WordNets: Principles and Applications
Summer School
in
Computational Linguistics and Represented Knowledge
Sozopol, Bulgaria
August 25 - September 8, 2000

Kiril Iv. Simov
Atanas K. Kiryakov
Linguistic Modelling Laboratory
Bulgarian Academy of Sciences
Acad. G. Bonchev Str. 25A, 1113 Sofia, Bulgaria
kivs@bgcict.acad.bg
naso@sirma.bg400.bg

Slides

September 1, 2000

Lexical Semantics and Lexical Relations

A Contextual Approach to Lexical Semantics

Semantic properties of a lexical item are fully reflected in actual and potential contexts.

Only linguistic contexts are consider because of:

1. The relation between a lexical item and extra-linguistic context is often crucially mediated by the purely linguistic contexts;
2. Any aspect of an extra-linguistic context can be mirrored linguistically;
3. Linguistic context is more easily controlled and manipulated.

Lexical semantics is principally concerned with words containing open set elements.

Every aspect of the meaning of a word is reflected in a characteristic pattern of semantic normality (and abnormality) in grammatically appropriate contexts.

Principal Varieties of Semantic Anomaly

A. **Pleonasm**

Kick it with one of your feet.

A female mother.

He was murdered illegally.

B. **Dissonance**

Arthur is a married bachelor.

Let us drink time.

C. **Improbability**

The kitten drank a bottle of claret.

The throne was occupied by a pipe-smoking alligator.

Arthur runs faster than the wind.

D. **Zeugma**

They took the door off its hinges and went through it.

Arthur and his driving licence expired last Thursday.

He was wearing a scarf, a pair of boots, and a look of considerable embarrassment.

Word Meaning

Two definitions:

1. The meaning of a word is fully reflected in and constituted by its **contextual relations**.
2. The meaning of a word is constituted by its **semantic traits**.

Contextual Relations

The full set of normality relations which a lexical item contracts with all conceivable contexts will be referred to as its **contextual relations**.

The contextual relations of a word can be considered as determined by a pattern of affinities and disaffinities with all the other words in the language with which this word is capable of contrasting semantic relations in grammatical contexts.

Affinities are of two kinds, **syntagmatic** and **paradigmatic**.

Syntagmatic affinity

is established by a capacity for normal association in an utterance;

and it always presupposes a particular grammatical relationship.

For instance, there is a syntagmatic affinity between *dog* and *barked*, since *The dog barked* is normal.

Paradigmatic affinity:

is defined between grammatically identical words;

is established by a capacity for normal association in an utterance;

Paradigmatically, a semantic affinity between two grammatically identical words is the greater the more congruent their patterns of syntagmatic normality.

For instance, *dog* and *cat* share far more normal and abnormal contexts than, say, *dog* and *lamp-post*:

Arthur fed the dog/cat/?lamp-post.

The dog/cat/?lamp-post ran away.

The ?dog/?cat/lamp-post got bent in the crash.

We painted the ?dog/?cat/lamp-post red.

Semantic Traits

A model of the meaning of a word in which the meaning of the word is viewed as being made up, at least in part, of the meanings of other words.

A particular word-meaning which participates in this way in the meaning of another word will be termed a **semantic trait** of the second word.

Statuses (degrees of necessity) of semantic traits: **criterial**, **expected**, **possible**, **unexpected** and **excluded**.

"animal" is a criterial trait of *dog*

"fish" is an excluded trait of *dog*

"can bark" is an expected trait of *dog*

"can stay on two legs" is an possible trait of *dog*

"can sing" is an unexpected trait of *dog*

The two definitions of meaning of words are not incompatible.

They highlight different aspects of meaning.

The affinity between *dog* and *cat* reveals itself in the number of equi-status or near-equi-status traits they have in common; and the differences between *dog* and *cat* appear more sharply when the affinity patterns are articulated in greater detail by means of diagnostic frames.

Semantic traits whose absence is regarded as a defect will be called **canonical traits**.

Consequences of the Contextual Approach

- Arbitrariness of the difference between the meaning of a word and 'encyclopaedic' facts concerning its extra-linguistic referents.
- There is no motivation for isolating 'pragmatic meaning' as a separate domain of lexical meaning.
- Most importantly, it would seem that we have no grounds for believing that the meaning of a word, when viewed in this fashion, is finitely describable.
- This conception of word-meaning has the advantage of being intuitively plausible.

The Syntagmatic Delimitation of Lexical Units

An ordinary dictionary characterises a lexical item in three distinct, though intimately inter-connected, ways: its form (graphic and phonological); second, its grammatical function; and, third, its meaning.

The basic syntagmatic lexical units of a sentence is defined as the smallest parts which satisfy the following criteria:

(i) a lexical unit must be at least one semantic constituent;

(ii) a lexical unit must be at least one word.

A preliminary illustration of the points:

- the prefix *dis-* of *disobey* is not a lexical unit because, although it is a semantic constituent, it is smaller than a word.
- The *pulled* of *Arthur pulled a fast one* is not a lexical unit because, although it is a word, it is not a semantic constituent.

Any constituent part of a sentence that bears a meaning which combines with the meaning of the other constituents to give the overall meaning of the sentence will be termed a **semantic constituent**.

A semantic constituent which cannot be segmented into more elementary semantic constituents will be termed a **minimal semantic constituent**.

Lexical Units and Lexemes

Lexical units are those form-meaning complexes with (relatively) stable and discrete semantic properties which

stand in meaning relations such as antonymy and hyponymy, and which

interact syntagmatically with contexts in various ways.

A particular lexical unit expresses its semantic identity through its contextual relations, but its essence cannot be exhaustively characterised in terms of any determinate set of such relations.

The meaning aspect of a lexical unit will be termed a **sense**.

Lexemes, are the items listed in the lexicon, or 'ideal dictionary', of a language.

Lexical units and *lexemes* have different functions:

Senses need to represent unitary 'quanta' of meaning, but they do not need to be finite in number.

A lexeme, on the other hand, may well be associated with indefinitely many senses, but the set of lexemes must be finite enumerable.

It seems that there is a high degree of creativity in the lexicon which we must take account of.

Selection and Modulation of Senses

One of the basic problems of lexical semantics is the apparent multiplicity of semantic uses of a single word form (without grammatical difference).

The meaning of any word form is in some sense different in every distinct context in which it occurs.

There are two distinct types of variation in the semantic contribution that a word form makes to different sentences.

The first types of variation involves the selection, by the context, of different units of sense.

The other type is a matter of contextual modification of a single sense.

1. Sue is visiting her cousin.
2. We finally reached the bank.

Here the word form *cousin* is **general** with respect to the distinction "male cousin"/"female cousin";

bank is said to be **ambiguous** with respect to the sense distinction "financial institution"/"side of river".

Selection and Modulation of Senses

Context modifies the semantic contribution of a word in two ways:

1. First, a single sense can be modified in an unlimited number of ways by different contexts, each context emphasising certain semantic traits, and obscuring or suppressing others. This effect of a context on an included lexical unit will be termed **modulation**. Continuous.
2. The second manner of semantic variation concerns the activation by different contexts of different senses associated with ambiguous word forms. This will be termed **contextual selection** (of senses). Discrete.

Modulation - Promotion and Demotion

Promotion and demotion represent changes in the status of semantic traits along the dimension of necessity:

3. A nurse attended us.

   A pregnant nurse attended us.

4. Arthur poured the butter into a dish.

Modulation - Highlighting and Background

Some part of an object may be thrown in relief relative to other parts: *The car needs servicing* and *The car needs washing*.

What is highlighted or backgrounded is an attribute, or range of attributes, of entity referred to: *We can't afford that car*, *The car crushed Arthur's foot*.

Selection

A lexical form may well be associated with an unlimited number of possible senses, but these are not all of equal status -**potential** and **established** senses.

The difference between established senses and potential senses is not merely one of frequency of use.

Passive selection

The context acts merely as a kind of filter and the selection is from among the pre-established senses.

Productive selection

The selected sense is not established.

The context acts rather as a stimulus for a productive process, namely, the activation of a set of rules or principles which 'generate' the sense in question.

Lexemes

A **lexeme** is a family of lexical units.

An ideal dictionary would be expected to define all the established senses within each lexeme.

Lexical form is a set word forms which differ only in respect to inflections.

Lexical unit is used to refer to a **lexical form** together with a single distinguished sense.

A lexeme which has a number of senses is **polysemous** (or as manifesting the property of **polysemy**).

A lexical form which realises lexical units belonging to more than one lexeme as **homonymous**.

Lexical Relations

Semantic relation which hold between lexical units.

Sense relations are of two fundamental types: paradigmatic and syntagmatic.

Paradigmatic relations reflect the way infinitely and continuously varied experienced reality is apprehended and controlled through being categorised, subcategorised and graded along specific dimensions of variation. They represent systems of choices a speaker faces when encoding his message.

Syntagmatic aspects of lexical meaning, on the other hand, serve discourse cohesion, adding necessary informational redundancy to the message, at the same time controlling the semantic contribution of individual utterance elements through disambiguation, for instance, or by signalling alternative – e.g. figurative – strategies of interpretation.

Congruence

The four basic relations between classes furnish a model not only for establishing a fundamental group of sense relations, but also for defining a set of systematic variants applicable to virtually all other paradigmatic sense relations. The basic lexical relations will be referred to collectively as **congruence relations**, and the variants as **congruence variants**.

Congruence

The relations between classes are as follows:

1. identity: class A and class B have the same members

2. inclusion: class B is wholly included in class A

3. overlap: class A and class B have members in common but each has members not found in the other

4. disjunction: class A and class B have no members in common

Cognitive synonymy

The lexical relation which parallels identity in the membership of two classes is synonymy.

There are different degrees of synonymity; the relation defined in terms of truth-conditional relations will be distinguished as **cognitive synonymy**:

X is a cognitive synonym of Y if

(i) X and Y are syntactically identical, and

(ii) any grammatical declarative sentence S containing X has equivalent truth-conditions to another sentence S', which is identical to S except that X is replaced by Y.

Hyponymy

The lexical relation corresponding to inclusion of one class in another is **hyponymy**.

X will be said to be a **hyponym** of Y (and, by the same token, Y a **superordinate** of X) is A *is* $f(X)$ entails but is not entailed by A *is* $f(Y)$ (where *f(X)*):

This is a DOG unilaterally entails *This is an ANIMAL*

Compatibility

The lexical relation which corresponds to overlap between classes will be given the name **compatibility**.

The defining characteristics of compatibles are two:

There are no systematic entailments between sentences differing only in respect of compatibles in parallel syntactic positions.

A pair of compatibles must have a common superordinate.

Compatibles, therefore, have some semantic traits in common, but differ in respect of traits that do not clash.

Strict compatibility and **contingent compatibility**.

X and Y are strict compatibles if they have at least one shared hyponym or hyponymous expression which is independently characterisable.

Incompatibility

The sense relation which is analogous to the relation between classes with no members in common is **incompatibility**.

Two items X and Y are incompatibilities if a sentence of the form A *is* $f(X)$ can be found which entails a parallel sentence of the form A *is not* $f(Y)$:

It's a cat entails *It's not a dog*

Congruence variants

Suppose some lexical unit X stands in a lexical relation R to another lexical unit Y. (R must be some relation other than one of the primary congruence relations.) If every occurrence of X stands in the relation R to Y, and every occurence of Y stands in the relation R (or its converse, if R is asymmetric) to X, then we shall say that X is a **congruent** R to Y.

If every occurrence of X stands in the relation R to Y, but there are occurrences of Y which do not stand in the relation R to X, then we shall say that X is a **hypo-R** of Y, and Y is a **super-R** of X.

If some, but not all, occurrences of X stand in the relation R to Y, and some, but not all, occurrences of Y stand in the relation R to X, then we shall say that X and Y are **semi-Rs**.

The following are examples of the three congruence variants: *finger* is a congruent meronym of *hand*; *doctor* is a hypo-converse of *patient*, and *patient* a super-converse of *doctor*, *index* is a semi-meronym of *book*.

Syntagmatic relations

Every word in a sentence interacts semantically with every other word.

Contextual relevance:

A: I need to cash a cheque.

B: You'd better make straight for the bank, otherwise you'll be too late.

The grammatical structure of the sentence is a series of nested constructions forming a hierarchy.

Semantic head is an element of construction which interacts directly with an element or elements outside the construction.

Bi-directional and directional properties of semantic co-occurrence restrictions.

Head-modifier construction - the head alone can play a grammatical role in the sentence identical to that of the whole construction.

We drank red wine. We drank wine.

Head-complement construction is typically not reducible syntactically to the head.

Arthur stroked the cat.

Syntagmatic relations

Two sorts of directional property.

I. **Selector** and **Selectee**

In a head-modifier construction, the modifier is the selector.

In a head-complement construction, the head which is the selector.

II. Relationship between the head and any dependent item or items.

A dependent item is expected to bring to a construction semantic traits not already prefigured in the head.

The two sorts of directional property described above work in opposite directions in head-modifier constructions, but in parallel in head-complement construction.

A set of syntagmatic relation can be based on the results of putting grammatically appropriate lexical units together in a construction.

If the combination is normal, we shall say that the lexical units involved are **philonyms**.

If the combination is pleonastic, we shall speak of head and **tautonym**.

If dissonance results, the lexical units will be called **xenonyms**.

Collocational and Selectional Restriction

Syntagmatic and paradigmatic relations of sense can be used to define degrees of dissonance.

Three such grades are **inappropriateness, paradox** and **incongruity**.

Inappropriateness is diagnosed by the fact that there exists a cognitive synonym of the selector for which the selectee is a philonym.

Those presuppositions of a selector, which, if not satisfied by the selectee, give rise to inappropriateness, will be termed the **collocational restrictions** of the selector.

A paradox appears when (a) there is no possibility of resolving dissonance by synonymous substitution, but (b) there exists a (not too remote) superordinate of either xenonym which is a philonym of the other.

Incongruity is characterised by the fact that there is no superordinate of either xenonym which can restore normality.

Those presuppositions of a selector whose non-satisfaction leads to paradox or incongruity will be called its **selectional restriction**.

WordNet - History

The initial idea was to provide an aid to use in searching dictionaries conceptually, rather than merely alphabetically—it was to be used in close conjunction with an online dictionary of the conventional type.

WordNet was seen as a program that would allow users to explore an on-line dictionary on the basis of semantic, rather than alphabetic, similarities. Browsing.

One of the project's original presuppositions was the *separability hypothesis:* that the lexical component of language can be isolated and studied in its own right.

Another presupposition was the *patterning hypothesis:* that people could not master and have readily available all the lexical knowledge needed to use a natural language unless they could take advantage of systematic patterns and relations among the meaning that words can be used to express.

A third presupposition was the *comprehensiveness hypothesis:* that computational linguistics, if it were ever to process natural languages as people do, would need to have available a store of lexical knowledge as extensive as people have.

WordNet version 1.0 was released in June 1991.

WordNet - Basic Relations

WordNet assumes the relational model of lexical semantics.

WordNet represents the lexical semantics by means of semantic network based on several lexical relations.

WordNet divides the lexicon into four separate semantic nets, one for each open word class.

WordNet encodes only paradigmatic lexical relations.

The main lexical relation in WordNet is **synonymy**.

Synonymy relation organizes the vocabulary of WordNet as a set of units represented sets of synonyms called **synsets**.

Each synset is defined by a number of words (or phrases) which share a sense. Additionally an informal gloss is given for most of the synsets in WordNet.

A synset represents a concepts.

The hierarchical structure of WordNet is defined over the set of synsets.

The hierarchical structure of the different parts of WordNet is given by different lexical relations for the four part of speeches.

Some synsets in WordNet are represented by short phrases, such as *bad person*, that are not paraphrasable by a single word.

These phrases reflect lexical gaps, which sometimes are lexicalized in other languages, and they reveal conceptual structures as distinct from lexical structures.

Nouns

The main structural relations in the noun semantic network is **hyponymy**.

The nouns in WordNet form a lexical inheritance system. The semantic traits of a noun synset are part of the set of semantic traits of its hyponyms. This inheritance is only implicit in WordNet and in principle will depends on the statuses of the traits.

WordNet divides the nouns into several hierarchies.

The top synset of each such hierarchy is called **unique beginner**.

These multiple hierarchies correspond to relatively distinct semantic field, each with its own vocabulary.

The design of the set of unique beginners have to meet one important criterion: they should provide a place for every noun.

WordNet has 25 unique beginners:

{act, activity}	{animal, fauna}	{artifact}	{attribute}
{body}	{cognition, knowledge}	{communication}	{event, happening}
{feeling, emotion}	{food}	{group, grouping}	{location}
{motivation, motive}	{natural object}	{natural phenomenon}	{person, human being}
{plant, flora}	{possession}	{process}	{quantity, amount}
{relation}	{shape}	{state}	{substance}
{time}			

The hierarchies below the unique beginners vary in size and are not mutually exclusive, but on the whole they cover distinct conceptual and lexical domains.

Tennis Problem

The specialized vocabulary of tennis is distributed in several noun hierarchies in WordNet:

Tennis players are under {person} unique beginner.

Tennis equipment is under {artifact} unique beginner.

Tennis court is under {location} unique beginner.

The strokes are under {act} unique beginner.

Co-occurrence relations are not encoded in WordNet because it is not clear how to encode contexts in synsets.

Parts and Meronymy

Meronymy relation is asymmetric and (with reservations) transitive. Meronymy and hyponymy relations are intertwined. Meronyms are distinguished features that can be inherited.

Three kinds of meronymy are coded in WordNet:

'is a component part of'

'is a member of'

'is the stuff that X is made from'

Antonymy

Antonymy relation is not a fundamental organizing relation between nouns.

Some of the semantic oppositions are encoded in WordNet, but they are not inherited along the hyponymy relation.

Adjectives

WordNet divides adjectives loosely into two categories:

Descriptive adjectives - big, beautiful, possible,...

Relational adjectives - electrical,...

Descriptive adjectives typically ascribe to nouns a value of an attribute.

The package is heavy presupposes $\texttt{WEIGHT}(package) = heavy$.

The basic semantic relation among descriptive adjectives is antonymy.

Direct and indirect antonyms:

heavy is a direct antonyms of *light*

ponderous is an indirect antonyms of *light*.

Gradation - *Astronomical, huge, large, small, tiny, infinitesimal*

Markedness - attribute orientation. *deep/shallow*.

Gradation and markednes are not implemented in WordNet.

Verbs

The verb lexicon of WordNet is divided in several hierarchies on the base of semantic domains.

A general division on verbs denoting actions and events and verbs denoting states is made.

The verbs denoting actions and events are subdivided into 14 more specific semantic domains as: verbs of motion, perception, contact, communication, ...

Verbs *to be* and *to do* are excluded as unique beginners because they are very polysemous, very abstract and the semantic relations between them and some of the other verbs are not observed psycholinguistically.

Some of semantic domains can be represented only by several independent trees.

For example, verbs of possession go upward to three concepts: {*give, transfer*}, {*take, receive*}, and {*have, hold*}.

Such independent hierarchies form a coherent semantic field because most of the verbs in these hierarchies select for the same kind of noun arguments.

For example, the verbs of bodily care and functions.

Verbs

The basic semantic relation for verbs is **lexical entailment**.

Entailment between two verbs V_1 and V_2 holds when the sentence *Someone* V_1 logically entails the sentence *Someone* V_2.

For example, *snore* entails *sleep*.

Lexical entailment is a unilateral relation: if a verb V_1 entails another verb V_2, then it cannot be the case that V_2 entails V_1.

The hyponymy relation over verbs in WordNet distinguishes a 'verb hyponym' from its superordinate by the means of a manner relation. This relation is called **troponymy**.

The troponymy relation relation between two verbs can be expressed by the formula:

To V_1 *is to* V_2 *in some particular manner.*

Subsets of particular kinds of manners tend to cluster within a given semantic field.

For example, among competition verbs many troponyms are conflations of the basic verb *fight* with nouns denoting the occasion for, or form of, the fight: *battle, war, tourney, joust*, an so on.

Verbs

Troponymy is a particular kind of entailment. Every troponym V_1 of a more general verb V_2 also entails V_2. Also the activity referred to by a troponym and its more general superordinate are always temporally coextensive.

The verbs that are related only by entailment and proper temporal inclusion cannot also be related by troponymy.

Verb hierarchies constructed by means of the troponymy relation tend to have a more shallow structure than nouns.

Almost every verb hierarchy has a distinct level with far more verbs than the other levels in the same hierarchy.

As one descends in a verb hierarchy, the variety of nouns that the verbs on a given level can take as potential argument decreases.

Inheritance along the lines of troponymy relation is also defined for some kinds of knowledge.

For example, the troponyms can inherit their argument structure from their superordinate.

Similarly, although troponyms of *speak* may differ widely in the particular manners of speaking, they all share the aspects of meaning associated with *speak* - vibration of the vocal cords, etc.

Verbs

There are several subtypes of **semantic opposition relation** encoded in WordNet. These include:

Converses that are not associated with common superordinate or entailed verb: *give/take, teach/learn*. They occur within the same semantic field.

Stative or change-of-state verbs: *exclude/include, wake/sleep, lengthen/shorten*.

Co-troponyms: *rise/fall*

Common entailed verb: *fail* and *succeed* entail *try*. Backward presupposition.

The **cause relation** picks out two verb concepts, one causative (like *give*), the other what might be called "resultative" (like *have*).

WordNet connects causative, transitive verbs to the corresponding anticausative (inchoative), intransitive sense of the same word: *break*.

The glass door broke. - The storm broke the glass door.

The cause relation also shows up systematically among the motion verbs: *bounce, blow*.

She blew a soap bubble in his face.

The soap bubble blew in his face.

The cause relation is a specific kind of entailment: if V_1 necessary causes V_2, then V_1 also entails V_2.

Lexical Chains

If a text is cohesive and coherent, successive sentences are likely to refer to concepts that were previously mentioned and to other concepts that are related to them.

The words of the text that make such references can be thought of as forming **cohesive chains** in the text.

Each word in the chain is related to its predecessors by a particular **cohesive relation** such as identity of reference.

For example, in (1) the italicized words form a chain with this relation:

(1) The major potential complication of total joint replacement is *infection. It* may occur just in the wound or deep around the prosthesis. *It* may occur during the hospital stay or after the patient goes home ... . *Infections* in the wound area are generally treated with antibiotics.

But the relation need not be identity.

(2) The major potential *complication* of total joint replacement is *infection.*

(3) The evening prior to admission, take a *shower* or *bath*, *scrubbing* yourself well. Rinse off all the *soap.*

A lexical chain is a cohesive chain in which the criterion for inclusion of a word is that it bear a cohesive relation of one kind or another to a word that is already in the chain.

Morris and Hirst (1991; Morris 1988) suggested that the discourse structure of a text may be determined by finding lexical chains in the text.

Two words could be considered to be related if they are "connected" in the thesaurus in one (or more) of five possible ways:

1. Their index entries point to the same thesaurus category or to adjacent categories.
2. The index entry of one contains the other,
3. The index entry of one points to a thesaurus category that contains the other.
4. The index entry of one points to a thesaurus category that in turn contains a pointer to a category pointed to by the index entry of the other.
5. The index entries of each point to thesaurus categories that in turn contain a pointer to the same category.

Relations between Words

Three kinds of relation are defined: extra-strong, strong, and medium-strong.

If a relation is not any of these, it is said to be **weak** and is not used in the creation of lexical chains.

The definitions of these relations use a classification of WordNet synset relations into the directions **upward**, **downward**, and **horizontal**:

Relation	**Direction**
Also see	Horizontal
Antonymy	Horizontal
Attribute	Horizontal
Cause	Down
Entailment	Down
Holonymy	Down
Hypernymy	Up
Hyponymy	Down
Meronymy	Up
Pertinence	Horizontal
Similarity	Horizontal

An **extra-strong** relation holds only between a word and its literal repetition; such relations have the highest **weight** of all relations.

There are three kinds of **strong** relations:

1. The first occurs when there is a synset common to two different words, such as *human* and *person*.
2. The second occurs when there is a horizontal link between synsets associated with two different words, such as *precursor* and *successor*.
3. The third occurs when there is any kind of link at all between a synset associated with each word if one word is a compound word or a phrase that includes the other, such as *school* and *private school*.

A **medium-strong** relation between two words occurs when there is an **allowable path** connecting a synset associated with each word.

For example, *apple* and *carrot*.

Unlike extra-strong and strong relations, medium-strong relations have different weights. The weight of a path is given by

weight =

*C - path length - k * number of changes of direction*

The longer the path and the more changes of direction, the lower the weight.

The rationale for the allowable patterns is as follows: If a multilink path between two synsets is to be indicative of some reasonable semantic proximity, the semantics of each lexical relation must be taken into consideration.

An upward direction corresponds to generalization.

A downward link corresponds to specialization.

Horizontal links are less frequent than upward and downward links; a synset rarely has more than one. But such links are usually highly indicative of meaning.

Two rules have been stated to define which patterns are allowable:

(R1) No other direction may precede an upward link.

(R2) At most one change of direction is allowed.

(R2') It is permitted to use a horizontal link to make a transition from an upward to a downward direction.

Creating and Managing Chains

First, an empty chain is created.

Then, a chain word record is allocated, initialized with the word *economy*, and inserted into the new chain.

Next, to insert *sectors*, another word record is constructed and inserted into the chain.

The kind of relation (extra-strong, strong, or medium-strong) between the new word and its related word (or words) in the chain is also stored in the word record.

Thus, the word order in a chain corresponds only to insertion order, not necessarily to relations between words.

Synsets that are not involved in the current word connection are removed.

Identifying Words and Relations

Only noun chains.

Each word that is found in the noun hierarchy is considered.

This decision is based on the assumption that most words in other grammatical categories that have a nominal from are semantically close to that from.

Compound words and phrases are preferred .

For instance, *private school*, which is listed in the noun index as *private_school*, is more indicative than *private* and *school* taken separately.

The stop-word list contains closed-class words and many vague high-frequency words that tend to weaken chains by having little content (e.g., *one, two, dozen, little, relative, right*).

If a word is potentially chainable:

an extra-strong relation is sought throughout all chains;

if not, strong relations are sought, but the search scope is seven sentences back;

in case of no strong relation, medium-strong relations are sought, but the search scope is three sentences back.

All medium-strong relations are found and one with highest weight is added.

Testing the Lexical Chainer

Testing is difficult.

Problems:

1. limitation in the set of relations in WordNet, or a missing connection;
2. inconsistency in the measure of semantic proximity that is implicit in links in WordNet; and
3. incorrect or incomplete disambiguation.

Examples:

(A) Nasdaq volume has been burgeoning daily, and yesterday hit 146.1 million shares.

first problem

(B) Prices of over-the-counter stocks surged yesterday ...

first problem

(C) The cost means no holiday trips and more *stew* than *steak*, but she is satisfied that her children, now in grades 3 and 4, are being properly taught.

second problem

{stew} IS-A {dish} IS-A {aliment} INCLUDES {meat} INCLUDES {cut, cut_of_meat} INCLUDES {piece,slice} INCLUDES {steak}

(D) {public} IS-A {people} HAS-MEMBER {person} INCLUDES {adult} INCLUDES {professional}

second problem

(E) We suppose a very long *train* traveling along the *rails* with the constant *velocity v* and in the direction indicated ...

third problem

{sequence, succession, sequel, train} - events that are ordered in time.