



**LACUS  
FORUM  
XXXII**

*Networks*



Dartmouth

## YOUR RIGHTS

This electronic copy is provided free of charge with no implied warranty. It is made available to you under the terms of the Creative Commons Attribution-NonCommercial license version 3.0

(<http://creativecommons.org/licenses/by-nc/3.0/>)

Under this license you are free:

- **to Share** — to copy, distribute and transmit the work
- **to Remix** — to adapt the work

Under the following conditions:

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial** — You may not use this work for commercial purposes.

With the understanding that:

- **Waiver** — Any of the above conditions can be waived if you get permission from the copyright holder.
- **Other Rights** — In no way are any of the following rights affected by the license:
  - Your fair dealing or fair use rights;
  - The author's moral rights;
  - Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

**Notice:** For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to the web page cited above.

For inquiries concerning commercial use of this work, please visit

<http://www.lacus.org/volumes/republication>

---

**Cover:** The front cover of this document is licensed under the Creative Commons Attribution-No Derivative Works 3.0 license (<http://creativecommons.org/licenses/by-nd/3.0/>) and may not be altered in any fashion. The LACUS “lakes” logo and Dartmouth University logo on the cover are trademarks of LACUS and Dartmouth University respectively. The Dartmouth University logo is used here with permission from the trademark holder. No license for use of these trademarks outside of redistribution of this exact file is granted. These trademarks may not be included in any adaptation of this work.

# AUTOMATING THE IMPORTATION OF LEXICAL INFORMATION INTO A RELATIONAL NETWORK

IAN C. CHOW  
*City University of Hong Kong*

THIS PAPER DISCUSSES the implementation of a lexical knowledge base on the basis of a conceptualization of the problem domain of linguistic information in terms of computational Relational Network Notation as proposed in Webster (2005). In particular, the focus here is on importing linguistic lemmas into the knowledge base as instances from WordNet 2.0.

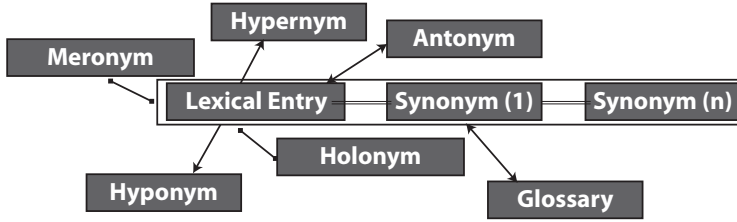
Relational Network Notation (RNN) 'offers a simple yet powerful means for representing lexicogrammatical, semantic and sememic information' (Webster 2005). RNN is driven by Relational Network Theory (Lamb 2004) and incorporates developments in the theory which have been shown not only to describe but also to explain linguistic phenomena in a neurologically plausible manner.

RNN is used to represent lexical information in a knowledge base comprised of instances of lexical entries drawn from WordNet. WordNet is a lexical system whose design is inspired by psycholinguistic theories of human lexical memory. English words (lemmas) are organized into synonym sets (synsets), each representing one concept. There are different relations linking the synsets. This organization of lexical information provides a clear specification of lexical relations which makes it ideally suited to representation in RNN.

Frame Logic (FLogic), a database modeling language, is used to implement WordNet. FLogic allows a systematic concept hierarchy for the structural aspects of domain knowledge with various knowledge base modeling features such as object identity, inheritance, methods, axioms, etc. One can also apply axioms to draw inferences needed to compute new knowledge and relations, thus enlarging the database. FLogic WordNet (FWN) is constructed on the basis of the prolog implementation of WordNet2.0 (<http://www.cogsci.princeton.edu/2.0/>), consisting of over 200,000 lemmas with indications of their part-of-speech and lexical relations and other properties.

FWN, then, serves as a source for the automatic mapping of lemmas into RNN. The mapping involves the following steps: 1) unification of identical concepts, i.e. classes referring to the same concept should share the same object type, and 2) application of mapping axioms to populate the relational network with instances from FWN.

1. RELATIONAL NETWORK THEORY. Relational Network Theory (RNT) suggests a neurologically plausible organization of information applicable to both linguistic and non-linguistic information. RNT explains the access of linguistic information throughout multiple strata, from a phonemic stratum to a conceptual stratum. The linguistic information system is represented as a 'network of interrelationships' (Lamb 1999: *passim*). A



**Figure 1.** Framework of WordNet.

stimulus, or a sequence of stimuli (e.g. a string of phonemes), triggers a node in the network, activating corresponding pathways and the intersecting nodes, and consequently activating the lexeme and the corresponding sememe. Stimulation of a conceptual node can activate other conceptual nodes and connectivity is restricted in neither a bottom-up nor a top-down direction; thus lexemes can be activated by stimulation from both conceptual and phonemic nodes. Information is accessed in a bi-directional and stratum-independent manner.

2. RELATIONAL NETWORK NOTATION. The computational Relational Network Notation offers a simple solution for conceptualizing linguistic interrelationship based on Relational Network. Exploring a domain involving phonemic, lexemic, and sememic strata, RNN only requires the declaration of one class of objects—(node)—and one type of parameterized relation between nodes—( to@ ( threshold,identifier,{sequence} ) ). As I implement RNN with FLogic, a small knowledge base can be enlarged under appropriate rules with the application of axioms.

In fact, this simple network schema is sufficient to represent both linguistic and non-linguistic information without reference to objects. In RNT, linguistic information is realized through spreading activation along inter-connecting paths. Similarly, in RNN, details of linguistic information, such as phoneme, lexeme, and sememe, function as parameters (**identifiers**) in the relations between nodes.

3. WORDNET. WordNet is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. It was developed by the Cognitive Science Laboratory at Princeton University in 1985. English nouns, verbs, adjectives, and adverbs are organized into synonym sets (synsets), each representing one underlying lexical concept, a sememe. Between the synsets, there are different relations, such as antonym, hyponym, meronym, pertainym, entailment, etc. WordNet not only serves as a word-sense reference system but also as an English dictionary referring the usage of the language and the consensus of the interrelationship of English words according to its psycholinguistic basis. In WordNet, a lexical entry is referred to as a **lemma**, and its store of lexical concepts is referred to as a **glossary**. **Figure 1** depicts a small portion of a WordNet synset with its general connections.

	Synsets	Glossary
1 (Noun)	father, male parent, begetter	a male parent (also used as a term of address to your father) ...
2 (Noun)	forefather, father, sire	the founder of a family; ...
3 (Noun)	Father, Padre	'Father' is a term of address for priests in some churches ...
4 (Noun)	don, father	the head of an organized crime family
5 (Verb)	beget, engender, father, mother, sire, generate, bring forth...	make children; 'Abraham begot Isaac'; 'Men often father children but don't recognize them'

**Table 1.** Synsets of father from WordNet 2.0.

One method of instantiating RNN with lexical entries in WordNet is to convert the WordNet database into the formalism of RNN. FLogic WordNet is constructed for the express purpose of automatically importing lexical data from WordNet into RNN representation. Axioms are then employed to draw inferences and thus facilitate the mapping of WordNet data into RNN.

4. POPULATING LEXICAL DATA INTO RNN. In WordNet, a group of lemmas is included in a synset sharing an entry in the glossary of meaning. A synset is represented in RNN as a group of nodes. Each of them is associated with a lexeme, and these lexemic nodes are each connected to a sememic node, which is associated with a sememe. See **Table 1** for several synset examples with the word *father* from WordNet and **Figure 2** (overleaf) for these synsets in RNN representation.

Using the above mapping model, over 200,000 lemmas have been imported into RNN from WordNet.

Before enriching the RNN by importing lexical relations such as hypernyms or meronyms in WordNet, it is necessary to declare the appropriate schema governing the connections in the network. Although these connections are called **lexical relations**, the connection path for them in terms of RNT should run between sememic nodes rather than lexemic nodes. That is, access to such information is achieved through activation from sememe to sememe rather than from lexeme to lexeme.

**Figure 3** (overleaf) shows the relational network between synsets with a hypernym relation in RNN representation. In **Figure 3**, a synset is represented as a group of nodes taking the same node as their sememic head-node. The single pathway connecting the sememic nodes or lexemic node denotes its associated sememe or lexeme respectively. Focusing on the sememic stratum, a sememic node connects to other sememic nodes through the relation of hyponym or hypernym. Such relations drawn from WordNet are therefore re-modeled into layers of connection between sememic nodes in RNN. It should be noted that each sememic node is also a representation of a WordNet synset. Activation of a sememe can be triggered via either a sememe or a lexeme.

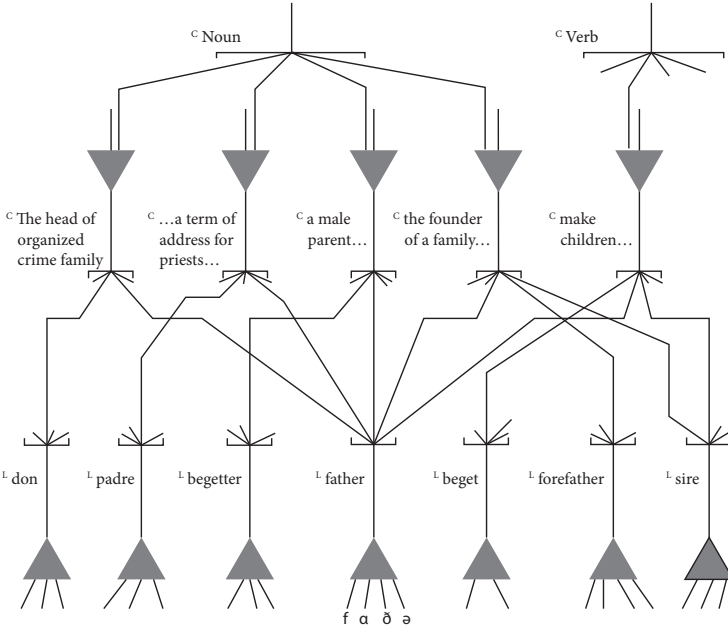


Figure 2. 'father' from WordNet in RNN representation.

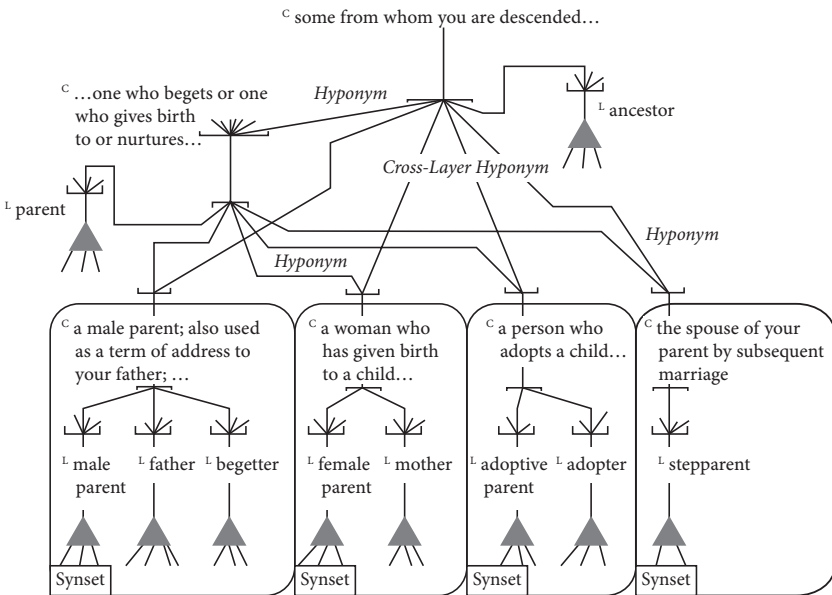


Figure 3. Hyponym relation in RNN representation.

5. MAPPING PROCESS. In WordNet, each synset carries an identity number; in the following FLogic source, a concept synsetId refers to a synset: it has offset, hasword, gloss, etc. as attributes. The following is the partial WordNet database for *father* in FLogic:

```
109434163:synsetId[hasword->>W1_109434163; hasword->>W2_109434163;
  offset->"o09434163"; pos->NOUN; ssType->NOUN;
  gloss->"a male parent (also used as a term of address to
  your father); \"his father was born in Atlanta\""].
W1_109434163:word[lemma->"father"; pos->NOUN].
W2_109434163:word[lemma->"begetter"; pos->NOUN].
FORALL S,Wd,L ( S:synsetId[haslemma->>L] ) <-
  ( S[hasword->>Wd] and Wd[lemma->L] ).
```

Four axioms were used to import the WordNet data into a RNN relational network:

- (1) FORALL X,Y Y:lexeme <- X:synsetId[haslemma->>Y].
- (2) FORALL X,Y Y:sememe <- X:synsetId[gloss->Y].

Axiom (1) creates lexeme from lemma and axiom (2) creates sememe from gloss. The next step is declaring the existence of **nodes** and their **connections**. The process uses the attributes lemma and offset of synsetId to name the nodes; the lexeme and sememe created are applied as identifiers in the relation **...to@[ (threshold, identifier)->>...]**.

- (3) FORALL A,B,S,L,G,O **A:node**[to@(1,G)->>(B:node)] <-  
 S:synsetId[offset->>O; gloss->G; haslemma->>L] AND G:sememe  
 AND concat("node\_",O,A) AND concat("node\_",L,B).
- (4) FORALL A,B,L,G **B:node**[to@(1,(L:lexeme))->>A] <-  
 A:node[to@(1,G)->>B] AND G:sememe AND concat("node\_",L,B).

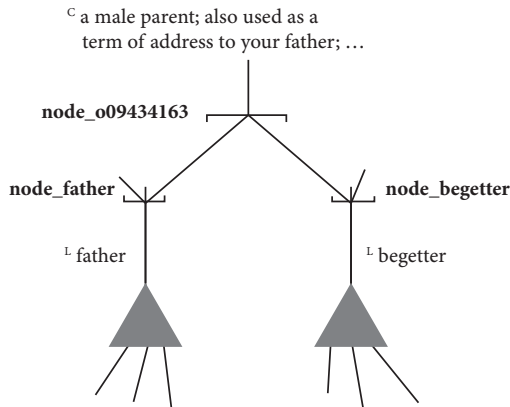
Axiom (3) creates nodes (A) named with offset and nodes (B) named with lemma. Nodes (A) connect to nodes (B) via the identifier sememe.

Axiom (4) refers to the opposite direction of connection. i.e. that nodes (B) named with lemma connect to nodes (A) named with offset via the identifier lexeme.

With regard to the partial WordNet database for *father*, the axioms import to the RNN with

```
2 lexemes : 'father' and 'begetter';
1 sememe  : 'a male parent...';
3 nodes   : node_o09434163, node_father and node_begetter.
```

These nodes are connected as in **Figure 4** (overleaf) and the following FLogic statements:



**Figure 4.** RNN with lexical data import from WordNet—'father'.

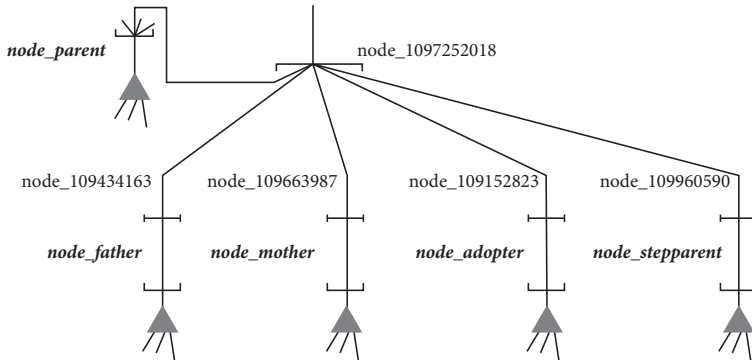
```
//FLogic statement of Figure 4
node_o9434163 [ to@(1, "a male parent...") ->> node_father ;
               to@(1, "a male parent...") ->> node_begetter ].
node_father[to@(1, "father") ->>node_o09434163].
node_begetter[to@(1, "begetter") ->>node_o09434163].
```

We can declare more axioms to further conceptualize the building of nections, i.e. the connections between nodes, and at the same time import more data. The following axiom is applied for the importation of hypernym relations from WordNet:

- (5)   FORALL A, B, P, Q, Ga, Gb, Sa, Sb A:node [to@(1, Hyper) ->>B] <-  
           Sa [gloss->Ga; hyper->>Sb] AND Sb [gloss->Gb]  
           AND A:node [to@(1, Ga) ->>P] AND B:node [to@(1, Gb) ->>Q].

Axiom (5) infers new connections by the **to@** relation with a new identifier, **Hyper**. **Hyper** labels the hypernym relations in WordNet: **Figure 5** shows the RNN conceptualization according to the WordNet instances in the following FLogic network:

```
//synset instances and hypernym relation from WordNet database
109663987:synsetId[haslemma->"mother"].
109152823:synsetId[haslemma->"adopter"].
109434163:synsetId[haslemma->"father"].
109960590:synsetId[haslemma->"stepparent"].
109725018:synsetId[haslemma->"parent"].
109663987[hyper->>109725018].   109434163[hyper->>109725018].
109152823[hyper->>109725018].   109960590[hyper->>109725018].
```



**Figure 5.** RNN with lexical data import from WordNet – hypernym of parent.

```
//FLogic statements for Figure 5
node_109434163 [to@(1,Hyper) ->>node_1097252018] .
node_109663987 [to@(1,Hyper) ->>node_1097252018] .
node_109152823 [to@(1,Hyper) ->>node_1097252018] .
node_109960590 [to@(1,Hyper) ->>node_1097252018] .
```

6. CONCLUSION. Lexemes and sememes do not exist as objects in the human brain. What exist in human brain are millions of interconnected neurons, represented here as interconnected nodes in a Relational Network. ‘The cognitive system does not have nor does it need places to store symbols like those of analytical linguistics... Since its information is in the connectivity of the network it requires no storage space other than the network itself’ (Lamb 1999:120).

Lexemes and sememes are identifying parameters rather than objects; they are comprehended when corresponding nodes and pathways are activated. At the same time, these parameters govern the connections, act as a function in the *to@* method in the computation of node activation in the RNN knowledge base. An activation starts up the connection, triggering both the parameters and the nodes. The totality of activated parameters in sum accounts for the realization of information.

The declaration of axioms creates the connection of nodes which represents the building of nections in Relational Network Theory. The inference power of axioms simulates the increase of connection strength in the generation of established connections from latent connections.

RNN offers a logical schema for the representation of a complex lexicographical database which allows the modeling of various types of linguistic information. With Lamb’s Relational Network as a support to the schema of the model, a computerized ontology of phonemic, lexemic, and sememic information is created. To understand and investigate topics in neurocognitive linguistics, it should be useful to have a linguistic information model with lexicographical information imported. Upon the development of knowledge

base modeling, reuse, importation, and mapping of ontology, it will be a powerful tool for information organization.

## REFERENCES

- LAMB, SYDNEY M. 1999. *Pathways of the brain: The neurocognitive basis of language*. Amsterdam: John Benjamins.
- . 2004. *Language and reality* (selected papers, ed. by Jonathan J. Webster). London: Continuum Books.
- . *et al.* 2000. Langbrain language and brain: Neurocognitive linguistics. <http://www.ruf.rice.edu/~lngbrain/> (accessed 1 August 2005).
- WEBSTER, JONATHAN J. 2005. Ontology engineering for linguistic information: Implementing relational networks. *LACUS forum* 31:337–45.
- WORDNET. 2003. <http://www.cogsci.princeton.edu/~wn/> (accessed on 1 August 2005).

