridge Studies in Advanced Mathematics 3. Cambridge University Press, Cambridge.

Achim Jung. 1990. Cartesian closed categories of algebraic cpos. *Theoretical Computer Science*, 70:233–250.

Paul John King. 1994. An expanded logical formalism for head-driven phrase structure grammar. Technical Report 59 – 1994, Sonderforschungsbereich 340, Tübingen.

Marcus Kracht. 1995. Is there a genuine modal perspective on feature structures? *Linguistics and Philosophy*, 18:401–458.

Karel Lambert and Gordon G. Brittan. 1987. *An Introduction to the Philosophy of Science*. Ridgeview, Atascadero, CA, 3. edition.

Saunders Mac Lane and Ieke Moerdijk. 1992. *Sheaves in Geometry and Logic*. Springer, New York.

Michael Makkai and Gonzalo E. Reyes. 1977. *First Order Categorical Logic*. Lecture Notes in Mathematics 611. Springer, Berlin.

Ernest G. Manes and Michael A. Arbib. 1986. *Algebraic Approaches to Program Semantics*. Springer, New York.

M. Andrew Moshier and Carl J. Pollard. 1994. The domain of set-valued feature structures. *Linguistics and Philosophy*, 17:607–631.

M. Andrew Moshier. 1993. On completeness theorems for feature logics. *Annals of Mathematics and Artificial Intelligence*, 8:175–213.

Reinhard Muskens, Johan van Benthem, and Albert Visser. 1997. Dynamics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 587–648. North-Holland, Amsterdam.

Fernando C. N. Pereira and Stuart M. Shieber. 1984. The semantics of grammar formalisms seen as computer languages. In *Proceedings of the 10th International Conference on Computational Linguistics*, pages 123–129.

Carl J. Pollard and M. Andrew Moshier. 1990. Unifying partial descriptions of sets. In P. H. Hanson, editor, *Information, Language, and Cognition*, Vancouver Studies in Cognitive Science, Vol. 1, pages 285–322. Oxford University Press, Oxford.

Carl J. Pollard and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Vol. 1*. CSLI Lecture Notes, No. 13. CSLI Publications, Stanford, CA.

Carl J. Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Carl J. Pollard. 1999. Strong generative capacity in HPSG. In Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol, editors, *Lexical and Constructional Aspects of Linguistic Explanation*, pages 281–297. CSLI Publications, Stanford, CA.

W. V. Quine. 1960. *Word and Object*. MIT Press, Cambridge, MA.

W. V. Quine. 1969. Ontological relativity. In *Ontological Relativity and Other Essays*, pages 26–68. Columbia University Press, New York.

W. V. Quine. 1976. Algebraic logic and predicate functors. In *The Ways of Paradox and Other Essays*, pages 284–307. Harvard University Press, Cambridge, MA, 2. edition.

W. V. Quine. 1982. *Methods of Logic*. Harvard University Press, Cambridge, MA, 4. edition.

W. V. Quine. 1995. *From Stimulus to Science*. Harvard University Press, Cambridge, MA.

William C. Rounds. 1997. Feature logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 475–533. North-Holland, Amsterdam.

Stuart M. Shieber. 1992. *Constraint-Based Grammar Formalisms*. MIT Press, Cambridge, MA.

Gert Smolka. 1992. Feature-constraint logics for unification grammars. *Journal of Logic Programming*, 12:51–87.

Johan van Benthem. 1996. *Exploring Logical Dynamics*. Studies in Logic, Language, and Computation. CSLI Publications, Stanford, CA.

Bas C. van Fraassen. 1980. *The Scientific Image*. Clarendon Press, Oxford.

Steven Vickers. 1989. *Topology via Logic*. Cambridge Tracts in Theoretical Computer Science 5. Cambridge University Press, Cambridge.

Steven Vickers. 1993. Geometric logic in computer science. In G. L. Burn, S. J. Gay, and M. D. Ryan, editors, *Proceedings of the First Imperial College Department of Computing Workshop on Theory and Formal Methods*, pages 37–54.

Steven Vickers. 1999. Topology via constructive logic. In Jonathan Ginzburg, Lawrence S. Moss, and Maarten de Rijke, editors, *Logic, Language, and Computation, Vol. 2*. CSLI Publications, Stanford, CA.

Wolfgang Wechler. 1992. *Universal Algebra for Computer Scientists*. Springer, Berlin.

Guo-Qiang Zhang. 1992. Disjunctive systems and L-domains. In *Proceedings of the 19th International Colloquium on Automata, Languages, and Programming*, Lecture Notes in Computer Science 623, pages 284–295, Berlin. Springer.

Guo-Qiang Zhang. 1994. A representation of SFP. *Information and Computation*, 110(1):233–263.