Integration of Semantic and Syntactic Constraints for Structural Noun Phrase Disambiguation*

Stefan Wermter
Department of Computer and Information Science
University of Massachusetts
Amherst, Massachusetts 01003

Abstract

A fundamental problem in Natural Language Processing is the integration of syntactic and semantic constraints. In this paper we describe a new approach for the integration of syntactic and semantic constraints which takes advantage of a learned memory model. Our model combines localist representations for the integration of constraints and distributed representations for learning semantic constraints. We apply this model to the problem of structural disambiguation of noun phrases and show that a learned connectionist model can scale up the underlying memory of a Natural Language Processing system.

1 Introduction

The structural and semantic understanding of noun phrases and prepositional phrases is one of the most important tasks for natural language processing systems. Lately issues of prepositional phrase attachment have been addressed in different systems for sentence understanding (e.g. [Wilks et al. 85], [Schubert 86], [Dahlgren and McDowell 86], [McClelland and Kawamoto 86], [St. John and McClelland 88]). These systems focus on deciding whether a prepositional phrase attaches to a verb phrase or a noun phrase, for instance [Wilks et al. 85]:

The woman wanted the dress on the rack.
The woman positioned the dress on the rack.

In the first example "on the rack" attaches to the noun phrase "the dress", in the second example to the verb "positioned".

All these referenced systems emphasize prepositional phrase attachment in sentences of the form <S><VP><NP><PP>, and concentrate on the attachment of a *single* prepositional phrase based on predictive verbal knowledge. However, attachment decisions for *multiple* prepositional phrases have to rely on syntactic and semantic knowledge associated with nouns and prepositions as well. The importance of this knowledge about nouns and prepositions is very obvious for the attachment decisions in isolated noun phrases, as for example in titles of scientific articles. In this paper we restrict our efforts to prepositional attachment in noun phrases using a corpus of titles and scientific articles from the physical sciences, for instance:

Forces on charged particles of a plasma in a cavity resonator.
Irregularities in the drag effects on sputniks.

We describe a two-level architecture for integrating syntactic and semantic constraints to disambiguate PP-attachment in noun phrases. The bottom level consists of backpropagation networks using distributed representations for the semantic relationships between nouns and prepositions. The backpropagation networks are trained with examples of these prepositional relationships for each preposition, so that the backpropagation networks learn the underlying semantic constraints. The top level consists of a relaxation network using localist representations for the integration of syntactic constraints with the learned semantic constraints. This approach allows the disambiguation of noun phrases which the system has not been trained on.

2 Noun Features for Prepositional Relationships

Prepositional relationships depend on domain-specific features of the involved nouns. The noun phrases for our experiments were taken from the NPL corpus [Sparck-Jones 76] which contains article titles for scientific and technical domains. Typical examples in the corpus are:

Pulse techniques for probe measurements in gas discharges.
The influence of the radiation intensity on discharges in the Van-Allen-belt.

For each of the 10 most frequent prepositions in the corpus, 100 noun phrases were extracted which contained the specific preposition. The typical structure of the considered noun phrases is a sequence of up to five noun

*This research was supported by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract #N00014-87-K-0238, the Office of Naval Research, under a University Research Initiative Grant, Contract #N00014-86-K-0764 and NSF Presidential Young Investigators Award NSFIST-8351863

groups each separated by a preposition. The head noun in the noun group was characterised with semantic features. We found the following 16 features useful as a basic representation for the noun groups in this domain (see Figure 1).

Features	Examples
MEASURING-EVENT	Observation
CHANGING-EVENT	Amplification
SCIENTIFIC-FIELD	Mechanics
PROPERTY	Intensity
MECHANISM	Experiment
ELECTRIC-OBJECT	Transistor
PHYSICAL-OBJECT	Earth
RELATION	Cause
ORGANIZATION-FORM	Layer
GAS	Air
SPATIAL-LOCATION	Antarctic
TIME	June
ENERGY	Radiation
MATERIAL	Aluminium
ABSTRACT-REPRESENTATION	Note
EMPTY	Cavity

Figure 1: Features of the Nouns and Examples

Most nouns have a clear preference for one of the 16 features, for example "June" for TIME. Although prepositional relationships could be defined with one feature class [Herskovits 86], nouns can have more than one feature, for example "radiation" can be a form of ENERGY, and a CHANGING-EVENT. To account for these multiple features of single nouns each noun is represented as a binary vector of length 16.

3 The Structural Disambiguation of Noun Phrases

The disambiguation of noun phrases relies on two types of knowledge: first, semantic, domain-dependent constraints for the plausibility of prepositional relationships and second, syntactic, domain-independent constraints for crossing dependencies and locality.

3.1 The Bottom Level: Learning Semantic Constraints with Backpropagation Networks

Semantic constraints based on the plausibility of prepositional relationships determine how different prepositional phrases in a noun phrase can attach to one another. In many systems semantic constraints are formulated as rules (e.g. [Wilks et al. 85] [Dahlgren and McDowell 86]). We describe a different approach for learning the semantic constraints in prepositional relationships.

Learning prepositional relationships for different prepositions is defined as learning to differentiate between plausible prepositional relationships and implausible prepositional relationships. Plausible prepositional relationships are relationships which can be true. For instance, the prepositional relationship "radiation in atmosphere" is plausible. Implausible prepositional relationships are relationships which violate semantic restrictions. For instance, the prepositional relationship "symposium in ionosphere" violates semantic restrictions because meetings are not supposed to take place in the upper atmosphere.

Backpropagation networks are useful to learn the plausibility of prepositional relationships between two nouns and to generalize the regularities for the plausibility of pairs of nouns with which the network has not been trained. We used the backpropagation algorithm as described in [Rumelhart et. al 86]. One backpropagation network is used for representing the prepositional relationships for one preposition. Each network consists of 32 input units, 12 hidden units and one output unit (see figure 2). The input units represent the binary features of the two nouns. The output unit is a real value between 0 and 1 representing the plausibility of the prepositional relationship between two nouns. The hidden units represent the learned mapping between the noun features and the plausibility value.

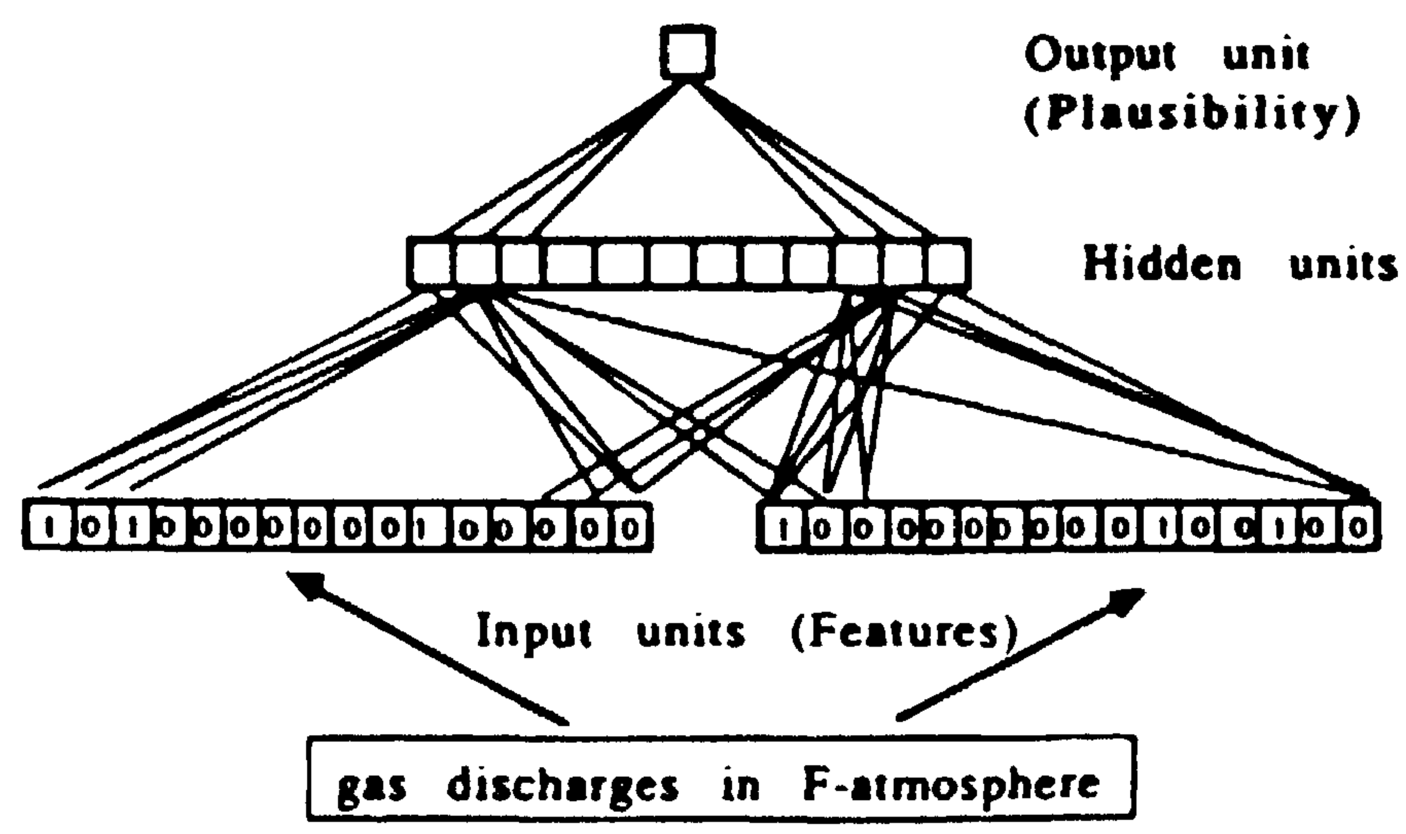


Figure 2: Bottom Level: Backpropagation Network for Learning Prepositional Relationships for the Preposition "in" (only some connections shown)

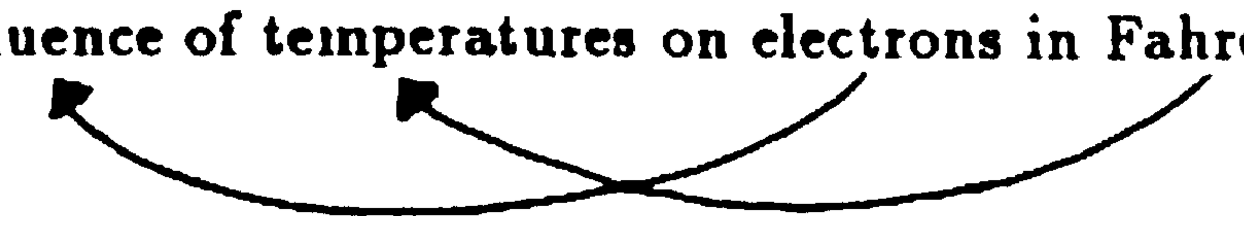
The backpropagation networks were trained by presenting about 200 training examples for each specific preposition. A training example consisted of the feature representations for the two nouns together with the plausibility value " 1 " for "plausible" or "0" for "implausible". After the backpropagation networks were trained for 1600 epochs with the training set, each network was tested with the training set and a testing set. The testing set consisted of 30 examples of prepositional relationships (each characterized by 32 features) which the network had not been trained on. A prepositional relationship was considered correct on a scale from 0 to 1 if the value of the output unit was higher than 0.5 for a desired plausible relationship and smaller than 0.5 for a desired implausible relationship. The testing results for three examined prepositions [Wermter 89] showed that the backpropagation networks learned almost all prepositional relationships in the training set and most of the relationships in the testing set. For instance, the network for the preposition "in" got 93% of the 248 training examples correct and 83% of 30 unknown testing examples.

3.2 The Bottom Level: Representing Syntactic Constraints

The first form of syntactic knowledge considered for noun phrase disambiguation is the locality constraint. The Locality constraint models the heuristic that a prepositional phrase in a noun phrase is more likely to attach to a close preceding noun than to a distant preceding noun. For instance in a noun phrase like "techniques for measurements in discharges" the prepositional phrase "in discharges" tends to attach to "measurements", although "in discharges" could attach to "techniques" as well. The locality constraint can be interpreted as a generalisation of Right Association [Kimball 73]. While Right Association for a noun phrase states that a prepositional phrase attaches to the directly preceding noun, the locality constraint claims that there is only a strong tendency for a local attachment to directly preceding nouns. This tendency decreases with the distance between noun and prepositional phrase.

The second form of syntactic knowledge is the No-crossing constraint. The no-crossing constraint states that the prepositional phrase attachment in a noun phrase does not show crossing branches (see e.g. [Tait 83]). The following constructed example illustrates a violated no-crossing constraint. Although "influence on electrons" and "temperatures in Fahrenheit" are plausible prepositional relationships, this structural interpretation is considered wrong due to the crossing attachment.

Influence of temperatures on electrons in Fahrenheit.



3.3 The Top Level: Integrating Syntactic and Semantic Constraints in a Relaxation Network

The semantic constraints for the prepositional relationships and the syntactic constraints for no-crossing and locality are integrated in a relaxation network to allow parallel interactions between these different constraints. In the past, relaxation networks have been shown to be successful for integrating different constraints in a variety of natural language tasks like sentence processing [Waltz and Pollack 85], word sense disambiguation [Bookman 87], attachment decisions [Lehnert 89] and lexical access [Cottrell 88). These approaches depend on the initialisation of the input nodes with suitable values bul this decision is not based on a memory model. In our new approach we demonstrate that (1) trained back propagation networks supply a more powerful underlying model for the input of a relaxation network and (2) relaxation networks are extremely useful for integrating different constraints for structural noun phrase disambiguation.

First we will describe the interface between our two levels, then we will outline the overall architecture of the relaxation network at the top level. This description is illustrated with an example of a noun phrase with three prepositions:

The influence of the radiation intensity on discharges in the Van-Allen-belt.

3.3.1 The Interface between the Top and Bottom Levels

The interface between the two levels is represented with three types of nodes: semantic nodes, locality nodes, and no-crossing nodes (see figure 3). In our example there are six Semantic nodes representing the semantic constraints for the six possible prepositional relationships: influence of intensity, influence on discharges, influence in Van-All en-belt, intensity on discharges, intensity in Van-Allen-belt, discharges in Van-Allen-belt.

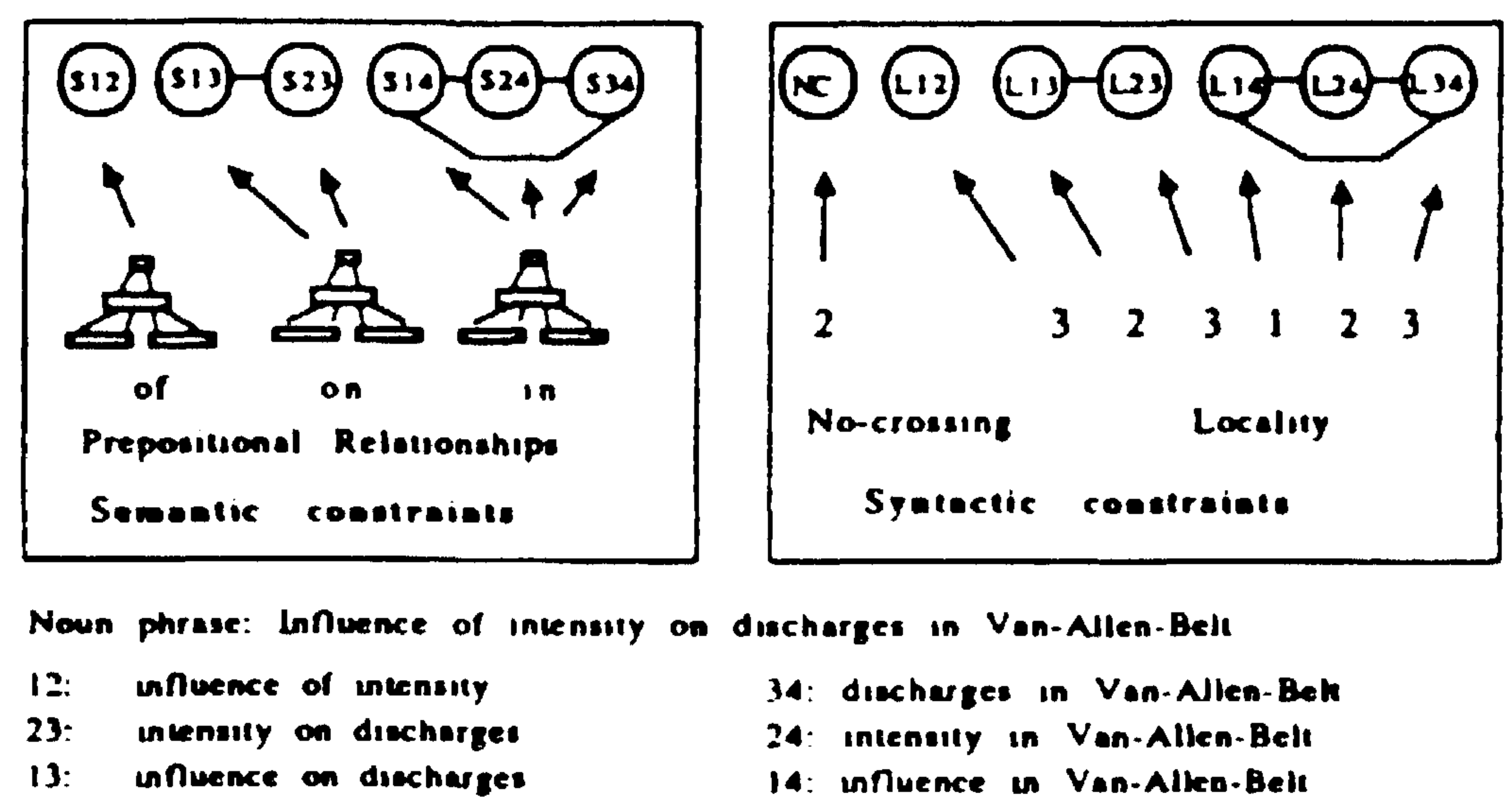


Figure 3: The Interface between Top Level and Bottom Level: Input Nodes for Semantic and Syntactic Constraints

The input potential for the six semantic nodes in the relaxation network is based on the output units of the back propagation networks described in section 3.1. The semantic nodes representing "influence of intensity", "influence on discharges", "influence in Van-Allen-belt", "intensity in Van-Allen-belt", and "discharges in Van-Allen-belt" get high input potential, because these relationships are plausible. The semantic node for "intensity on discharges" gets a low input potential, because that relationship is implausible.

In addition to the semantic nodes there are seven syntactic nodes representing the syntactic constraints for locality and crossing dependencies. The potential of the six Locality nodes reflects the distance between the nouns of a prepositional relationship: the closer the nouns of a prepositional relationship in the noun phrase, the higher the potential of the node. For instance, "influence of intensity" gets a higher value than "influence in Van-Allen-belt" because the nouns in the first prepositional relationship are closer. The one No-crossing node prevents crossing attachments, so that in noun phrases with three prepositions the third noun cannot attach to the first noun while the fourth noun attaches to the second. The connections of all nodes are described in the next section.

3.3.2 The Top Level: Architecture of the Relaxation Network

The relaxation network (see figure 4) consists of nodes connected via inhibitory and excitatory connections and can be generated for noun phrases with different lengths. For noun phrases with three prepositions there are 13 input nodes and six output nodes. The input nodes for the semantic constraints and locality constraints are connnected via inhibitory connections if the two prepositional relationships have the same noun in the second position of the prepositional relationship and a different noun in the first position. For example "influence on discharges" and "intensity on discharges" are connected via inhibitory connections, because "influence" competes with "intensity" for "discharges".

The output nodes represent the six possible structural interpretations of the noun phrase. Therefore the output nodes will be referred to as Structure nodes. One structure node can be described as a triple of number pairs. Each number stands for the position of a noun in a noun phrase, for instance the triple "1-2,2-3,3-4" is the representation for "influences of intensity", "intensity on discharges" and "discharges in Van-Allen-belt". All structure nodes are in competition and connected via inhibitory connections.

The semantic nodes and the locality nodes are connected with the structure nodes via excitatory connections if the prepositional relationship of the input node occurs in the structure node. The no-crossing node is inhibitorily connected to the structure node "1-2,1-3,2-4" which represents crossing dependencies.

3.3.3 Processing in the Relaxation Network

The nodes in the relaxation network are initialized with a potential between 0 and 10. The semantic nodes receive input based on the output of the backpropagation networks. They obtain a high start potential of 10 for a plausible prepositional relationship and a low start potential of 2 for an implausible prepositional relationship. The initialisation values of the locality nodes depend on the distance between nouns in a noun phrase. For instance, if the attachment is over 1 preposition we initialise with 3, for attachment over 2 prepositions with 2, and for attachment over 3 prepostions with 1. This ensures that local attachment gets more reinforcement than distant attachment. The rest of the nodes, the no-crossing node and the structure nodes, are initialized with low values of 2.

Once the relaxation algorithm [Feldman and Ballard 82] is started, nodes update their potential. Incoming excitatory connections increase the potential of a node, incoming inhibitory connections decrease the potential. One cycle consists of updating every node once. Although our implementation of this process is sequential, the actions within one cycle could be processed in parallel. After about 30 cycles the network converges to a stable state in which the potentials do not change any more. The structure node with the highest potential represents the preferred structural interpretation of the noun phrase.

In our example "The influence of the radiation intensity on discharges in the Van-Allen-belt" the following structure node had the highest potential of 8.9 at the end of the relaxation (the other structure nodes had values around 0.9):

Influence of intensity
Influence on discharges
Discharges in Van-Allen-belt

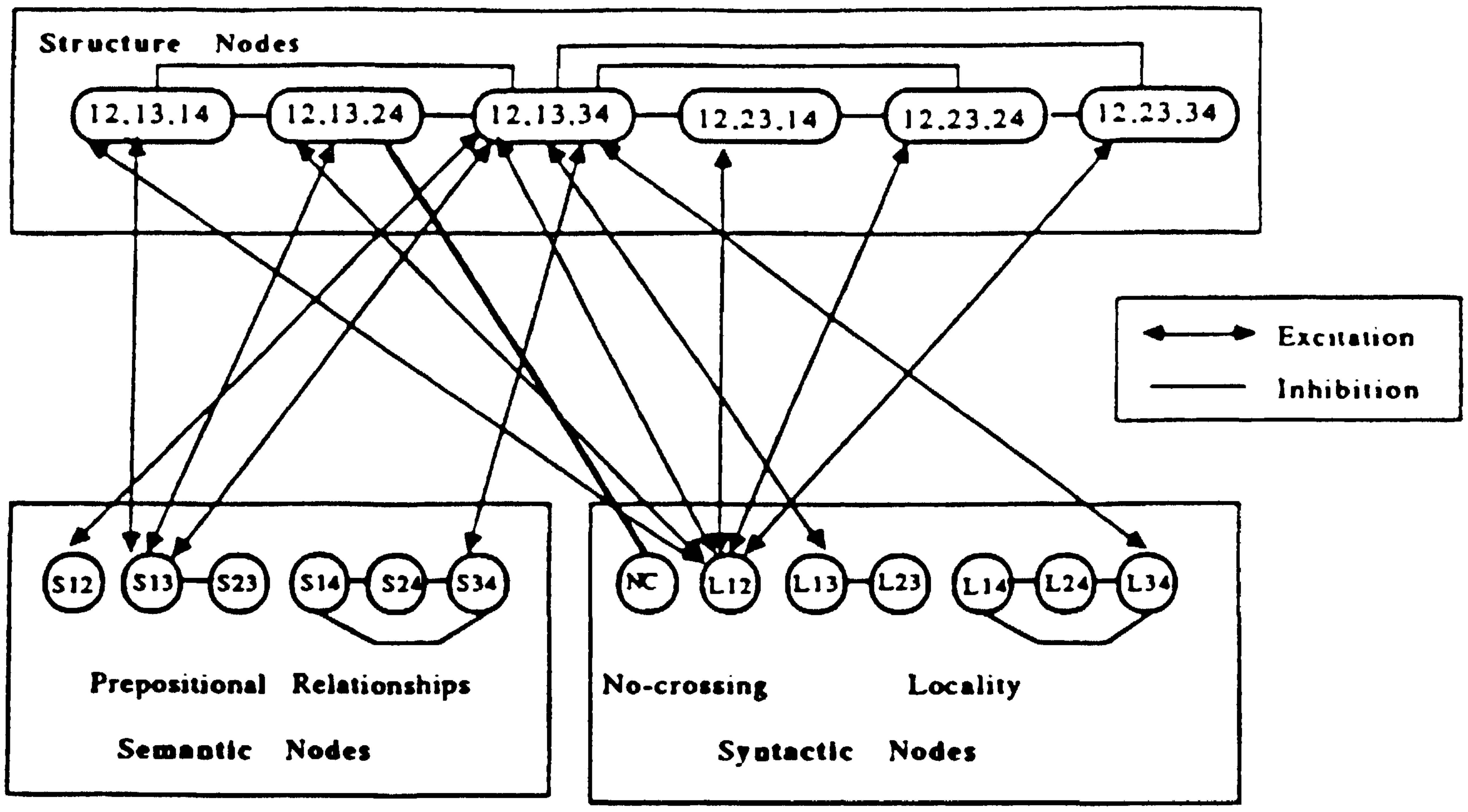
The network integrated the syntactic and semantic constraints: the semantic constraint "intensity on discharges" is implausible and therefore the semantic constraint "influence on discharges" is found as the preferred attachment for "discharges", although the syntactic locality constraint prefers the local attachment "intensity on discharges" compared to "infuence on discharges". This example shows how semantic constraints can overrule locality constraints.

Looking at the noun "Van-Allen-belt" we notice the syntactic influence. "Van-Allen-belt" could attach to all three preceding nouns, because all these prepositional relationships are plausible. At the same time the locality constraint imposes a preference for a local attachment, so that "discharges in Van-Allen-belt" is preferred to "influence in Van-Allen-belt" and "intensity in Van-Allen-belt".

4 Discussion

We use two different mechanisms at two levels for the task of structural noun phrase disambiguation. At the domain-dependent bottom level we use distributed representations and backpropagation networks for each preposition to learn the semantic relationships. At the domain-independent top level we use localist representations and a relaxation network to integrate syntactic and semantic constraints. Although work on related relaxation networks has to rely on some initial setting of the start activation (e.g. [Waltz and Pollack 85], [Bookman 87], [Lehnert 87], [Cottrell 88]), our model bases its initialization on learned memory. While other work on PP-attachment has mostly concentrated on the attachment of single prepositional phrases in sentences ([Wilks et al. 85], [Schubert 86], [Dahlgren and McDowell 86], [McClelland and Kawamoto 86], [St. John and McClelland 88]) we have concentrated on the attachment of multiple prepositional phrases in noun phrases.

Our approach demonstrates progress over related connectionist work [Cosic and Munro 88] by using distributed representations for nouns, by integrating semantic and syntactic constraints and by allowing for noun phrases with arbitrary length. We must also point out that our underlying memory model of prepositional relationships can be used as part of a full sentence ana-



Noun phrase: Influence of intensity on discharges in Van-Allen-Belt

12: influence of intensity
23: intensity on discharges
13: influence on discharges
34: discharges in Van-Allen-Belt
24: intensity in Van-Allen-Belt
14: influence in Van-Allen-Belt

Connections: Only the excitatory connections for the semantic node SI3. for the syntactic node L12 and for the structure node 12.13.34 are shown completely

Only the inhibitory connections for structure node 12.13.34 are shown completely

Figure 4: Top Level: Relaxation Network for the Integration of Semantic and Syntactic Constraints (only some connections shown)

lyser as well. For example, the sentence analyser CIRCUS [Lehnert 89] can combine our semantic memory model with predictive knowledge during sentence processing.

5 Conclusions

We have described an approach for learning and integrating semantic and syntactic constraints. Backpropagation networks and distributed representations are used to learn the plausibility of semantic relationships and to generalise the learned regularities to semantic constraints. Relaxation networks and local is t representations are used to integrate these semantic constraints with syntactic constraints. We have demonstrated that a connectionist model supplies a powerful memory model for the learning and integration of constraints for structural noun phrase disambiguation. Since the problem of learning and integrating constraints occurs in many other language tasks like word sense disambiguation or compound noun interpretation, our memory model is of importance for many Natural Language Processing problems.

Acknowledgements

I would like to thank Wendy Lehnert, Andy Barto, Steven Bradtke, David Lewis and the NLP group for discussions and comments on earlier drafts of this paper.

References

1. Bookman L.A. 1987. A microfeature based scheme for modelling semantics. *Proceedings of the IJCAI*
2. Cottrell G.W. 1988. A Model of Lexical Access of ambiguous words. In: Small, S.I., Cottrell,G.W., Tanenhaus,M.K. *Lexical Ambiguity Resolution*
3. Cosic C, Munro P. 1988. Learning to represent and understand locative prepositional phrases. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society.*
4. Dahlgren K., McDowell J. 1986. Using Commonsense Knowledge to disambiguate Prepositional Phrase Modifiers. *Proceedings of the AAAL*
5. Feldman J.A., Ballard D.H. 1982. Connectionist Models and their properties. *Cognitive Science 6*

6. Herskovits A. 1986. *Language and spatial cognition.* Cambridge University Press.

7. Kimbail J. 1973. Seven principles of surface structure parsing in natural language. *Cognition 2.*

8. Lehnert W.G. 1987. Learning to integrate syntax and semantics. *Proceedings of the Fourth International Workshop on Machine Learning.*

9. Lehnert W.G. 1989. Symbolic/Subsymbolic Sentence Analysis: Exploiting the Best of Two Worlds. In: Barnden,J., Pollack,J.: *Advances in Connectionist and Neural Computation Theory* Vol 1, Ablex Publishers, in press (also available as COINS Technical Report 88-99, Department of Computer and Information Science, University of Massachusetts, Amherst, MA)

10. McClelland J.L., Kawamoto A.H. 1986. Mechanisms of Sentence Processing: Assigning Roles to Constituents. In: McClelland,J.L., Rumelhart,D.E. (Ed.) *Parallel Distributed Processing Vol 2.* MIT Press.

11. Rumelhart D.E., Hinton G.E., Williams R.J. 1886. Learning Internal Representations by Error Propagation. In: McClelland,J.L., Rumelhart,D.E. (Ed.) *Parallel Distributed Processing Vol 1.* MIT Press.

12. Schubert L.K. 1986. Are there preference tradeoffs in Attachment decisions? *Proceedings of AAAI.*

13. Sparck-Jones K., Van Rijsbergen C.J. 1976. Progress in Documentation. *Journal of Documentation, vol 32, no 1*

14. St. John M.F., McClelland J.L. 1988. Applying contextual constraints in sentence comprehension. *Proc. of the Tenth Annual Conference of the Cognitive Science Society.*

15. Tait J.I. 1983. Semantic parsing and syntactic constraints (Mark IV). In: Sparck Jones K., Wilks Y.: *Automated natural language processing.* Ellis Horwood Limited.

16. Waltz D.L., Pollack J.B. 1985. Massively Parallel Parsing: A strongly interactive model of natural language interpretation. *Cognitive Science 9*

17. Wermter S., Lehnert W.G. 1989. Noun phrase analysis with connectionist networks. In: Sharkey N., Reilly R.: *Connectionist Approaches to Language Processing,* (in press).

18. Wilks Y., Huang X., Fass D. 1985. Syntax, Preference and Right Attachment. *Proceedings of the IJCAJ.*