Learning Grammatical and Taxonomic Knowledge in Tandem

Udo Hahn and **Kornél G. Markó**

Linguistische Informatik / Computerlinguistik
Albert-Ludwigs-Universität Freiburg, Werthmannplatz 1, D-79085 Freiburg, Germany
http://www.coling.uni-freiburg.de/

Abstract. We introduce a methodology for automating the maintenance of domain-specific concept taxonomies and grammatical class hierarchies simultaneously, based on incremental natural language text understanding. The acquisition process is centered around the linguistic and conceptual 'quality' of various forms of evidence underlying the generation, assessment and on-going refinement of lexical and concept hypotheses. On the basis of the quality of evidence, hypotheses are ranked according to plausibility, and the most reasonable ones are selected for assimilation into the given lexical class hierarchy and domain taxonomy.

1 Introduction

Intelligent systems require knowledge-intensive resources to reason on. As their creation is usually delegated to human experts who are slow and costly, these systems face the often deplored knowledge acquisition bottleneck. The knowledge supply challenge is even more pressing when *multiple* knowledge sources have to be provided within the framework of a single system, all at the same time. This is typically the case for knowledge-intensive natural language processing (NLP) systems which require simultaneous feeding with a lexical inventory, morphological and syntactic rules or constraints, and semantic as well as conceptual knowledge.

Each of these subsystems embodies an enormous amount of specialized component knowledge of its own. Much emphasis has already been put on providing machine learning support for single of these components — acquiring morphological [JT97], lexical [RJ99,SM00], syntactic [Bri95,Cha93], semantic [GSH97] and conceptual knowledge [HL94,SL94]. But only Cardie [Car93] has made an attempt so far to combine these different streams of knowledge within a uniform approach, i.e., to learn *different* types of relevant NLP knowledge *in tandem*.

We also propose such an approach, a learning system capable of acquiring syntactic and conceptual knowledge simultaneously. We build on already specified grammar and domain knowledge but these resources are continuously enhanced as a by-product of text understanding processes. New concepts are acquired and positioned in the concept taxonomy, as well as the grammatical status of their lexical correlates is learned taking three knowledge sources into account. *Domain knowledge* serves as a comparison scale for judging the plausibility of newly derived concept descriptions in the light of prior knowledge. *Grammatical knowledge* contains a type hierarchy of lexical classes according to which increasingly restrictive grammatical constraints are made available.

Structural linguistic patterns, finally, are used to assess the strength of the interpretative force that can be attributed to the grammatical construction in which unknown lexical items occur. Our model makes explicit the kind of qualitative reasoning that is behind such a multi-threaded learning process [SH98].

2 A Learning Scenario

Consider a learning scenario as depicted in Figure 1 from a grammatical perspective and in Figure 2 from a conceptual one. Suppose, your knowledge of the information technology domain tells you nothing about *Itoh-Ci-8*. Imagine, one day your favorite technology magazine features an article starting with *"The Itoh-Ci-8 has a size of . . . "*. Has your knowledge increased? If so, what did you learn from just this phrase?

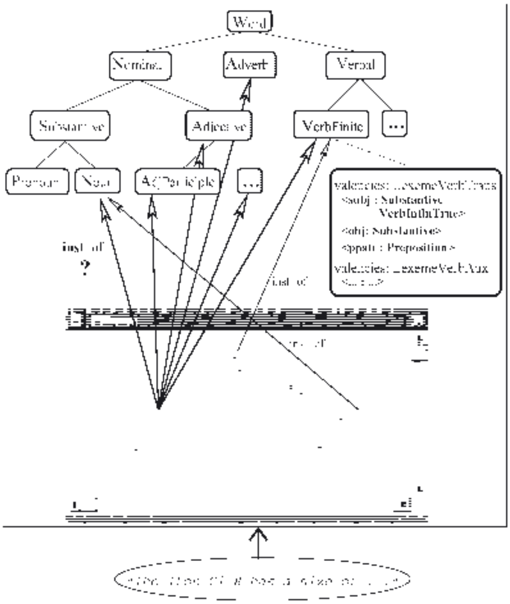

Fig. 1. Sample Scenario: Grammatical Learning

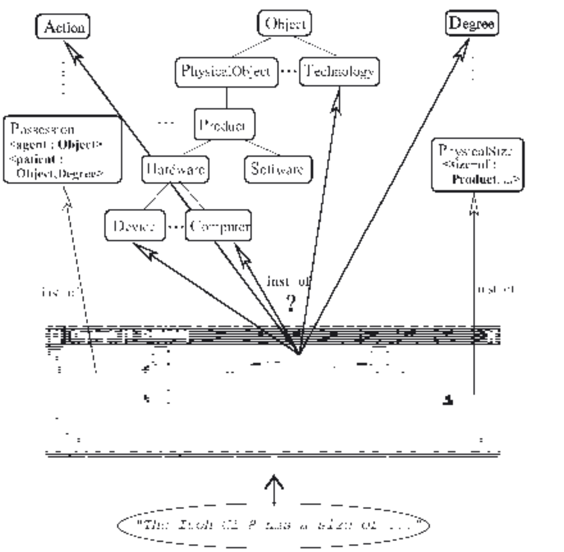

Fig. 2. Sample Scenario: Conceptual Learning

The learning process starts upon the reading of the unknown word *"Itoh-Ci-8"*. In this initial step, the corresponding hypothesis space incorporates all the top level concepts available in the ontology for the new lexical item *"Itoh-Ci-8"*. So, the concept ITOH-CI-8 may be an OBJECT, an ACTION, a DEGREE, etc. (cf. Figure 2). Similarly, from a grammatical viewpoint (cf. Figure 1), the lexical item *"Itoh-Ci-8"* can be hypothesized as being an instance of one of the top-level part-of-speech categories, e.g., a NOMINAL, an ADVERB or a VERBAL.[1] Due to grammatical constraints, however,

[1] While the distinction between NOMINAL and VERBAL should be obvious, the prominent role of ADVERB at the top level of word class categories might not be. However, NOMINAL as well

(*"Itoh-Ci-8"* directly follows *"The"*) the VERBFINITE hypothesis and related ones can be immediately rejected (cf. the darkly shaded box in Figure 1).

While processing the noun phrase *"The Itoh-Ci-8"* as the subject of the verb *"has"*, the ADVERB hypothesis, as well as alternative ADJECTIVE hypotheses, including participles acting as adjectives, ADJPARTICIPLE (cf. the grey shaded boxes in Figure 1), become invalid (none of the instances of any of these word classes must intervene a determiner and a finite verb), still leaving the SUBSTANTIVE hypothesis intact. Additional supportive evidence for the latter comes from the part-of-speech constraints imposed by the `subj` (or, alternatively, by the `obj`) dependency relation (cf. the valency requirements attached to the verb *"has"* in Figure 1), while the equally possible VERBINFIN alternative is ruled out due to violating syntactic evidence. Since *"Itoh-Ci-8"* is not a pronoun (it does not match this closed list), we hypothesize it to be a NOUN, finally.

From an ontological perspective (cf. Figure 2), the concept ITOH-CI-8, at this stage of analysis, is related via the AGENT role to the ACTION concept POSSESSION, the concept denoted by *"has"* (lexical ambiguities, e.g., for the verb *"has"*, lead to the creation of alternative hypothesis spaces). Since POSSESSION requires its AGENT to be an OBJECT, ACTION and DEGREE are no longer valid concept hypotheses for ITOH-CI-8. Their cancellation (cf. the darkly shaded boxes in Figure 2) yields a significant reduction of the huge initial hypothesis space. The learner then aggressively specializes the remaining single hypothesis to the immediate subordinates of OBJECT, *viz.* PHYSICALOBJECT and TECHNOLOGY, in order to test more restrictive hypotheses which due to more specific constraints are easier falsifiable.

In addition, the linguistic constraints for the verb *"has"* indicate that the grammatical direct `object` relation is to be interpreted in terms of a conceptual PATIENT role. Accordingly, the phrase *"... has a size of ... "* is processed such that *size*.1 is the PATIENT of the POSSESSION relationship. Subsequent conceptual interpretation steps combine the fillers of the AGENT and PATIENT roles such that the following terminological expressions are asserted:

(P1) *size*.1 : PHYSICALSIZE
(P2) *Itoh-Ci*-8.1 HAS-SIZE *size*.1

Assertion (P1) indicates that *size*.1 is an instance of the concept class PHYSICALSIZE and (P2) relates *size*.1 and *Itoh-Ci*-8.1 via the binary relation HAS-SIZE.

Given the conceptual roles attached to PHYSICALSIZE, the system recognizes that all specializations of PRODUCT can be related to the concept PHYSICALSIZE (via the role SIZE-OF), while for TECHNOLOGY no such relation can be established. So, we prefer the conceptual reading of ITOH-CI-8 as a kind of a PRODUCT over the TECHNOLOGY hypothesis (cf. the grey-shaded box in Figure 2). At this intial stage, we come up with two assumptions – grammatically, we consider the lexical item *"Itoh-Ci-8"* as a NOUN, while conceptually, we interpret ITOH-CI-8 as a PRODUCT.

---

as VERBAL carry grammatical information such as `case`, `gender`, `number`, or `tense`, `mood`, `aspect`, respectively, none of which is shared by ADVERBs.

3 The Learning Model

The system architecture for elicitating conceptual and grammatical knowledge from texts is summarized in Figure 3. It depicts how linguistic and conceptual evidence are generated and combined to continuously discriminate and refine the set of concept hypotheses (the unknown item yet to be learned is characterized by the black square).

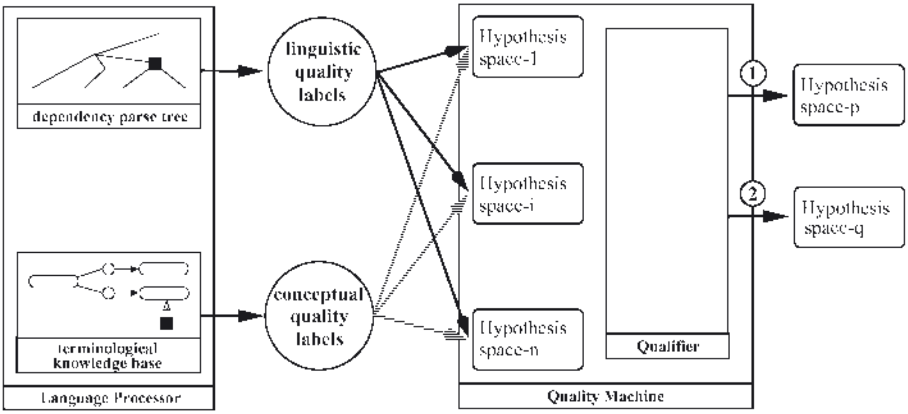

Fig. 3. Architecture for Text-Based Concept and Grammar Learning

Grammatical knowledge for syntactic analysis is based on a fully lexicalized dependency grammar [HSB94]. Such a grammar captures binary valency constraints between a syntactic head (e.g., a noun) and possible modifiers (e.g., a determiner or an adjective). These include restrictions on word order, compatibility of morphosyntactic features and semantic integrity conditions. For a dependency relation $\delta \in \mathcal{D} := \{$*specifier, subject, dir-object, ...*$\}$ to be established between a head and a modifier, all valency constraints must be fulfilled. In this approach, lexeme specifications form the leaf nodes of a lexicon tree which are further abstracted in terms of a hierarchy of word class specifications at different levels of generality. This leads to a word class hierarchy, which consists of word class names $\mathcal{W} = \{$VERBAL, VERBFINITE, SUBSTANTIVE, NOUN, ...$\}$ and a subsumption relation $isa_{\mathcal{W}} = \{$(VERBFINITE, VERBAL), (NOUN, SUBSTANTIVE), ...$\} \subset \mathcal{W} \times \mathcal{W}$, which characterizes specialization relations between word classes.

The *language processor* [HBN00] yields structural dependency information from the grammatical constructions in which an unknown lexical item occurs in terms of the corresponding *parse tree*. The kinds of syntactic constructions (e.g., genitive, apposition, comparative), in which unknown lexical items appear, are recorded and assessed later on relative to the credit they lend to a particular hypothesis.

Conceptual knowledge is expressed in terms of a KL-ONE-like knowledge representation language [WS92]. A domain ontology consists of a set of concept names $\mathcal{F}$ $:= \{$COMPANY, HARD-DISK, ...$\}$ and a subsumption relation $isa_{\mathcal{F}} = \{$(HARD-DISK, STORAGEDEVICE), (IBM, COMPANY), ...$\} \subset \mathcal{F} \times \mathcal{F}$. The set of relation names $\mathcal{R} := \{$HAS-PART, DELIVER-AGENT, ...$\}$ contains the labels of conceptual relations which are also organized in a subsumption hierarchy $isa_{\mathcal{R}} = \{$(HAS-HARD-DISK, HAS-PHYSICAL-PART), (HAS-PHYSICAL-PART, HAS-PART), ...$\}$.

The semantic interpretation of parse trees [RMH99] involving unknown lexical items and their conceptual correlates in the *terminological knowledge base* forms the basis for the derivation of *concept hypotheses*, which are further enriched by conceptual annotations reflecting structural patterns of consistency, mutual justification, analogy, etc. This kind of initial evidence, in particular its predictive "goodness" for the learning task, is represented by corresponding sets of linguistic and conceptual quality labels.

Linguistic quality labels reflect structural properties of phrasal patterns or discourse contexts in which unknown lexical items occur — we 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 therefrom. Appositive constructions (*"the laser printer X"*), e.g., constrain the conceptual status of the unknown item much more than, e.g., genitives (*"X's price"*).

Conceptual quality labels result from comparing the representation structures of a concept hypothesis with those of alternative concept hypotheses or a priori representation structures in the underlying domain knowledge base from the viewpoint of structural similarity, compatibility, etc. The closer the match with given knowledge, the more credit is lent to a hypothesis. For instance, a very positive conceptual quality label, M-DEDUCED, is assigned to multiple derivations of the *same* concept hypothesis in *different* hypothesis (sub)spaces, a definitely negative one is illustrated by INCONSISTENT, which annotates contradictory assertions.

Multiple concept hypotheses for each unknown lexical item are organized in terms of corresponding *hypothesis spaces*, each one holding a different or a further specialized concept hypothesis. The *quality machine* estimates the overall credibility of single concept hypotheses by taking the assembled 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 plausible concept hypotheses (for a formal specification of the underlying qualification calculus, cf. [SH98]).

4 The Learning Scenario Revisited

Depending on the type of the syntactic construction in which the unknown lexical item occurs, different hypothesis generation rules may fire. Genitives, such as *"The switch of the Itoh-Ci-8 ..."*, place by far fewer constraints on the item to be acquired than, say, appositives like *"The laser printer Itoh-Ci-8 ..."*. In the following, let *target* be the unknown item (*"Itoh-Ci-8"*) and *base* be the known item (*"switch"*), whose conceptual relation to the target is constrained by the syntactic relation in which their lexical correlates appear. The main constraint for genitives, e.g., says that the target concept fills (exactly) one of the n roles attached to the base concept. Since the correct role cannot yet be decided upon without additional evidence, n alternative, equally likely hypotheses have to be posited (unless additional constraints apply). Following on that, the target concept is assigned as a tentative filler of the i-th role of base in the corresponding i-th hypothesis space. As an immediate consequence, the classifier derives a suitable concept hypothesis by specializing the target concept (initially TOP, by default) according to the value restriction of the base concept's i-th role (cf. [HL94] for a similar constraint propagation mechanism). Additionally, the hypothesis generation rule assigns a syntactic quality label to each i-th hypothesis indicating the type of syntactic construction in which target and base co-occur (here, `Genitive`).

After the processing of *"The Itoh-Ci-8 has a size of ..."*, the target ITOH-CI-8 is already predicted as a PRODUCT. Prior to continuing with the phrase *"The switch of the Itoh-Ci-8 ..."*, consider a fragment of the conceptual representation for SWITCHes:

$$
\begin{aligned}
&(\text{P3})\ \text{SWITCH-OF} \doteq {}_{\text{SWITCH}}|\text{PART-OF}|_{\text{HARDWARE}}\\
&(\text{P4})\ \text{SWITCH} \doteq\\
&\quad \forall \text{HAS-PRICE}.\text{PRICE} \sqcap\\
&\quad \forall \text{HAS-WEIGHT}.\text{WEIGHT} \sqcap\\
&\quad \forall \text{SWITCH-OF}.\begin{pmatrix}\text{OUTPUTDEV} \sqcup \text{INPUTDEV} \sqcup\\ \text{STORAGEDEV} \sqcup \text{COMPUTER}\end{pmatrix}
\end{aligned}
$$

The relation SWITCH-OF is defined by (P3) as the set of all PART-OF relations which have their domain restricted to SWITCH and their range restricted to HARDWARE. In addition, (P4) reads as "all fillers of HAS-PRICE, HAS-WEIGHT, and SWITCH-OF roles must be concepts subsumed by PRICE, WEIGHT, and the disjunction (OUTPUTDEV ⊔ INPUTDEV ⊔ STORAGEDEV ⊔ COMPUTER), respectively". So, three roles have to be considered for relating the target ITOH-CI-8, as a tentative PRODUCT, to the base concept SWITCH. Two of them, HAS-PRICE and HAS-WEIGHT, are ruled out due to the violation of a simple integrity constraint (PRODUCT does not denote a unit of measure). Therefore, only the role SWITCH-OF must be considered. Due to the definition of SWITCH-OF (cf. P3), ITOH-CI-8 is immediately specialized to HARDWARE by the classifier. Since the classifier aggressively pushes hypothesizing to be maximally specific, four distinct hypotheses are immediately created due to the specific range restrictions of the role SWITCH-OF expressed in (P4), *viz.* OUTPUTDEV, INPUTDEV, STORAGEDEV and COMPUTER, and they are managed in four distinct hypothesis spaces, h_1, h_2, h_3 and h_4, respectively. Within h_1, h_2, and h_3, DEVICE, their common superconcept, is *multiply* derived by the classifier, too. Accordingly, this hypothesis is assigned a high degree of confidence by issuing the conceptual quality label M-DEDUCED.

5 Evaluation

The knowledge base on which we performed our experiments contained 1,215 concepts and 1,721 conceptual relations. We randomly selected from a corpus of information technology magazines 39 texts, with a total amount of 75 unknown words (= **TestSet**, see below) from a wide range of word classes (excluding VERBALS). For each of them, up to 16 learning steps were considered. A *learning step* captures the final result of all semantic interpretations being made after new textual input, usually a clause or sentence in which the item to be learned occurs, has been processed.

5.1 Offline Performance

In a first series of experiments, we neglected the incrementality of the learner and evaluated our system in terms of its *bare off-line* performance. By this we mean its potential to determine the correct concept description at the end of each text analysis, considering the outcome of the final learning step only. Following previous work on evaluation measures for learning systems [HL94], we distinguish here the following parameters:

- **Hypothesis** denotes the set of concept or grammatical class hypotheses derived by the system as the final result of the text understanding process for each target item;
- **Correct** denotes the number of cases in the test set in which **Hypothesis** contains the correct concept or grammatical class description for the target item;
- **OneCorrect** denotes the number of cases in the test set in which **Hypothesis** is a singleton set, i.e., contains only the correct concept or grammatical description;
- **ConceptSum** denotes the number of concepts generated by the system for the target item considering the entire test set.

Measures were taken under four experimental conditions (cf. Table 1). In the second column (indicated by –), we considered the contribution of a plain terminological reasoning component to the concept acquisition task, the third column contains the results of incorporating linguistic quality criteria only (denoted by **TH**), while the fourth column mirrors linguistic as well as conceptual quality criteria (designated by **CB**). The fifth column contains data from the grammar learner.

In an attempt to relate these results of the quality-based concept learner to a system close in spirit to our approach, we here consider CAMILLE [HL94], whose reasoning mode is similar to the terminological classifier we use. CAMILLE (cf. Table 1, column one), for the noun interpretation task, outperforms our system with respect to recall (44% *vs.* 29%), as well as with respect to precision (67% *vs.* 9%/19/%30%) under all three test conditions. Still, the data is hard to compare (and, even harder, to generalize) given the few number of learning cases in CAMILLE and in our system, diverging tasks (learning concepts denoted by nouns only in CAMILLE *vs.* learning concepts denoted by non-verbal items), and the different evaluation frameworks (amount and specificity of the background knowledge available). Two interesting observations, however, can be made. First, learning without the qualification calculus, just relying on terminological reasoning, leads to particularly disastrous precision results (9%). Second, in 27% of all learning cases our system derived a single and valid concept hypothesis.

We have no reasonable comparison data for the grammar learner right now. The success rate of to-day's best performing POS taggers (ranging on the order of 97-99% [Vou95,Bri95] should, nevertheless, be taken with caution in comparison to our framework, since the diversity and specificity of the word classes we employ is much higher

	CAMILLE	–	TH	CB	Grammar
TestSet	(9 * 2) = 18	75	75	75	75
Correct	8	22	22	22	66
OneCorrect	*	0	0	20	16
ConceptSum	12	248	114	74	169
RECALL := $\frac{\textbf{Correct}}{\textbf{TestSet}}$	44%	29.3%	29.3%	29.3%	88.0%
PRECISION : $\frac{\textbf{Correct}}{\textbf{ConceptSum}}$	67%	8.7%	19.3%	30.2%	39.1%

Table 1. Performance Measures for Concept and Grammar Learning

(on the order of 80) than that in PennTree bank-style grammars (with 36 POS tags). The data in Table 1 (column five) lists simply all word class hypotheses generated at all by the parser. Still they indicate a considerable recall (88%), while precision is low (39%). That number will substantially increase when we add a simple heuristics, *viz.* to count multiple derivations and to select the most frequent one(s) as the preferred hypothesis.

5.2 Online Performance: Learning Accuracy

In the context of class or type hierarchies, a prediction may be more or less precise, i.e., it may approximate the target concept at different levels of specificity. Varying degrees of the precision of hypotheses are captured by a measure of *learning accuracy*, which takes into account the topological distance of a hypothesis to the goal concept of an instance, i.e., the degree to which it correctly predicts the concept class which subsumes the target concept to be learned. Learning accuracy (LA) is formally defined as (n being the number of concept hypotheses for the target):[2]

$$LA := \sum_{i \in \{1 \ldots n\}} \frac{LA_i}{n} \quad \text{with} \quad LA_i := \begin{cases} \frac{CP_i}{SP_i} & \text{if } FP_i = 0 \\ \frac{CP_i}{FP_i - DP_i} & \text{else} \end{cases}$$

SP_i specifies the length of the *shortest path* (in terms of the number of nodes being traversed) from the TOP node of the concept hierarchy to the maximally specific concept subsuming the instance to be learned in hypothesis i; CP_i specifies the length of the path from the TOP node to that concept node in hypothesis i which is *common* both to the shortest path (as defined above) and the actual path to the predicted concept (whether correct or not); FP_i specifies the length of the path from the TOP node to the predicted (in this case *false*) concept ($FP_i = 0$, if the prediction is correct), and DP_i denotes the node *distance* between the predicted false node and the most specific common concept (on the path from the TOP node to the predicted false node) still correctly subsuming the target in hypothesis i.

Figure 4 depicts the learning accuracy curve for concept learning for the entire data set (75 items). We also have included the graph depicting the growth behavior of hypothesis spaces (Figure 5). For both data sets, we distinguish again between three measurements — **LA CB** gives the accuracy rate for the full qualification calculus including linguistic and conceptual quality criteria, **LA TH** for linguistic criteria only, while **LA –** depicts the accuracy values produced by the plain terminological reasoning component without incorporating any quality criteria. In Figure 4, we start from LA values in the interval between 31% to 34% for **LA –**/**LA TH** and **LA CB**, respectively, in the first learning step, whereas the number of hypothesis spaces (**NH**) range between 0.6 and 0.5. The latter values are due to the fact that an unknown lexical item cannot always be syntactically related when it occurs because it forms extragrammatical input, and therefore the generation of concept hypotheses fails. In the course of the analysis more and

[2] Note that 'learning accuracy', as we define it, is different from the notion of 'accuracy' in [HL94], in which the number of hypotheses which contain the correct interpretation is divided by the number of hypotheses generated at all.

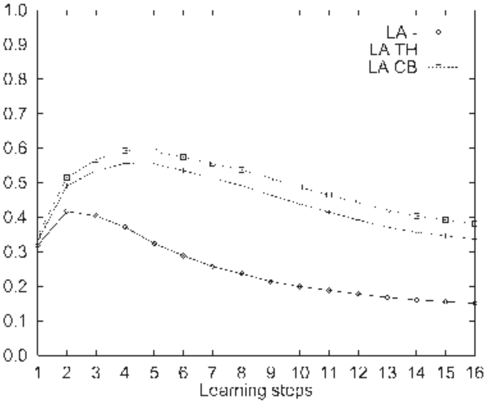

Fig. 4. Learning Accuracy (LA)

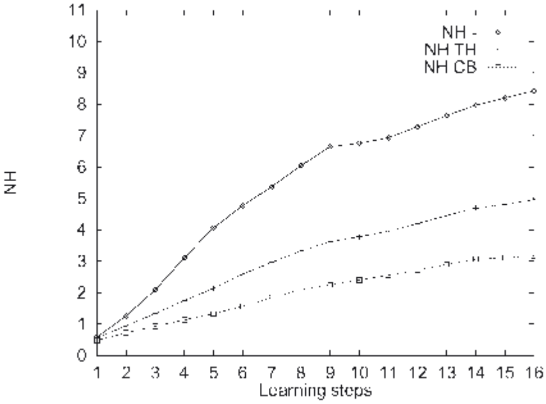

Fig. 5. Number of Hypotheses (NH)

more evidence can be collected for an appropriate grammatical and conceptual integration of the unknown item, though the number of hypotheses increases. With a maximum of 16 learning steps, learning accuracy approximates 15%, 34% and 39% for **LA –**, **LA TH** and **LA CB**, respectively, and the **NH** values rise to 8.4, 4.9 and 3.2 for each of the three criteria, respectively. Peak values are reached after 4 and 5 learning steps (60%) using the full qualification system.

This data diverges considerably from that generated under more 'friendly' experimental conditions [HS98], where we achieved LA values ranging between 79% (for **LA –**) and 87% (for **LA CB**). These superior results were due to perfect parses (we here included runs with partial parses, too), and the focus on one word class only (*viz.* nouns), while in the recent experiments we refrained from such restrictions in order to generate realistic data (similar decline effects are mentioned though not measured in [HL94]). The main findings from the previous study could be replicated, however. The pure terminological reasoning machinery consistently achieves an inferior level of learning accuracy and generates more hypothesis spaces than the learner equipped with the qualification calculus.

6 Conclusions

Knowledge-based systems provide powerful forms of reasoning, but it takes a lot of effort to equip them with the knowledge they need by means of manual knowledge engineering. In this paper, we have introduced an alternative solution based on an automatic learning methodology in which concept and grammatical class hypotheses emerge as a result of the incremental assignment and evaluation of the quality of linguistic and conceptual evidence related to unknown words. No specialized learning algorithm is needed, since learning is a (meta)reasoning task carried out by the classifier of a terminological reasoning system. This distinguishes our methodology from Cardie's case-based approach also combining conceptual and grammatical learning [Car93].

The work closest to ours has been carried out by [RJZ89] and [HL94]. They also generate concept hypotheses from linguistic and conceptual evidence. Unlike our approach, their selection of hypotheses depends only on an ongoing discrimination process based on the availability of this data but does not incorporate an inferencing scheme for reasoned hypothesis selection. The crucial role of quality considerations becomes obvious when one compares plain and quality-annotated terminological reasoning for the learning task. In the light of our evaluation study (cf. Fig. 4, final learning step) the difference amounts to 24%, considering between **LA -** (plain terminological reasoning) and **LA CB** values (terminological metareasoning based on the qualification calculus [SH98]).

References

[Bri95] E. Brill. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21:543–565, 1995.

[Car93] C. Cardie. A case-based approach to knowledge acquisition for domain-specific sentence analysis. In *Proceedings of the AAAI'93*, pages 798–803, 1993.

[Cha93] E. Charniak. Tree-bank grammars. In *Proceedings of the AAAI'96*, pages 1031–1036, 1993.

[GSH97] F. Gomez, C. Segami, and R. Hull. Determining prepositional attachment, prepositional meaning, verb meaning and thematic roles. *Computational Intelligence*, 13(1):1–31, 1997.

[HBN00] U. Hahn, N. Bröker, and P. Neuhaus. Let's PARSETALK: Message-passing protocols for object-oriented parsing. In H. Bunt and A. Nijholt, editors, *Advances in Probabilistic and other Parsing Technologies*, pages 177–201. Dordrecht: Kluwer, 2000.

[HL94] P. Hastings and S. Lytinen. The ups and downs of lexical acquisition. In *Proceedings of the AAAI'94*, pages 754–759, 1994.

[HS98] U. Hahn and K. Schnattinger. Towards text knowledge engineering. In *Proceedings of the AAAI'98*, pages 524–531, 1998.

[HSB94] U. Hahn, S. Schacht, and N. Bröker. Concurrent, object-oriented natural language parsing: The PARSETALK model. *International Journal of Human-Computer Studies*, 41(1/2):179–222, 1994.

[JT97] C. Jacquemin and E. Tzoukerman. Guessing morphology from texts and corpora. In *Proceedings of the SIGIR'97*, 1997.

[RJ99] E. Riloff and R. Jones. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the AAAI'99*, pages 474–479, 1999.

[RJZ89] L. Rau, P. Jacobs, and U. Zernik. Information extraction and text summarization using linguistic knowledge acquisition. *Information Processing & Management*, 25(4):419–428, 1989.

[RMH99] M. Romacker, K. Markert, and U. Hahn. Lean semantic interpretation. In *Proceedings of the IJCAI'99*, pages 868–875, 1999.

[SH98] K. Schnattinger and U. Hahn. Quality-based learning. In *Proceedings of the ECAI'98*, pages 160–164, 1998.

[SL94] S. Soderland and W. Lehnert. WRAP-UP: A trainable discourse module for information extraction. *Journal of Artificial Intelligence Research*, 2:131–158, 1994.

[SM00] B. Schiffman and K. McKeown. Experiments in automated lexicon building for text searching. In *Proceedings of the COLING 2000*, pages 719–725, 2000.

[Vou95] A. Voutilainen. A syntax-based part-of-speech analyser. In *Proceedings of the EACL '95*, pages 157–164, 1995.

[WS92] W. Woods and J. Schmolze. The KL-ONE family. *Computers & Mathematics with Applications*, 23(2/5):133–177, 1992.