What are the Basic Units of Mandarin Chinese Discourse?

Li Rong[1] and Wang Hongjun[2]

[1]Xi'an International Studies University
No.437, Changannan Road, Yata District, Xi'an, China
lirong1210@126.com

[2]Peking University
No.5, Yiheyuan Road, Haidian District, Beijing, China
fthjww@pku.edu.cn

ABSTRACT. *The discourse segmentation plays a very important role in discourse analysis, especially on defining the basic units. But there is no consensus in the literature on what a basic discourse unit is and how it should be identified. This process is more difficult for Chinese discourse due to the wildly use of zero anaphora and lack of morphological markers. We compare the minimal units in Chinese discourse and English discourse, and claim that basic unit of Chinese discourse is "Dou" (clause) which has no special grammatical markers but have certain prosody features which consist with the comma in written language. We also discuss the special "Liushu" (flowing water) sentence in Chinese discourse and the reason why it is widely used. In addition, we discuss the need of define the basic units of discourse and how to prove it is a natural unit inside the native speaker's brain while reading.*
Keywords: Basic Unit; Sentence; Dou (Clause); Mandarin Chinese Discourse

1. **Introduction.** In discourse analysis it is commonly acknowledged that discourse is (hierarchically) structured. Please refer to the work of [13, 16, 26, 8, 9, etc]. There are two kinds of discourse structures, one is micro-structure, and the other is macro-structure. What are the basic units of the discourse? The answers fall to the research of the micro-structure of discourse. The assumption is that the discourse is built up from smaller "building blocks" related to one another in a coherent way. What these building blocks actually look like differs according to the discourse models. Determining the basic units of the discourse is a very important issue, especially for anaphora resolution. Because it always influence the results and bring questions the conclusion.

However, many authors avoid the problem of defining the minimal segments. The work of [26] is content with stating that the minimal units are "typically clauses". Others just work this definition in view of (automated) discourse segmentation[1].

---

[1] For a detailed overview, see [30].

This problem is even bigger in Chinese discourse. Since Chinese is lack of morphological markers, it is difficult to determine when the sentence stops. For example,

那辆车 $_i$ 价钱太贵， $Ø_i$ 颜色也不好， 我 $_j$ 不喜欢 $Ø_i$， $Ø_j$ 也不想买 $Ø_i$。[2]　　(1)

(a) *Na-liang chei jiaqian tai gui, $Ø_i$ yanse ye bu hao, wo $_j$ bu xihuan $Ø_i$, $Ø_j$ bu xiang mai $Ø_i$.*
(b) that-CL[3] car $_i$, price too high, $Ø_i$ color also not good, I $_i$ not like $Ø_i$, $Ø_j$ not want buy $Ø_i$.
(c) That car is too expensive. The color is not good either. I don't like it and (I) [4] don't want to buy it.

In this example, there are 4 zero amphora which "i" refer to "that car", while "j" refer to "i". In Chinese, these four clauses are in same sentences which marked with a full stop while in the English translation, it is corresponded to 4 sentences as in line (c).

This Chinese sentence even has two topics, "the car" and "I". So if you want to do the research about topic chain or anaphora, it need to be divided into two parts otherwise it will be a problem. One may wonder: Why so much information can fit into one sentence? Or should we still use the term "sentence" to call discourse fragments like example (1) since they are so different from the sentence in English? Why the Chinese native speakers seem to have no problem with so blur boundary between sentences and clause?

This all leads to the same question: what are the basic units of the Chinese discourse? We will give a brief review about the formal research about this issue in next section.

2. **Overview about the basic units of Chinese discourse.** There is a long history of research about the basic units of Chinese discourse, we presents you the most influential four theories and compared their differences with examples.

2.1. Phrase as the basic unit. It is mainly propose by [44, 45], which also called "Cizu(word group)" in Chinese. He believes that any phrases in Chinese plus intonation will become a sentence. In other words, the phrases in Chinese have the basic elements of a sentence, but it must have the intonation.

2.2. Minor sentence as the basic unit. According to [43], sentence is a piece of utterance which is paused in both ends which can be divided into full sentence and minor sentence. A full sentence has both subject and predicate while minor sentence only has subject or predicate. A full sentence in Chinese is made up of two minor sentences[5], which are normal and frequently used in daily life. And it is the basic units of Chinese spoken language. This is the first time the term of "the basic units of spoken language" has been put forward which is the same as the basic units of discourse which we discuss here.

The interesting thing is that minor sentence does not have certain prosodic markers, only syntactic definitions (only subject or predicate). So it there is a subject with no pause

---

[2] This example is modified from an example in [28]. Line (a) is pinyin which shows the pronunciation of the example; line (b) is a word-to-word translation while line (c) is the actual translation of the whole sentence.
[3] CL means measure word which is a special part of speech in Chinese, usually appears between numbers and nouns.
[4] ( ) means this is zero anaphora in Chinese and we add it in the English translation according to its context.
[5] The minor here means "piece" or "not complete".

in both ends, is it still a minor sentence, such as example,

    (a) 饭啊，还没好呢！   (2)

         Fan a, hai mei hao ne!

         Meal a-particle, still not good ne-particle！

         The meal is not ready yet.

    (b) 饭还没好呢，Ø等等再走。[6]

         Fan hai mei hao ne, Ø deng deng zai zou.

         Meal still not good ne-particle, Ø wait wait then go.

         The meal is not ready yet, then (we) wait and go.

In example (a), "fan a" only has subject, Chao define it as minor sentence. We also agree because it not only has the subject but also have pause or extension, which is a sign of prosodic boundary. The work of [27] questioned that, if we don't consider the prosodic features, then in the first clause of example (b), "fan (meal)"is also the subject, but it don't have the pause in the end or have the particle "a". Then is this subject also minor sentence? Is that all the subject and predicate in SVO should be treated as two minor sentences? There are no explicit answers in [43]. But from the analysis of the examples, we can conclude that the research of [44] actually only regards the subject which have pauses in both ends as the minor sentence. We think the definition of minor sentence should fit two criterions: (1) the sentences which are not syntactically complete, only have the subject or the predicate. (2) The sentences which have the prosodic pause or extension.

Because of the primacy of minor sentences, the grammatical meaning of the subject is literally the topic, often loosely related to the predicate，and the division of verbal and nominal predicates is less relevant．Because a succession of two minor sentences does not necessarily make a full sentence，a Chinese discourse is often made up of a series of "flowing sentences" as defined by [22].

2.3. Clause as the basic unit. This theory was proposed by Xing. The reader can refer to [40, 41]. He pointed the nature of the phrase and clause is different. The clause is a minimal grammatical unit which is independent and expressing concepts. It can formulate meaning, which means states intentions or reflects a specific purpose of the speaking. Every clause has its own mood. A clause is independent. That means this clause in not embedded in another clause. In conclusion, a clause is a pragmatic unit which is formulated, independent and with intonation. But phrase is a grammatical unit. His work does not discuss the question we proposed before: if a discourse fragment only has subject or predicate but it also has intonation, should we categorize it into clause or some other units?

Another advocator of the idea that clause is the basic unit of Chinese discourse is Lu, see [23]. He considers that clause is a grammatical unit, and the units which are above clause are discourse units. A clause is one subject-predicate construction which both components can be omitted. One subject predicate construction is corresponded with a event, which can be related to verbs and its arguments. We can only discuss the

---

[6] This example is from [27].

relationship of the events above the event level. As for the units which are higher than clause, he believes that a clause can become a sentence or even a "Juqun"[7] (group of sentence) by itself. He does not distinguish a clause and a phrase.

The work of [42] favors using "discourse clause "to define the basic unit of Chinese discourse. In his work, a discourse clause is basically a subject predicate structure, which includes the ones omitted subjects. He employs the sub-standard of pauses and functions to divide a discourse clause. The research of [5] defines a clause as a sentence which has at least a different form of predicate. In the book *A Discourse Grammar of Mandarin Chinese,* see [5] he uses clause as the basic units of discourse instead of sentence.

2.4. "Punctuation sentence" as the basic unit. As we discussed before, it seems difficult for the segment of clauses in Chinese discourse. While some research argued that the Chinese native speakers don't appear to have this problem. So why don't we use the punctuation as the standard of discourse segmentation? The writer or speaker usually has their own consideration while the readers or listeners may have a different understanding. This may be the origin of the problem. Another benefit of this view is that it makes discourse processing much earlier. Since one don't have to divide the discourse by using other standards, such as subject predicate structure or prosodic features, they can just use the punctuation features which is already exists in discourse.

The interesting work of [34] proposed this method mainly for the convenience of computer processing discourse. He argues that there is not a generally accepted, operated definition of clause, sentence, and group of closely related sentences. If go beyond phrases, the main formal labels should be punctuations. In the light of easier operation, we should use "punctuation sentences" as the basic units of the discourse which beyond words and phrases. In fact, many papers targeted in discourse analysis are using punctuation sentences as the main standards combined with some adjustments to divide the basic unit. For example, the research of [20] is about Chinese top-chains. And the work of [29] is about Chinese third person pronoun and zero anaphora. In reference [1], he explained that the paper is using punctuation as the features, treating the fragments which are separated by comma, stop, and question marks as clauses[8].

Although many papers are using punctuation as the method to determine the basic units of Chinese discourse, is this a reliable way or not? We should tend to some related experiment to find the answer.

In the work of [36], Tsao carried a very interesting experiment called "adding punctuations". During this experiment, he has 18 Chinese native speakers who are from Taiwan Normal University to put the right punctuations according to their understanding of two short texts, one is in Chinese, and the other is in English. The material is listed below. [9]

曾经是历史最光辉的拳王阿里，近年来胜利以后，总是说要退休，但总未退休，

---

[7] "Jun qun" is a Chinese grammatical unit, which means a group of closely related sentence. The clauses are in causal relation, successive relation, suppositional relation etc.
[8] This is quoted from [2] pp.182.
[9] This is the punctuation is from the original texts. When doing the experiment, the experimenter removes the original punctuations and let the subjects put what they think is right.

结果败在初出道的史宾斯科手下，本来可以光荣退休，却不料落下这一个下场。(3)

cengjing shi lishi zui guanghui de quanwang Ali₍ᵢ₎, jinnian lai Øᵢ shengli yihou, Ø zongshi shuo yao tuixiu, dan Øᵢ zongwei tuixiu, jieguo Øᵢ baizai chu chudao de Shibingsike shouxia, benlai Øᵢ keyi guangrong tuixiu, que Øᵢ buliao luoxia zheyige₍ⱼ₎ xiachang.

Once is history the most brilliant de-particle box champion Ali $_i$, in recent years $Ø_i$ victory after, $Ø_1$ always said to retire, but $Ø_i$ has not retired, the result $Ø_i$ is defeated at the beginning of Sibinsike's debuts, $Ø_i$ could retire in glory, but $Ø_i$ unexpected to fall this one.

Ali used to be the most brilliant box champion in history. After recent years of victory, (he) always said want to retire, but turned out not. At last, (he) was defeated by the Sibinsike on his first debuts. (He) could retire in glory, but fall on this unexpected one.

It seldom rains. The geography books credit this portion of Utah nine to ten inches of precipitation. Actual rainfall and snowfall vary widely from year to year. There are a few perennial springs hidden in secret places knows only to the deer and coyotes, to myself and a few friends, but the water does not flow far before vanishings into the air.   (4)

In the experiment, he removes the original punctuations and let the 18 subjects put what they think is right. Notice the subjects are Chinese native speakers and second language learners of English. The results are in table 1:

TABLE 1: Results on adding punctuations in English and Chinese texts

Items	Chinese		English	
The original numbers of clause	1	2	6	5
The average of the clause numbers which the subjects determined	2.53	3.82	5.49	4.94
How many clauses are included in a sentence	1-4	2-5	4-6	3-6
The number of the subjects put the same punctuation as the original text	1	1	8	6

As we can see from table1, although 18 subjects who participant the experiment are Chinese native speakers, their punctuations have huge difference with the original writer. Even on how many clauses are included in a sentence, the views is apparently inconsistency (see 1 vs. 2.53). However, in spite of their half mastery of English as second language, they manage to have the good agreement with the original writer (see 5 vs. 4.96).   The finding of [36] can explain these results. English is a sentence prominent language, which sentence is a grammatical unit with complete structures. The relationship between subjects, predicates and objects have clear marker. The boundary of the sentences is quite clearly-cut. While Chinese is a sentence group prominent language, in which sentence is difficult to define from the grammatical perspective.

In [7], the researcher carried the adding punctuations experiment, and she discovered the differences between "real stop" and "stop-to-be". In some places of the discourse, 90% of the subjects tend to use the full stop, which means this is the boundary of real stop. But in some other places, only 30% of the subjects are willing to put a full stop.

This should be the "stop-to-be" boundaries. The results shows that there are two discourse levels in the native speaker's mind: one is clause, and the other is sentence. These two units closely connect with top chains.

2.5. Summary. As we discussed before, these four influential theoretical frameworks about what is the basic unit of Chinese discourse actually have something in common. They both meet the difficulty of cut the boundary between sentence and clause, in other words, they all try to define two concepts: clause and sentence. Some believes clause is a pragmatic unit, such as [41], [42], [5] etc. while other think clause is a grammatical unit just difficult to define, see [1] and [23]. Although their beliefs are far apart, they tack the theoretical issue of defining basic units of Chinese discourse acknowledge their bidimensional status either with a linguistic appearance (in the form of syntactic clauses) or a textual appearance (in the form of contextualized information units). More importantly the maximal units at the micro-syntactic level (clauses) do not necessarily correspond to the minimal discourse units at the macro-level. In other words, there is no clear-cut boundary between the two levels. More recently, there has been a tendency to consider basic units of discourse as being multidimensional. That is to say they are defined in terms of different criteria: the realization of a mood, a conceptual content, syntactic dependency relations and a prosodic intonation (refer to [28]). "Typical" basic discourse units would amalgamate these four dimensions, while "non typical" discourse units may lack one of these dimensions.

As for Chinese discourse, due to the widely use of the zero anaphora and lack of the markers for the agreement of the subject and predicate, to determine a clause in context becomes more difficult. For example,

这个人 ¡！ Ø¡ 也不跟朋友打招呼！　　　　　　　　　　　　　　　　　　(5)

Zhe ge ren! Ø$_i$ ye bu gen pengyou dazhoahu!

This person! Ø$_i$ Also not with friends say hello!

This person! (He) does not say hello to friends!

This example can easily alter to the next example by remove the exclamation mark and two clauses become a sentence.

这个人也不跟朋友打招呼！　　　　　　　　　　　　　　　　　　　　　(6)

Zhe ge ren ye bu gen pengyou dazhoahu!

This person also not with friends say hello!

This person does not say hello to friends!

The difference between example (5) and (6) is only the pause between the subject and the predicate. The content and mood is the same. The comparison tells us if we want to determine the basic units of Chinese discourse, we need to count the prosodic features.

3. **The Traditional Chinese linguistic View "*Dou*" and related issues.** People who know about the history of Chinese will understand this problem with a completely different view. In traditional Chinese, there are no punctuations. People who go to school are learning how and when to pause in a passage. We can not give the statistical proof of the diachronic influence of the mess about determining the clause in modern Chinese. But we are sure can inspire from the traditional Chinese linguistic view of dealing with this

complex by using the concept "*Dou*". We think "*Dou*" can be the mark of Chinese clause in discourse. The form of "*Dou*" equals to a comparatively longer pause in prosody that is the flat-postpone intonation. The prosodic features are: the vowel of the last characters postponed along with that the pith shows a bit rising or falling intonation. In written language, it shows as comma ",". "*Dou*" means comma in Chinese. The pragmatic dimension of "*Dou*" is the intention or mood of the interaction between speaker and listener. Pay attention that this prosodic feature is different from the sentence which contain a complete event or a complete intention that writer or speaker wants to express. The prosodic feature of the sentence is that he vowel of the last characters become short along with the pith shows a comparatively bigger rising or falling intonation, then have a much longer silence which is the real stop. [10] In other words, from the prosodic features, the Chinese discourse is similar to the English discourse that is they both have the minimal and less minimal two levels of units. But the grammatical features are different. Example (5) and (6) fits the prosodic features of Chinese discourse, however, only example (6) fits the grammatical features.

In the book [43], Chao said, the researchers who study modern linguistics all agree that if we are study a one language, we should not try to look for the familiar things that we happened to know in a language we speakers, but the efforts should be made to determine what actually encountered and give appropriate names to it. So the content of the name not the name itself is more important. As for this issue, if we want to focus on the two level units, we can use clause or sentence. If we want to emphasis the grammatical differences between Chinese and English discourse, we should use another name, like the traditional Chinese linguists called "*Dou*" (clause) and "Ju" (sentence).

Compared to clauses in English discourse, "*Dou*" as the minimal unit in Chinese discourse has its own distinguish features: (1) there is no certain grammatical form of "*Dou*". It does not need to contain one subject and predicate structure. Also the need of having one finite verb is not necessary. In fact, we can also see this from the four theories we introduced in 2, although they emphasis different aspects. (2) There is no limit of the connections of the units. Despite juxtaposed relation, transitional relation, conditional relation, causal relations, it can be all kinds of explanation relations. Therefore, usually there is only one or two comma in English discourse while there are usually 7 or 8 commas in the Chinese discourse. See the comparison of example (7) and its translation in line c.

他ᵢ下了决心1，Øᵢ不跟她j吵2，Øᵢ不跟她j闹3，Øᵢ倒头就睡4，　　　　(7)
Øᵢ明天照旧出来拉车5，她j爱怎样就怎样6！

(a) **ta**ᵢ *xia le juexin 1,* Øᵢ*bu gen ta chao 2,* Øᵢ*bu gen ta*j *nao 3,* Øᵢ*daotou jiu shui 4,*
   Øᵢ*mingtian zhaojiu chulai lache 5, ta*j *ai-zenyang-zenyang 6!*

(b) **ta**ᵢ made le-particle decision 1, Øᵢnot with herj quarrel 2, Øᵢnot with herj fight 3,
   Øᵢ lie-down then sleep 4, Øᵢ tomorrow same come-out pull-cart 5,
     shej do-what-she-want 6!

(c) Heᵢ had made up his mind 1. (He) would not quarrel with her 2, would not be angry

---

[10] More detailed information is in section 3.

with her 3, but would just lay his head down and go off to sleep 4. Tomorrow (he) would go out as he had today to pull his rickshaw 5. She could do what she liked 6.[11]

When we compare the Chinese text in example (7) and its English translation, one can easily discover the Chinese discourse use only one sentence while in its English translation, there are four sentences. The Chinese clauses have a very rich and colorful joint ways, as clause 2-6 is the explanation of his decision in clause1. And clause 2 and clause 3 can be seen as parallel relation while they form a transitional relation with clause 4 and 5. In addition, clause 6 is the condition. In spite of her action, he will just do the same thing. In English translation, we find clause 2.3 and 4 are in one sentence. But the other relations have to form different sentences.

Chafe in work [3] argues that the clause appears to be the prototypical intonation unit type of English discourse, from which most other types are derived, or are derivations. And in [4] he claims that Some intonation units are parts of clauses and some may contain more than a single clause, but the mean proportion of single-clause substantive intonation units in the socially varied (interactional conversation) sample was 60%. If it is a discourse using formal language, the present will be much higher.

The research in [32] finds out that clauses in different language have different corresponding relationship to intonation units. See Table 2 for the detailed result.

TABLE 2：The relationship between clause and intonation units across languages[12]

	Clause=intonation units	Others units=intonation units
English	53.6%	46.4%
Japanese	45.4%	54.6%
Chinese	39.8%	60.2%

From Table2, we can see in Chinese discourse, more than half (60.2%) intonation units are equivalent to other units but not clauses. So this may be the origin of the problem. Some times the intonation units equals to a clause which makes the typical basic units in discourse, see example (6). Though in other circumstances, the intonation unit equals to a complex sentence like example (7) or equals to just part of the clause, such as the subject in example (5).

When we compared the percent whether the intonation units is a complete sentence or not, we get table 3.

TABLE 3: sentence completion intonation vs. non-sentence completion intonation across languages[13]

	Sentence completion intonation	Non-sentence completion intonation
English	82.5%	17.5%
Japanese	24.2%	75.8%
Chinese	36.7%	63.3%

---

[11] This example came from a famous Chinese novel. Please refer to [20].
[12] This table is from [18].
[13] This table is from [18].

As we can see from Table 3, English usually have a sentence completion intonation. So it is very easy to cut boundary. While Chinese is more difficult, only 36.7% of the intonation units are complete sentences. But the most difficult one is Japanese, only 24.3% of the Japanese intonation units are complete sentence. However, Japanese has very rich morphology markers to help the language users to draw a line in discourse. [14] The only clues we have in Chinese may be the prosodic features and some unique grammatical structures due to the lack of morphology markers.

The paper of [2] gives a very detailed description about intonation units in Chinese narrative spoken language. They give the definition of intonation units as well as purposing the standards. They use the corpus of *the pear story* to research on the correspondence of grammatical units and intonation units. Their have listed all the results in table 4.

TABLE 4: The grammatical units vs. Intonation units in Taiwan Guoyu[15]

Intonation units		Numbers	percent
Clause		373	47.2%
	Complete clauses	114	14.4%
	Non complete clauses	259	32.8%
	Parts of clause	81	10.3%
	Zero anaphora	167	21.1%
	Other anaphora	11	1.4%
Compound sentences		25	3.2%
	Complete sentences	10	1.3%
	Non complete sentences	15	1.9%
NP structures		217	27.4%
	Independent ones	45	5.7%
	dependent ones	172	21.7%
Other grammatical structures		156	19.7%
Non grammatical structures		20	2.5%
Total		791	100%

From table 4, we can see very few complete clauses realize as intonation units (14.4%). Most of the intonation units take the form of NP structures (27.4%) or other grammatical structures (19.7%). In Chinese, the subject-predicate structure can be a predicate or subject of the sentence, consider example (8), (9) and (10) are the other grammatical structures listed in table 4.

电影我看报了 1，没什么不好的 2。[16]   (8)

_____

[14] For example, Japanese have a topic marker wa to make the topic. Chinese don't.
[15] "Taiwai guoyu" is a dialect of mandarin Chinese spoken in Taiwan. They are very much alike but also bear some differences.
[16] Examples (8)-(11) are our own example.

Dianying wo kanle1, mei shem buhaode 2.

Movie I read newspaper le-particle 1, no what bad de-particle 2.

I read about the movie in newspaper 1, and it was not bad 2.

In this example, "wo kanbao le (I read the newspaper)" is a subject-predicate structure, but it works at the modifier of movie. Clause 2 was the comment of the movie.

你游泳学会了吗？   (9)

Ni youyong xuehui le ma?

You swim learned le-particle ma-particle?

Have you learned how to swim?

In this example, "ni youyong" which means "you swim" in Chinese is a subject-predicate structure which usually form an independent clause in English discourse. However, here it only functions as the subject of the verb "have learned".

We can also see in example (10), in which "xuexiao yijing you guiding" (The school has already established the rules) is a subject-predicate structures served as a predicate here.

作弊处罚的事情学校已经有规定了。   (10)

Zuobi chufa de shiqing xuexiao yijing you guiding le.

Cheating punishment de-particle thing school already have rules le-particle

The rule about punishment the cheating in exams has already well established.

From example (8), (9) and (10) we can conclude that sometimes the intonation units are bigger than one subject-predicate unit. On the contrary, sometimes it is smaller than one subject-predicate unit, like in example (5) and example (11):

工作吧，找不到；老婆吧，闹离婚；孩子吧，不听话。   (11)

Gongzuo ba, zhao budao; laopo ba, nao lihun, haizi ba, butinghua.

Work ba-particle, can't find; wife ba-particle, have divorce; children ba-particle, disobedient.

(He) can't find a job, and his wife is filing a divorce while his children are disobedient.

In this example, "gongzuo(work)"and "zhaobudao(can't find)" are both intonation units and they are the object-verb relation. But "laopo(wife)" and "naolihun(file a divorce)" are both intonation units but they are subject-verb relation. Same as "haizi(children)" and "butinghua(disobedient)".

In a word, Chinese has a very loose Noun-Verb relationship, they can move comparatively freely in the sentence. The results in table 4 are consist with the Chao's argument that minor sentence is the basic units of Chinese spoken language. For more discussion on this topic, please refer to [43]. These are spoken data. As for written language, we always spotted this discourse phenomenon: a Chinese sentence has several commas inside it. In other words, a complex sentence which included many clauses is very common in Chinese discourse. However, English sentence often have around two commas. If we translate Chinese into English, we usually need to change some commas to stops or add commas in some places. See example (7).

According to these findings, we argue that Chinese discourse has two basic units equal to the clauses in English discourse. They have 3 dimensions. (1) Prosodic features: they exhibit either the flat-postponed or dramatic rising or falling intonation in spoken

language. And can be signify by comma, stop or semicolon etc in written language. (2) Pragmatic features: They all show the interaction intentions or information exchange between the speaker/writer and listener/reader. (3) The minimal units can form the upper units independently, or combine several minimal units to constitute the upper units. The deference lies in the way of the inner structure and the external structure.

In conclusion, the minimal unit of Chinese discourse is "Dou" or clause. It has no special grammatical markers, but has some prosodic features in spoken language which consist with the comma in written language.

4. **Further Discussion.** In this section, we will discuss three important questions: the first one is why we do need to define the basic unit of discourse. To answer this question, we will compare the basic units in English discourse and Chinese discourse.

Take English as an example, the basic units of English discourse is clause. It has two definitions: one is a structure which has a finite verb or several parallel verbs which functions are the same. The other is a structure which has a verb or several parallel verbs which functions are the same. The difference is weather the verb is finite? If we agree with the verb should be finite, such as showing the difference in tense (past/present tense), or the difference is modality, in other words, the formal definition. Then most clauses can be used alone to form a sentence or discourse independently. Because it possess these elements, such as the intention of the speaker (mood), the communication scene (space-time positioning, etc.) associated with dynamic elements involved in the process of discourse changing into a communicative context. Thus we can call these clauses with finite verbs "dynamic grammar basic units". "Clause"which discourse studies uses is with this kind of definition. Under it, the units have different kinds of grammatical relations. Above it, the units have the discourse structure relations, like the role/intention of the speaker, the commutation scene, the changing of the topic, etc. The counting units of the distance which a pronoun anaphora resolves or the length or persistency of the topic chain are usually clauses too. [17]

In sum, the clause in English discourse has two basic definitions: (1) a subject-predicate structures in an unmarked mood. (2) The structures which contain a verb and its arguments and the verb can be finite or infinite. Generally speaking, we usually use definition 1 to determine the clause as the basic unit of the discourse analysis. The clause with an infinitive verb is usually a lower unit, which amount to a phrase. As in example (12), (13) and (14),

The next train to arrive is from Beijing.   (12)

In this example, "the next train to arrive "is a clause with an infinitive verb. We usually do not see this kind of structure as the clause for discourse analysis. We treat this sentence as a basic unit to analysis. This analysis can apply to "be invited to Mary's birthday party"in example (13) and "be eating something "in example (14).

It's a great honor to be invited to Mary's birthday party.   (13)
He seems to be eating something.   (14)

---

[17] More detailed information please refers to [10].

Now let us explore the magic in Chinese discourse. Chinese has no morphology markers, so we can not define "finite verb "grammatically. That leaves us the way to using the more basic semantic or pragmatic angles, such as the role or intention speaker played or expressed and the communication scene, observing all the possible form exhibitions related including phonology, grammar and writing features, and research its similarity. Then we use the forms which fits the similarities to determine the basic units of the Chinese discourse. We have reached the definition of basic units in Chinese discourse in section 3.

More specifically, we believe that the basic units of discourse which all human language shared should have the following features: (1) it should have the cross-level correlation. It must play the multirole of prosodic units, the grammatical units in use, and the grammatical units in inventory. When divided these three kinds of units, they should be roughly the same. (2) It should have certain forms. Prosodic features: it takes the form of the intonation units (IU, intonation unit). In written language it can be represented as commas, semicolons, full stop. Its Grammar role should be subject- predicate structure, finite verb, etc. With these two features of the basic unit can become speaker psychological natural unit and the computer automatic segmentation is relatively easy.

The cross-level units are very important in all aspects of the language study. For instance, the Chinese "character" and English "word "is the basic fundamental grammar unit respectively in their language system from the synchronic perspective. The Chinese syllable/intonation (prosody), Chinese characters (in written language), morpheme (minimum combination of sound and meaning) the units in these three different levels are roughly consistent. As for English, word stress (prosody), text word (in written language), grammar words (the minimal free units to make a sentences) the units belong to these three different levels are roughly the same. Therefore, in modern Chinese, one syllable (speech character) = one combination of sound and meaning (grammatical character) = a square shape[18] (text word). This makes the research on "Zi" (character) is much more important than "Ci" (word). However, in English, the role of word is more important than morpheme. Since English word stress (speech word) = a free minimal units to make a sentence (grammatical word) = letters separated by space (text word).

Discourse is the speech unit produced in the process of using language. It is deeply involved with the context, speaker/writer, the listener/readers pragmatic factors, but meanwhile discourse is generated by the grammar mechanism. The internal small units in discourse have to conform to the rules of grammatical level. Therefore, we need to find the discourse units which are used across the speech (prosody) - grammar - pragmatic level. This basic unit of multi-level will become the interface of the syntactic and discourse.

The second question is how to analysis "*Liushui*" (flowing water) sentence? "*Liushui*" sentence is a unique sentence pattern in Chinese discourse. In the work of [22], he used the term of "Liushu sentence" which means a series of sentence which can be joint together or divided freely just like the flowing of water. He claimed that the clause not

---

[18] Chinese characters are pictographic usually take the form of a square.

the sentence should be seen as a basic unit of Chinese discourse, for it is better adapted to the Chinese context. Because there is a lot of "*Liushui*" sentence in Chinese spoken language, one after another, and it can be break or connects. For example,

他ᵢ 这时已是将近六旬的人，Øᵢ 一表人才，Øᵢ 高个儿，Øᵢ 眉目清秀， (14)
Øᵢ 头发 j 又多又黑，Øⱼ 略带花白，Ø 恰好衬出他那堂堂的仪表。

(a) *Ta ᵢ zheshi yishi jiangjin liuxun de ren, Ø ᵢyibiaorencai, Ø ᵢgaoge er, Ø ᵢmeiqingmuxiu, Ø ᵢtoufa*ⱼ *youduo you hei, Ø ⱼ lue dai huabai, Ø qiahao chenchu ta na tangtang de yibiao.*

(b) He ᵢ this time is nearing sixty de-particle person, Ø ᵢhandsome, Ø ᵢtall, Ø ᵢhaving beautiful features, hair much and black, Ø ⱼa little gray, Ø just add to his elegant appearance.

(c) He was at this time nearing sixty, a tall, elegant man with good features and thick dark hair only sufficiently graying to add to the distinction of his appearance.

This example has 7 clauses which all describes his appearance. It used 6 zero anaphora which the first 4 all point to he while the 5th one means his hair. And the last zero anaphora means all the clause before which is his appearance.

As we mentioned before there is no punctuation in traditional Chinese, so it is very common that there are different interpretations of the ancient novels. Some uses comma while other uses stops. Why there are so many "*Liushui*" sentence in Chinese discourse? As we discussed before, Chinese is lacking of grammatical limitations of clause, such as the agreement of the subject-predicate structure. So the relationship of the clause is very loose. The clauses can be combined in all kinds of relationships and even changing different topic during one sentence. See example (5). The "Dou" or clause is majority in language, and they can form a sentence independently or combine into a more complex sentence. There are no formal markers between sentences except for the pause and intonation units. The connectives can not be used as a criterion for it seldom uses and the relationship between sentences always inferred from the context. Some linguist called this process of guessing the relation between clauses "*Yihe*" (connects by meaning) which is on the contrary of "*Xinghe*"(connects by visible words, such as connectives). The work of [19] introduced a test about the pauses between "*Liushui*" sentences and found that different people who are reading the same text tend to have different places to pause and also the length of the pauses is not the same. The work of [33] introduced a very interesting utterance in spoken language.

老王ᵢ 呢? Øᵢ 又生病了吧! Øᵢ 也该请个假呀! Øᵢ 走不动了吧! Øᵢ 儿子女儿ⱼ 呢?
Øⱼ 上班忙吧? Øᵢ/ⱼ 请个保姆呗! Øᵢ/ⱼ 工资低呀! Øᵢ/ⱼ 先借点呢? 犟脾气一个呀! (15)

(a) *Laowang*ᵢ *ne? Ø*ᵢ *you shengbing le ba! Ø ᵢ Ye gai qingge jia ya-particle! Ø ᵢZoubudong le ba! Ø*ᵢ *Erzi nver ne? Ø*ᵢ *shangban mang ba? Ø ᵢ/ⱼ Qingge baomu bei! Ø ᵢ/ⱼ Gongzi di ya!*

*Ø ᵢ/ⱼ Xian jie dian ne! Ø*ᵢ *qiang piqi yige ya!*

(b) Old Wangᵢ ne-particle? Øᵢ Again sick le-particle ba-particle! Øᵢ Should also ask for leave ya-particle! Øᵢ Walk not move ba-particle! Øᵢ Son and daughterⱼ ne-particle? Øⱼwork busy ba-particle? Øᵢ/ⱼ hire a nanny bei-particle! Øᵢ/ⱼ the low wages ya-particle! Øᵢ/ⱼ first borrow some ne-particle? Øᵢ Stubborn temper ya-particle!

(c) How is Laowang? (He) is sick again! (He) should ask for a leave? (He) can not walk! What about (his) son and daughter? (They) are busy working! (He or they) should hire a nanny! (His or their) wages is low! What about borrow a bit? (He) has such a stubborn temper. [19]

We can understand this example in two ways: the first is a dialogue flow that makes up by a series of independent zero-sentence (no subject or predicate).The other is a person's monologue. Every clause is next clause's trigger. And every clause is the former clause's answer, such as "is ill again," is to question "How is Laowang?" and at the same time triggers the following "should he ask for a leave" and this clause trigger the "he can not walk ". Every clause is the cause of the next clause. However there is no explicit "yinwei"or "youyu" (because), this relationship is drawing from the context. This example is a bit extreme but it illustrated the characteristics of the "*Liushui*" sentence very well. In fact, the reason of the wildly use of "*Liushui*" sentence is the special features of Chinese "*Dou*".

The third question is how can we prove that "*Dou*" is a basic unit native speaker has in their mind? As we discussed before, punctuations are the markers of prosody, so we can prove this through a reading experiment. It is believed that the native speakers may subconsciously use the language sense while reading or listening. One can refer to psychological linguistics research about this hypothesis. We know that in the eye tracking experiments, the places have a fixation point corresponding to advanced cognitive activities, such as people will pause at the end of the clause to pack the information they absorbed. The work of [37] introduced a Dutch reading experiment. If we compare the reading time of Clause1, and Clause2 and Clause1 and Clasues2, the result will be like this: when there is a comma in the end of the first clause reading speed will slow down, but the second small sentence will grow fast. There will be a fixation point at the end of first clause. If there is no comma, the reading of the second clause will slow down. Visible commas are like traffic signals which guide the reader to slow down and processing the information collected before. The second clause is a new unit which needs access to some knowledge reserved in the brain.

From these reading experiment results, we can conclude that "*Dou*" can be used as the units while reading. But unfortunately we didn't see the results of the related Chinese eye-tracking experiment. So we can not say for sure. However, this is definitely the good direction of the follow-up work on this topic. And to get a solid conclusion, the experiment must cover all kinds of combination of different clause.

The goal of this paper is to find the basic unit of Chinese discourse. At last, we reached to a conclusion that basic unit of Chinese discourse is "Dou" or clause which has no special grammatical markers but have certain prosody features which consist with the comma in written language. The follow-up work about this topic should focus on: (1) how many patterns are commonly used for join clause into sentences? (2) The differences between "Dou" and sentence, and their relationship with pragmatic categories in different

---

[19] This clause means "he" (Laowang) will not borrow money from other people because he is too stubborn to do so.

discourse levels.

REFERENCES

[1] Chen, Ping. *Referent Introducing and Tracking in Chinese Narrative.* Unpublished UCLA Ph.D. dissertation, 1986.

[2] Chen & Hongyin Tao. *A Preliminary Study of Patterns of Syntax-Prosody Mismatch in Mandarin Narrative Discourse.* In S. F. Huang, ed., Selected Papers from the Second Conference on Languages in Taiwan. Taipei: Crane Publishing Company, pp.183-205, 1988.

[3] Chafe, W. L. The flow of thought and the flow language. In Givon, T. (Ed.), *Discourse and Syntax of Syntax and Semantics, Vol.12.* New York Academic Press, pp.38-39, 1979.

[4] Chafe, W.L. *Discourse, Consciousness, and Time.* Chicago: The University of Chicago Press, pp.76-78, 1994.

[5] Chu, Chauncey C. *A Discourse Grammar of Mandarin Chinese,* Beijing: Beijing Language and Culture University Press, 1998.

[6] Dafang, Wang. *An Optimizing Model of Anaphora Processing in the Light of Discourse Structure,* Unpublished Peking University Ph.D. dissertation, 2005.

[7] Du. *the Ellipses of Subjects in Chinese Discourse,* Unpublished Peking University Mater Dissertation, 1998.

[8] Fox, B. *Discourse Structure and Anaphora.* Cambridge: Cambridge University Press, 1987.

[9] Fox, B. (Ed.). *Studies in Anaphora.* Amsterdam; Philadelphia: J. Benjamins Pub, 1996.

[10] Givón,T.(Ed.). *Topic Continuity in Discourse: A Quantitative Cross-Language Study.* Amsterdam; Philadelphia: J. Benjamin's Pub, 1983.

[11] Grosz, B., & Sidner, C.. Attention, Intentions and the Structure of Discourse. *Computational Linguistics* no.12, pp. 175-204, 1986.

[12] Grosz, B., Joshi, A., & Weinstein, S. providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21ˢᵗ Annual Meeting of the Association for Computational Linguistics*, pp.44-50, 1983.

[13] Grosz, B., Joshi, A., & Weinstein, S. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, Vol.21, no. 2, pp, 203-226, 1995.

[14] Henk Pander Maat & Ted Sanders. How grammatical and discourse factors may predict the forward prominence of referents: two corpus studies .*linguistics*, Vol, 47, no, 6, pp.1273-1319, 2009.

[15] Halliday M.A.K & R Hasan. *An Introduction to Function Grammar.* Arnold, London, 1994.

[16] Hobbs, J.Coherence and coreference. *Cognitive Science*, no, 3, pp.67–90, 1979.

[17] Hongyin, Tao. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar,* Amsterdam and Philadelphia: John Benjamin, 1996.

[18] Tao H. A Corpus-Based Investigation of Absolutely and Related Phenomena in Spoken

American English. *Journal of English linguistics*, no, pp. 5-29, 2007.

[19] Mingyang Hu & Song Jing. Liushu Ju chu tan. (a Preliminary Study on Liushu sentence). Yuyan jiaoxueyu yanjiu,(*Language Teaching and Linguistic Studies*),no.4,pp.42-54, 1989.

[20] Jing Wang .*The study of semantic-syntactic structure of Chinese text,* Unpublished Peking University Ph.D. dissertation, 2002.

[21] Lao She. *Luotuo xiangzi.*(Rickshaw Boy).Beijing: Haita, 1936.

[22] Lv (吕叔湘). *The Issues of Chinese Grammar Analysi*s, Beijing: Shangwu Press, 1979.

[23] Lu (陆镜光). Lun Xiaoju zai Hanyu Yufa zhong de Diwei (The Discussion of the Role of Clause in Chinese Grammar), *Hanyu Xubao*, no,15, pp.2-14, 2006.

[24] Li, C. N., & Thompson, S. A. Third-person Pronoun and Zero-anaphora in Chinese Discourse. In T. Givón (Ed.) *Syntax and Semantics 12: Discourse and Syntax*. New York: Academic Press, 1979.

[25] Mann, W., & Thompson, S. Relational propositions in discourse. *Discourse Processes,* Vol.9, no,1 ,pp.57-90, 1986.

[26] Mann, W & Thompson, S. Rhetorical structure theory: toward a functional theory of text organization. *Text, Vol.8, no.*3, pp. 243-281, 1988.

[27] Wang （王洪君）.*The Minimal Units of the Modern Chinese Discourse*. Reported in the Chinese Processing Seminar, Beijing Peking University, 2010.

[28] Wendan, Li. Topic Chains in Chinese Discourse, *Discourse processing*, Vol, 37, no.1, pp.25-45, 2004.

[29] Rong Li. *the Multidisciplinary Research about the Thematic Third Person Anaphora in Chinese Texts,* Unpublished Peking University Ph.D. dissertation, 2012.

[30] Passoneau, R. J. & Litman, D. J. Discourse Segmentation by Human and Automated Means. *Computational Linguistics,* no.23, pp. 103-139, 1997.

[31] Passoneau, R. Interaction of discourse structure with explicitness of discourse anaphoric noun phrases. In Marilyn Walker and Aravind K. Joshi, and Ellen Prince (Eds.), *Centering Theory in Discourse*. Oxford: Clarendon Press, pp. 327–358, 1998.

[32] Shoichi Iwasaki & Hongyin Tao. *A Comparative Study of the Structure of the Intonation Unit in English, Japanese, and Mandarin Chinese*, the Annual Meeting of the Linguistic Society of America, Los Angeles, CA. Jan. 9, 1993.

[33] Sheng (沈家煊). On minor sentences and flowing sentences in Chinese: In commemoration of the 120th birthday of Yuen Ren Chao. *Zhongguo yuwen* (Studies of the Chinese Language), no, 5, pp.403-415, 2012.

[34] Song(宋柔). A research on the properties of syntactic relation between P-clauses in modern Chinese, *Zhongwen Xinxi Xuebao(Journal of Chinese Information Processing)*,no, 3.pp.62-68, 1992.

[35] Taso.F. *A functional study of topic in Chinese: The first step towards discourse analysis.* Taipei, *Linguistics*, no.2, pp. 65-93, 1979.

[36] Tsao .F. *Sentence and clause structure in Chinese: A functional perspective*. Taipei, Taiwan, student book, 1990.

[37] Sanders, Ted J. & Noordman, G. M. L. The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *Vol,*29 ,no,1, pp.37-60, 2000.

[38] Sanders, Ted J., Wilbert P. M. Spooren, & Leo G. M. Noordman. towards a taxonomy of coherence relations. *Discourse Processes*, no, 15, pp.1-35, 1992.

[39] Sperber, D & D. Wilson. *Relevance: Communication and Cognitio*n . Beijing: Foreign Language Teaching and Research Press, 2001.

[40] Xing(邢福义)..Lun Xiandai Hanyu Juxing Xitong (The sentence patterns in modern Chinese), in *the Exploration and Research about Grammar*. Vol. 1. Beijing, Peking University Press, 1993.

[41] Xing. *Xiaoju Zhongshu Shuo(The Clause as the Center)*，*zhongguo yuwe(Studies of the Chinese Language),no,* 6, pp.420-428, 1995.

[42] Xu (徐赳赳). Xushu wen zhong de "ta" de huayu fenxi .(*The Discourse Analysis of "ta(He)" in Narratives*)，*zhongguo yuwe(Studies of the Chinese Language),* no,5.pp.325-337, 1990.

[43] Yuan Ren Chao, *A Grammar of Spoken Chinese*, University of California Press, Los Anglesis, 1962.

[44] Zhu(朱德熙).Yuda jiangyi(the Lecture Lotes of Grammar)，*<Zhudeixi Wenji>* Vol.1 Beijing, Shangwu Press, 1999.

[45] Zhu. Yufa Fenxi he Yufa Tixi. (The System of Chinese Grammar Analysis), *<Zhudeixi Wenji>* Vol.2.eijing: Shangwu Press, 1999.