Qualia Modification in Noun-Noun Compounds:

A Cross-Language Survey

Chih-yao Lee, Chia-hao Chang, Wei-chieh Hsu, Shu-kai Hsieh
Department of English
National Taiwan Normal University
chihyaolee@gmail.com

Abstract

In analyzing the formation of a given compound, both its internal syntactic structure and semantic relations need to be considered. The *Generative Lexicon Theory (GL Theory)* provides us with an explanatory model of compounds that captures the qualia modification relations in the semantic composition within a compound, which can be applied to natural language processing tasks. In this paper, we primarily discuss the qualia structure of noun-noun compounds found in Chinese as well as a couple of other languages like German, Spanish, Japanese and Italian. We briefly review the construction of compounds and focus on the noun-noun construction. While analyzing the semantic relationship between the words that compose a compound, we use the *GL Theory* to demonstrate that the proposed qualia structure enables compositional interpretation within the compound. Besides, we attempt to examine whether or not for each semantic head, its modifier can fit in one of the four quales. Finally, our analysis reveals the potentials and limits of qualia-based treatment of composition of nominal compounds and suggests a path for future work.

1. Introduction

Compounding is regarded as one means of word formation especially productive for analytic languages such as Chinese that want agglutinative inflectional markers [2]. Morpho-syntactically, within each compound word there is an internal structure comparable to those found on the sentential level of a given language. In Chinese, for example, the verb-noun construction of *xún-qiú-zhèn-gjù* 'to look for proof' can also be morphologically encoded as a discrete unit *qiú-zhèn* 'seek-proof'. According to *Headedness Principle* [3], in addition to the foregoing verb-noun structure, compounds in Chinese can also be composed of noun-noun, noun-verb and verb-verb. Such categorization is essentially based on the parts of speech of the individual words within a compound.

Packard [3] has a different system of classifying Chinese compounds by taking grammatical relations into consideration. Under his classification, the compound *qiú-zhèn* 'seek-proof' is decomposed into a verb along with its object, *zhǔ-gù* 'master-care' a subject plus its predicate. However, very few, if any, work has taken into account the semantics of the separate constitutive words that, when combined, would derive the meaning of the compound. This present study is set to look at the traditionally labeled noun-noun compounds found across a couple of dominant languages in addition to

Chinese under the framework of the *Generative Lexicon*, and to see whether such mechanism would apply across different languages.

2. Semantic Classification in Noun-Noun Compounds

In English and Chinese noun-noun compounds, the most common construction is modifier-head with the head on the right. For example, in *bookstore*, the modifier is *book* and the head is *store*; in *shā-táng* (sand-sugar) 'sugar', the modifier is *shā* and the head noun is *táng*. However, such analysis is derived on the basis of syntax rather than semantics. Since this present work focuses on the meaning relationship between the components of a given compound, we may encounter noun-noun compounds in which the head noun can be both the word on the left and the word on the right.

When choosing examples, we eliminate compound-like words which are composed of the construction of a word and an affix. For instance, some Chinese noun-noun compounds may be composed of a noun and an affix, such as *zhuō-jiǎo* (table-leg) 'table leg', in which *zhuō* is an affix rather than a word. This sort of compound is not taken into consideration due to the fact that an affix, which cannot stand alone as an individual word, is not a noun.

Moreover, in other languages, especially modern Romance languages, we may also find different constructions of noun-noun compounds from those found in English and Chinese. For example, in Italian, a noun-noun compound is composed as head-preposition-modifier, such as *succo di limone* (juice-preposition-lemon) 'lemon juice'; in French, *café au lait* (coffee-preposition-milk) 'white coffee' is also composed as head-preposition-modifier; Spanish has *cuchillo de cocina* (knife-preposition-kitchen) 'kitchen knife'. In the Romance languages mentioned above, the semantic head noun is usually located before the preposition, which means that the preposition in a noun-noun compound can function as an indicator for the head noun, while Chinese does not have this linguistic property.

In the semantic analysis of nominal compounds, Gagne and Shoben [1] first proposed a set of 'thematic relations', which are claimed to cover the majority of semantic relations between modifier and head in English noun-noun compounds:

relation	example
head *causes* modifier	flu virus
modifier *causes* head	college headache
head *has* modifier	picture book
modifier *has* head	lemon peel
head *makes* modifier	milk cow
head *made of* modifier	chocolate bird
head *for* modifier	cooking toy
modifier *is* head	dessert food
head *uses* modifier	gas antiques
head *about* modifier	travel magazine
head *located* modifier	mountain cabin
head *used by* modifier	servant language
modifier *located* head	murder town
head *derived from* modifier	oil money

Table 1: Thematic relations (Gagne et al. 1997)

However, upon close scrutinization some deviance and discrepancies can be found. The *enumerative* approach to the compositionality of compounds would easily lose the power in facing with usage in the novel context. As an alternative approach, in the next section, we will introduce a formally elaborated lexical semantic theory of a *generative* approach to compound meaning.

3. Qualia Structure in the Generative Lexicon Theory

3.1 Overview

The *Generative Lexicon Theory* (*GL Theory*) gives a new interpretation of the traditional qualia structure mentioned in the previous section. As pointed out by Pustejovsky [4], the qualia structure of lexical items can be explained as follows:

a. CONSTITUTIVE: the relation between an object and its constituents, or proper parts
    i. Material
    ii. Weight
    iii. Parts and component elements

b. FORMAL: that which distinguishes the object within a larger domain
    i. Orientation
    ii. Magnitude
    iii. Shape
    iv. Dimensionality
    v. Color
    vi. Position

c. TELIC: purpose and function of the object
    i. Purpose that an agent has in performing an act
    ii. Built-in function or aim which specifies certain activities

d. AGENTIVE: factors involved in the origin or "bringing about" of an object
    i. Creator
    ii. Artifact
    iii. Natural Kind
    iv. Causal Chain

Let us examine some English examples that can demonstrate the qualia structure in noun-noun compounds, which in English is composed in the modifier-head structure. In *glass door*, *chocolate cake*, and *oil painting*, the modifier is a sort of material of the head. Here, the examples demonstrate the CONSTITUTIVE quale, while in *history book*, *history* distinguishes *book* from other sorts of books, such as math books or chemistry books; in *horror movie*, *horror* distinguishes *movie* form other sorts of movies, such as action movies; in *noun phrase*, *noun* distinguishes *phrase* from other sorts of phrases, such as verb phrases; in *college student*, *college* distinguishes *student* from other sorts of students, such as high school students. Here, the examples demonstrate the FORMAL quale.

In *jewelry box*, *jewelry* indicates the function of *box*, which means that the box is used to contain jewelry; in *bookstore*, the function of *store* is to sell books; in *operation knife*, *knife* is with the function of surgical operation; in *drinking water*, *water* is with the function of being safely drunk by people. Here, the examples demonstrate the TELIC quale. In *adenovirus pneumonia*, *adenovirus* is a sort of virus that causes *pneumonia*; in *steamboat*, *steam* is the power for *boat*; in *turtle egg*, *turtle* is the producer of *egg*. Here, the examples demonstrate the AGENTIVE quale.

3.2 The Base Modes of the Qualia Structure

In this research, we use a lexical entry of the *GL Theory* in order to represent the structure of noun-noun compounds as follows: (α as the term itself; TYPESTR as type structure; ARG as argument; ARGSTR as argument structure; EVENTSTR as event structure)

$$
\begin{bmatrix}
\alpha \\
\text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \mathbf{x}:\ \tau \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix}
\text{CONSTITUTIVE} = \\
\text{FORMAL} = \mathbf{x} \\
\text{AGENTIVE} = \mathbf{R(e',x)} \\
\text{TELIC} = \mathbf{R(e,x)}
\end{bmatrix}
\end{bmatrix}
$$

Figure 1. The Base Mode of the Qualia Structure

The representation of the CONSTITUTIVE quale is as the following:

$$
\begin{bmatrix}
\mathbf{glass\ door} \\
\text{ARGSTR} = \begin{bmatrix} ARG_1 = \mathbf{phys\_obj} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{CONSTITUTIVE} = \mathbf{glass} \end{bmatrix}
\end{bmatrix}
$$

Figure 2 The Mode of the CONSTITUTIVE Quale

For the FORMAL quale, Pustejovsky [4] interprets the FORMAL quale with the *GL Theory* into two structures: (a). *Simple Typing*: value of FORMAL role is identical to sortal typing of the argument. (b). *Complex Typing*: value of FORMAL role defines the relation between the arguments of different types. The base mode of the FORMAL quale is as the following form:

$$
\begin{bmatrix}
\mathbf{college\ student} \\
\text{ARGSTR} = \begin{bmatrix} ARG_1 = \mathbf{human} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \mathbf{college\_level} \end{bmatrix}
\end{bmatrix}
$$

Figure 3. Types of the FORMAL Quale

The *GL Theory* interprets the TELIC quale into two base modes as described below:

$$\begin{bmatrix} \textbf{drinking water} \\ \text{ARGSTR} = & \begin{bmatrix} ARG_1 = \textbf{liquid} \end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix} \text{FORMAL} = \textbf{drink\_act} \end{bmatrix} \end{bmatrix}$$

Figure 4. Direct TELIC: something which one acts on directly

$$\begin{bmatrix} \textbf{operation knife} \\ \text{ARGSTR} = & \begin{bmatrix} ARG_1 = \textbf{tool} \end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix} \text{TELIC} = \textbf{operate(surgery)\_act} \end{bmatrix} \end{bmatrix}$$

Figure 5. Purpose TELIC: something which is used for facilitating a particular activity

Also, the AGENTIVE quale can be interpreted by the *GL Theory* with the base mode as the following:

$$\begin{bmatrix} \textbf{steamboat} \\ \text{ARGSTR} = & \begin{bmatrix} ARG_1 = \textbf{phys\_object} \end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix} \text{AGENTIVE} = \textbf{power\_act} \end{bmatrix} \end{bmatrix}$$

Figure 6. Types of the FORMAL quale

The next section is going to discuss the modification of the qualia structure in the *GL Theory* with cross-language data for comparison.

4. Qualia Modification

In previous sections, we discuss the morphology of Chinese compounds and the *GL Theory*. In this section, we are going to discuss the mapping of qualia in Chinese compounds. The proposed qualia structure in GL theory can be used to provide the lexical connection which binds semantic contributions of modifying nouns and the head noun in the compound [5].

In the *Headedness Principle* Packard [3] noted that there are two kinds of headedness, one is structurally (syntactically), the other is semantically. For example, *gāng-qín-jiàn-pán*, which means piano keyboard, structurally the head noun must be *jiàn-pán*, and thus the qualia structure is FORMAL. On the contrary, if we analyze the compound semantically, both *gāng-qín* and *jiàn-pán* can be the head noun. While we consider *gāng-qín* the head noun, the qualia structure is CONSTITUTIVE; however, if we consider *jiàn-pán* the head noun, the qualia structure is FORMAL. In short, the form of the semantic relation between the head and the modifier is not as specific as viewed from a syntactic perspective.

Therefore, we choose the word that serves the semantic content of the compound to be the head noun.

4.1 TELIC Qualia Modification

To illustrate qualia modification, we first discuss the TELIC role in which the modifying noun describes the purpose of the head noun. In the Chinese compound *cài-dāo* 'cleaver', *cài* modifies *dāo*'s purpose, which is to cut vegetable.
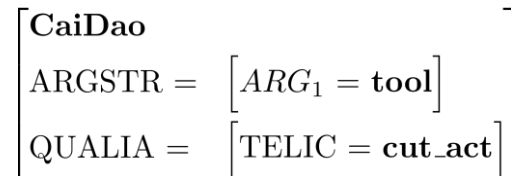
$$
\begin{bmatrix}
\textbf{CaiDao} \\
\text{ARGSTR} = \begin{bmatrix} ARG_1 = \textbf{tool} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{TELIC} = \textbf{cut\_act} \end{bmatrix}
\end{bmatrix}
$$

Figure 7. The TELIC Mode

Table 2

N1 is served as N2's function:

cài-dāo	菜刀	vegetable-knife	'cleaver'
shuǐ-guǒ-dāo	水果刀	fruit-knife	'fruit knife'
fàn-wǎn	飯碗	rice-bowl	'rice bowl'
yóu-jǐng	油井	oil-well	'oil well'
yǎn-jìng-hé	眼鏡盒	eyeglasses-box	'glasses case'

The TELIC quales of the compounds are also shown in other languages:

Table 3

ITALIAN			
	coltello da pane	knife-bread	'bread knife'
	bicchiere da vino	glass-wine	'wine glass'
JAPANESE			
	うつるえかん 映畫館	movie-building	'cinema'
	でんわちょう 電話帳	telephone-notebook	'telephone book'
	ぶんぐてん 文具店	stationery-store	'stationery'
	たくたまだい 桌球台	ping-pong-table	'pingpong table'
	どうろひょうしき 道路標識	road-sign	'road sign'
	サングラス	sun-glasses	'sunglasses'
	ほんばこ 本箱	book-box	'book box'
FRENCH			
	couteau de cuisine	knife-kitchen	'kitchen knife'
	boîte à bijoux	box-jewelry	'jewelry box'
	salle de bain	room-bath	'bathroom'

GERMAN			
	Lebensmittelgeschäf	foodstuff-store	'grocery store'
	Briefmarke	letter-mark	'stamp'
	Buchhandlung	book-action	'book store'

SPANISH			
	cuchillo de cocina	knife-kitchen	'kitchen knife'
	plato de arroz	plate-rice	'rice plate'
	vaso de vino	glass-wine	'wine glass'
	manguera de agua	hose-water	'water hose'
	bus escolar	bus-school	'school bus'

For example, the Italian complex nominal *coltello da pane* 'bread knife' shows the quale of *coltello* and *pane*, the preposition *da* is used as a connector. There are also examples from other languages.

4.2 AGENTIVE Qualia Modification

The AGENTIVE quale explains how something comes into being. In compounds, the modifier brings out the head noun. In the Chinese compound *jī-dàn* 'egg', the modifier *jī* brings out the head noun *dàn.*
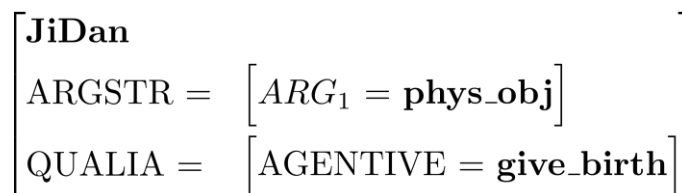
$$
\begin{bmatrix}
\textbf{JiDan} \\
\text{ARGSTR} = \begin{bmatrix} ARG_1 = \textbf{phys\_obj} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{AGENTIVE} = \textbf{give\_birth} \end{bmatrix}
\end{bmatrix}
$$

Figure 8. The AGENTIVE Mode

Table 4

N2 is produced by N1:

niú-nǎi	牛奶	cow-milk	'milk'
zhí-wù-yóu	植物油	plant-oil	'oil from plants'
yú-luǎn	魚卵	fish-egg	'roe'
hǎi-yán	海鹽	sea-salt	'sea salt'
hǎi-ní	海泥	sea-mud	'mud'
gōng-chǎng-fèi-qì	工廠廢氣	factory-exhaust	'exhaust from a factory'
huǒ-chē	火車	fire-car	'train'
diàn-chē	電車	electric-car	'tram'
zhēngqì-chuan	蒸汽船	steam-boat	'steam boat'
níngméng-zhī	檸檬汁	lemon juice	'lemon juice'
jǐng-shuǐ	井水	well-water	'water form a well'

In the Italian complex nominal *succo di limone,* which means lemon juice, *limone* serves the AGENTIVE role, which modifies *succo*'s origin. The appropriate preposition for the Italian form appears to be *di*.

Table 5

ITALIAN			
	foro di pallottola	hole-bullet	'bullet hole'
	succo di limone	lemon-juice	'lemon juice'
JAPANESE			
	蜂蜜	bee-honey	'honey'
	烏魚子	mullet-roe	'mullet roe'
FRENCH			
	jus de citron	juice-lemon	'lemon juice'
	arc-en-ciel	arch-sky	'rainbow'
	pomme de terre	apple-earth	'potato
GERMAN			
	Hühnerei	chicken-egg	'egg'
	Meersalz	sea-salt	'sea salt'
	Dampfschiff	steam-boat	'steam boat'
SPANISH			
	jugo de limón	juice-lemon	'lemon juice'
	carne de cerdo	meat-pig	'pork'
	aceite de oliva	oil-olive	'olive oil'

4.3 CONSTITUTIVE Qualia Modification

In complex nominals or compounds, there are modifiers used to specify a subpart of the denotation of the head noun or the material of which it is composed. The Chinese compound *pí-xié* 'leather shoes' shows that the modifier *pí* is something of which the head noun *xié* is composed.
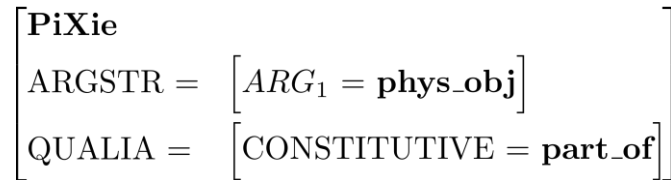
$$
\begin{bmatrix}
\textbf{PiXie} \\
\text{ARGSTR} = \begin{bmatrix} ARG_1 = \textbf{phys\_obj} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{CONSTITUTIVE} = \textbf{part\_of} \end{bmatrix}
\end{bmatrix}
$$

Figure 9 The CONSTITUTIVE Mode

Table 6

Head noun is made or composed of the modifier:

tiě-lù	鐵路	iron-road	'railroad'
cǎo-méi-dàn-gāo	草莓蛋糕	strawberry-cake	'strawberry cake'
bōlí-mén	玻璃門	glass-door	'glass door'

The CONSTITUTIVE quale in words can also be seen in other languages. In the Italian complex nominal *porta a vetri,* means glass door, *verti* is the modifier of *porta*, and the appropriate preposition here is *a*.

Table 7

ITALIAN			
	porta a vetri	door-glass	'glass door'
	seni al silicone	breast-silicon	'silicon breast'
JAPANESE			
	味噌汁	miso-soup	'miso soup'
	花 瓣	flower-petal	'petal'
	咖哩飯	curry-rice	'curry rice'
FRENCH			
	café au lait	coffee-milk	'white coffee'
GERMAN			
	Käsekuchen	cheese-cake	'cheese cake'
	Türknopf	door-knob	'door knob'
SPANISH			
	pastel de queso	cake-cheese	'cheese cake'
	puerta de vidrio	door-glass	'glass door'
	sopa de maíz	soup-corn	'corn soup'
	tenedor de plastico	fork-plastic	'plastic fork'
	casa de madera	house-wood	'wooden house'
	casa de ladrillo	house-brick	'brick house'

4.4 FORMAL Qualia Modification

In the FORMAL quale, the modifier distinguishes the head noun within a larger domain. For example, in the Chinese compound *lán-huā*, *lán* distinguishes different types of *huā*.
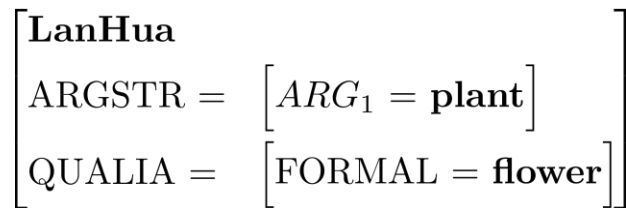
$$
\begin{bmatrix}
\textbf{LanHua} \\
\text{ARGSTR} = \begin{bmatrix} ARG_1 = \textbf{plant} \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \textbf{flower} \end{bmatrix}
\end{bmatrix}
$$

Figure 10 The FORMAL Mode

Table 8

N1 is a type or subclass of N2:

shǒu-biǎo	手錶	hand-watch	'watch'
huǒ-chē-biàn-dang	火車便當	train-lunchbox	'lunchbox sold on trains'
pí-zhěn	皮疹	skin-rash	'rash'
xīn-bìng	心病	heart-disease	'mental disorder'

wèi-ái	胃癌	stomach-cancer	'stomach cancer'
shā-táng	砂糖	sand-sugar	'granulated sugar'
chá-zhuān	茶磚	tea-brick	'brick tea'
shuǐ-niǎo	水鳥	water-bird	'aquatic bird'
diàn-nǎo-yíng-mù	電腦螢幕	computer-screen	'monitor'
hǎi-niú	海牛	sea-cow	'manatee'
méi-kuàng	煤礦	coal-mine	'coal mine'

The FORMAL quale also appears in other languages to distinguish an object from a larger set.

Table 9

ITALIAN			
	cibo spazzatura	food-junk	'junk food'
	fermata del taxi	stop-taxi	'taxi stop'
JAPANESE			
	中學校教師	middle-school-teacher	'middle school teacher'
	模型飛行機	model-aircraft	'model aircraft'
	螢光燈	fluorescent light	'daylight lamp'
	指輪	finger-ring	'finger ring'
	庭石	garden-rock	'garden rock'
FRENCH			
	professeur de lycée	teacher-middle school	'middle school teacher'
GERMAN			
	Samstagnachmittags	Saturday-afternoon	'Saturday afternoon'
	Familienname	family-name	'last name'
	Ladentisch	store-table	'counter'
	Düsenflugzeug	nozzle-airplane	'jet'
	Rindfleisch	cow-meat	'beef'
	Schweineflesich	pig-meat	'pork'
	Fleischfresser	meat-eater	'carnivore'
	Mittagessen	noon-food	'lunch'
	Abendessen	evening-food	'dinner'
	Arbeitszeit	work-time	'working time'
SPANISH			
	comida chatarra	food-scrap	'junk food'
	partido de futbol	party-football	'soccer game'
	cancha de tennis	court-tennis	'tennis field'
	arbol de manzanas	tree-apples	'apple tree'

5. Discussion

In previous sections, we choose the word that serves the primary semantic content of the compound to be the head noun. Although the syntactic head noun and semantic head noun are the same in our examples, there are words that the syntactic head noun is not the same as semantic head noun. For example, in Chinese *làng-huā* (wave-flower) 'surf', if we assume that the syntactic head noun is *huā*, there is a semantic confusion due to the fact that intuitively, *làng-huā* is not a kind of *huā*, but the shape of *làng*, and therefore *làng* serves the primary semantic content. Thus we say the head noun is *làng* rather than *huā*. Similar examples are *huǒ-huā* 'sparkle', *dàn-huā* 'egg flower', *xuě-huā* 'snowflake' and *shuǐ-huā* 'water spray'.

Still, there are some words that cannot fit into anyone of the four quales, even though they are individual words. For example, in Chinese *shān-jiǎo* (mountain-leg) 'the base of a mountain', we cannot judge which quale in which the modifier *shān* fits if we take *jiǎo* as the semantic head noun.

6. Conclusion and Future Works

In this paper, we only discuss noun-noun compounds. However, in the study of Chinese morpho-syntax, there are other sorts of construction within compounds, such as noun-verb, verb-noun, verb-verb, etc. For further study on the qualia structure of compounds, other constructions can also be taken into consideration.

Moreover, large-scale statistics can also enhance similar research. With appropriate statistics of the distribution of the four quales within compounds, we can examine whether the four quales are used with apparently different frequencies. Also, this research can be applied to some important NLP and IR/IE tasks such as automatic classification of semantic relations between nominals. With manually annotated training data at hand, we are working on the system that can automatically predict the meaning of a compound by matching the qualia structures of the nominals from a given compound.

References

[1] C. L. Gagne and E. J. Shoben. "Influence of thematic relations on the comprehension of modifier-noun combinations". *Journal of Experimental Psychology: Learning, Memory and Cognition.* 1997.

[2] G. F. Arcodia, "Chinese: A Language of Compound Words?" *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse, ed. Fabio Montermini, Gilles Boyé, and Nabil Hathout,* 79-90, Somerville, Giorgio Francesco, 2007.

[3] J. L. Packard, *The Morphology of Chinese*, New York, Cambridge University Press, 2000.

[4] J. Pustejovsky, *The Generative Lexicon*, Cambridge, The MIT Press, 1995.

[5] M. Johnston and F. Busa, "Qualia Structure and the Compositional Interpretation of Compounds", Waltham, Brandeis University, 1996.