

Cognitive linguistics approach to Vietnamese noun compounds

DINH Dien

National University of HCM City, VIETNAM

Noun Compound (NC for short) in this paper is a combination of one head noun and its modifier(s) to form a compound which functions as a noun with its own meaning as a whole, such as: hệ thống máy tính (*computer system*), cáp máy in (*printer cable*),.....

NC is a very popular phenomenon in many languages (especially in materials of Science & Technology) and it plays a very important role in understanding natural languages. So this paper will present the Vietnamese Noun Compound in the aspects of word order, syntactic structures, semantic relations and automatic processing models.

We make use of the cognitive linguistics approach with the research on the conceptual hierarchical systems, particularly the well-known conceptual system WordNet and the case grammar of Fillmore.

We have applied the above-mentioned approach to build a semantic processing model for the Vietnamese Noun Compound, in which the implicit semantic relationships between the head noun and modifiers will be explicitly shown.

From the basis of this NC model, we can deduce many possible interpretations of the Vietnamese Noun Compound, which is the initial foundation for processing the Vietnamese language automatically by computers.

1. Introduction

Noun Compound is a phenomenon that has a very simple form (surface structure) but very complicated contents (deep structure) with its ambiguous syntactic and complex semantic relations. In order to understand

the meaning of NC, we need not only linguistic knowledge but also extra-linguistic knowledge (i.e., of the real world, living experience, etc.).

NC is a very popular phenomenon in most languages. Diachronic research by Leonard on English texts of the last two centuries (Lauer 1995, p. 26) shows that the use of NC has been increasing, especially in materials of Science and Technology.

Understanding NC plays a very important role in understanding natural languages, so a sophisticated research on NC is a necessary and meaningful task, especially in computational linguistics. For human beings, understanding NC is very simple and natural, but for computers this task is so complicated that so far it hasn't been solved perfectly.

In studying English NC, many sophisticated works have been carried out with various approaches, among which were the two following typical works: first, MacDonald (1982) made use of the semantic relational model approach of cognitive linguistics based on the conceptual representation NETL and Fillmore's case grammar (MacDonald, 1982). Second, Lauer (1995) studied NC by using the "Statistical Language Learning" model. This paper will apply a part of the first work and combine it with the conceptual system WordNet.

In studying Vietnamese, Nguyen Tai Can (1998) briefly mentioned NC when writing about the structure of the Vietnamese Noun Phrase (p. 236-239). He considered NC as the short-form of Noun Phrase having only a head Noun and its modifier(s). He also showed some semantic and syntactic relations between the head noun and its modifier(s), but did not suggest a model or system of those NC-relations.

Hence, the aim of this paper is to apply the cognitive linguistics approach and Fillmore's case grammar similar to those of MacDonald but basing ourselves on the conceptual system WordNet in order to build the NC model for Vietnamese.

The NC in this paper will be understood as a combination of one head Noun and its modifier(s) to form a compound which functions as a noun with its own meaning as a whole, such as: *hệ thống máy tính* (*computer system*), *cáp máy in* (*printer cable*),....

We won't examine NCs that have such idiomatic meanings as: *kỷ luật sắt* (*iron discipline*), *tóc rể tre* (*bristled hair*), etc...

The word order, syntactic structure, and semantic relations of NC in this paper will be the normal, central, unmarked forms (stylistics means in

literature). These NCs will be restricted to materials of Science & Technology, in which these NCs are used very often.

This paper has the 4 following parts:

- **Noun Phrase (NP) and Noun Compound (NC) in Vietnamese: NC as a special case of NP; various forms of NC.**
- **Aspects of cognitive linguistics: conceptual system WordNet, Vietnamese conceptual tree.**
- **Vietnamese NC Model: in order to recognize automatically the semantic relation between NC components. This is the main part of this paper.**
- **Conclusion, application and development: remarks on NC, its applications and development in the future.**

2. Vietnamese noun phrase and noun compound

In order to study the Vietnamese NC, firstly, we must start from the organization of the Vietnamese NP, and then we determine the boundary of the Vietnamese NC.

2.1 Vietnamese noun phrase

Referring to the research results of Vietnamese linguists and my own study, I would like to suggest the organization of modifiers (relatively stable) in Vietnamese NP as follows:

Table 1. The structure of Vietnamese NP

No		Quantifier	Cardinal	Specifier	Classifier	Head Noun	Modifier	Ordinal	Demonstrative
		-3	-2	-1	-0	0	+1	+2	+3
1	All those first 6 computer systems	Tất cả All	6 6		cái (the)	hệ thống system	vi tính computer	đầu tiên first	đó those
2	Areca garden					Vườn garden	cau areca		
3	Brick yard					Sân yard	gạch brick		
4	Teacher's book					Sách book	giáo viên teacher		
5	Morning class					Lớp học Class	buổi sáng morning		
6	Study table					Bàn table	học learning		
7	Long table					Bàn table	dài long		
8	Front gate					Cổng Gate	trước front		
9	Chapter 5					Chương chapter	5 5		

No		Quantifier	Cardinal	Specifier	Classifier	Head Noun	Modifier	Ordinal	Demonstrative
		-3	-2	-1	-0	0	+1	+2	+3
10	Very good chair					<i>Ghế</i> chair	<i>rất tốt</i> very good		
11	My blue silk shirt			cái (the)		áo shirt	lụa; màu xanh; của tôi silk; blue; my		
12	The pupil who came yesterday			người (the)		học sinh pupil	đến sáng hôm qua (who) came yesterday		
13	The book which you lent me			cuốn (the)		sách book	mà anh cho tôi mượn which you lent me		

After observing the above NP organization, we realize that the Vietnamese NC is considered as a truncation of NP which has only 2 columns: the head noun (column 0) and modifiers (column+1). These modifiers may be a word (examples 1 → 9), a phrase (examples 10, 11, 12) or clause (example 13). For the sake of simplicity, in this paper we only examine the case where the modifier is only one word, but in fact we are also able to extend it to a phrase or clause with the same generality of this model.

2.2 Vietnamese noun compound

From the lexical aspect, a modifier (in cases of a single word) in NC may be a noun (examples 1 → 5), a verb (example 6), adjective (example 7), preposition (example 8), or number (example 9).

From the syntactic aspect, the modifier always plays a modifying role to the head noun.

In word-order aspect, the head noun is always in the first position, before the positions of modifiers.

From the semantic aspect, the modifier is used to name a feature of the thing identified by the head noun (Ex 2: “vườn cau”: areca garden), or to point out something which is related to the thing identified by the head noun, such as the following:

- Relation of possession, e.g.: *sách (của) giáo viên*. (book of teacher)
- Relation of material, e.g.: *sân (bằng) gạch*. (yard made of brick)
- Relation of content, e.g.: *hệ thống (về) vi tính*. (system of computer)
- Relation of time, space: *lớp học (vào) buổi sáng; cổng (ở phía) trước* (class in the morning; gate at the front)
- Relation of attributes (volume, shape,..) : *bàn (thì) dài*. (table is long)
- Relation of reason, purpose: *bàn (để) học*. (table to learn)

In order to recognize these implicit semantic relationships automatically, we need to build a model of the above NCs and from it, we can extract its possible relations.

3. Aspect of cognitive linguistics

Researchers in cognitive linguistics have achieved successes in investigating the cognitive ability of human beings, discovering the understanding mechanism of our brains and applying that mechanism to computational linguistics in order to instruct computers how to process our natural languages automatically. One of those successes is the system of primitive concepts in the real world.

Cognitive linguists assume that although we use different languages with different forms in lexicon and syntax, we have the same content, i.e., we have common primitive concepts. Each word in different languages can be represented by these primitive concepts, and they build systems of such primitive concepts.

One of those systems is WordNet which was built by psychologists in Princeton University (USA) in 1985. Now this system contains about 70,000 concepts which are organized in hierarchical trees and there are relationships between these concepts.

This paper will take advantage of the conceptual tree in WordNet to construct the Vietnamese NC. The WordNet system, however, was built for English words, so we must re-compile it into a Vietnamese conceptual tree with the same original ideas.

The conceptual system here is a knowledge-based system constructed on a hierarchical model consisting of nodes and relations. The model consists of 2 main types: concept and relation.

Concept

A concept is the meaning of a word. These concepts are organized as a hyponym tree where daughter concepts are able to inherit the mother concepts' relationships, properties and functions. These concepts may be notions of the real world, e.g.: *biology, animal, botany, human, living, non-living,...* (Miller, 1996).

As an example of a hyponym: the concept "biology" contains the concept "animal", we symbolize as *animal*→*biology*, similarly: *animal*→*biology*→*living,...*

Relation

There are many kinds of relation between concepts. Each relation initiates a proper sense. These relations are uni-directional. The main relations are as follows: *Has_Attribute* (points to Adjectives), *Has_Function* (points to Verbs), *Has_part* (points to Nouns).

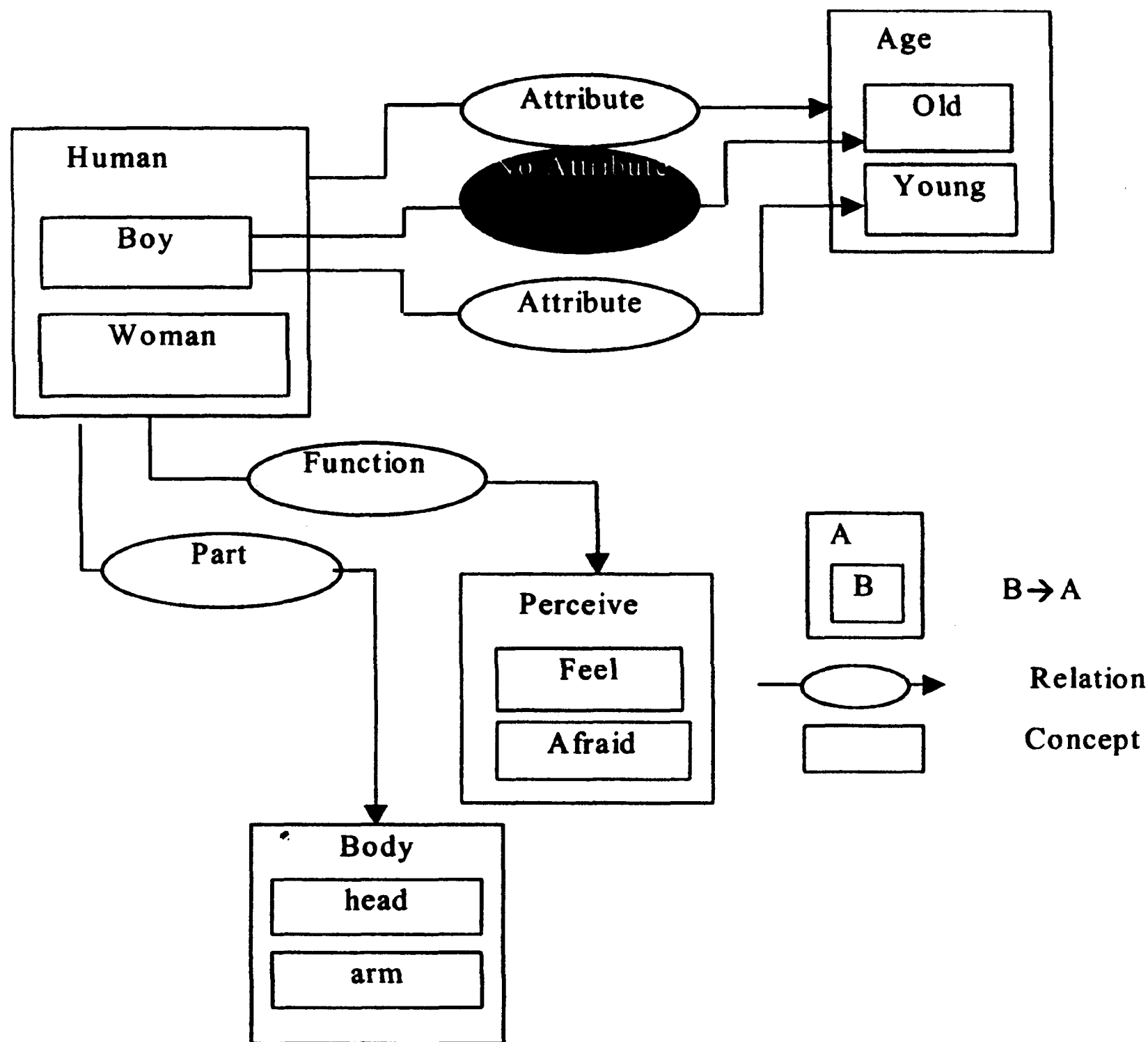


Figure 1. Semantic network of concepts and relations.

Example: Examining the semantic network in figure 1, we see the concepts: *Human, Boy, Woman, Perceive, Feel, Afraid, Age, Old, Young*.

+ Hyponym concepts:

- Boy → Human; Woman → Human; Feel → Perceive; Afraid → Perceive; Old → Age; Young → Age

+ Semantic Network having relationships between concepts as follows:

- The relationship between Human and Age is Has_Attribute
- The relationship between Human and Body is Has_Part
- The relationship between Human and Perceive is Has_Function
- The relationship between Boy and Young is Has_Attribute
- The relationship between Boy and Old is HasNo_Attribute

+ And based on the above relationships, we can deduce the following relationships:

- The relationship between Woman and Age is Has_Attribute
- The relationship between Woman and Perceive is Has_Function
- The relationship between Woman and Afraid is Has_Function
- The relationship between Woman and Young is Has_Attribute

4. Vietnamese noun compound model

In this model, we first organize the conceptual dictionary as follows: each word will be listed with all its possible parts-of-speech, in each of which will be distinguished meanings which will be linked (pointed) to an object – its corresponding semantic frame structure. Then, following the fields in each object, we determine possible semantic relationships between 2 objects and finally we choose the most reasonable relationship between them.

4.1 *Semantic frame structure of objects*

First, we build objects with their semantic frame structure for each meaning of each word of each part-of-speech:

4.1.1 *Noun objects*

Each meaning of a noun is an object with the following structure:

Name of field	Example 1	Example 2
Name of object	bird	animal
Hypernym	animal	biology
Hyponym	sparrow, dove,...	human, bird, ...
Holonym	flock of..	population, community
Meronym	swing, beak, ...	organs, ...
Function / ability	fly, sing, ...	eat, act, ...
Material attributes	small, light, color...	size, ...
Spiritual attributes
Properties (optional)

Table 2. Frame structure of Noun Object

- **Hypernym field:** indicating its parent object. Based on this link, it will inherit their parents' attributes, functions (except prohibited cases mentioned in "Properties" field).
- **Hyponym field:** indicating child object.
- **Holonym field:** indicating "organization", "class",.. to which this object belongs.
- **Meronym field:** indicating parts that directly constitute this object.
- **Function field:** links to verb-network to show actions, abilities which this object is able to perform or traditional purposes of this object.
- **Attribute field:** linking verb-network to show attributes, materials, and states of this object. These fields are slots like OAV (Object-Attribute-Value) which require contextual words to fill it and which are also satisfied with its default value.
- **Property field:** showing abnormal or special attributes of this object (non-inherit, non-principle,..).

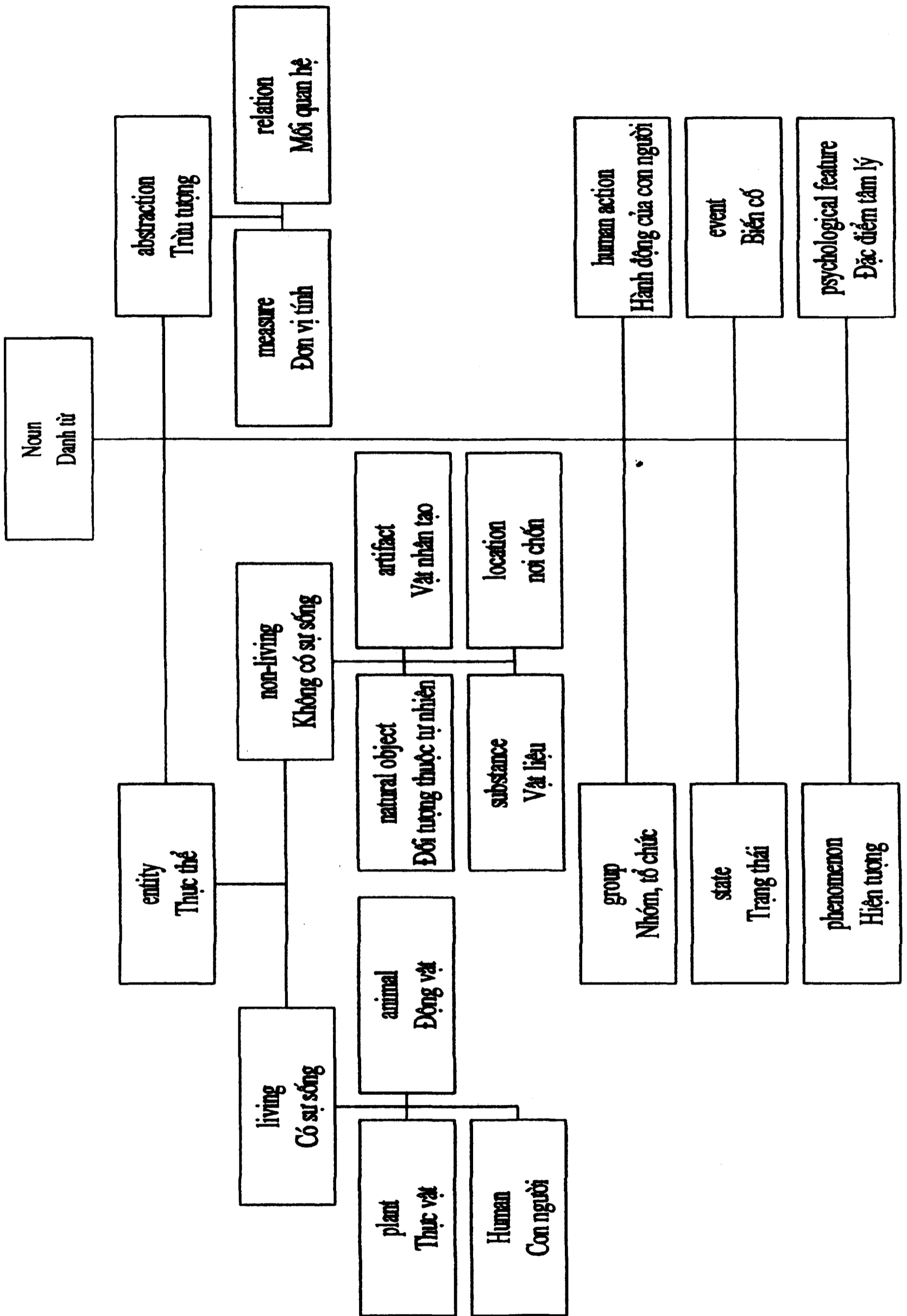


Figure 2. Schema of Nouns

4.1.2 *Verb objects*

Each meaning of a verb is an object with the following structure:

Name of field	Example 1	Example 2
Name of action	write	give
Hypernym	make words	abstract motion (possession)
Agentive role	writer	donor
Objective role	writings, works, ...	thing
Instrument role	pen, pencil,	facility
Patient role	to whom	receiver
Benefactive role	for whom	
Location, Spatial role		
Properties (optional)

Table 3. Frame structure of Verb Object.

Example: "Mr. Ba writes a letter to the teacher for me with a pen"

Agent V Obj Patient Beneficiary Instrument

In the relation to other words in a sentence, a verb plays a central role and requires nouns (subject, object,...) to meet the semantic role which its semantic frame has specified.

4.1.3 *Adjective objects*

Each meaning of an adjective is an object with the following structure:

Name of field	Example 1	Example 2
Name of attribute	long	long
Parent Class	shape	speed
Synonym cluster	size	time
Antonym cluster	short	fast
Properties (optional)

Table 4. Frame structure of Adjective Object.

Adjectives are organized into the classes shown in Fig. 3.

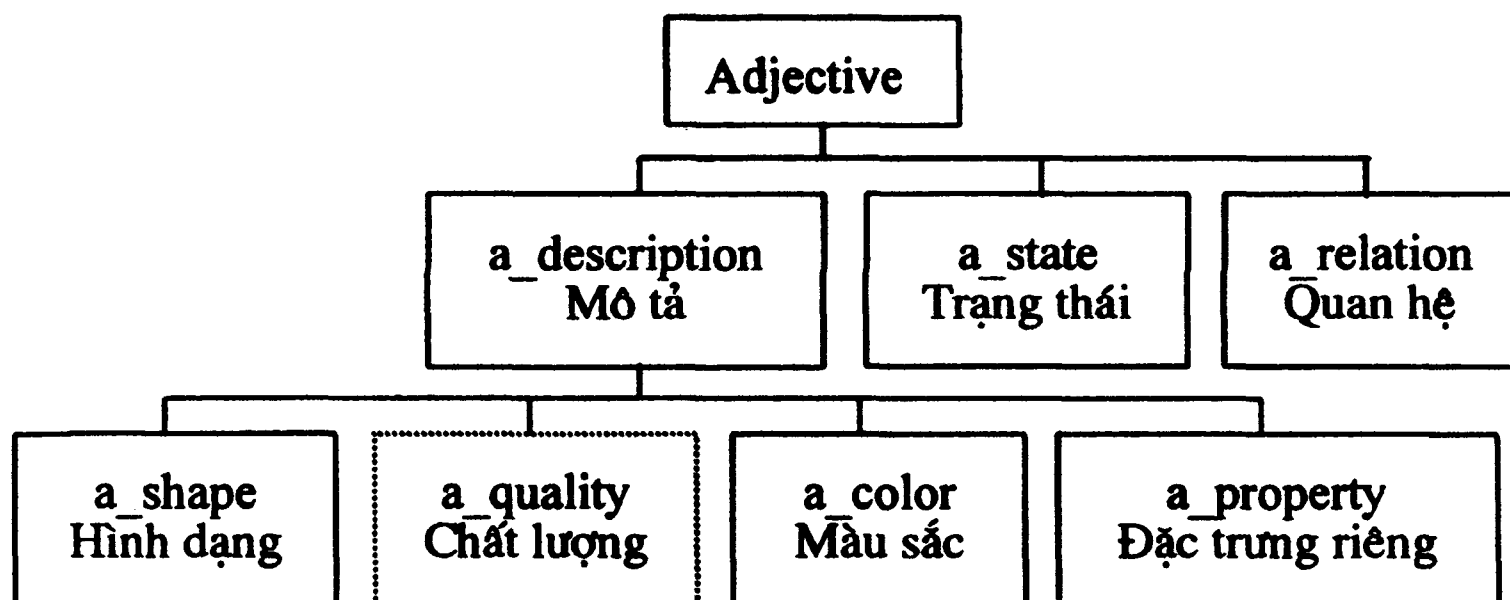


Figure 3. Schema of Adjectives.

*Each meaning of an adjective is classified separately according to its properties. These different meanings depend on the head noun which they modify.

*We need to distinguish between the descriptive adjectives and the relational adjectives:

- **Descriptive adjectives:** describing the attributes of the noun which it modifies, e.g.: “*Người bạn trẻ*” (*young friend*); “*Người bạn già*” (*old friend*). Descriptive adjectives are possible in predicative form, e.g.: “*Người bạn của tôi già*” (*my friend is old*).
- **Relational adjectives:** describing the relation between the head noun and the attribute which that adjective shows. e.g.: “*Người bạn mới*” (*new friend*) ; “*Người bạn cũ*” (*old friend*). Descriptive adjectives are impossible in predicative form, e.g.: **“Người bạn của tôi cũ”* (**my friend is old*). In this case, we must understand that *the relation (between me and my friend) is old*, rather than *my friend is old*.
- We may consider the descriptive adjectives as reference modifying and the relational adjectives as referent modifying.
 - Examples of descriptive adjectives: *cuộc bầu cử thú vị* (*interesting election*), *cuốn sách mới* (*new book*),...
 - Examples of relational adjectives: *cuộc bầu cử tổng thống* (*presidential election*), *sách vật lý* (*physics book*),...

4.2 Determining the semantic relation between objects

Example: Consider the NC: “sân gạch” consisting of two words: “sân” and “gạch”. We will go through the three following steps:

Step 1: look up a Vietnamese dictionary to get its part(s)-of-speech and its meaning(s):

- The word “sân” in Vietnamese has only one possible part-of-speech, Noun, with the meaning: “*the land around the house or place for amusement, sports, ...*” (Hoang, 1998:823).
- The word “gạch” has two possible parts-of-speech, Noun or Verb. As Noun there are two main meanings: 1. *Material for constructing* 2. *Matter of (liver, egg) in crabs*. As a Verb there is only one meaning: *to make a line*.
- Basing ourselves on the syntactic parsing of the Vietnamese NounCompound (section 2.2), we know that the first word of an NC must be a Noun and the head Noun (grammatically and semantically), and the second word could be Noun, Verb, Adjective, Pronoun, Preposition, Number... So for *sân* and *gạch* we only examine above-mentioned cases.

Step 2: examine the semantic frame of words in the NC:

- The semantic frame of Noun “sân” and “gạch” (meaning 1 and 2):

Name of field	Head Noun	Modifier	Modifier
Name of object	sân_dt	gạch_dt1	gạch_dt2
Hypernym	flat space region	material	matter
Hyponym
Holonym	house, stadium	construct material	crab
Meronym	surface, base,...		
function	to contain	construct	biological activity
Material attributes	size, ...	size, weight,...	amount, color,...

- The semantic frame of Verb “gạch”:

Name of field	Modifier
Name of action	gạch_dt
Hypernym	make a line
Agentive role	maker
Objective role	line
Instrument role	pen, pin-point thing,...
Patient role	surface,...

Step 3: examine the semantic relation of the words in the NC:

- We examine the semantic relation between the head noun “sân” (sân_dt) and the first meaning of the sub-noun “gạch” (gạch_dt1), then the second meaning of the sub-noun “gạch” (gạch_dt2), and finally the meaning of the verb “gạch”.
- If the relationship between the head noun and its modifiers is closer or clearer (i.e. its measurement of semantic distance is shorter), this relation will be given higher marks (higher potential). Finally, we choose the highest-mark relation. In this case, the material relation “yard made of brick,” has the highest mark so is chosen.
- The measurement of semantic distance is the number of links connecting one object to another, with its corresponding weight calculated from statistics.

4.3 Remarks on the semantic relation between objects

4.3.1 The relation between noun and noun:

- If the head noun matches the field “Meronym” or “Hyponym” of the sub-noun, we have the PART-OF relation. Ex: “chân bàn” (chân của cái bàn: leg of table), “đoàn viên công đoàn” (đoàn viên của công đoàn: member of community).
- If the sub-noun is related to the field “Function” of the head noun (e.g.: Objective, Reason, Purpose, ...), we could have the PURPOSE relation. Ex: “ly nước” (water glass) => “ly (glass) chứa (to contain) nước (water)”, “vườn cau” (areca garden => “vườn (garden) trồng (to plant) cau (areca), tai nạn xe cộ (vehicle accident) => “tai nạn (accident) do (by) xe cộ (vehicle), ...
- If the head noun belongs to the class “material object”, its sub-noun modifies the material attributes, external forms of the head noun, e.g.: “sân gạch” (brick yard) => “sân” (yard) làm bằng (made of) “gạch” brick;...
- If the sub-noun is a human noun or pronoun, its relation is often a POSSESSION relation, e.g.: “sách giáo viên” (“sách” của “giáo

viên”: book of teacher), “nhà tôi” (“nhà” của “tôi” : house of mine),
...

- In addition, there are other relations between two nouns, e.g.: “hệ thống máy tính” (computer system) => “hệ thống (system) bao gồm các (CONSISTING OF) máy tính” computer : the CONTENT relation, e.g.: “sách toán” (mathematics book) => “sách” (book) về (on) “toán” mathematics,...

4.3.2 The relation between noun and verb:

- If the verb matches the field “Function” of the head noun, we have the PURPOSE relation. Ex: “sân chơi” (sân để chơi : yard to play); “bàn học” (bàn để học : table to learn), ...
- If the head noun is the nominalization of a verb (“sự”, “việc”,...), the sub-noun will be related to the head noun through roles in Case Grammar of that nominalized verb. Ex: “việc sưu tầm tem” (stamp collection) => “tem” (stamp) is the object of “sưu tầm” (to collect); “việc cầu nguyện buổi sáng” (morning prayer) => “cầu nguyện” (to pray) in “buổi sáng” (morning) : TIME relation,...

4.3.3 The relation between noun and adjective:

- If the head noun belongs to the class “material object”, its adjective modifies the material attributes or the external forms of the head noun, e.g.: “cái bàn dài” (long table) => “cái bàn” (table) has size “dài” (long), “hoa vàng” (yellow flower) => “hoa” (flower) has color “vàng” (yellow),...
- If the head noun belongs to the class “abstract thing”, its adjective modifies the spiritual attributes or the internal content of the head noun, e.g.: “bộ phim hay” (interesting film) => “bộ phim” (film) has content “hay”(interesting),...
- If its adjective is Descriptive, it will describe the attributes of the noun which it modifies, e.g.: “sách đẹp” (pretty book). If its adjective is Relational, it will describe the relation between the head noun and the thing which that adjective indicates. e.g.: “Người bạn cũ” (old friend) => the relation (between me and my friend) is old.

Summary: The semantic relationship between the head noun and its modifier is determined by the linkages between the semantic frame of each meaning of each word. In each semantic frame, there are pointers indicating the related objects (e.g.: link to Hypernym, Hyponym, Meronym, Function, Agent, ...). We will use the measurement of distance of these linkages to determine whether its semantic relation is close or far in order to choose the most reasonable relation.

5. Conclusion, application and development

The above model is only a suggestion for the Vietnamese Noun Compound, which bases itself on the semantic relation model of English NC and the semantic database of English words (WordNet). In order to complete this model, however, we must solve many difficult issues, such as re-building a semantic database for Vietnamese words, which will consume much time and effort. We also must make statistical calculation for all forms of Vietnamese NC and determine the semantic relation for each type and limit ourselves to types that can be solved by this model.

Research on NC is a necessary task, especially in processing natural languages by computers. Once we understand Noun Compounds clearly, we will be able to solve the most difficult problem in natural languages – disambiguation of languages. We realize that NC includes most elements of syntax as well as semantics, so once we can analyze and understand NC, we will understand higher levels: Noun Phrase, Verb Phrase, Sentence,...

From the basis of NC models for English and Vietnamese, we can compare the similarities and differences between these two models and be further able to build the algorithm for translating automatically NC from English into Vietnamese and vice versa. This is the first step in moving toward the automatic textual translation by computers between English and Vietnamese.

REFERENCES

- Đỗ Hữu Châu. 1997. *Từ vựng - ngữ nghĩa tiếng Việt*. NXB ĐH QG Hà Nội.
- Downing, Angela and Locke, Philip. 1992. *A University course in English Grammar*. Prentice Hall. London.
- Hoàng Phê. 1998. *Từ điển tiếng Việt*. Trung tâm Từ điển học, NXB Đà Nẵng.
- Lauer, Mark. 1995. *Designing SLL: experiments on Noun Compounds*. Macquarie University, NSW, Australia.
- MacDonald. 1982. *Understanding Noun Compounds*. Carnegie-Mellon University Pittsburgh, PA, USA.
- Miller, George A. 1996. *Introduction to WordNet*. (Online lexical database) <http://www.cogsci.princeton.edu/~wn/>. Princeton.
- Nguyễn Tài Cẩn. 1998. *Ngữ pháp tiếng Việt*. NXB ĐH QG Hà Nội.

Received: 10 June 2001

National University of HCM City
Vietnam

<ddien@siagonnet.vn>

