Hybrid Connectionist Natural Language Processing

Stefan Wermter

Dept. of Computer Science

University of Hamburg

Vogt-Koelln-Str. 30

D-22527 Hamburg

Germany

Email: wermter@informatik.uni-hamburg.de

Preface

The objective of this book is to describe a new approach in hybrid connectionist natural language processing which bridges the gap between strictly symbolic and connectionist systems. This objective is tackled in two ways: the book gives an overview of hybrid connectionist architectures for natural language processing; and it demonstrates that a hybrid connectionist architecture can be used for learning real-world natural language problems. The book is primarily intended for scientists and students interested in the fields of artificial intelligence, neural networks, connectionism, natural language processing, hybrid symbolic connectionist architectures, parallel distributed processing, machine learning, automatic knowledge acquisition or computational linguistics. Furthermore, it might be of interest for scientists and students in information retrieval and cognitive science, since the book points out interdisciplinary relationships to these fields.

We develop a systematic spectrum of hybrid connectionist architectures, from completely symbolic architectures to separated hybrid connectionist architectures, integrated hybrid connectionist architectures and completely connectionist architectures. Within this systematic spectrum we have designed a system SCAN with two separated hybrid connectionist architectures and two integrated hybrid connectionist architectures for a scanning understanding of phrases. A scanning understanding is a relation-based flat understanding in contrast to traditional symbolic in-depth understanding. Hybrid connectionist representations consist of either a combination of connectionist and symbolic representations or different connectionist representations. In particular, we focus on important tasks like structural disambiguation and semantic context classification. We show that a parallel modular, constraint-based, plausibility-based and learned use of multiple hybrid connectionist representations provides powerful architectures for learning a scanning understanding. In particular, the combination of direct *encoding* of domain-independent structural knowledge and the *connectionist learning* of domain-dependent semantic knowledge, as suggested by a scanning understanding in SCAN, provides concepts which lead to flexible, adaptable, transportable architectures for different domains.

1

Introduction

1.1 Learning a Scanning Understanding

In the past there have been several approaches in natural language processing which aimed for an in-depth understanding of complete texts and dialogs. In contrast, we will pursue an approach of learning a *scanning understanding*[1] of separate natural language phrases. Natural language phrases are extremely frequent and important for describing objects, states and events, and they express the important contents efficiently. Examples of a scanning understanding of phrases include situations such as looking at the headlines in a newspaper and classifying book titles according to their semantic context. For instance, consider the following titles of books in a library.

Constructivist architecture in the USSR
Computer architecture and organization
Influence of the intensity on discharges in the Van Allen belt

A librarian has to learn to understand such phrases as quickly as possible, in most cases without a deep understanding of the book's field. Among other subtasks, contextual phrase analysis and structural phrase analysis have to be dealt with. For instance, the first phrase belongs to an architecture context, the second phrase belongs to a computer science context, and the third to a physical sciences context. Different lexical interpretations of the same word ('architecture' of a computer versus 'architecture' of a building) can be distinguished based on the semantic context. Furthermore, the third phrase shows a structural ambiguity since 'discharges' could refer

[1] The word 'scanning' is used here in the sense of 'glance at quickly' for the field of natural language processing. A scanning understanding is an understanding based on learned flat semantic relationships rather than encoded in-depth semantic relationships.

to 'intensity' and 'influence' if semantic knowledge is not considered. In a similar way, 'Van Allen belt' could refer to three different preceding nouns. So our subtask of *contextual phrase analysis also includes lexical disambiguation* between different contexts and the subtask of *structural phrase analysis includes structural disambiguation* in order to determine a desired structural interpretation. We take these two subtasks, contextual phrase analysis and structural phrase analysis, as examples of our overall task of a scanning understanding of phrases.

Contextual phrase analysis and structural phrase analysis are two subtasks for a scanning understanding of phrases but they are not independent. Rather, the structural phrase analysis interacts with the plausibility of semantic relationships within the semantic context of a phrase. That is, semantic relationships for structural phrase analysis depend on their semantic context: for instance, the underlying properties of the nouns of a semantic relationship 'architecture in USSR' are very different from the properties of the nouns in 'discharges in Van Allen belt'. We assume a parallel interleaved processing of contextual phrase analysis and structural phrase analysis. If contextual phrase analysis starts to interpret the first noun group of a phrase it makes predictions about the potential semantic context. However, only when the third noun group occurs can there be structural ambiguities based on prepositional phrase attachment or coordination. Consider, our example 'Influence of intensity on discharges in Van Allen belt'. When the word 'discharges' has been seen, this is the first possibility for a structural ambiguity. At that time contextual phrase analysis can make a prediction for a physical sciences context. Using the plausibility of semantic relationships for the physical sciences context, structural phrase analysis would attach 'discharges' to 'influence'. When the last noun group 'Van Allen belt' is processed the physical sciences context is confirmed. Using this context, finally the structural phrase analysis of the whole phrase can be performed, that is attaching 'discharges' to 'influence' and 'Van Allen belt' to 'discharges'.

Furthermore, there are phrases with semantic relationships whose plausibility depends on the semantic context. For instance, the semantic relationship 'tree in file' in an architecture or biology context would be implausible since a tree (plant) is too big to fit into a file (physical folder). However, if we consider the same phrase in a computer science context it could be plausible, since a tree (abstract data type) can be stored in a file (electronic medium). Now consider the examples 'Computer science notes on trees in files' and 'Architecture notes on trees in files'. In both cases the initial words 'computer science' and 'architecture' would initiate a prediction of these respective semantic contexts. In a computer science context, 'trees in files' would be plausible, and would lead to a structural attachment of 'files' to 'trees'. However, in the architecture context 'trees in files' would be implausible so that 'files' would be attached to 'notes'.

In these examples we can see the relationship between contextual phrase analysis and structural phrase analysis. Therefore, predictions about the current semantic context are important for structural disambiguation in order to reduce the search for applicable plausible semantic relationships for structural disambiguation. These subtasks, contextual phrase analysis and structural phrase analysis, belong to the overall task of a scanning understanding. While approaches for an in-depth understanding often use detailed encoded knowledge for processing some texts or dialogs, a scanning understanding of phrases aims at learning and processing language in a shallower manner for a wide variety of phrases. The underlying principle of our approach is to go as far as possible with as little knowledge as possible ('invest a little, get a lot').

1.2 The General Approach

In this section we will first focus on several properties of a scanning understanding of phrases before we motivate our general choice of representations. While interpreting a phrase, different *modular subtasks* can proceed in parallel but not all possible subtasks are required for all phrases[2]. Depending on the subtasks, different *constraints* interact in parallel to provide an overall phrase interpretation. For instance, for the subtask of structural phrase analysis, syntactic constraints may suggest a local attachment of a noun group while semantic constraints may suggest a distant attachment. Such constraints have to be integrated depending on their plausibility. Therefore, *graded plausibility* is another essential property for processing phrases. Since natural language concepts for phrases are inherently fuzzy rather than true or false, a graded plausibility is very important for a flexible interpretation of phrases. For instance, a semantic relationship 'person in house' may be plausible rather than implausible. Such plausible knowledge can be supported by *learning* regularities from different examples. In particular, for semantic knowledge about plausibility, learning is important to increase adaptability for different semantic domains.

So far we have pointed out the following properties as essential for a scanning understanding:

- Parallel modular subtasks
- Integration of different constraints
- Graded plausibility
- Learning and adaptability

These properties are essential for building computational models for

[2] For instance, metaphorical phrase analysis is not needed in many phrases, structural phrase analysis is not needed in one-word phrases, etc.

phrase analysis, but so far there has been no natural language architecture for phrasal analysis which focuses explicitly on these properties. Many systems for natural language understanding have an explicitly coded, fixed sequential or hierarchical architecture [von Hahn 1992] rather than a parallel subtask-oriented, constraint-based architecture. Furthermore, most previous systems have been strictly symbolic and they do not particularly support graded plausible interpretations, fault-tolerant analysis and learning.

These different properties motivated our choice of examining different hybrid connectionist representations. *Hybrid connectionist representations* denote connectionist representations which have been combined with other different connectionist representations or other symbolic representations. Connectionist networks support parallel modular subtasks, constraint integration, graded plausibility and learning. Therefore, we examine an architecture which focuses on connectionist representations, but we also do not rule out the use of symbolic representations, which can still play an important role, in particular for encoding some well-known domain-independent regularities. In this work we claim that a parallel modular, constraint-based, plausibility-based, and learned use of multiple hybrid connectionist representations provides a powerful new architecture for a scanning understanding of phrases. In particular, the combination of direct *encoding* of domain-independent structural knowledge and the *connectionist learning* of domain-dependent semantic knowledge provides concepts which lead to more flexible, adaptable, transportable architectures for different domains.

1.3 Towards a Hybrid Connectionist Memory Organization

One of the most influential proposals for a general framework of artificial intelligence architectures has been the approach of Marr [Marr 1982]. According to his proposal, the top-most level of *computational theory* contains the general strategies, goals and concepts. This computational theory level specifies *what* is processed and which general assumptions are made. In contrast, the level of *representation and algorithm* describes *how* the general strategies, goals and concepts of the computational theory level are represented[3]. Representative symbolic three-level architectures within Marr's framework are modular and sequential [Newell 1980] [Pylyshyn 1984]. In contrast, representative connectionist architectures are interactive and parallel, and rule-like symbolic knowledge can emerge naturally based on learning and generalization [Feldman and Ballard 1982] [McClelland et al. 1986]. We will argue that, at the representation level of a natural language system for phrase analysis, symbolic and

[3] The bottom-most level of the *hardware implementation* then provides the realization on a physical machine. In this work we do not focus on the hardware implementation.

connectionist representations are not as incompatible as has sometimes been argued, but they are mutually complementary.

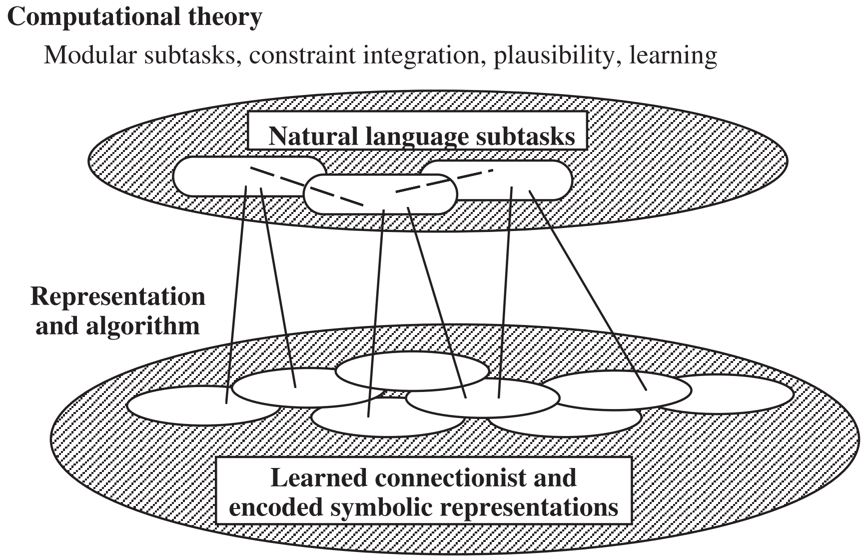

Figure 1.1 *Hybrid connectionist memory organization.*

Figure 1.1 describes a hybrid connectionist memory organization in the context of Marr's general principles for information processing. At the *computational theory level* subtask-dependent modularity, constraint integration, plausibility and learning are key issues for a scanning understanding. At the *representation level* connectionist and symbolic representations coexist based on a plausibility view which can support both plausible continuous connectionist representations as well as discrete symbolic representations.

Connectionist and hybrid connectionist models often use modularity for organizing memory more effectively (e.g. [Waltz and Pollack 1985] [Miikkulainen and Dyer 1989] [Jain and Waibel 1990]). Furthermore, there has been psychological evidence that human language processing is based on parallel but modular subtasks, for instance for structural and semantic processing (e.g. [Aaronson and Ferres 1986]). We will address this issue of modular and parallel design in the framework of a scanning understanding of phrases using the subtasks of contextual phrase analysis and structural phrase analysis. We will argue that an architecture for a scanning understanding can take advantage of modularity and parallel interaction between such subtasks. This subtask-oriented memory organization will be examined in different models using hybrid connectionist representations. In this context we will address several general issues: What are important properties and strategies for a scanning understanding of phrases? To what extent is it useful to combine symbolic and connectionist representations for natural language processing? How can we compare different hybrid (separated and integrated) connectionist models for subtasks like structural phrase analysis and contextual phrase analysis? How can connectionist learning of domain-independent semantic knowledge be combined with encoding domain-dependent knowledge for phrase analysis?

1.4 An Overview of the SCAN Architecture

Our goal is to examine the learning of a scanning understanding based on a Symbolic/Connectionist Approach for Natural language phrases (SCAN). Different natural language subtasks like structural phrase analysis and contextual phrase analysis can benefit from learning within a modular hybrid connectionist memory organization. In general, we take an approach which leads to a subtask-oriented hybrid connectionist memory organization based on encoded and learned representations. Rather than pursuing an architecture for a scanning understanding as a single connectionist network or a single symbolic structure, we examine the use of a learned semantic memory according to various modular subtasks.

The overall approach is illustrated in figure 1.2. The top-most computational theory level specifies the subtask-dependent modularity, constraint integration, plausibility and learning for the subtasks, while the representation level contains the actual underlying symbolic and connectionist representations. We will primarily focus on the subtasks of structural phrase analysis and contextual phrase analysis, but the general architecture could also be extended to other subtasks like case role analysis and metaphorical phrase analysis.

Central to this architecture is a connectionist *plausibility computation* for semantic relationships and semantic context. If the semantic context of a phrase is known *a priori*, for instance the physical sciences context, plausibility computation only needs to provide the plausibilities of semantic relationships within the known context. However, if the semantic context of a phrase is not known *a priori*, plausibility computation also provides hypotheses about the semantic context in order to determine the plausibility of semantic relationships within the semantic context. The plausibility computation interacts with the *structure computation* which generates structures and condensations based on symbolic and localist connectionist representations. Both plausibility computation and structure computation have access to the lexicon that contains syntactic categories, semantic features and semantic dimensions of words.

If SCAN is in learning mode a generated representation is compared with a desired representation. The evaluation of this comparison can then be used to trigger further learning in the plausibility computation. While the

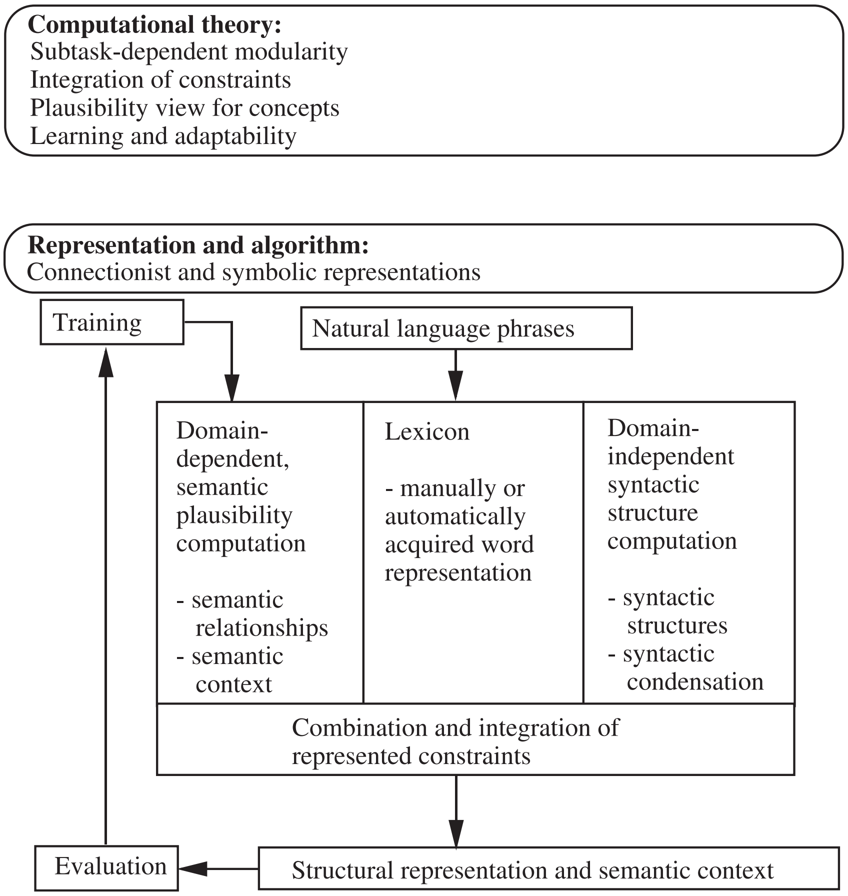

Figure 1.2 *Overview of the general architecture of SCAN.*

domain-independent structure computation is based on predefined knowledge, the domain-dependent plausibility computation can be learned inductively based on given training examples. One major issue for SCAN is to represent known domain-independent syntactic regularities directly and to learn unknown domain-dependent semantic regularities in connectionist networks.

1.5 Organization and Reader's Guide

We examine our approach for a scanning understanding for the two subtasks of structural and contextual phrase analysis. SCAN receives a phrase like a noun phrase and performs a structural disambiguation and semantic context classification of phrases.

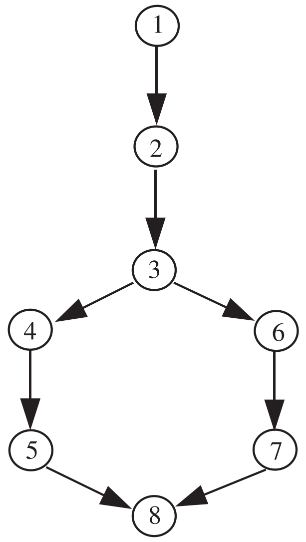

Figure 1.3 *Reader's guide.*

Figure 1.3 gives an overview of the dependencies between various chapters. After this introduction, chapter 2 gives an overview about related hybrid and connectionist architectures for learning natural language concepts. Then, the main phrasal constructions and ambiguities are outlined in chapter 3. Furthermore, chapter 3 provides the basic framework for the subsequent models, in particular connectionist plausibility networks. The following four chapters focus on our subtasks structural phrase analysis and contextual phrase analysis. For each subtask two possible models are designed in order to compare their mutual advantages and disadvantages. The first two models on structural phrase analysis assume that the semantic context of the phrase has already been identified, while the last two models on contextual phrase analysis focus on determining the incremental semantic context of a phrase. The separated disambiguation model (*SD model*) in chapter 4 combines a symbolic chart parser with connectionist plausibility networks for structural phrase analysis. The integrated disambiguation model (*ID model*) in chapter 5 is a combination of connectionist relaxation networks and plausibility networks for structural phrase analysis. While we emphasize structural phrase analysis in these two chapters, chapters 6 and 7 deal with contextual phrase analysis. The separated context model (*SC model*) in chapter 6 uses a symbolic chart parser for a condensation of noun phrases to compound nouns and learns their semantic context in a recurrent connectionist plausibility network. Finally, the integrated context model (*IC model*) in chapter 7 uses optional simple symbolic significance heuristics and learns semantic context classification in a recurrent connectionist plausibility network. In chapter 8 we discuss and evaluate our models and show how this architecture could be extended to other subtasks.

2

Order information from the publisher

.

Title: Hybrid Connectionist Natural Language Processing
Author: Stefan Wermter
Publisher: International Thomson Computer Press (Chapman and Hall)
ISBN: 0 412 59100 6
Pages: 190
Figures: 56
Price: 29.95 pounds sterling, 52.00 US dollars
Credit cards: all major credit cards accepted

For orders outside the US, please contact:

.

International Thomson Computer Press (Chapman and Hall)
Cheriton House
North Way
Andover
Hampshire, SP10 5BE, UK
United Kingdom

UK orders:
Tel: 01264 332424
Fax: 01264 342787
Internet: ITPUK@ITPS.CO.UK

Overseas orders:
Tel: +44 1264 332424
Fax: +44 1264 342787
Internet: ITPINT@ITPS.CO.UK

For orders in the US, please contact:

.

International Thomson Computer Press (Chapman and Hall)
7625 Empire Drive
Florence, KY 41042-2978
USA
Tel: 1-606-525-6600
Fax: 1-606-525-6600
Toll Free: 1-800-842-3636
Internet: ITPINT@ITPS.CO.UK

References

[Aaronson and Ferres 1986] Aaronson D., Ferres S. (1986) Reading strategies for children and adults: a quantitative model. *Psychological Review 93(1)*, pp. 89–112.

[Feldman and Ballard 1982] Feldman J. A., Ballard D. H. (1982) Connectionist models and their properties. *Cognitive Science 6*, pp. 205–254.

[Jain and Waibel 1990] Jain A. N., Waibel A. H. (1990) Incremental parsing by modular recurrent connectionist networks. In: Touretzky D. S. (Ed.) *Advances in Neural Information Processing Systems, 2*, pp. 364–371. Morgan Kaufmann, San Mateo, CA.

[Marr 1982] Marr D. (1982) *Vision.* Freeman, San Francisco, CA.

[McClelland et al. 1986] McClelland J. L., Rumelhart D. E., Hinton G. E. (1986) The appeal of parallel distributed processing. In: Rumelhart D. E., McClelland J. L. (Eds.) *Parallel Distributed Processing, Vol. 1*, pp. 3–44. MIT Press, Cambridge, MA.

[Miikkulainen and Dyer 1989] Miikkulainen R., Dyer M. G. (1989) A modular neural network architecture for sequential paraphrasing of script-based stories. *Proceedings of the International Conference on Neural Networks*, pp. II 49–56. Washington.

[Newell 1980] Newell A. (1980) Physical symbol systems. *Cognitive Science 4*, pp. 135–183.

[Pylyshyn 1984] Pylyshyn Z. (1984) *Computation and Cognition.* MIT Press, Cambridge, MA.

[von Hahn 1992] von Hahn W. (1992) Von der Verknüpfung zur Integration: Kontrollstrategie oder kognitive Architektur. *Proceedings of the KONVENS92*, pp. 1–10. Nürnberg, FRG.

[Waltz and Pollack 1985] Waltz D. L., Pollack J. B. (1985) Massively parallel parsing: a strongly interactive model of natural language interpretation. *Cognitive Science 9*, pp. 51–74.