Constructing Lexical Semantic Hierarchies from Binary Semantic Features

Rainer Osswald

Intelligent Information and Communication Systems
Fachbereich Informatik, FernUniversität in Hagen
http://pi7.fernuni-hagen.de/osswald/

Abstract. The paper employs and extends methods from Formal Concept Analysis for constructing semantic hierarchies from lexical classifications by binary semantic features. In particular, we show how the dichotomic character of binary features can be captured within an appropriate logical framework that makes recourse to intuitionistic logic. The investigation is motivated by the use of semantic features in a large-scale lexical semantic database.

1 Introduction

Binary semantic features have a long tradition in structural semantics (see e.g. [16]). Although of limited expressiveness, especially with respect to relational information, they are nevertheless useful for classificational purposes. The work presented in this paper is motivated by the use of binary semantic features within HaGenLex, a large-scale computer lexicon for German, where features such as animate and artificial are employed to classify the lexical entries as well as to specify selectional restrictions in the case frames of the entries. The semantic features used in HaGenLex are not necessarily independent of each other. For instance, animate:+ implies artificial:−. To prevent inconsistent choices by lexicographers, the possible feature combinations had therefore been predefined in a tree-shaped taxonomy. Such an approach, however, is problematic for two reasons: first, it usually overlooks possible combinations since it is not data-driven, and, second, it imposes an unmotivated ordering on features that are independent of each other.

In order to construct a more feasible semantic hierarchy of feature combinations, it is tempting to employ methods from Formal Concept Analysis. The most natural formalization of the data under consideration is a many-valued context, with lexical entries as objects and semantic features as many-valued attributes. The method of attribute exploration can then be applied to systematically (re)design the semantic hierarchy. However, when the standard version of attribute exploration is used on the derived context with respect to plain scaling, the well-known problem arises that scale information is also explored instead of being assumed as background knowledge. The inconsistency of animate:+ and animate:−, for example, should be treated as background knowledge without explicit affirmation of the user. Similarly, if the user has already accepted
that animate:+ implies artificial:− then he should not be asked whether artificial:+ implies animate:− since this follows by contraposition.

So, the key point of coping with binary features seems to be that F:− should be regarded as the logical negation of F:+. Classical logic, however, is too strong for our purposes because the law of excluded middle would require that every lexeme either satisfies F:+ or F:−, for every semantic feature F, whereas certain lexemes are inherently underspecified with respect to certain semantic features. We will argue in Section 4 that intuitionistic logic provides an appropriate framework for viewing F:− as the negation of F:+. Based on this insight, we show in Section 5 that Horn logic can be extended by a simple inference scheme that captures the dichotomic character of positive and negative feature values and covers contraposition without recourse to full classical logic.

2 Use Case: Lexical Semantic Classification

The theoretical investigation presented in the following has been initiated by the use of semantic features in HaGenLex (Hagen German Lexicon), a domain independent lexicon for German that currently comprises about 25,000 lexical entries [12]. All HaGenLex entries are semantically annotated, where the semantic description is based on the MultiNet paradigm, a knowledge representation formalism developed for the representation of natural language semantics [13]. MultiNet provides classificatory as well as relational means of representation, including a set of semantic roles for specifying semantic role frames in the lexicon.

In the context of the present paper, we restrict ourselves to the semantic classification of lexemes with respect to their ontological sort and semantic features. MultiNet provides a hierarchy of 45 ontological sorts such as d (discrete object), s (substance), and abs (situational object). In addition, lexemes are classified with respect to 16 binary semantic features such as human, movable, and artificial. These features are not independent of each other; for instance, human:+ implies animate:+ and artificial:−. To avoid inconsistencies the possible feature combinations are explicitly listed as named semantic classes, on which a natural specialization hierarchy is defined. All in all, there are more than 80 named semantic classes in HaGenLex.

Some of the semantic classes bear semantic features that are unspecified. Formally, this is captured by allowing features to have the value boolean, which generalizes + and −. The semantic class animate-object, for example, is unspecified with respect to the features animal, human, and movable. The purpose of such classes is to capture “natural” generalizations. Figure 1 depicts the hierarchy of all subclasses of con-object (short for concrete object) as used in HaGenLex — with features only displayed if their value is more restricted than that of the immediate superclass.

The hierarchy of semantic classes used in HaGenLex has been manually determined on the basis of lexicographic practice, with an eye on the semantic dependencies between feature values. No attempt has been made to systematically employ implications between feature values to constrain the possible feature-
Fig. 1. Semantic subclasses of con(crete)-object used in HaGenLex (see [11, p. 16])
value combinations nor to reveal the implications valid for the chosen system of semantic classes. The hierarchy of semantic classes therefore may implicitly contain assumptions about feature-value dependencies that are not intended by the designer of the hierarchy. This is problematic because, once adopted, the hierarchy gains normative impact when applied in classification tasks, say, of assigning semantic classes to lexemes, where the user is forced to choose from the given set of classes. Mnemonic designators for semantic classes such as natural discrete (for natural discrete object) make it even more difficult to detect defects in the hierarchy since they hide the actual distribution of feature values.

3 Formal Contexts and Concept Hierarchies

In the language of Formal Concept Analysis, the semantic classification of lexemes by binary semantic features can be seen as a many-valued formal context, where the lexemes are the formal objects, the semantic features are the (many-valued) formal attributes and $b(\text{olean})$, $+$, and $-$ are the attribute values. In attribute-value based approaches in computational linguistics, the values typically come along with an ordering, usually a semilattice, often called a type hierarchy [5]. Within Formal Concept Analysis, in contrast, the ordering on the values is usually defined as the concept lattice given by a certain formal context known as a scale. Nevertheless, there is a close connection between these two views via the general ordinal scale of the type hierarchy, whose concept lattice is the Dedekind-MacNeille completion of the hierarchy (see [10, p. 48]). In case the type hierarchy is a join semilattice, this construction just adds a least element. For the particularly simple case of binary features, the situation is illustrated in Figure 2.

Turning the many-valued context into a “one-valued” one by plain scaling essentially means to consider all feature value combinations of the form $F:b$, $F:+$ and $F: -$ as atomic attributes. Since we do not distinguish between underspecified and inappropriate features in this paper, we could equally well drop the attribute $F:b$; the information that a lexeme is underspecified with respect to $F$ is simply expressed by that fact that it does not satisfy the attribute $F:+$ nor $F: -$. There is, however, a technical problem with respect to plain scaling in this case.\(^1\) If

---

\(^1\) An interesting topic beyond the scope of this paper is the relation of underspecification to incompleteness; cf. [3].

\(^2\) See also the discussion in [10, p. 60].
we remove the row and column of the value \( b \) from the scale in Figure 2, we get the so-called dichotomic scale. Although the structure of the concept lattice generated by the scale remains the same, the method of plain scaling would now give rise to a dichotomic context, in which every lexeme is required to satisfy either \( + \) or \( - \). This would be an inappropriate requirement for lexical semantic classification because certain lexemes such as creature are inherently undetermined with respect to the feature human, for instance.

Therefore, the most suitable choice for formalizing the classification by binary features appears to be a formal context whose attribute set \( \Sigma \) is “dichotomized” in the sense that it has the form \( F \cup (-F) \), with \( -F = \{ -p \mid p \in F \} \), where no object satisfies both \( p \) and \( -p \), for all \( p \in F \). Lacking a better name, we speak of a semi-dichotomic context in this case. There are two key questions to be addressed in the rest of the paper:

1. Of which form are the concept hierarchies associated with semi-dichotomic contexts arising from lexical semantic classification by binary features?
2. What is the proper way of drawing logical inferences about dependencies between the attributes in such a context?

Before we go into more detail on these issues, let us review some facts about the interplay between attribute logic and the structure of the resulting concept hierarchies (see [7–9, 18]). Suppose \( \langle U, \Sigma, \models \rangle \) is a formal context.\(^3\) The extent \( p^\prec \) of an atomic attribute \( p \in \Sigma \) is \( \{ x \in U \mid x \models p \} \). Attributes can be inductively combined by means of conjunction, disjunction and negation, with conditional and biconditional defined as usual in classical logic. The extent of these compound attributes is defined compositionally: \((\phi \land \psi)^\prec = \phi^\prec \cap \psi^\prec\), \((-\phi)^\prec = U \setminus \phi^\prec\), etc. It is moreover convenient to introduce two additional attributes \( V \) and \( \Lambda \) with \( V^\prec = U \) and \( \Lambda^\prec = \emptyset \). We say that \( x \in U \) satisfies a compound attribute \( \phi \) (with respect to the given context) iff \( x \in \phi^\prec \).

We use \( \forall \) as an operator that turns compound attributes into what we call (universal) statements. If \( \phi \) is a compound attribute, we say that the statement \( \forall \phi \) holds or is true in the context if \( \phi^\prec = U \).\(^4\) Similarly, a set of statements is said to hold in the context if every of its statements holds. We write \( \phi \preceq \psi \) for \( \forall (\phi \rightarrow \psi) \) and \( \phi \equiv \psi \) for \( \forall (\phi \leftrightarrow \psi) \). Notice that \( \phi \equiv \psi \) holds in a context iff \( \phi^\prec = \psi^\prec \), in which case \( \phi \) and \( \psi \) are called extensionally equivalent with respect to the context. Two compound attributes are said to be globally equivalent if they are extensionally equivalent in every context. Similarly, we can define extensional and global implication. It is not difficult to see that global equivalence is equivalent to extensional equivalence in the so-called test context \( \langle \wp(\Sigma), \Sigma, \exists \rangle \) ([8, 9]).

We call a compound attribute positive if it is free of \( - \). A statement \( \phi \preceq \psi \) (\( \phi \equiv \psi \)) is said to have (bi)conditional form if \( \phi \) and \( \psi \) are positive. A statement

\(^3\) Throughout this paper, we consider only finite contexts.

\(^4\) With (compound) attributes seen as monadic predicates and \( \forall \phi \) as short for \( \forall x(\phi x) \), this definition of truth is just the standard definition used in first order logic, where \( ^\prec \) serves as an interpretation function with universe \( U \); see [18] for details.
\( \phi \preceq \psi \) is called \textit{implicational} if \( \phi \) is purely conjunctive or \( V \) and \( \psi \) is atomic; it is called a \textit{Horn statement} if \( \phi \) is conjunctive or \( V \) and \( \psi \) is either atomic or \( \Lambda \). If \( \Gamma \) and \( \Gamma' \) are sets of statements then \( \Gamma \) is said to (semantically) \textit{entail} \( \Gamma' \) if \( \Gamma' \) holds in every context in which \( \Gamma \) holds. A \textit{theory} is a set of statements closed under entailment. A \textit{base} of a theory \( \Gamma \) is a non-redundant subset of \( \Gamma \) that entails \( \Gamma \). The \textit{(Boolean) theory} of a context is the set of all statements that hold in the context, the \textit{Horn/implicational theory} of the context is the set of all Horn/implicational statements holding in the context.

Every set \( \Gamma \) of statements over \( \Sigma \) canonically determines a set \( C(\Gamma) \) of \textit{conceptual objects} as a subsystem of \( \wp(\Sigma) \) in the following way: \( X \subseteq \Sigma \) belongs to \( C(\Gamma) \) iff, for every \( (\forall \phi) \in \Gamma, X \) satisfies \( \phi \) in the test context.\(^5\) We speak of “conceptual objects” to emphasize their twofold character: On the one hand, we can regard the elements of \( C(\Gamma) \) as concepts by identifying concepts with sets of atomic attributes. From this perspective, we can refer to \( C(\Gamma) \) ordered by \( \supseteq \) as the \textit{concept hierarchy} generated by \( \Gamma \); a concept \( X \) is a subconcept of a concept \( Y \) in this hierarchy if \( X \) is more specific than \( Y \), i.e., if \( X \supseteq Y \). On the other hand, the elements of \( C(\Gamma) \) can be regarded as the formal objects of the formal context \( (C(\Gamma), \Sigma, \supseteq) \), henceforth called the \textit{generic context} associated with \( \Gamma \). It is not difficult to see that this context is the “largest object-clarified” context with respect to which \( \Gamma \) is true (cf. [18]).

In the terminology of Formal Concept Analysis, the relation between attribute logic and generated concept hierarchies can be described as follows ([9, 18]): The concept hierarchy generated by the Boolean theory of a formal context consists of the set of all object intents of that context. Moreover, the concept hierarchy generated by the implicational theory of the context coincides with the lattice of concept intents of the context. Finally, the concept hierarchy generated by the Horn theory of the context coincides with the concept intents possibly except for the bottom element of the lattice which is eliminated by the Horn theory if it represents just inconsistency. In the following, we always consider the concept hierarchy generated by the Horn theory unless indicated otherwise.\(^6\)

In order to develop a framework for attribute logic of many-valued contexts, [8] introduces the notion of a \textit{relative test context} for a fixed scaling. The idea is that the logic of the scales should be seen as background knowledge and not as part of the actual logic of the context. The relative test context thus intends to incorporate the logic of the scales. Formally, it is defined as the semidirect product of the scales. The objects of the relative test context for the dichotomic scale, henceforth called the \textit{dichotomic test context}, can be identified with all subsets of \( \Sigma = F \cup -F \) which contain either \( p \) or \( -p \), for every \( p \in F \). In a similar vein, we can take the objects of the \textit{semi-dichotomic test context} as those subsets of \( \Sigma = F \cup -F \) which do not contain both \( p \) and \( -p \), for every \( p \in F \). It is easy to see that both test contexts determine the same concept

\(^5\) In [9], \( C(\Gamma) \) is called the \textit{free extent} of \( \Gamma \); in [17–19], \( C(\Gamma) \) ordered by \( \subseteq \) is referred to as the \textit{information domain} of \( \Gamma \).

\(^6\) Since we started in Section 2 with a tree-shaped hierarchy, it should be mentioned that concept trees can be generated as well by appropriate statements; see [20, 4].
**Fig. 3.** Concept hierarchy of the (semi-)dichotomic test context over \{a, b\}

hierarchy.\(^7\) Figure 3 depicts the resulting hierarchy for \(F = \{a, b\}\). The objects of the dichotomic test context are determined by the four sets in the bottom row of the hierarchy whereas the objects of the semi-dichotomic test context are given by all nine sets of the hierarchy. In particular, the semi-dichotomic test context is object-reducible to the dichotomic test context.

Let us now look more closely at the attribute logic of negation and its relation to the (semi-)dichotomic scaling. In the dichotomic test context, \(\neg p\) is extensionally equivalent to \(-p\) because, by definition, every object satisfies either \(p\) or \(-p\). Thus \(V \preceq p \lor \neg p\) holds in the dichotomic test context. In the semi-dichotomic test context, in contrast, \(-p\) extensionally implies \(\neg p\), but not the other way around; for not satisfying \(p\) does not mean to satisfy \(\neg p\) in this context. We have argued above that with respect to constructing lexical semantic hierarchies from binary features we need underspecification, which means to work with semi-dichotomic contexts. At the same time, we need a logical framework that takes the dichotomy between positive and negative feature values into account. In other words, we look for some sort of attribute negation operator \(\sim\) such that \(\sim p\) is equivalent to \(-p\) in the semi-dichotomic test context. In the following, we will see that intuitionistic negation shows the desired behavior.

### 4 Intuitionistic Attribute Logic

The problem with classical negation in the semi-dichotomic test context is reflected by the fact that if a conceptual object satisfies \(\neg p\) we cannot conclude that a more specific object also satisfies \(\neg p\). In other words, classical negation is not persistent with respect to specialization. The “state of knowledge” interpretation of intuitionistic logic, in contrast, regards knowledge as persistent or permanent in the face of an increase of knowledge.\(^8\)

Let us define specialization and persistence more formally: Given a formal context \(\langle U, \Sigma, \vdash \rangle\), the specialization relation \(\sqsubseteq\) on \(U\) is defined such that \(x \sqsubseteq y\) if \(x \vdash p\) implies \(y \vdash p\) for all \(p \in \Sigma\). A subset \(S\) of \(U\) is called upwards closed with respect to \(\sqsubseteq\) if \(\uparrow S \subseteq S\), with \(\uparrow S = \{y \in U \mid \exists x \in S (x \sqsubseteq y)\}\). A (compound)

\(^7\) Turned upside down, the hierarchy associated with the (semi-)dichotomic test context is also known as the truth-value domain; see e.g. [14].

\(^8\) See [21] for an introduction to intuitionistic logic.
attribute is called (globally) persistent if its extent is upwards closed with respect to ⊑ in all contexts. Persistence can be characterized as follows (see e.g. [17]):

**Proposition 1.** A (compound) attribute is persistent iff it is globally equivalent to a positive attribute.  

There is a straightforward way to “persistify” any compound attribute φ by introducing an attribute operator □ such that

\[(□φ)^\mathfrak{a} = \{ x ∈ U \mid \uparrow\{ x \} ⊆ φ^\mathfrak{a} \} .\]

That is, x satisfies □φ just in case every y at least as specific as x satisfies φ. It is easy to see that □φ is persistent. We can now apply this method to define persistent versions of negation and the conditional: Let ~ and ⇒ be attribute operators such that

\[\sim φ = □\sim φ \quad \text{and} \quad φ ⇒ ψ = □(φ → ψ).\]

In particular, \sim φ = φ ⇒ Λ. Defining φ ⇒ ψ and ~φ along these lines means to equip them with a Kripke-style semantics in the sense of intuitionistic logic, where U ordered by ⊑ serves as the Kripke frame. To determine the extent of ~φ in practice it is useful to unravel definitions:

\[(~φ)^\mathfrak{a} = \{ x ∈ U \mid \uparrow\{ x \} ∩ φ^\mathfrak{a} = ∅ \} .\]

**Example 1.** Consider a convivial evening scenario involving five people, which, as the night wears on, can be classified as shown on the left of Figure 4. The corresponding concept hierarchy is shown on the right of the figure. A base for the Horn theory of the context is:

\[male \land female \preceq Λ \quad drunk \land sober \preceq Λ \quad sober \preceq female\]

\[Recall from Section 3 that a compound attribute is positive if it does not contain ¬, that is, positiveness is a syntactic property.\]

\[The definition of ⇒ and ~ by means of □ essentially resembles Gödel’s embedding of intuitionistic logic into the modal logic S4, which is known to be strongly complete with respect to the class of reflexive and transitive frames; see e.g. [2].\]
Let us determine the extent of $\sim\text{sober}$ in the associated generic context. (Notice that specialization in the generic context is set inclusion.) Since the extent of \textit{sober} is $\{\{\text{sober, female}\}\}$, the extent of $\sim\text{sober}$ consists of $\{\text{male}\}$, $\{\text{drunk}\}$, $\{\text{drunk, male}\}$, and $\{\text{drunk, female}\}$. Hence $\sim\text{sober} \equiv \text{male} \lor \text{drunk}$.

The logical operators for attribute combination within \textit{intuitionistic attribute logic}, as we call it, are conjunction and disjunction, with the same semantics as in Section 3, and the conditional $\Rightarrow$ with the semantics defined above. Moreover, $\sim \phi$ is defined as $\phi \Rightarrow \Lambda$. Consequently, all compound attributes of intuitionistic attribute logic are persistent. Notice that in intuitionistic attribute logic, $\phi \Rightarrow \psi$ is not globally equivalent to $\sim \phi \lor \psi$. In Example 1, for instance, $\{\text{female}\}$ is not in the extent of $\sim \text{sober} \lor \text{sober}$ whereas $\phi \Rightarrow \phi$ is globally equivalent to $\Lambda$.

The following proposition, whose easy proof can be found in [17], tells us how to do inferences in intuitionistic attribute logic:

**Proposition 2.** If $\phi$ is persistent, $\phi \land \psi \leq \chi$ is globally equivalent to $\phi \leq \psi \Rightarrow \chi$.

The corresponding inference schemes are stated as (I$_\Rightarrow$) and (E$_\Rightarrow$) in the Appendix, where a complete inference calculus for intuitionistic attribute logic is given including introduction and elimination schemes for $\land$ and $\lor$ as well as schemes for weakening and cut.

In the context of the present paper, we are primarily interested in extending Horn attribute logic by the rules of intuitionistic negation. In particular, we will make use of the following inference schemes:

\[
\frac{A \leq \sim B}{A \land B \leq A} \quad \frac{A \land B \leq C}{A \land \sim C \leq \sim B}
\]

The first of these schemes is just a special case of (E$_\Rightarrow$); the second scheme, whose straightforward derivation within the calculus given in the Appendix is left to the reader, will be referred to as \textit{generalized (weak) contraposition}. Notice that replacing $C$ by $A$ in the contraposition scheme leads to the reverse of the first scheme since $\sim A \equiv V$ and $A \land V \equiv A$.

Before we apply this framework to the study of semi-dichotomic contexts in Section 5, let us briefly describe how intuitionistic negation and conditional can be characterized algebraically by means of the so-called Lindenbaum construction, whose basic idea is to identify attributes if they are equivalent with respect to a theory. Given a theory $\Gamma$ over $\Sigma$ (in biconditional form), the \textit{Lindenbaum algebra} $L(\Gamma)$ of positive attributes of $\Gamma$ is the quotient $A[\Sigma]/\approx_\Gamma$, where $\phi \approx_\Gamma \psi$ iff $\phi \equiv \psi$ belongs to $\Gamma$, and $A[\Sigma]$ is the set of positive attributes over $\Sigma$. The Lindenbaum algebra of positive attributes is easily seen to be a \textit{finite distributive lattice with zero and unit} with $[\phi \land [\psi] = [\phi \land \psi], [\phi] \lor [\psi] = [\phi \lor \psi], 0 = [\Lambda]$, and $1 = [V]$. Figure 5 depicts the Lindenbaum algebra of the Horn theory of Example 1. Notice that the $\lor$-irreducible elements of the lattice, which are marked by shaded circles, stand in a one-to-one correspondence to the elements of the respective concept hierarchy; cf. Figure 4. This is just an illustration of the general fact that the concept hierarchy generated by a theory and its Lindenbaum algebra are connected by Birkhoff’s duality theorem [6]; see also [19]. (If the concept
Fig. 5. Computing intuitionistic negation algebraically

hierarchy generated by a theory is a join semilattice, as it is the case for Horn theories, then the Lindenbaum algebra of the theory is the distributive lattice freely generated by this join semilattice.) Moreover, in view of Proposition 2, defining \( \alpha \Rightarrow \beta = \bigvee \{ \gamma \mid \gamma \land \alpha \leq \beta \} \) gives rise to an operation \( \Rightarrow \) on \( L(\Gamma) \) such that \( [\phi \Rightarrow \psi] = [\phi] \Rightarrow [\psi] \).\(^{11}\) Figure 4 gives an example of how to compute the intuitionistic negation \( \sim \) sober of Example 1 in this algebra.\(^{12}\)

Returning to semi-dichotomic contexts, let us assume that in Example 1, male and female as well as drunk and sober are introduced as dichotomic pairs. To emphasize this, let us write \( -\text{male} \) for female and \( -\text{sober} \) for drunk. At the close of Section 3, we motivated the introduction of intuitionistic negation as an operation for representing dichotomic pairing. With respect to the context of Example 1, however, \( \sim \) sober is not equivalent to \( -\) sober. The reason is that the Horn theory of the context is not closed under contraposition: it contains the statement sober \( \leq -\) male but not male \( \leq -\) sober. In the following section, we will see that closure under contraposition with regard to dichotomic pairs is exactly what is needed for the Horn theory of a semi-dichotomic context to generate concept hierarchies with the desired properties.

\(^{11}\) The operation \( \Rightarrow \) turns the Lindenbaum algebra into a Heyting algebra; \( \alpha \Rightarrow \beta \) is known as the relative pseudo-complement of \( \alpha \) relative to \( \beta \).

\(^{12}\) A rather different approach towards defining negation in contextual logic is given in [15], where the focus is on appropriate algebraic operations on the concept lattice itself.
5 Dichotomic Concept Hierarchies

In the following, when we speak of the concept hierarchy of a context, we always mean the concept hierarchy generated by the Horn theory of the context. By a dichotomic concept hierarchy over $F$ we mean the concept hierarchy associated with a dichotomic context with attribute set $F \cup \neg F$. For any concept hierarchy $D$ let $m(D)$ be the set of those elements of $D$ that are maximal with respect to the specialization ordering, i.e., minimal with respect to the subconcept-superconcept ordering. The following simple observation is just a reformulation of the fact that the object intents of a dichotomic context are pairwise incomparable, or, put otherwise, that the context and thus its concept lattice are atomistic.

**Lemma 1.** Every element $X$ of a dichotomic concept hierarchy $D$ is the intersection of $\bigcap\{Y \in m(D) \mid X \subseteq Y\}$.

As discussed at the end of Section 4, a semi-dichotomic context does not necessarily support the equivalence of $\sim p$ and $\neg p$. The equivalence holds if the associated concept hierarchy is dichotomic:

** Proposition 3.** With respect to every dichotomic concept hierarchy over $F$ we have $\neg p \equiv \sim p$ (and $p \equiv \sim \neg p$), for all $p \in F$.

**Proof.** Notice that $\neg p \preceq \sim p$ follows from $p \land \neg p \preceq \bot$ by intuitionistic inference. It remains to verify $\sim p \preceq \neg p$. Let $D$ be a dichotomic concept hierarchy over $F$. We need to show that for every $X \in D$, if $X \models \sim p$ then $X \models \neg p$. Suppose $X \models \sim p$, that is, $\forall Y (X \subseteq Y \rightarrow Y \not\models p)$. By Lemma 1, $X = \bigcap\{Y \in m(D) \mid X \subseteq Y\}$. But for all $Y \in m(D)$, if $Y \not\models p$ then $Y \models \neg p$. Hence $X \models \neg p$. (By the same argument, it follows that $p \equiv \sim \neg p$.)
Example 2. The diagram in Figure 6 shows an example of a dichotomic concept hierarchy over \{a, b, c\}, whose corresponding dichotomic context is given by five objects that are classified as indicated in the bottom row of the hierarchy. The following list of statements is a base for the Horn theory of that context:

\[
\begin{align*}
    a \land -a & \preceq A \\
    b \land -b & \preceq A \\
    c \land -c & \preceq A \\
    a \land -b & \preceq c \\
    c \land -b & \preceq -a
\end{align*}
\]

Notice that if \(-\) is regarded as classical negation then contraposition applied to \(a \land -b \preceq c\) would give rise to \(-c \preceq -a \lor b\), which does not hold in the given concept hierarchy. (The example also illustrates the fact that within intuitionistic logic, \(\neg(\phi \land \psi)\) is not equivalent to \(\neg\phi \lor \neg\psi\); for \(\neg(a \land -b)\) is not equivalent to \(-a \lor -c\).)

Now let us conversely assume a concept hierarchy of a semi-dichotomic context over \(F\) such that \(-p \equiv \neg p\) and \(p \equiv \neg-\neg p\) for all \(p \in F\). Let \(\Gamma\) be the corresponding Horn theory over \(\Sigma = F \cup -F\). Then we know from Section 4 that \(\Gamma\) is closed with respect to the following bidirectional inference scheme, henceforth referred to as the dichotomic contraposition scheme:

\[
\frac{A \preceq l}{A \land -l \preceq A}
\]

Here, \(l\) stands for an atomic attribute and we use the notational convention that if \(l = -p\) then \(-l = p\). In particular, this inference scheme allows us to infer \(p \land -p \preceq A\) from \(p \preceq p\).

By a dichotomic theory \(\Gamma\) over \(F\) we mean a Horn theory over \(F \cup -F\) that is closed under dichotomic contraposition. A base of \(\Gamma\) (as a dichotomic theory) is defined as a non-redundant subset of \(\Gamma\) whose closure under Horn inference and dichotomic contraposition generates \(\Gamma\).\textsuperscript{13} In the case of Example 2, such a base is given by the two statements \(a \land -b \preceq c\) and \(c \preceq -b\).

Proposition 4. Dichotomic theories generate dichotomic concept hierarchies.

Proof. Let \(D\) be the concept hierarchy of a dichotomic theory \(\Gamma\). We need to show that every \(X \in D\) is the intersection of \(\mathcal{M}_X = \{Y \in m(D) | X \subseteq Y\}\). The nontrivial part is the claim that \(\bigcap \mathcal{M}_X \subseteq X\). Suppose \(l_0 \in \bigcap \mathcal{M}_X\) and \(X = \{l_1, \ldots, l_n\}\). Then \(-l_0 \notin Y\) for all \(Y \in \mathcal{M}_X\). Hence \((l_1 \land \ldots \land l_n) \cap (-l_0) = \varnothing\). It follows that \(l_1 \land \ldots \land l_n \land -l_0 \preceq A\) belongs to \(\Gamma\). Consequently \(l_1 \land \ldots \land l_n \preceq l_0\) belongs to \(\Gamma\) too, because \(\Gamma\) is dichotomic. Thus \(l_0 \in X\). \(\square\)

\textsuperscript{13} A sound and complete inference calculus for Horn logic is given by the \&-schemes listed in the Appendix, the axioms (R), (Q), and (U), and a cut rule similar to (C) with \(A \preceq B \lor C\) replaced by \(A \preceq B\).
6 Conclusion

In view of the results of Section 5, the answer to the first question addressed in Section 3 concerning the structure of the hierarchies arising from lexical semantic classification can be put briefly as follows: Lexical semantic hierarchies based on binary features are dichotomic. The answer to the second question is two-fold: Opposition in dichotomic pairs is best seen as intuitionistic negation and attribute inference of Horn statements must respect the dichotomic contraposition scheme. In order to apply these results for constructing or revising lexical semantic hierarchies, the next thing to do is to develop an algorithm for attribute exploration that takes the dichotomic contraposition scheme into account. Another future project is to use the resulting semantic hierarchy for improving the machine learning approach presented in [1] that has been developed for extracting semantic features from large text corpora.

References


**Appendix: A Calculus for Intuitionistic Attribute Logic**

\[
\begin{align*}
A \preceq A & \quad \text{(R)} & A \preceq A & \quad \text{(Q)} & A \preceq V & \quad \text{(U)} \\
\frac{A \preceq B}{A \preceq B \land C} & \quad \text{(I}_\land) & \frac{A \preceq B \land C}{A \preceq B} & \quad \text{(E}_1^\land) & \frac{A \preceq B \land C}{A \preceq C} & \quad \text{(E}_2^\land) \\
\frac{A \preceq C}{A \lor B \preceq C} & \quad \text{(I}_\lor) & \frac{A \lor B \preceq C}{A \preceq C} & \quad \text{(E}_1^\lor) & \frac{A \lor B \preceq C}{B \preceq C} & \quad \text{(E}_2^\lor) \\
\frac{A \preceq B}{C \land A \preceq B} & \quad \text{(W}_\land) & \frac{A \preceq B}{A \preceq B \lor C} & \quad \text{(W}_\lor) \\
\frac{A \preceq B \lor C}{A \land B \preceq C} & \quad \text{(C)} \\
A \land B \preceq C & \quad \text{(I}_\Rightarrow) & A \preceq B \Rightarrow C & \quad \text{(E}_\Rightarrow) & A \land B \preceq C & \quad \text{(E}_\Rightarrow)
\end{align*}
\]