Reasoning over Semantic Networks – A Typology of Axioms for Natural Language Inference

Hermann Helbig
Intell. Information & Comm. Systems
FernUniversität in Hagen, Germany
hermann.helbig@fernuni-hagen.de

Ingo Glöckner
Intell. Information & Comm. Systems
FernUniversität in Hagen, Germany
iglockner@web.de

Abstract

There are problems related to question answering which obviously call for semantically based logical inferences. For example, a question-answering system should produce the same answer if asked with logically equivalent paraphrases of a question. Moreover, an answer should not only be found for direct matches of the query with the knowledge base but also if the answer is entailed by the latter. A meaning representation of NL texts and logical axioms defined on this representation must account for these aspects. Starting with logical properties of axioms, we define a typology based on five main characteristics which affect the validity and efficiency of natural language inference.

1 Introduction

Only few systems for open-domain question answering (QA) rely on a semantic analysis of texts and subsequent logical inference. These systems usually solve the problem of representing unconstrained language by resorting to a lossy semantic representation which only captures part of the meaning of the sentences. A typical example is COGEX (Moldovan et al., 2003), an open-domain QA system which utilizes a form of logical inference on a shallow (i.e. syntax-oriented) semantical representation. The axiom base of COGEX was automatically generated from Wordnet glosses, containing a lot of challengeable entailments.

In this paper, we discuss a different formalism specifically designed for the meaning representation of unrestricted language, viz MultiNet (Helbig, 2006), a modern semantic network formalism which is successful in NLP applications like NL interfaces (Leveling and Helbig, 2002) and question answering. The InSicht QA system (Hartrumpf, 2004), which was successfully evaluated in CLEF 2004, operates on a MultiNet knowledge base (KB) generated from 4,9 Million sentences.

Such large QA systems need thousands of axioms for interrelating the meanings of words and for reproducing the expected textual entailments. The XWordNet KB used by the COGEX system, for example, comprises 130,000 non-differentiated axioms. In order to cope with such large axiom sets and control their effects on the validity and efficiency of inferences, we employ extra information for each axiom capturing its main characteristics for knowledge processing. Specifically, we define classificatory criteria for guiding axiom selection in inferences and for describing the reliability of conclusions from a particular axiom.

The MultiNet approach is characterized by a tight coupling of knowledge modelling and the computational lexicon, which serves as a central concept repository. The CYC KB (Lenat and Guha, 1990), by contrast, was not developed in parallel with a computational lexicon supporting NLP applications to access the full contents of the KB. OntoSem (McShane et al., 2004) is similar to MultiNet in that it also assumes a computational lexicon as the concept repository. However, there is no axiom classification for OntoSem comparable to the typology discussed in our paper.

2 A Meaning Representation for Unconstrained Natural Language

Before explaining the axiom classification, we will briefly describe the MultiNet formalism (Helbig, 2006) used as a semantic interlingua and knowledge representation formalism. As shown by the
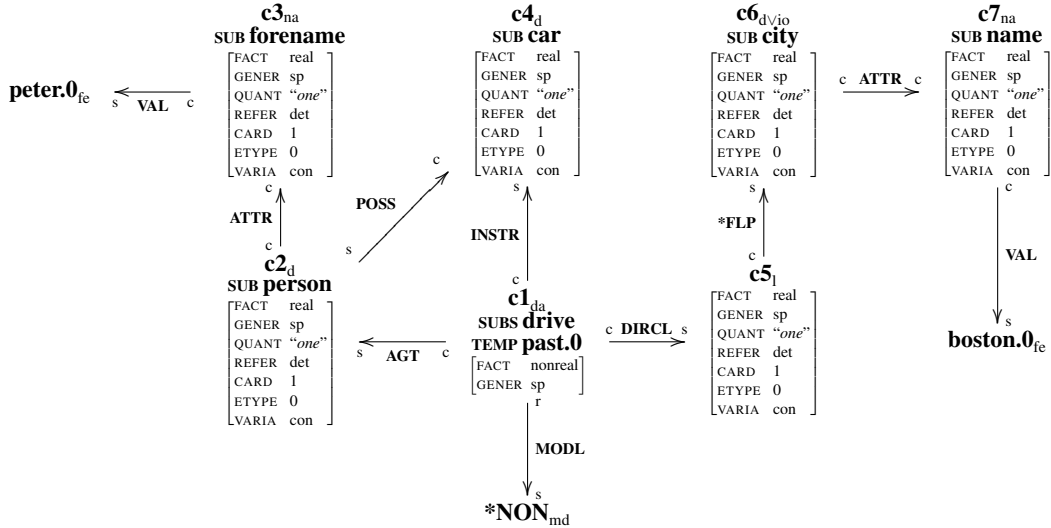
Figure 1

$c3_{na}$ — SUB **forename**
FACT real
GENER sp
QUANT "one"
REFER det
CARD 1
ETYPE 0
VARIA con

peter.0$_{fe}$ — s **VAL** c

$c2_d$ — SUB **person**
FACT real
GENER sp
QUANT "one"
REFER det
CARD 1
ETYPE 0
VARIA con

ATTR, **POSS**

$c4_d$ — SUB **car**
FACT real
GENER sp
QUANT "one"
REFER det
CARD 1
ETYPE 0
VARIA con

INSTR

$c1_{da}$ — SUBS **drive** TEMP **past.0**
FACT nonreal
GENER sp

s **AGT** c

c **DIRCL** s

$c5_l$
FACT real
GENER sp
QUANT "one"
REFER det
CARD 1
ETYPE 0
VARIA con

*FLP

$c6_{d\lor io}$ — SUB **city**
FACT real
GENER sp
QUANT "one"
REFER det
CARD 1
ETYPE 0
VARIA con

c **ATTR** c

$c7_{na}$ — SUB **name**
FACT real
GENER sp
QUANT "one"
REFER det
CARD 1
ETYPE 0
VARIA con

VAL

boston.0$_{fe}$

MODL

*NON$_{md}$

Figure 1: *"It is not true that Peter didn't drive to Boston with his car"*

MultiNet example in Fig. 1, the formalism extends ordinary semantic networks by the following features: (1) Every node is labeled by a sort from a predefined ontology of sorts and by bundles of layer attributes. (2) MultiNet admits functions and relations of arbitrary arity. (3) The arcs (relations) are formally characterized by associated axioms. (4) Subnetworks can be encapsulated to form concepts of higher order, which again can be connected to other concepts by relations and functions. (5) The relationships in the network are marked as categorically valid (c), prototypically valid (p), modally restricted (r), or situationally bounded (s) with regard to each argument.

MultiNet introduces a fixed set of about 140 relations and functions. Table 1 sketches the relations and functions which are used in the following. The formalism distinguishes 45 sorts used to define signatures of relations and functions (Helbig, 2006, Sect. 17.1). Some of these sorts, explained at the bottom of Table 1, also serve to constrain logical rules. Thus, the law of double negation holds for semantically total properties (sort [tq]) like *dead* and its negation *alive*, where *not alive* means the same as *dead*. Another sort [gq] is used for gradable properties like *friendly* and *unfriendly*. Though *unfriendly* means *not friendly*, the law of double negation does not hold here, i.e. if someone is *not unfriendly* this does not mean that the person is *friendly*.

MultiNet not only classifies conceptual entities by their sorts but also by the values of six so-called 'layer attributes'. Here we are concerned only with two of these attributes:

Facticity. We discern three kinds of facticity: [FACT=*real*] for existing entities (*Eiffel tower*), [FACT=*non*] for non-existing entities (*the light ether*), and [FACT=*hypo*] for hypothetical entities (*quarks*). Apart from the *extensional negation* expressed by a non-existing situation with [FACT=*non*], MultiNet supports the *intensional negation* of a situation s, expressed by the relation $\text{MODL}(s, *\text{NON})$. Both types occur in the example shown in Fig. 1. Facticity must be anchored in the logical language since special inference rules apply to hypothetical and non-existing objects.

Genericity. The GENER attribute (type of generality) divides the world of concepts into generic objects with [GENER=*ge*] (*house*) and specific objects with [GENER=*sp*] (⟨*my house*⟩). Note that the different degrees of generality in a conceptual hierarchy are expressed by the subordination relation SUB. Thus, assertions about a generic concept can be clearly separated from assertions about instances. Generic concepts are also needed to model prototypical knowledge. Consider "*Lions feed on antelopes*". A modeling by a universal quantifier ranging over lions is inadequate since the sentence expresses only default knowledge.

3 A Typology of Axioms

We use a first-order language to formulate axioms, which are usually implications $L_1 \land \cdots \land L_n \to (\exists x_1, \ldots, x_k) R_1 \land \cdots \land R_m$ where the L_i and R_j are (possibly negated) literals which correspond to the edges of the subnetworks described by premise

Relation	Signature	Short Characteristics
AFF	$si \times [o \cup si]$	C-Role – Affected object
AGT	$si \times o$	C-Role – Agent
ANTE	$[t \cup si] \times [t \cup si]$	Temporal successorship
ATTR	$o \times at$	Specification of an attribute
AVRT	$si \times o$	C-Role – Averting/Turning away from an object
CAUS	$si' \times si'$	Relation between cause and effect (Causality)
CIRC	$si \times si$	Relation between situation and circumstance
COMPL	$p \times p$	Complementarity relation
COND	$si \times si$	Hypothetical if-then relation
DIRCL	$[si \cup o] \times l$	Relation specifying a direction
EXP	$si \times o$	C-Role – Experiencer
FIN	$si \times [t \cup si]$	Relation between a situation and its temporal end
HSIT	$si \times si$	Relation between enclosing situation and embedded situation
IMPL	$si \times si$	Entailment relation
LOC	$[o \cup si] \times l$	Relation specifying the location
MIN	$qn \times qn$	Smaller-than relation
MODL	$si \times md$	Relation specifying a restricting modality
OBJ	$si \times [o \cup si]$	C-Role – Neutral object of a situation
ORNT	$si \times o$	C-Role – Orientation of a situation toward something
PARS	$co \times co$	Part-whole relationship
PROP	$o \times p$	Relation between object and property
SCAR	$st \times o$	C-Role – Carrier of a state
SSPE	$st \times [o \cup qn]$	C-Role – State-specifying entity
SUB	$o \times \overline{o}$	Relation of conceptual subordination (for objects)
SUBS	$si \times \overline{si}$	Relation of conceptual subordination (for situations)
TEMP	$si \times [t \cup si]$	Relation specifying the temporal embedding of a situation
VAL	$at \times [o \cup qn \cup p \cup fe]$	Relation between an attribute and its value

Table 1: Strongly abbreviated description of relations used in this paper. Explanation of sorts: objects o including, among other things, concrete objects co (*house*) and attributes at (*height*); situations si (*write*) including states st (*be awake*); locations l (*here*), times t (*now*), modal descriptors md (*impossible*); properties p (*dead*), quantificators and measurements qn (*many*, *two litres*), formal entities fe (figures or names). The notation si' demands [FACT = *real*], and the notation $\overline{si}$ demands [GENER = *ge*].

and conclusion. Layer features and sorts can also be used in these axioms to further constrain the admissible variable bindings.

The way in which such axioms will be used in inferences over MultiNet representations is demonstrated by an example given by two sentences S1 = *"The laptop weighs 1.6 kilograms"*, S2 = *"The laptop has a weight of 1.6 kilograms"* and their corresponding (linearized) MultiNet representations M1 and M2, respectively. M1 = [SUBS$(c_1, weigh) \wedge$ SCAR$(c_1, c_2) \wedge$ SUB$(c_2, laptop) \wedge$ SSPE$(c_1, q_1) \wedge q_1 = $ *QUANT$(1.6, kg)$] M2 = SUB$(c_4, laptop) \wedge$ ATTR$(c_4, c_5) \wedge$ SUB$(c_5, weight) \wedge$ VAL$(c_5, q_2) \wedge q_2 = $ *QUANT$(1.6, kg)$], where the *QUANT function serves to construct a measurement from a given numerical value and measurement unit.

Given the question *"What is the weight of the laptop?"*, a question pattern will be generated from the MultiNet representation which comprises the conjunction of edge literals, i.e. SUB$(x, laptop) \wedge$ ATTR$(x, y) \wedge$ SUB$(y, weight) \wedge$ VAL$(y, ?)$. The pattern can be directly matched

with the network M2, based on the substitution $\{x/c_4, y/c_5, ?/q_2\}$, giving ? = QUANT$(1.6, kg)$ as an answer. However, an axiom is needed to prove the question from the network M1. The relationship between weighing and having a weight is

$$\text{SUBS}(w, weigh) \wedge \text{SCAR}(w, k) \wedge \text{SSPE}(w, q) \rightarrow$$
$$(\exists a)\, \text{ATTR}(w, a) \wedge \text{SUBS}(a, weight) \wedge \text{VAL}(a, q).$$

In order to prove "x has a weight of v" from the semantic network for "x weighs v", a backward chaining step must then be carried out which reduces the representation of the former sentence to the representation of the latter sentence.

From a natural language point of view, the standard first-order logic (FOL) is too rigid with regard to the validity of expressions. While a logical expression is either true or false in FOL, a semantic formalism dealing with NL cannot assume this even for the basic assertions. Moreover, logical calculi normally do not give a clue how to use the axioms in an effective inference strategy. The following cross-classification of the axioms by five basic criteria captures additional informa-

tion about the axioms which might be useful to control inference and to judge the validity of computed answers.

3.1 Conceptual Boundedness of Axioms

R-axioms. From a syntactical point of view, there are two types of expressions describing axiomatic knowledge. The first type contains no lexical constants but only relation and function symbols (apart from logical signs). These expressions are called *conceptually non-bound* or *R-Axioms*. The following R-Axiom connects causality and time, saying that effects never take place before the cause: $\text{CAUS}(x,y) \rightarrow \neg\,\text{ANTE}(y,x)$.

Other examples are given by the IMPL-axioms below. Axioms which are conceptually not bound must be treated with care by the reasoner, since an R-axiom for relation R can be applied in inferences wherever R is involved (global effect).

B-axioms. Axioms containing the representative of at least one concept are called *conceptually bound* or B-axioms. Thus, with every selling act s there is a buying act b entailed by s. This relationship is given by the following axiom:

$$\text{SUBS}(s, sell) \land \text{AGT}(s,a) \land \text{OBJ}(s,o) \land \text{ORNT}(s,d) \rightarrow$$
$$(\exists b)\,\text{SUBS}(b, buy) \land \text{OBJ}(b,o) \land \text{AVRT}(b,a)$$
$$\land\,\text{AGT}(b,d) \land \text{IMPL}(s,b)\,.$$

Such B-axioms have only a local effect, i.e. they are applied in those cases where one concept has to be connected to another during inference.

3.2 Core Axioms vs. Frame Axioms

In a B-axiom like the sell/buy example, only the change of participant roles (like AGT, AFF, AVRT, and OBJ) is specified, but nothing is said about the local, temporal and circumstantial embedding of the main situation (mainly represented by LOC, TEMP, and CIRC, resp.) In order to solve this form of the **frame problem** in AI, we distinguish *core axioms*, i.e. B-axioms for lexico-semantical associations, from auxiliary *frame axioms*, which handle the transfer of temporal or local context, of situational embeddings etc.

Different B-axioms can be connected with very different sets of frame axioms. For example, the temporal specification of the selling act s in the above axiom transfers unchanged to b, since selling and buying take place at the same time. However, there is no such transfer of $\text{TEMP}(s_1,t_1)$ of a sending act s_1 to the corresponding receiving act

s_2. Here we need a frame axiom like this:

$$\text{SUBS}(s_1, \langle send\text{-}act\rangle) \land \text{TEMP}(s_1,t_1) \land \text{IMPL}(s_1,s_2) \land$$
$$\text{SUBS}(s_2, \langle receive\text{-}act\rangle) \land \text{TEMP}(s_2,t_2)$$
$$\rightarrow \text{ANTE}(t_1,t_2)\,.$$

However, frame axioms are not necessarily B-axioms tied to specific lexicalized concepts. Consider the buy/sell example. Some important transfer properties are already encoded in the IMPL relation which asserts that the buying act is entailed by the selling act:

$$\text{IMPL}(s_1,s_2) \land \text{IMPL}(s_2,s_3) \rightarrow \text{IMPL}(s_1,s_3)$$
$$\text{IMPL}(s_1,s_2) \land \text{FACT}(s_1){=}real \rightarrow \text{FACT}(s_2){=}real$$
$$\text{IMPL}(s_1,s_2) \rightarrow \text{GENER}(s_1){=}\text{GENER}(s_2)\,.$$

The first axiom merely asserts the transitivity of entailment relationships. The second and third axioms state that the implied situation and the assumed situation share the same facticity and the same degree of genericity, respectively.
The invariance of spatio-temporal embedding can also be handled on the abstract level R-axioms. Thus a selling act and the implied buying act take place at the same time in the same local context:

$$\text{SUBS}(s, sell) \land \text{AGT}(s,a) \land \text{OBJ}(s,o) \land \text{ORNT}(s,d)$$
$$\rightarrow (\exists b)\,\text{SUBS}(b, buy) \land \text{OBJ}(b,o) \land \text{AVRT}(b,a)$$
$$\land\,\text{AGT}(b,d) \land \text{HSIT}(s,b)\,.$$

The HSIT relation expresses the embedding of situations.[1] It is transitive like IMPL, but also has

$$\text{HSIT}(s,p) \land \text{LOC}(s,\ell) \rightarrow \text{LOC}(p,\ell)$$
$$\text{HSIT}(s,p) \rightarrow \text{TEMP}(p,s)$$

These axioms express local and temporal invariance of situational embeddings, resp. Thus if the soccer world championship takes place in Germany, then its finale also takes place in Germany (like all other situational parts). In the sell/buy example, the second axiom ensures that the buying and selling acts happen at the same place. The analogue holds for the temporal specification.

3.3 Categorical vs. Prototypical Axioms

Categorically Valid Axioms. It seems to be a contradiction to speak of axioms of restricted validity. The next axiom states that one of two complementary properties must hold: $\text{COMPL}(p_1,p_2) \rightarrow \text{PROP}(o,p_1) \lor \text{PROP}(o,p_2)$. This

---

[1]Two situations can mutually embed each other. Such situations describe different perspectives on a state of affairs.

rule has no exceptions. But, for NL semantics, one also needs *prototypical* regularities.

Prototypically Valid Axioms. The inheritance of the part-whole relationship has only the status of default (or prototypically valid) knowledge:

$$\text{SUB}(d_1, d_2) \wedge \text{PARS}(d_3, d_2) \rightarrow (\exists d_4) \, \text{SUB}(d_4, d_3) \wedge \\ \text{PARS}(d_4, d_1) \,.$$

Thus, while ships normally have a keel, there are also ships which have not. Categorical axioms lead to monotonic reasoning, while prototypical axioms involve non-monotonicity. In MultiNet, deductions with a default produce default knowledge, which is checked for *local* contradictions in a neighborhood of the involved concepts.

3.4 Deductive Axioms vs. Destructive Axioms

Deductive Axioms. This type of axioms permits derive new knowledge, given by the conclusion, provided that the premise be fulfilled. The important feature is that no piece of knowledge in the knowledge base must ever be retracted.

Destructive Axioms. There are axiomatic regularities which also cancel earlier knowledge, like deriving the temporal end of a situation *s*:

$$\text{SUBS}(e, end) \wedge \text{AFF}(e, s) \wedge \text{TEMP}(e, t) \rightarrow \\ \text{FIN}(s, t) \,|\, \text{DEL TEMP}(s, \_) \,.$$

Thus if an activity *e* ends a situation *s* at time *t*, then a new relation FIN for *s* must be added and the earlier specification of *s* by the relation TEMP must be deleted. While the first type of axioms can be treated by symbolic derivations, the latter type requires actions on the KB like deleting arcs.

3.5 Epistemic Restrictedness

Epistemically Restricted Axioms. There are axioms which are epistemically restricted in the sense that their validity is only warranted within a certain epistemic or cognitive context:

$$\text{CAUS}(k_1, k_2) \wedge \text{CAUS}(k_2, k_3) \rightarrow \text{CAUS}(k_1, k_3)$$

A fading effect prevents infinite prolongation of causality chains by a presumed (but not strongly valid) transitivity of CAUS. This happens because persons stating a causal relation focus on a single cause and neglect other necessary conditions.

Epistemically Non-restricted Axioms. For most axioms no epistemically motivated restriction can be observed. The transitivity axioms for subordination and spatial inclusion hold unconditionally:

$$\text{SUB}(o_1, o_2) \wedge \text{SUB}(o_2, o_3) \rightarrow \text{SUB}(o_1, o_3)$$

$$\text{LOC}(o, *\text{IN}(m)) \wedge \text{LOC}(m, *\text{IN}(n)) \rightarrow \text{LOC}(o, *\text{IN}(n))$$

The first axiom concludes from pigeons being birds and birds being animals that pigeons are also animals, while the second axiom concludes from Marc's sun glasses being in his car and the car being in the garage that Marc's sun glasses are in the garage. These inferences are always valid and not tied to a certain epistemic level. For *epistemically restricted* axioms, we propose the use of built-in procedures which treat borderlines of epistemic levels by special parameters (e.g. by a recursion limit). This treatment might pay off even when the transitivity is not epistemically restricted.

4 Conclusion

We have proposed enriched descriptions of logical axioms in order to cope with the large number of axioms in open-domain natural language inference. The introduction of auxiliary 'frame axioms' in addition to regular meaning postulates was necessary to describe the transfer of contextual embedding between the situations involved in a meaning postulate. The success of a MultiNet prover (Glöckner, 2006) in the CLEF 2006 evaluation campaign, which scored best in the logical answer validation track for German, demonstrates the tractability of the proposed representational means and their typology in applications.

References

I. Glöckner. 2006. University of Hagen at QA@CLEF 2006: Answer validation exercise. In *Working notes of the CLEF06 workshop*, Alicante, Spain.

Sven Hartrumpf. 2004. Question answering using sentence parsing and semantic network matching. In *Proc. of CLEF 2004*, pages 385–392.

Hermann Helbig. 2006. *Knowledge Representation and the Semantics of Natural Language*. Springer.

D. B. Lenat and R. V. Guha. 1990. *Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley.

Johannes Leveling and Hermann Helbig. 2002. A robust natural language interface for access to bibliographic databases. In *Proceedings of SCI 2002*, pages 133–138, Orlando, Florida, July.

M. McShane, M. Zabludowski, S. Nirenburg, and S. Beale. 2004. OntoSem and SIMPLE: Two multilingual world views. In *ACL-2004 Workshop on Text Meaning and Interpretation*, Barcelona, Spain.

D. Moldovan, C. Clark, S. Harabagiu, and S. Maiorano. 2003. COGEX: A logic prover for question answering. In *Proc. NAACL-HLT 2003*, pages 87–93.