

The Cleft of Disyllable in Mandarin Oriented to NLP

Xue Hongwu

College of arts, Chongqing Normal University
Chongqing, China
E-mail: xuehongwu386@sina.com.cn

Received May 2013; revised July 2013

ABSTRACT. *Chinese lack rigid morphological limits on its words, so disyllable words are often be cleft by inserting other words. In order to improve the efficiency and accuracy of word segmentation and tagging in a dynamic discourse and semantic computation basing on rules, this paper study the complex phenomena of the cleft in disyllable words. It analyzes the two types of cleft phenomena in disyllable words systematically and describe their grammatical motivation, forming conditions and different meaning in their forming process and their syntactic representation of constituents clefted from them. So we get some systematic and multi-level conclusions on the cleft phenomena of disyllable words in Mandarin.*

Keywords: disyllable, cleft, syntactic feature, semantic meaning, Mandarin

1. Introduction. One of the starting points of the semantic identification of computer is able to continually segment words and tag the property of their part-of-speech in the NLP. But at present, there are the two problems which the validity and accuracy are not up to standard. One of the formed reasons for them is Mandarin are strictly lack of morphology-bound as Indo-European languages, so the deformation or degeneration of the words in sentence grow universally. One of the most typical representations is the double syllable words are separated due to inserting the related to syntactical components. How to make a segmenting and tagging to the cleft words in their quantity and quality, the current actual operation is not consistent in NPL. At the same time, after the word is separated, whether their semantic meaning can change or not, more attention do not pay to. On this condition, the computer is difficult in semantic understanding and computing semantic orientation based on the rules.

At present, the research of the separated words can be summarized as the description of facts or phenomena are not adequate and systematical in denotation, i.e. the current research is only confined to the so-called Lihe-word which can be separated and reunited syntactically. It is coded into IP in the paper. for example, the word “休息” [xiuxi] (*to rest*), “上班” [shangban] (*to go to work*) and “头痛” [toutong] (*headache*) so on. The Ip is naturally continuous category between words and phrases, is not typical word form. Without examining the typical disyllabic words (code it as Iw), it is not adequate. We believe, only

the cleft of typical disyllabic words are explored systematically, can Mandarin information processing make a great and deep progress.

This paper will from the static and dynamic grammatical hierarchy respectively, systematically study the cleft of the IP and IW, and their composition features and semantic meanings. There are the two purposes in it, one is to segment the words and tag their features in sentences quantitatively and qualitatively, the other serves the semantic orientation computation. For the expression convenience and concentration, the constituents of Iw and Ip are all coded as the “m1m2”, the constituents inserted in them are all coded into (x) and y. the parentheses of x is a optional symbol. As for the cleft of multi-syllable, we can be inferred by the paper conclusion.

2. The form and meaning of the Iw cleft. Every kinds of feature and different syntactical matching pattern of the Iw in sentences, can universally be inserted by the other constituents and form the word cleft. The syntactic derivation formula is following as:

$$D \{\lambda [x, y] (m1m2)\} \rightarrow (x) m1ym2 \quad (1)$$

The D refers to the pragmatic factors. The sufficient or necessary conditions which the word structure of the “m1m2” is disintegrated as the “(x) m1ym2”, are that inserting $\lambda[x, y]$ into m1m2 must get D-license. Based the effects of D, the deformation or degeneration of the “m1m2” which is due to the constituent of x or y in sentence, can be defined to as the pragmatic Configuration. (Xue Hongwu 2009)

2.1. The cleft of compound word. The noun cleft. Please observe:

(1) 老李头说，一块儿工作的，和我有点关系。

Lao Li tou shuo, yikuaier gongzuode, hewo youdian guanxi.

The old man whose surname is Li said: “ he works together with me and has a little relationship with me.”

The “老李头” [LaoLitou] (old man Li) can be analyzed as the “李”[Li](surname) is inserted in the noun “老头”[Laotou] (old man). “李”is an attributive of the noun“老头”. Mean while, it also can be analyzed as the “老” attributes the “李头”[Litou](surname), thus the “老” is an attributive. However, focusing on Mandarin syntactic system, only the former analysis is appropriate, because of the “李头” in Mandarin is not a grammatical unit. See the example below:

(2) a. 重庆北路 [Chongqing beilu](North Chongqing Road)

b. 化两小时妆[hua liangxiaoshi zhuang](making up for two hours)

The (2a)(2b) are formed by the way that the "north" and "two hours" is inserted in the same structure of the m1m2, about this Xiao Guozheng (2001:160-161) has been demonstrated. “老李头” in the (1) is strictly same to the (2b) by analogy, although the later is a verb phrase.

In short, the cleft word of “老李头” express the interpersonal feelings of speaker contrasting to the “李老头”[Lilaotou](old man Li), which are lovely and warm meaning. The (2a) and (2b) are the rhythm, syntactic and semantic requirement, this will be discussed in Section 3.1. Now observe another example:

(3) 他老师常说：“艺术”“艺术”，有艺还得有术。

Ta laoshi changshuo : “yishu” “yishu”, youyi hai de youshu.

His teacher often says: there should be both the skills and arts for an artist.

Because the subject “艺术”[yishu](art) of the second clause in (3) is emphasized by a repeated way, in order to keep the information distribution rules which the former “subject” is light and the latter “predicator” is heavy, the predicate use the cleft-structure, i.e. the parallel verb structure “有艺有术” [youyi youshu](to have both skills and art). Through this syntactical operation, the (3) achieves to emphasize and comment the subject “艺术”, meanwhile, also to maintain the balance of it.

the cleft of adjective word. For example:

(4) 看来你不但认了真了，而且认真得一塌又糊涂。

Kanlai ni budan renle zhenle, erqie renzhen de yita you hutu.

It seems that you are not only conscientious, but also are in an awful mess state.

The adjective “认真”[renzhen](conscientious) is inserted the “了”[le](perfect aspect marker) and form a verb-object phrase“认了真”[ren le zhen](be serious). “认了真” can emphasize the object “真”[zhen](real) of the “认”[ren](recognize), comparing to it, “认真” has not the emphatic function. The formal mark is that “认了真” does not have its complement any more, such as we can not say “认了真了一塌糊涂”[renlezhenle yitahutu](be very serious). On the contrary, “认真” can do, for example the “认真得一塌糊涂”[renzhen de yitahutu](be especially serious) in the second clause. In the same, the cleft form “一塌又糊涂” [yita you hutu](very) is to emphasize the negative state of the “awful state” which “认真” has in syntax, and the “一塌又糊涂” in the semantic function is equivalent to the two words of “一塌糊涂”.

The cleft of verb.

(5) 出路还在改革，非改革不可……得真正依靠工人阶级来改，来革。

Chulu haizai gaige, fei gaige buke,……dei zhenzheng yikao gongren jieji laigai, laige.

Its way out lies in reforming, we must reform, …… and we must really rely on the working class to reform.

(6) 他当这个小女婿，后老悔了。

Ta dang zhege xiao nvxu, hou laohui le.

He deeply regretted that he became his son-in-law.

The "reform" in (5) is inserted by the“来”[lai](in order to) and the pause (comma) and form the cleft formation “来改,来革” [laigai, laige](to renovate). It is the synthesis predicate of the defaulted subject “我们”[women](we), in the same, its semantic is equivalent to two “改革”[gaige](to reform) which emphasize the intention and belief of the speaker in "reform". the “后老悔”[hou lao hui](to regret deeply) in (6), is the cleft form of the “后悔”, and change into a V-O structure. Its function also emphasized the high degree of the "后悔".

2.2. The cleft of mono-morphemic word. The disyllable mono-morphemic word also can be separated by inserting the constituents in sentence, and its two syllables serve as two

syntactic constituents, and its grammatical meanings are emphatic too. For example:

(7) 以为把人逗笑的东西就是幽默，其实它既不幽也不默。

Yiwei bar en douxiao de dongxi jiushi youmo, qishi ta ji bu you ye bu mo.

Someone think that humor is something that can make people laugh, however, it is not the humor in a real sense.

The “既不幽也不默”[*ji bu you ye bu mo*](*It is not humor at all*) is the serial predicate of the “它”[*ta*](*it*), its function is to emphasize the feature of the “humor”. Both “幽”[*you*](*morpheme*) and “默”[*mo*](*morpheme*) are the indispensable syllables (see the section 3.2), they all are equivalent to the adjective “幽默”[*youmo*](*humor*) in the semantic and grammatical function.

2.3. The cleft of derivational word. For example, after a sleepless wife heard the noisy, he pushed up his sleeping husband and made such a dialogue as:

(8) a. 好像有老鼠，你去看去。

Haoliang you laoshu, ni kankanqu.

It seems that there are mice, you go and have a look.

b. 老什么鼠，快睡吧!

Lao shenme shu, Kuai shuiba!

You are talking nonsense, go to sleep now.

The“老什么鼠”[*lao shenme shu*](*mouse*)in (9b)is the cleft form of the derivational word “老鼠” [*laoshu*](*mouse*)which is formed by inserting the “什么”[*shenme*](*what*). It expresses the strong impatient moods of speaker (the husband) to his wife. Comparing it with the (8b1) below, we will find out this point. Please observe:

(8) b1. mice? Go to sleep now!

Apparently, the emotional intensity of the (8b1) is weaker than the (8b). It should be pointed out that the “老”[*lao*](*noun prefix*)in the (8b) is the mention of reflexivity of the “老鼠” rather than the its “use”, and the semantic meaning of the“老什么鼠” is same that don't talk nothing of mice any more. The prefix “老” has become a verb, and its function substitutes for the “老鼠”. Accordingly, “鼠” [*shu*](*mouse*)is the object to the "use". So the syntactical form of the (9b) actually should be:

(8) b2. 老鼠什么老鼠!

Laoshu shenme laoshu!

Don't talk of mice any more!

The (8b) is different it in that the former retained the morpheme “老” and omitted the morpheme “鼠” in the first “老鼠”, the second one was reversed. This syntactic operation is decided by the the disyllable of the “老鼠”. This syntactical configuration can be systematically explained by the "mention" usage of the monosyllable word in the rhetorical question. Please observe the example (9b) as following.

(9) a. 爸爸，我的书呢?

Baba, wo de shu ne?

Dad, where are my books?

b. 书什么书，你学吗?

Shu shenme shu, ni xue ma?

Don't talk of them, do you read?

The analysis above show, the grammatical changed from the word “m1m2” to the phrase “(x) m1ym2” can be summarized as the three respects in morphologies, including its length, internal structure or feature. According to them, the semantic changes were from the expressing concept to the expressing the interpersonal mood. All these conformed to the Iconicity Principles which refers to if linguistic entities represent the unusual unique things, so its grammatical form code corresponding to the complex and multidimensional. (Givón 1990:968)

3. The difference between the clefts of IP and IW

3.1. **The grammatical device of the IP separation.** Please observe the example (10):

(10) 洗澡(xizao, to have a shower)

→a.洗一次澡(xi yicizao, to have a shower)

→b.洗十分钟澡(xi shifenzhong zao, having a shower for ten minutes)

→c.洗热水澡(xi reshui zao, to have a hot water shower)

The “洗澡”[xizao] (shower) which is a typical Lihe-word, can be separated as a phrase in a static syntactic environment. Its property is not only unchanged, but the syntactical and semantic relations between the “洗”[xi](to have) and the “澡”[zao](shower) are not also changed. The inserted constituents in semantic all are oriented to the “洗” or the “澡”, such as in the (10a) and(10b) the inserted constituents “一次” [yici](once) and “十分钟”[shifenzhong](ten minutes) all oriented to “洗”, in (10c), “热水” [reshui](hot water) is oriented to the“澡”. In a word, that Lihe-word is separated is to express the different concept-semantic from its cleft forms, the inserted constituents are obligatory in syntax. For example, “洗澡” is different from (10)a (10)b and (10)c in their concept and semantics, and the differences are caused by the inserting of constituents “一次”, “十分钟” and “热水”.

Based on the discussion above, we can draw a conclusion the motivation of IP cleft has two aspects. One is that express the concept-semantic, such as the denotation of (10)a“洗一次澡” is small and specific comparing to the word “洗澡”. The other is decided by the syntactical proximity iconicity (Givón 1990:970), for example, “洗澡” and “一次” “十分钟” “热水” have two syntactical orders in theory as followings:

(10') a.洗一次澡(xi yicizao, to have a shower) | *洗澡一次(xizao yici, to have a shower)

b.洗十分钟澡(xi shifenzhong zao, to have a shower for ten minutes) | *洗澡十分钟(xizao shifenzhong, to have a shower for ten minutes)

c.洗热水澡(xi reshui zao, to have a hot water shower) | *洗澡热水(xi zao reshui, to have a hot water shower)

Because “一次”, “十分钟” and “热水” are predicates and are similar with the “洗” in semantic property, they are in close proximity to the “洗” and form the effective syntactical structure. These unacceptable phrases with asterisk are due to violating the proximity iconicity. Givón (1990:970) Obviously, the motivation which the IP is separated, is in the syntactical and semantic needs.

Further observing the cleft of “洗澡” in sentence, we will find out the “澡” can not only be topicalized, but also be defaulted in the fully contexts, For examples.

(12) 一个月他也不洗一次澡。

Yige yue ta ye buxi yici zao.

For one month, he haven't have a shower .

a. 澡, 他一个月也不洗一次。

Zao, ta yige yue ye buxi yici.

For one month, he haven't have a shower .

b. 一个月, 他连一次也不洗。

Yige yue ta lian yici zaobuxi .

For one month, he haven't have a shower .

Based on the analysis and description above, the conditions which the Ip is separated, are free. As for the motivation, can be represented the below (the letters “G/S” stands for the syntax and semantics):

$$G/S\{\lambda [x, y] (m1m2)\} \rightarrow (x) m1ym \quad (2)$$

3.2. The difference between the IP and IW. There mainly are the four different respects between the IW and IP.

Firstly, the Iw can be operated only in sentence or clause, in the static levels , it cannot be did. Such as:

(13)*认了真 (renle zhen,) | *一塌又糊涂(yita you hutu)

*来改来革(laigai laige) | *既不幽也不默(jibuyou yebumo)

*老什么鼠(lao shenshu) | *后老悔(hou laohui)

The “认了真” is grammatical, but it has become a V-O structure, and is not a word any more. semantic also became the dominant relation. “老李头” in (1) is grammatical, because it is has been lexicalized (lexicalized) in Mandarin, which the motivation is the high frequency of use and special interpersonal expression function. Although “有艺还得有术” also is grammatical in the (4), the syntactical sum of “艺” and “术” is not equal to the reference of the “艺术” .

Secondly, the m1 and m2 in the cleft Iw, are bound in syntactical distribution, and are no grammar activities. Such as regardless of the changes in the “认了真” syntactical structure and grammatical properties comparing to the “认真”, so far as the “真” , it can never be topicalized, and can only be bound after the “认” .

(4') *真你不但认了, 而且认真得一塌又糊涂。

* Zhen ni budan renle, erqie rende yita you hutu.

You are not only be serious but also be very serious.

Its cause which the “真” can not be topicalized, will be further explained in section 3.2.

Thirdly, all kinds of constituents in grammatical properties are different about the cleft Iw. Please look at the inserted x and y first. Comparing to the y in the cleft of Ip ((x) m1ym2), the y or (x) in the cleft Iw are redundancy. Omitting the x and y of all the examples in the section 2.1, the truth-condition and the syntactical structures of the

sentences are not changed. Cases (6) and transform (6') for comparison:

(5') 出路还在改革，非改革不可……真正依靠工人阶级改革。

Chuluhaizai gaige, fei gaige buke...zhenzheng yikao gongrenjieji gaige.

Its way out lies in reforming, we must reform and really rely on the working class to reform.

It is the same in truth-condition as the (5). The predicate centre in the (5) is a series verb phrase, i.e. [真正[依靠工人阶级[来改, 来革]]], the predicate centre in the (5') is the word[真正[依靠工人阶级[改革]]]([really [to rely on the working class [to reform]]]), although the two are the adverb- verb structure. It is noted there are some inserted constituents, they are not strong, such as the “老” in the “后老悔”, this is caused by the “后悔” which is tendency to the phrase in grammatical units.

The properties in the m1 and m2 in the cleft IW are that some are no semantic in static grammatical environments, such as the “一塌” [yita] and “幽” [you] and “默” [mo] are syllables. some are affixes, such as the “老” [lao] in the “老鼠” [laoshu]. some have semantic, but they act as syntactic constituents independently, such as the “革” and “术” [shu]. Besides these, there are some morphemes which can form a word, for example the “艺” [yi], but its reference is not equal to the “艺术” [yishu]. In brief, the m1 and m2 of the cleft Iw are not grammatical units independently in a static environment, only in the dynamic sentences, can they act as the constituents in syntactic and semantic meaning. Based on these, they can be called the dynamic syntactic constituents. Semantically, the m1 and m2 are all equal to the IW. In syntax, the m1 and m2 must in the co-occurrence states, by which they consist of the IW in order. Now, we represent their semantic and syntactical properties as followings.

$$m1(s)=m1m2, m2(s)=m1m2; \quad m1(g) \wedge (g)m2 \quad (3)$$

The “s” and “g” stand for the semantic and syntactical needs. Undoubtedly, the semantic and grammatical functions of the m1 and m2 are in the proliferation and degeneration, when the Iw was separated in sentences. About these, contrasting the (6') to (6) will find out the m1 and m2 as the dynamic syntactical constituents must simultaneously satisfy the two syntactic the conditions as following:

The one is the sentence which the m1 and m2 distributes, need the precedent sentences, or the isolated "(x) m1ym2" sentence can have problems, making a comparison between the (8') and (8), and we will see it:

(6') 逗笑人的东西既不幽也不默。

Douxiaoren de dongxi ji bu you ye bu mo.

The thing which can make people laugh is not necessarily the humor.

It is very unacceptable in the native spoken. the (8) is grammatical and can emphasize the “幽默” [youmo], because it has the precedent sentence “以为把人逗笑的东西就是幽默” [yiwei ba ren douxiao de dongxi jiushi youmo] as a pave the way. For example, If we extract the “老什么鼠” [laoshenmoshu] from the (9) as a independent sentence, it is almost unintelligible.

The other is the m1 and m2 must be coordinate with each other, and their syntactical

order in sentence must be parallel to the order which they are in the Iw. The "coordinate" refers to the m1 and m2 must be adjacent and echo each other. The (4') is not grammatical, because the “真” is too far away from the “认” to lose the coordinate and parallel syntactical links. The same condition is dependent each other between the m1 and m2, if anyone is lack in sentence, the sentence in which they distribute, can be acceptable very weakly. For example:

(6') 他当这个小女婿，老悔了。

He dang zhege xiao nvxu, laohui le.

(8') a. 好像有老鼠，你看看去。

Haoxiang you laoshu, ni kankanqu.

You are talking nonsense, go to sleep now.

b. 什么鼠！睡吧。

Shenme shu! Kuai shui ba.

You are talking nonsense, Go to sleep now.

The syntactical order of the m1 and m2 is parallel to their order as morphemes in IW, this refers to that the m1 and m2 cannot be translocation, such as the“一塌又糊涂”can be not expressed as the“糊涂又一塌”. In the example (4) “a collapse and confused” sense of language is generally not as "confused and a collapse".

The “(x)m1ym2” of IW must be licensed in sentence by the *D* which specifically refers to the modality, including the epistemic, denotic and dynamic. Epistemic modality is of the status attitude which the speaker owns the faith in the action, event or affair-state. For example, the “看来” in the (4) and the “真正”[*zhenren*] in the (5) and the “了” in (6) “and the “认为……就是”[*renwei...jiushi*] in(7) and the “什么” in (8)”. Deontic modality is that the speaker hold to stop a behavior or propose an event or action in accordance with the sort of moral responsibility, for example the “得” [*de*] In (3) and the “非……不可” in (5). Dynamic modality is associated with the ability and intention of speaker. The behavior and affair-state is generally applied in the speaker's subjective based on the intervention and power, as the “要”[*yao*] and “来”[*lai*] in (5) (orientation to the speaker's willingness), and the modal particle “吧” which express the command or impatient tone in (8). Finegan (1995) think that the speaker’s mood, intention, attitude or emotion is an important aspect of the sentence subjectivity. Lyons (1977:102) has defined subjectivity as ‘the way in which natural languages, in their structure and their normal manner of operation, provide for the locutionary agent’s expression of himself and of attitudes and beliefs’. In other word, the subjectivity refers to the structures and strategies that languages evolve in the linguistic realization of subjectivity or to the relevant processes of evolution themselves. In view of this, the cleft of Iw is subjective structure (Finegan 1995).

It’s worth noting the way which express subjective category by the cleft of IW and IP, is not found out in Indo-European. Up to now, we can conclude that apply the cleft form of

⁶Xiao Guozheng(1999) pointed out that the “了”(le, equal to the perfect aspect)has such the semantic function as emphases, deviation of anticipation and euphemistic. They all reflect the epistemic modality in nature. meanwhile, Xue Hongwu(2012) looked the “了” as a epistemic word which can soften the illocutionary act of sentence in pragmatics .

the word in sentence to transmitting the speaker's subjective stances or attitudes, is the Mandarin's grammar properties in some degree. Its precondition is that interacts with the modal constituents in sentence, this includes the two respects; one is not only the formation of "(x) m1ym2" accompanied by the modal constituents, and also its distribution must be done by it, the other is the subjective form and semantic meaning of "(x) m1ym2" in Mandarin are quite extensive and prominent, such as the "有艺还得有术" in (3) comment on the "艺术". For another example, the (4) is a qualitative comment on the "你" by the cleft form of "认了真" and "一塌糊涂". So, what is the formation mechanism of the cleft IW's subjective meaning? Please observe and make a comparison among the three clauses in the (5):

(5) 出路还在改革 i, 非改革 i 不可……真正依靠工人阶级来改 m1i, 来革 m2i。

Chulu haizai gaige, fei geige buke, ……(women)dei zhenzheng yikao gongren jieji laigai, laige.

Its way out lies in reforming, (we) must reform…… and we must really rely on the working class to reform.

The first clause emphasizes the "改革" by the adverb "在" and the nature of stress, on the one hand the "改革" is a natural focus with the [+prominent] and [+contrast] semantic features, on the other hand, its focus of identity is marked by the "在" again. i.e. the "改革" in the first clause have the double focus identity. The second clause emphasizes the "改革" by its natural focus identity and the lexical semantic meaning of the "非……不可"(deontic modal word). When the third emphasizes the "改革", it must depend on the "改革" above the two, we can express it as "i=m1i+m2i". Only based on it, together with the "intention" semantic meanings of the inserted constituent of "来" and the natural focus identity of the "改""革", can we construe their complete semantic and grammatical properties of the "来改, 来革", otherwise it is puzzled.

It is clear the separated constituents of "改" and "革" as same two words, are to emphasize the speaker's subjective intention. The formation of their semantic meaning includes the syntactic operation and pragmatic inference, the former has already been discussed above, and the latter refers to that their emphatic meaning cannot do without two antecedent clause's index, i.e. there are contextual factors in them. Because of this, it will have the force of expression which the two antecedent clauses do not have. Reasoning by this, the subjective emphatic meanings of the invariants of the "改革" in (5), are increased one by one, and the "来改来革" is the information peak in the (5). In the same, the syllables of "幽" and "默" can not originally act as the words in static levels, because of they are the two syllables of the precedent cause's focus, in its index we can construe them as the two words which their semantic and grammatical function are all equal to the "幽默" (humor) . By all these, we can explain completely why the m1 and m2 as the dynamic syntactical constituents must have the two conditions mentioned in section 3.2.

4. conclusion and future work. All the disyllable content words can be separated in a sentence. According to the motivation, conditions and semantic difference, they can be divided into the two kinds of cleft forms, i.e. the IP and IW. We conclude them as

followings:

$$Iw: D \{ \lambda_{[x, y]} + m1m2 \} \rightarrow (x) m1ym2$$

$$Ip: G/S \{ \lambda_{[x, y]} + m1m2 \} \rightarrow (x) m1ym2 \quad (4)$$

The equations (4) can simply be explained by the theory of markedness as that the cleft of Ip is an unmarked static syntactic phenomenon, and its semantic is objective; on the contrary, the cleft of Iw is marked dynamic syntactic phenomenon, its semantic is subjective. The two must be paid great attention to the NPL, especially to the two following aspects; one is the semantic orientation computation based on the grammatical rules, and the other important issue related to this is the automatic segmentation and part-of-speech tagging. At present, there are the discrepancies of operation in both the word -segmenting and word-tagging of the cleft form of IP and IW. From this paper perspective, after the Iw and Ip are separated, they are not words but phrases. The differences between them lie only when their intra-word components are separated as the syntactic constituents, they have the different syntactical level requirements respects. So, they can be unified in full to segmenting words in sentence and making a quantitative and qualitative tagging for them. Of course these kinds of operation have the different treatments from the normal text. For examples (marked only the discussing object):

(5)出路还在改革，非改革(i+j)不可……真正依靠工人阶级来/v 改(i+j)/v，来 v/革(i+j)/v。

Chulu haizai gaige, fei geige buke,……women dei zhenzheng yikao gongren jieji laigai, laige.

Its way out lies in reforming, we must reform, …… and we must really rely on the working class to reform.

(7)以为把人逗笑的东西就是幽默(i+j)，而其实它既不幽/a(i+j)也不默/a(i+j)。

Yiwei ba ren douxiao de dongxi jiushi youmo, qishi ta ji bu you ye bu mo.

Looking the humor as which can make people laugh, in fact it is not.

(8)a.好像有老鼠(i+j)，你~~看看~~去。

Hoaxing you laoshu, ni kankanqu.

There seem mice, you have a look.

(8)b.老(i+j)/v 什么鼠/n(i+j)！快睡吧。

Lao shenme shu! Kuai shuiba.

You are talking nonsense, Go to sleep now.

(10) 一个月他也不洗/v 一次澡/n。

Yige yue ta ye buxi yici zao.

For one month, he haven't take a shower once.

(10) a. 澡/n，他一个月也不洗/v 一次。

Zao, ta yige yue ye buxi yici.

For one month, he haven't take a shower once.

This paper's deeply illustrated the theories and methods rooted from the Indo-European are extremely difficult in processing the word cleft of Mandarin. Such as when expressing

the same subjective emphatic semantic meaning as the cleft form of Ip and Iw, Indo-European language represents it only in accordance with the hard syntactic rules. The causes are the Indo-European grammar units and their syntactic activity are bound by the strict forms, and separating the disyllable word is extremely limited. Mandarin words have not dead morphology, so the pragmatic factors easily play a positive role on the grammar units, separating the disyllable words are the representations of the respect. As a result, the flexible syntactic forms and their according to semantic grammatical meanings of the all kinds of hierarchy and units, are formed. These not only pay attention to the language study, also can it be paid more in NPL.

Inferences from the conclusions above, the function word of the “m1m2” pattern can also be separated. Such as it at least can be separated as the "m1 什么 m2", for examples "所什么以"[*suo shenme yi*](*so*)and “随什么着”[*sui shenme zhe*](*along with*) etc.. Of course, the inserted word of x or y are extremely limited in paradigms.

Acknowledgment. The research work is funded by National Natural Science Foundation of China, “Construction and Automatic Analysis of Chinese Semantic Resource of Connective Structure”, under Grant No.61202193, and “A study on the Mandarin pragmatic configuration” by the Doctor Fund of Chongqing Normal University under Grant NO. 11XWB012.

REFERENCES

- [1] Finegan, E. Subjectivity and Subjectivisation: an Introduction. Subjectivity and Subjectivisation, Edited by Dieter Stein and Susan Wright. Cambridge University Press, (1995).
- [2] Givón, T. Syntax: A Functional typological Introduction (vol.II), Amsterdam/Philadelphia: John Benjamin Publishing Company,(1990): 968.
- [3] Lyons. J. Semantic. Cambridge: Cambridge University Press, (1977).
- [4] Xiao Guozheng. A study on Chinese grammar theory. Wuhan: Huazhong Normal University press, (2001).
- [5] Xiao Guozheng. Analyzing the semantic meanings of the “le” in sentence end, Facing the challenges of the new century of modern Chinese grammar study. edited by Lu Jianming. Ji'nan: Shandong Education Press, (1999).
- [6] Xue Hongwu. A study on the Mandarin pragmatic configuration. Shanghai: Chinese Department of Fudan University, (2009).
- [7] Xue Hongwu. The semantic grammatical function of "youzhe" and the "zhe". Chinese learning, (5), (2012): 62~72.