

A Text Understander that Learns

Udo Hahn & Klemens Schnattinger

© Computational Linguistics Lab, Freiburg University
Werthmannplatz 1, D-79085 Freiburg, Germany
{hahn,schnattinger}@coling.uni-freiburg.de

Abstract

We introduce an approach to the automatic acquisition of new concepts from natural language texts which is tightly integrated with the underlying text understanding process. The learning model is centered around the ‘quality’ of different forms of linguistic and conceptual evidence which underlies the incremental generation and refinement of alternative concept hypotheses, each one capturing a different conceptual reading for an unknown lexical item.

1 Introduction

The approach to learning new concepts as a result of understanding natural language texts we present here builds on two different sources of evidence — the prior knowledge of the domain the texts are about, and grammatical constructions in which unknown lexical items occur. While there may be many reasonable interpretations when an unknown item occurs for the very first time in a text, their number rapidly decreases when more and more evidence is gathered. Our model tries to make explicit the reasoning processes behind this learning pattern.

Unlike the current mainstream in automatic linguistic knowledge acquisition, which can be characterized as quantitative, surface-oriented bulk processing of large corpora of texts (Hindle, 1989; Zernik and Jacobs, 1990; Hearst, 1992; Manning, 1993), we propose here a *knowledge-intensive* model of concept learning from *few*, positive-only examples that is tightly integrated with the non-learning mode of text understanding. Both learning and understanding build on a given core ontology in the format of terminological assertions and, hence, make abundant use of terminological reasoning. The ‘plain’ text understanding mode can be considered as the instantiation and continuous filling

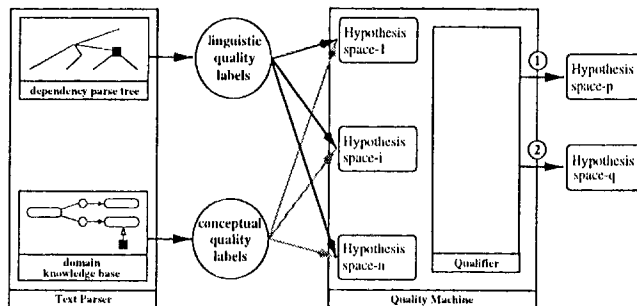


Figure 1: Architecture of the Text Learner

of roles with respect to *single concepts* already available in the knowledge base. Under learning conditions, however, a *set of alternative concept hypotheses* has to be maintained for each unknown item, with each hypothesis denoting a newly created conceptual interpretation tentatively associated with the unknown item.

The underlying methodology is summarized in Fig. 1. The *text parser* (for an overview, cf. Bröker et al. (1994)) yields information from the grammatical constructions in which an unknown lexical item (symbolized by the black square) occurs in terms of the corresponding *dependency parse tree*. The kinds of syntactic constructions (e.g., genitive, apposition, comparative), in which unknown lexical items appear, are recorded and later assessed relative to the credit they lend to a particular hypothesis. The conceptual interpretation of parse trees involving unknown lexical items in the *domain knowledge base* leads to the derivation of *concept hypotheses*, which are further enriched by conceptual annotations. These reflect structural patterns of consistency, mutual justification, analogy, etc. relative to already available concept descriptions in the domain knowledge base or other hypothesis spaces. This kind of initial evidence, in particular its predictive “goodness” for the learning task, is represented by corresponding sets of *linguistic* and *conceptual qual-*

Syntax Semantics	Axiom Semantics
$C \sqcap D \quad C^{\mathcal{I}} \cap D^{\mathcal{I}}$	$A \doteq C \quad A^{\mathcal{I}} = C^{\mathcal{I}}$
$C \sqcup D \quad C^{\mathcal{I}} \cup D^{\mathcal{I}}$	$a : C \quad a^{\mathcal{I}} \in C^{\mathcal{I}}$
$\forall R.C \quad \{d \in \Delta^{\mathcal{I}} \mid R^{\mathcal{I}}(d) \subseteq C^{\mathcal{I}}\}$	$Q \doteq R \quad Q^{\mathcal{I}} = R^{\mathcal{I}}$
$R \sqcap S \quad R^{\mathcal{I}} \cap S^{\mathcal{I}}$	$a R b \quad (a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$
$c R \quad \{(d, d') \in R^{\mathcal{I}} \mid d \in C^{\mathcal{I}}\}$	
$R c \quad \{(d, d') \in R^{\mathcal{I}} \mid d' \in C^{\mathcal{I}}\}$	

Table 1: Some Concept and Role Terms Table 2: Axioms for Concepts and Roles

ity labels. Multiple concept hypotheses for each unknown lexical item are organized in terms of corresponding *hypothesis spaces*, each of which holds different or further specialized conceptual readings.

The *quality machine* estimates the overall credibility of single concept hypotheses by taking the available set of quality labels for each hypothesis into account. The final computation of a preference order for the entire set of competing hypotheses takes place in the *qualifier*, a terminological classifier extended by an evaluation metric for quality-based selection criteria. The output of the quality machine is a ranked list of concept hypotheses. The ranking yields, in decreasing order of significance, either the most plausible concept classes which classify the considered instance or more general concept classes subsuming the considered concept class (cf. Schnattinger and Hahn (1998) for details).

2 Methodological Framework

In this section, we present the major methodological decisions underlying our approach.

2.1 Terminological Logics

We use a standard terminological, KL-ONE-style concept description language, here referred to as *CDL* (for a survey of this paradigm, cf. Woods and Schmolze (1992)). It has several constructors combining *atomic* concepts, roles and individuals to define the terminological theory of a domain. *Concepts* are unary predicates, *roles* are binary predicates over a domain Δ , with *individuals* being the elements of Δ . We assume a common set-theoretical semantics for *CDL* – an interpretation \mathcal{I} is a function that assigns to each concept symbol (the set \mathbf{A}) a subset of the domain Δ , $\mathcal{I} : \mathbf{A} \rightarrow 2^{\Delta}$, to each role symbol (the set \mathbf{P}) a binary relation of Δ , $\mathcal{I} : \mathbf{P} \rightarrow 2^{\Delta \times \Delta}$, and to each individual symbol (the set \mathbf{I}) an element of Δ , $\mathcal{I} : \mathbf{I} \rightarrow \Delta$.

Concept terms and *role terms* are defined inductively. Table 1 contains some constructors and their semantics, where C and D denote concept terms, while R and S denote roles. $R^{\mathcal{I}}(d)$ represents the set of role fillers of the individual d , i.e., the set of individuals e with $(d, e) \in R^{\mathcal{I}}$.

By means of *terminological axioms* (for a subset, see Table 2) a symbolic name can be introduced for each concept to which are assigned necessary and sufficient constraints using the definitional operator ‘ \doteq ’. A finite set of such axioms is called the *terminology* or *TBox*. Concepts and roles are associated with concrete individuals by *assertional axioms* (see Table 2; a, b denote individuals). A finite set of such axioms is called the *world description* or *ABox*. An *interpretation* \mathcal{I} is a model of an ABox with regard to a TBox, iff \mathcal{I} satisfies the assertional and terminological axioms.

Considering, e.g., a phrase such as ‘*The switch of the Itoh-Ci-8 ..*’, a straightforward translation into corresponding terminological concept descriptions is illustrated by:

- (P1) *switch.1* : SWITCH
(P2) *Itoh-Ci-8* HAS-SWITCH *switch.1*
(P3) HAS-SWITCH \doteq
(OUTPUTDEV \sqcup INPUTDEV \sqcup |HAS-PART|SWITCH
STORAGEDEV \sqcup COMPUTER)

Assertion P1 indicates that the instance *switch.1* belongs to the concept class SWITCH. P2 relates *Itoh-Ci-8* and *switch.1* via the relation HAS-SWITCH. The relation HAS-SWITCH is defined, finally, as the set of all HAS-PART relations which have their domain restricted to the disjunction of the concepts OUTPUTDEV, INPUTDEV, STORAGEDEV or COMPUTER and their range restricted to SWITCH.

In order to represent and reason about concept hypotheses we have to properly extend the formalism of *CDL*. *Terminological hypotheses*, in our framework, are characterized by the following properties: for all stipulated hypotheses (1) the same domain Δ holds, (2) the same concept definitions are used, and (3) only different assertional axioms can be established. These conditions are sufficient, because each hypothesis is based on a unique discourse entity (cf. (1)), which can be directly mapped to associated instances (so concept definitions are stable (2)). Only relations (including the ISA-relation) among the instances may be different (3).

Axiom	Semantics
$(a : C)_h$	$a^{\mathcal{I}} \in C^{\mathcal{I}_h}$
$(a R b)_h$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}_h}$

Table 3: Axioms in \mathcal{CDL}^{hypo}

Given these constraints, we may annotate each assertional axiom of the form ‘ $a : C$ ’ and ‘ $a R b$ ’ by a corresponding hypothesis label h so that $(a : C)_h$ and $(a R b)_h$ are valid terminological expressions. The extended terminological language (cf. Table 3) will be called \mathcal{CDL}^{hypo} . Its semantics is given by a special interpretation function \mathcal{I}_h for each hypothesis h , which is applied to each concept and role symbol in the canonical way: $\mathcal{I}_h : \mathbf{A} \rightarrow 2^\Delta$; $\mathcal{I}_h : \mathbf{P} \rightarrow 2^{\Delta \times \Delta}$. Notice that the instances a, b are interpreted by the interpretation function \mathcal{I} , because there exists only one domain Δ . Only the interpretation of the concept symbol C and the role symbol R may be different in each hypothesis h .

Assume that we want to represent two of the four concept hypotheses that can be derived from (P3), viz. *Itoh-Ci-8* considered as a storage device or an output device. The corresponding ABox expressions are then given by:

$$\begin{aligned} & (Itoh-Ci-8 \text{ HAS-SWITCH } switch.1)_{h_1} \\ & (Itoh-Ci-8 : STORAGEDEV)_{h_1} \\ & (Itoh-Ci-8 \text{ HAS-SWITCH } switch.1)_{h_2} \\ & (Itoh-Ci-8 : OUTPUTDEV)_{h_2} \end{aligned}$$

The semantics associated with this ABox fragment has the following form:

$$\begin{aligned} \mathcal{I}_{h_1}(\text{HAS-SWITCH}) &= \{(Itoh-Ci-8, switch.1)\}, \\ \mathcal{I}_{h_1}(\text{STORAGEDEV}) &= \{Itoh-Ci-8\}, \\ \mathcal{I}_{h_1}(\text{OUTPUTDEV}) &= \emptyset \\ \mathcal{I}_{h_2}(\text{HAS-SWITCH}) &= \{(Itoh-Ci-8, switch.1)\}, \\ \mathcal{I}_{h_2}(\text{STORAGEDEV}) &= \emptyset, \\ \mathcal{I}_{h_2}(\text{OUTPUTDEV}) &= \{Itoh-Ci-8\} \end{aligned}$$

2.2 Hypothesis Generation Rules

As mentioned above, text parsing and concept acquisition from texts are tightly coupled. Whenever, e.g., two nominals or a nominal and a verb are supposed to be syntactically related in the regular parsing mode, the semantic interpreter simultaneously evaluates the conceptual compatibility of the items involved. Since these reasoning processes are fully embedded in a terminological representation system, checks are made as to whether a concept denoted by one of these objects is allowed to fill a role of the other one. If one of the items involved is unknown, i.e., a lexical and conceptual gap is

encountered, this interpretation mode generates initial concept hypotheses about the class membership of the unknown object, and, as a consequence of inheritance mechanisms holding for concept taxonomies, provides conceptual role information for the unknown item.

Given the structural foundations of terminological theories, two dimensions of conceptual learning can be distinguished — the taxonomic one by which new concepts are located in conceptual hierarchies, and the aggregational one by which concepts are supplied with clusters of conceptual relations (these will be used subsequently by the terminological classifier to determine the current position of the item to be learned in the taxonomy). In the following, let *target.con* be an unknown concept denoted by the corresponding lexical item *target.lex*, *base.con* be a given knowledge base concept denoted by the corresponding lexical item *base.lex*, and let *target.lex* and *base.lex* be related by some dependency relation. Furthermore, in the hypothesis generation rules below variables are indicated by names with leading ‘?’; the operator **TELL** is used to initiate the creation of assertional axioms in \mathcal{CDL}^{hypo} .

Typical linguistic indicators that can be exploited for *taxonomic* integration are appositions (‘. the printer @A@ ..’), exemplification phrases (‘. printers like the @A@ ..’) or nominal compounds (‘. the @A@ printer ..’). These constructions almost unequivocally determine ‘@A@’ (*target.lex*) when considered as a proper name¹ to denote an instance of a **PRINTER** (*target.con*), given its characteristic dependency relation to ‘printer’ (*base.lex*), the conceptual correlate of which is the concept class **PRINTER** (*base.con*). This conclusion is justified independent of conceptual conditions, simply due to the nature of these linguistic constructions.

The generation of corresponding concept hypotheses is achieved by the rule **sub-hypo** (Table 4). Basically, the type of *target.con* is carried over from *base.con* (function **type-of**). In addition, the syntactic *label* is asserted which characterizes the grammatical construction figuring as the structural source for that particular hy-

¹Such a part-of-speech hypothesis can be derived from the inventory of valence and word order specifications underlying the dependency grammar model we use (Bröker et al., 1994).

<pre>sub-hypo(<i>target.con</i>, <i>base.con</i>, <i>h</i>, <i>label</i>) ?<i>type</i> := type-of(<i>base.con</i>) TELL (<i>target.con</i> : ?<i>type</i>)_{<i>h</i>} add-label((<i>target.con</i> : ?<i>type</i>)_{<i>h</i>}, <i>label</i>)</pre>
--

Table 4: Taxonomic Hypothesis Generation Rule

pothesis (h denotes the identifier for the selected hypothesis space), e.g., APPOSITION, EXEMPLIFICATION, or NCOMPOUND.

The *aggregational* dimension of terminological theories is addressed, e.g., by grammatical constructions causing case frame assignments. In the example ‘. @B@ is equipped with 32 MB of RAM ..’, role filler constraints of the verb form ‘equipped’ that relate to its PATIENT role carry over to ‘@B@’. After subsequent semantic interpretation of the entire verbal complex, ‘@B@’ may be anything that can be equipped with memory. Constructions like prepositional phrases (‘. @C@ from IBM ..’) or genitives (‘. IBM’s @C@ ..’) in which either *target.lex* or *base.lex* occur as head or modifier have a similar effect. Attachments of prepositional phrases or relations among nouns in genitives, however, open a wider interpretation space for ‘@C@’ than for ‘@B@’, since verbal case frames provide a higher role selectivity than PP attachments or, even more so, genitive NPs. So, any concept that can reasonably be related to the concept IBM will be considered a potential hypothesis for ‘@C@’, e.g., its departments, products, Fortune 500 ranking.

Generalizing from these considerations, we state a second hypothesis generation rule which accounts for aggregational patterns of concept learning. The basic assumption behind this rule, *perm-hypo* (cf. Table 5), is that *target.con* fills (exactly) one of the n roles of *base.con* it is currently permitted to fill (this set is determined by the function *perm-filler*). Depending on the actual linguistic construction one encounters, it may occur, in particular for PP and NP constructions, that one cannot decide on the correct role yet. Consequently, several alternative hypothesis spaces are opened and *target.con* is assigned as a potential filler of the i -th role (taken from *?roleSet*, the set of admitted roles) in its corresponding hypothesis space. As a result, the classifier is able to derive a suitable concept hypothesis by specializing *target.con* according to the value restriction of *base.con*’s i -th role. The function *member-of*

<pre>perm-hypo(<i>target.con</i>, <i>base.con</i>, <i>h</i>, <i>label</i>) ?<i>roleSet</i> := perm-filler(<i>target.con</i>, <i>base.con</i>, <i>h</i>) ?<i>r</i> := ?<i>roleSet</i> FORALL ?<i>i</i> := ?<i>r</i> DOWNTO 1 DO ?<i>role</i>_{<i>i</i>} := member-of(?<i>roleSet</i>) ?<i>roleSet</i> := ?<i>roleSet</i> \ {?<i>role</i>_{<i>i</i>}} IF ?<i>i</i> = 1 THEN ?<i>hypo</i> := <i>h</i> ELSE ?<i>hypo</i> := gen-hypo(<i>h</i>) TELL (<i>base.con</i> ?<i>role</i>_{<i>i</i>} <i>target.con</i>)?<i>hypo</i> add-label((<i>base.con</i> ?<i>role</i>_{<i>i</i>} <i>target.con</i>)?<i>hypo</i>, <i>label</i>)</pre>
--

Table 5: Aggregational Hypothesis Generation Rule

selects a role from the set *?roleSet*; *gen-hypo* creates a new hypothesis space by asserting the given axioms of h and outputs its identifier. Thereupon, the hypothesis space identified by *?hypo* is augmented through a TELL operation by the hypothesized assertion. As for *sub-hypo*, *perm-hypo* assigns a syntactic quality label (function *add-label*) to each i -th hypothesis indicating the type of syntactic construction in which *target.lex* and *base.lex* are related in the text, e.g., CASEFRAME, PPATTACH or GENITIVENP.

Getting back to our example, let us assume that the target *Itoh-Ci-8* is predicted already as a PRODUCT as a result of preceding interpretation processes, i.e., *Itoh-Ci-8* : PRODUCT holds. Let PRODUCT be defined as:

PRODUCT \doteq
 \vee HAS-PART.PHYSICALOBJECT \sqcap \vee HAS-SIZE.SIZE \sqcap
 \vee HAS-PRICE.PRICE \sqcap \vee HAS-WEIGHT.WEIGHT

At this level of conceptual restriction, four roles have to be considered for relating the target *Itoh-Ci-8* – as a tentative PRODUCT – to the base concept SWITCH when interpreting the phrase ‘The switch of the Itoh-Ci-8 ..’. Three of them, HAS-SIZE, HAS-PRICE, and HAS-WEIGHT, are ruled out due to the violation of a simple integrity constraint (‘switch’ does not denote a measure unit). Therefore, only the role HAS-PART must be considered in terms of the expression *Itoh-Ci-8* HAS-PART *switch.1* (or, equivalently, *switch.1* PART-OF *Itoh-Ci-8*). Due to the definition of HAS-SWITCH (cf. P3, Subsection 2.1), the instantiation of HAS-PART is specialized to HAS-SWITCH by the classifier, since the range of the HAS-PART relation is already restricted to SWITCH (P1). Since the classifier aggressively pushes hypothesizing to be maximally specific, the disjunctive concept referred to in

the domain restriction of the role HAS-SWITCH is split into four distinct hypotheses, two of which are sketched below. Hence, we assume *Itoh-Ci-8* to denote either a STORAGEDEV or an OUTPUTDEV or an INPUTDEV or a COMPUTER (note that we also include parts of the IS-A hierarchy in the example below).

$$\begin{aligned} & (Itoh-Ci-8 : STORAGEDEV)_{h_1}, \\ & (Itoh-Ci-8 : DEVICE)_{h_1, \dots}, \\ & (Itoh-Ci-8 \text{ HAS-SWITCH } switch.1)_{h_1} \\ & (Itoh-Ci-8 : OUTPUTDEV)_{h_2}, \\ & (Itoh-Ci-8 : DEVICE)_{h_2, \dots}, \\ & (Itoh-Ci-8 \text{ HAS-SWITCH } switch.1)_{h_2, \dots} \end{aligned}$$

2.3 Hypothesis Annotation Rules

In this section, we will focus on the quality assessment of concept hypotheses which occurs at the knowledge base level only; it is due to the operation of hypothesis annotation rules which continuously evaluate the hypotheses that have been derived from linguistic evidence.

The M-Deduction rule (see Table 6) is triggered for any repetitive assignment of the same role filler to one specific conceptual relation that occurs in different hypothesis spaces. This rule captures the assumption that a role filler which has been *multiply* derived at different occasions must be granted more strength than one which has been derived at a single occasion only.

EXISTS $o_1, o_2, R, h_1, h_2 :$
 $(o_1 R o_2)_{h_1} \wedge (o_1 R o_2)_{h_2} \wedge h_1 \neq h_2 \implies$
TELL $(o_1 R o_2)_{h_1} : \text{M-DEDUCTION}$

Table 6: The Rule M-Deduction

Considering our example at the end of subsection 2.2, for *Itoh-Ci-8* the concept hypotheses STORAGEDEV and OUTPUTDEV were derived independently of each other in different hypothesis spaces. Hence, DEVICE as their common superconcept has been multiply derived by the classifier in each of these spaces as a result of transitive closure computations, too. Accordingly, this hypothesis is assigned a high degree of confidence by the classifier which derives the conceptual quality label M-DEDUCTION:

$$\begin{aligned} & (Itoh-Ci-8 : DEVICE)_{h_1} \wedge (Itoh-Ci-8 : DEVICE)_{h_2} \\ \implies & (Itoh-Ci-8 : DEVICE)_{h_1} : \text{M-DEDUCTION} \end{aligned}$$

The C-Support rule (see Table 7) is triggered whenever, within the same hypothesis space, a hypothetical relation, R_1 , between two instances can be justified by another relation, R_2 ,

involving the same two instances, but where the role fillers occur in ‘inverted’ order (R_1 and R_2 need not necessarily be semantically inverse relations, as with ‘buy’ and ‘sell’). This causes the generation of the quality label C-SUPPORT which captures the inherent symmetry between concepts related via quasi-inverse relations.

EXISTS $o_1, o_2, R_1, R_2, h :$
 $(o_1 R_1 o_2)_h \wedge (o_2 R_2 o_1)_h \wedge R_1 \neq R_2 \implies$
TELL $(o_1 R_1 o_2)_h : \text{C-SUPPORT}$

Table 7: The Rule C-Support

Example:

$$\begin{aligned} & (Itoh \text{ SELLS } Itoh-Ci-8)_h \wedge \\ & (Itoh-Ci-8 \text{ DEVELOPED-BY } Itoh)_h \\ \implies & (Itoh \text{ SELLS } Itoh-Ci-8)_h : \text{C-SUPPORT} \end{aligned}$$

Whenever an already filled conceptual relation receives an additional, yet different role filler in the same hypothesis space, the Add-Filler rule is triggered (see Table 8). This application-specific rule is particularly suited to our natural language understanding task and has its roots in the distinction between mandatory and optional case roles for (ACTION) verbs. Roughly, it yields a negative assessment in terms of the quality label ADDFILLER for any attempt to fill the same mandatory case role more than once (unless coordinations are involved). In contradistinction, when the same role of a non-ACTION concept (typically denoted by nouns) is multiply filled we assign the positive quality label SUPPORT, since it reflects the conceptual proximity a relation induces on its component fillers, provided that they share a common, non-ACTION concept class.

EXISTS $o_1, o_2, o_3, R, h :$
 $(o_1 R o_2)_h \wedge (o_1 R o_3)_h \wedge (o_1 : \text{ACTION})_h \implies$
TELL $(o_1 R o_2)_h : \text{ADDFILLER}$

Table 8: The Rule AddFiller

We give examples both for the assignment of an ADDFILLER as well as for a SUPPORT label:

Examples:

$$\begin{aligned} & (produces.1 : \text{ACTION})_h \wedge \\ & (produces.1 \text{ AGENT } Itoh)_h \wedge \\ & (produces.1 \text{ AGENT } IBM)_h \\ \implies & (produces.1 \text{ AGENT } Itoh)_h : \text{ADDFILLER} \end{aligned}$$

$$\begin{aligned} & (Itoh-Ci-8 : \text{PRINTER})_h \wedge (Itoh-Ct : \text{PRINTER})_h \wedge \\ & (Itoh \text{ SELLS } Itoh-Ci-8)_h \wedge (Itoh \text{ SELLS } Itoh-Ct)_h \wedge \\ & (Itoh : \neg \text{ACTION})_h \\ \implies & (Itoh-Ci-8 : \text{PRINTER})_h : \text{SUPPORT} \end{aligned}$$

2.4 Quality Dimensions

The criteria from which concept hypotheses are derived differ in the dimension from which they are drawn (grammatical *vs.* conceptual evidence), as well as the strength by which they lend support to the corresponding hypotheses (e.g., apposition *vs.* genitive, multiple deduction *vs.* additional role filling, etc.). In order to make these distinctions explicit we have developed a “quality calculus” at the core of which lie the definition of and inference rules for quality labels (cf. Schnattinger and Hahn (1998) for more details). A design methodology for specific quality calculi may proceed along the following lines: (1) *Define the dimensions* from which quality labels can be drawn. In our application, we chose the set $\mathcal{LQ} := \{l_1, \dots, l_m\}$ of linguistic quality labels and $\mathcal{CQ} := \{c_1, \dots, c_n\}$ of conceptual quality labels. (2) *Determine a partial ordering ρ among the quality labels from one dimension* reflecting different degrees of strength among the quality labels. (3) *Determine a total ordering among the dimensions.*

In our application, we have empirical evidence to grant linguistic criteria priority over conceptual ones. Hence, we state the following constraint: $\forall l \in \mathcal{LQ}, \forall c \in \mathcal{CQ} : l \succ_{\rho} c$

The dimension \mathcal{LQ} . *Linguistic* quality labels reflect structural properties of phrasal patterns or discourse contexts in which unknown lexical items occur² — we here assume that the type of grammatical construction exercises a particular interpretative force on the unknown item and, at the same time, yields a particular level of credibility for the hypotheses being derived. Taking the considerations from Subsection 2.2 into account, concrete examples of high-quality labels are given by APPPOSITION or NCOMPOUND labels. Still of good quality but already less constraining are occurrences of the unknown item in a CASEFRAME construction. Finally, in a PPATTACH or GENITIVENP construction the unknown lexical item is still less constrained. Hence, at the quality level, these latter two labels (just as the first two labels we considered) form an equivalence class whose elements cannot be further discriminated. So we end up with the following quality orderings:

NCOMPOUND	$=_{\rho}$	APPPOSITION
NCOMPOUND	$>_{\rho}$	CASEFRAME
APPPOSITION	$>_{\rho}$	CASEFRAME
CASEFRAME	$>_{\rho}$	GENITIVENP
CASEFRAME	$>_{\rho}$	PPATTACH
GENITIVENP	$=_{\rho}$	PPATTACH

The dimension \mathcal{CQ} . *Conceptual* quality labels result from comparing the conceptual representation structures of a concept hypothesis with already existing representation structures in the underlying domain knowledge base or other concept hypotheses from the viewpoint of structural similarity, compatibility, etc. The closer the match, the more credit is lent to a hypothesis. A very positive conceptual quality label, e.g., is M-DEDUCTION, whereas ADDFILLER is a negative one. Still positive strength is expressed by SUPPORT or C-SUPPORT, both being indistinguishable, however, from a quality point of view. Accordingly, we may state:

M-DEDUCTION	$>_{\rho}$	SUPPORT
M-DEDUCTION	$>_{\rho}$	C-SUPPORT
SUPPORT	$=_{\rho}$	C-SUPPORT
SUPPORT	$>_{\rho}$	ADDFILLER
C-SUPPORT	$>_{\rho}$	ADDFILLER

2.5 Hypothesis Ranking

Each new clue available for a target concept to be learned results in the generation of additional linguistic or conceptual quality labels. So hypothesis spaces get incrementally augmented by quality statements. In order to select the most credible one(s) among them we apply a two-step procedure (the details of which are explained in Schnattinger and Hahn (1998)). First, those concept hypotheses are chosen which have accumulated the greatest amount of high-quality labels according to the linguistic dimension \mathcal{LQ} . Second, further hypotheses are selected from this linguistically plausible candidate set based on the quality ordering underlying \mathcal{CQ} .

We have also made considerable efforts to evaluate the performance of the text learner based on the quality calculus. In order to account for the incrementality of the learning process, a new evaluation measure capturing the system’s on-line learning accuracy was defined, which is sensitive to taxonomic hierarchies. The results we got were consistently favorable, as our system outperformed those closest in spirit, CAMILLE (Hastings, 1996) and SCISOR (Rau et

²In the future, we intend to integrate additional types of constraints, e.g., quality criteria reflecting the degree of completeness *vs.* partiality of the parse.

al., 1989), by a gain in accuracy on the order of 8%. Also, the system requires relatively few hypothesis spaces (2 to 6 on average) and prunes the concept search space radically, requiring only a few examples (for evaluation details, cf. Hahn and Schnattinger (1998)).

3 Related Work

We are not concerned with lexical acquisition from *very large* corpora using surface-level collocational data as proposed by Zernik and Jacobs (1990) and Velardi et al. (1991), or with hyponym extraction based on entirely syntactic criteria as in Hearst (1992) or lexico-semantic associations (e.g., Resnik (1992) or Sekine et al. (1994)). This is mainly due to the fact that these studies aim at a shallower level of learning (e.g., selectional restrictions or thematic relations of verbs), while our focus is on much more fine-grained conceptual knowledge (roles, role filler constraints, integrity conditions).

Our approach bears a close relationship, however, to the work of Mooney (1987), Berwick (1989), Rau et al. (1989), Gomez and Segami (1990), and Hastings (1996), who all aim at the automated learning of word meanings from context using a knowledge-intensive approach. But our work differs from theirs in that the need to cope with *several competing* concept hypotheses and to aim at a *reason-based selection* in terms of the *quality* of arguments is not an issue in these studies. Learning from real-world texts usually provides the learner with only sparse and fragmentary evidence, such that multiple hypotheses are likely to be derived and a need for a hypothesis evaluation arises.

4 Conclusion

We have introduced a solution for the semantic acquisition problem on the basis of the automatic processing of expository texts. The learning methodology we propose is based on the incremental assignment and evaluation of the quality of linguistic and conceptual evidence for emerging concept hypotheses. No specialized learning algorithm is needed, since learning is a reasoning task carried out by the classifier of a terminological reasoning system. However, strong heuristic guidance for selecting between plausible hypotheses comes from linguistic and conceptual quality criteria.

Acknowledgements. We would like to thank our colleagues in the CLIF group for fruitful discussions, in particular Joe Bush who polished the text as a native speaker. K. Schnattinger is supported by a grant from DFG (Ha 2097/3-1).

References

- R. Berwick. 1989. Learning word meanings from examples. In D. Waltz, editor, *Semantic Structures.*, pages 89–124. Lawrence Erlbaum.
- N. Bröker, U. Hahn, and S. Schacht. 1994. Concurrent lexicalized dependency parsing: the PARSETALK model. In *Proc. of the COLING'94. Vol.1*, pages 379–385.
- F. Gomez and C. Segami. 1990. Knowledge acquisition from natural language for expert systems based on classification problem-solving methods. *Knowledge Acquisition*, 2(2):107–128.
- U. Hahn and K. Schnattinger. 1998. Towards text knowledge engineering. In *Proc. of the AAAI'98*.
- P. Hastings. 1996. Implications of an automatic lexical acquisition system. In S. Wermter, E. Riloff, and G. Scheler, editors, *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, pages 261–274. Springer.
- M. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of the COLING'92. Vol.2*, pages 539–545.
- D. Hindle. 1989. Acquiring disambiguation rules from text. In *Proc. of the ACL'89*, pages 26–29.
- C. Manning. 1993. Automatic acquisition of large subcategorization dictionary from corpora. In *Proc. of the ACL'93*, pages 235–242.
- R. Mooney. 1987. Integrated learning of words and their underlying concepts. In *Proc. of the CogSci'87*, pages 974–978.
- L. Rau, P. Jacobs, and U. Zernik. 1989. Information extraction and text summarization using linguistic knowledge acquisition. *Information Processing & Management*, 25(4):419–428.
- P. Resnik. 1992. A class-based approach to lexical discovery. In *Proc. of the ACL'92*, pages 327–329.
- K. Schnattinger and U. Hahn. 1998. Quality-based learning. In *Proc. of the ECAI'98*, pages 160–164.
- S. Sekine, J. Carroll, S. Ananiadou, and J. Tsujii. 1994. Automatic learning for semantic collocation. In *Proc. of the ANLP'94*, pages 104–110.
- P. Velardi, M. Pazienza, and M. Fasolo. 1991. How to encode semantic knowledge: a method for meaning representation and computer-aided acquisition. *Computational Linguistics*, 17:153–170.
- W. Woods and J. Schmolze. 1992. The KL-ONE family. *Computers & Mathematics with Applications*, 23(2/5):133–177.
- U. Zernik and P. Jacobs. 1990. Tagging for learning: collecting thematic relations from corpus. In *Proc. of the COLING'90. Vol.1*, pages 34–39.