

# Chunks in Information Flow: a Corpus-based Analysis of Legal Discourse

**Abstract:** Chunks promoting discourse development appear in the discourse information flow as they do in the sentences. The present study, guided by Discourse Information Theory (DIT) and discourse information flow, analyses some legal data from “the Corpus for the Legal Information Processing System (CLIPS)”. Starting with information levels, information knots, key words and sharing categories, the paper explores the structure of information chunks and information chunking in the discourse development, and analyses the features of information chunks and their effects on information flow.

**Keywords:** discourse information, information flow, information chunks, tree information structure

---

\***Corresponding author: Jinshi Chen**, Guangdong University of Foreign Studies, China, E-mail: chenjinshi@gdufs.edu.cn.

## 1 Introduction

Chunk is an information-based meaning unit and closely-connected information integration (Gobet et al 2001). Chunk analysis can be made to go deep into the inner of words, for example, ‘anticipation’ can be divided into five chunks ‘an-’, ‘ti-’, ‘ci-’, ‘pa-’ and ‘-tion’. Lexical chunks in a sentence can also be analysed, which are kernel words surrounded by groups of functional words (Abney 1991), such as noun chunks, verb chunks, noun phrase chunks and so on (Zhao & Huang 1999; Zhou et al 2000). The theoretical exploration and practical application of chunk theory involve information processing (Miller 1956; Rubensson & Rudberg 2014), language acquisition and teaching (Ellis 2003; Song 2002), and psychological research (Cowan 2001, 2011; Gilchrist et al 2009). Therefore, chunk has become one of hot topics in language studies with much importance attached to lexical chunks in a sentence, but further research on chunk itself is still of necessity (Huang & Wang 2011). However, few researches on chunks at the discourse level have been conducted so far, particularly from the perspective of discourse information. Thus, this paper, guided by Discourse Information Theory (DIT hereafter) (Du 2007, 2014), focuses on the study of chunks in discourse information flow.

## 2 Theoretical framework

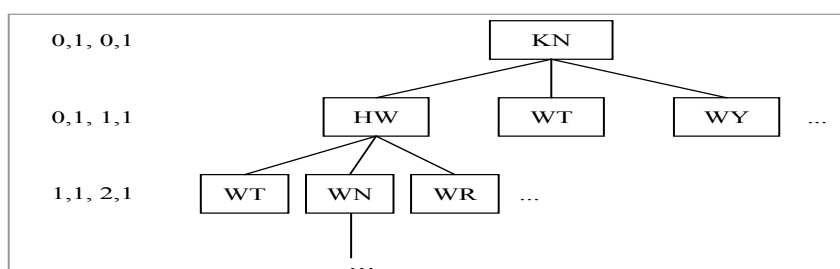
### 2.1 Discourse information theory

Different from the binary approach, e.g. old and new information in discourse (see Halliday 1985) or other nonlinguistic information theories, DIT (Du 2007, 2014) takes information processing as the core of a discourse. Discourse information is based on cognition and expressed through language. Discourse information structure refers to the mechanism in which the overall discourse information system in a discourse is formed by the embodiment of discourse information, basic information units, relationships and connections between information units. Discourse information structure is a tree network with information units at various levels, serving the kernel information in a discourse. Discourse information shares the common information features in that discourse information can be counted, stratified, identified, processed, transmitted and shared (Du 2007; Zhao 2011), which is of convenience to identify, classify, analyse and count discourse information units. From this new perspective, a discourse information unit has been endowed with different values or properties, like information knots, information levels, sharing categories and key words, making it easier to further the exploration of chunks in discourse information flow.

Based on the application of DIT in legal discourse studies (see Du 2007, 2015; Chen 2011, 2015; Ge 2014; Guan 2015), information units are represented by key words, which can be used to analyse the information content in a legal discourse. There are hierarchical relationships between information units in which the information at the lower levels can be developed to that at the higher levels. Those relations are called information knots, labelled by 15 *Wh*- phrases (Du 2007), for example, WT (what thing), WN (when), WF (what fact), WY (why), HW (how), WR (where), WI (what inference), WJ (what judgment), WB (what basis), WA (what attitude) and so on, which actually denote from what aspect subordinate information units specify their superordinate ones in a discourse, such as thing, time, fact, reason... (Chen 2011).

In each discourse, the kernel proposition (KN) and its subordinate information units can be identified, which correspond to information knots and level codes in the tree diagram (see Figure 1). Level codes are composed of level number and position number of the information unit or information knot and its superordinate ones. In Figure 1, the information focus KN is at the top of the tree structure, with the level code '0,1,0,1'. It is developed into subordinate information knots HW, WT and WY at the first lower level with the level codes '0,1,1,1', '0,1,1,2' and '0,1,1,3' respectively. HW has its own subordinate knots WT, WN and WR with the level codes '1,1,2,1', '1,1,2,2' and '1,1,2,3'. '1,1,2,3' means knot WR takes up the third position (tagged as 3) at level 2. It is subordinate to HW at the first position (tagged as 1) at level 1 (Chen 2011: 75-76).

Figure 1 Discourse Information Levels (from Chen 2011: 75)



Information sharing categories reflect the degree to which participants in the communication share the information to some degree for information transmission. The more information the participants share, the fewer details they need to provide in the communication. In DIT, there are six types of sharing categories, including A-events (Known to A, but not B), B-events (Known to B, but not A), C-events (Known to both A and B), E-events (Known to neither), O-events (Known to everyone) and D-events (Known to be disputable) (Du 2007). The use of sharing categories, together with concrete legal context, is helpful to study legal discourse from the cognitive and social aspects in that particular shared knowledge of participants forms the basic mechanism of requests, rejections and even rules (Chen, 2011).

## 2.2 Discourse information flow

Du (2009) believes that the core issue of discourse information flow is information development, information transmission, information flowing conditions and various phenomena in the process. Information development ranges from such local-discourse levels as inner-sentence level and inter-sentence level to global-discourse level. The more levels the discourse has, the more deeply the information develops. Discourse mainly serves as information transmission which refers to information processing between information senders and information receivers, involving a series of factors like participants, transmission purposes, transmission means, information content and so on. Information flow occurs in information development and information transmission, with one important condition as information potential energy. The amount of such potential energy relies on the degree of information surplus between information source and information destination, i.e. the information gap between information surplus and information vacancy. The larger the information gap is, the greater information potential energy is, and the more smoothly the information flows. Information flow takes on such phenomena as information hyperplasia, information loss, slow information, information vortex, etc.

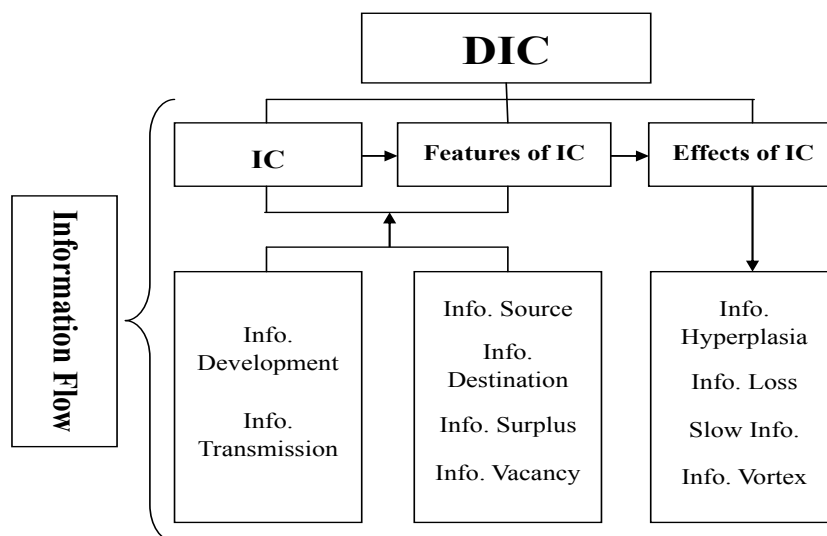
## 2.3 Analytical framework

In accordance with the exploration of lexical chunk (Abney 1991) and phrase chunks (Zhao & Huang 1999; Zhou et al 2000), a working definition in this paper will be

given as follows: Discourse Information Chunk (DIC hereafter) takes a grouped information units as a chunk unit, i.e. a combination of no less than two information units in discourse information flow, and it is a information cluster surrounding the head information unit, with relatively independent meaning and regular flow.

This paper, guided by the theory mentioned above, focuses on the study of DIC at both the inter-sentence level and the discourse level. A framework has been put forward for the chunk analysis in discourse information flow (see Figure 2). Three questions will be discussed in details. (1) How are information chunks formed in discourse information flow? (2) What are the features of discourse information chunks? (3) What is the influence of information chunks on discourse development?

Figure 2 An Analytical Framework for Chunks in Discourse Information Flow<sup>1</sup>



Discourse information flow, including information development and information transmission, is realised under the conditions that there have been clear information source and clear information destination, and that information surplus and information vacancy have already been formed. Accordingly, information chunks will be formed, taking on their various features. Then information chunks, in turn, affect such phenomena as information hyperplasia, information loss, slow information and information vortex in discourse information flow. This paper will, taking oral courtroom discourse and written legal English discourse from ‘Corpus of Legal Information Processing System’ (CLIPS) as examples, conduct a discourse information analysis to answer the aforementioned research questions. The materials from the corpus, having been analysed and labelled in accordance with tree structure of discourse information in DIT (see Appendix), are suitable for the present research.

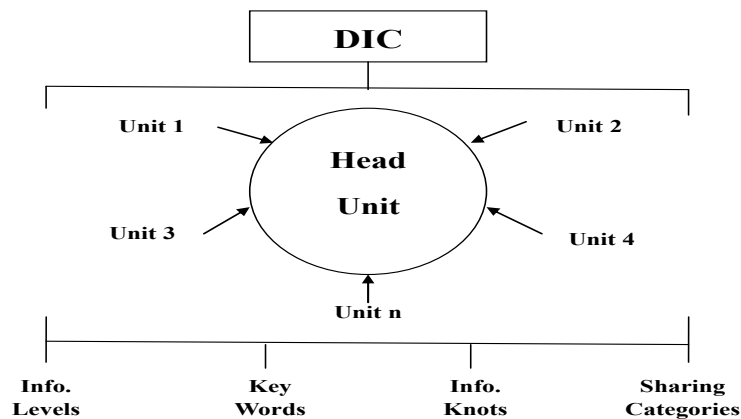
### 3 Formation of discourse information chunks

#### 3.1 Structures of discourse information chunks

<sup>1</sup> IC: information chunk; Info.: information

According to the working definitions of tree structure of discourse information and information chunks, a discourse information chunk consists of a head information unit and some subsidiary ones. Information chunks can be explored via such main properties of information units as information levels, key words, information knots and sharing categories. On those grounds, a diagram for structure analysis of information chunks has been established (see Figure 3):

Figure 3 A Model for the Analysis of DIC Structures



When analysing the macro structures of information chunks, the positions or the deepness of head unit and its subsidiary ones in a discourse, or their hierarchical relationship can be investigated from the perspective of information levels, and detailed contents of head unit and its subsidiary ones will be obtained from the point of key words. As for the micro structures, the logic relationship between information units can be explored from the aspect of information knots, and the sharing degrees of information units and their influence on chunks will be analysed from the angle of sharing categories. For example:

Extract 1

<2,7,3,7,WF,A,implied warranties>Implied warranties are such warranties which do not need to be expressed but which the law implies.<3,7,4,1,WF,A,scope of implied warranties>Some of these types of warranties would include warranties of title, fitness for a particular purpose, and quality or merchantability.<4,1,5,1,WF,A,application of the latter two>Many times the application of the latter two types of warranty depends upon the type of sale and whether the seller is a merchant acting in the course of business.

In Extract 1, three information units form an information chunk ‘Implied Warranties’. The properties of every information unit reflect the basic structure of a chunk. Thus the position <2,7,3,7> in the discourse is the head information unit in the chunk ‘Implied Warranties’, in which the superordinate information of the head unit is the seventh unit at Level 2 in the discourse, the head unit itself is the seventh unit at Level 3. The two subordinate information units of the head are <3,7,4,1> and <4,1,5,1>, belonging to the first ones of Level 4 and Level 5 respectively, which means that there exists a hierarchical relation, i.e. the superior and the subordinates.

Therefore, in this chunk three-level units have been formed, with the information becoming deep level by level. From the definition to the classification and then to the application, the three key words <implied warranties>, <scope of implied warranties> and <application of the latter two> make the contents become increasingly concrete and specific. As this discourse is a written introduction to some legal knowledge, the three information knots in this chunk are all <WF>, which refers to some facts. Without any particularly complex logic among them, only an introduction of ‘implied warranties’ has been made to readers. Sharing categories in this chunk is <A> (known to the author himself) with quite low information sharing degree (Du 2007, 2014), which means that the three information units all are new information and the author hopes that they can be digested and absorbed by the readers.

### 3.2 Chunking of discourse information

During the process of information transmission and information development, information destination exists due to some communicative needs. One party in the interaction consciously integrates scattered information units into a larger meaningful one, i.e. an information chunk, which gradually reaches the state of information surplus to prepare for the flow of surplus information to the information vacancy. That is the process of information chunking. Chunking is a dynamic process to adjust or organise some new information, and a process of activating chunks. For example:

Extract 2

审判长: <3,59,4,11,WA,b,是否  
有意见>被告人是否有意见?

被告人: <3,59,4,12,WA,b,质证  
意见>有, <5,1,6,1,WF,b,不是  
实情>她说的不是实情,  
<4,12,5,1,WI,b,不合理>不是  
很合理。<6,1,7,1,WF,b,知道姓  
名>她知道我真实姓名。  
<6,1,7,2,WF,b,不在那天>不是  
在 1 月 20 日。<5,1,6,2,WF,b,  
不真实>还有她说的我母亲和  
爷爷的情况, 不是真实的。

审判长: <3,59,4,13,WA,b,是否  
有意见>辩护人是否有意见?

Judge: <3,59,4,11,WA,b, any  
objections>  
Defendant, do you have any objections?

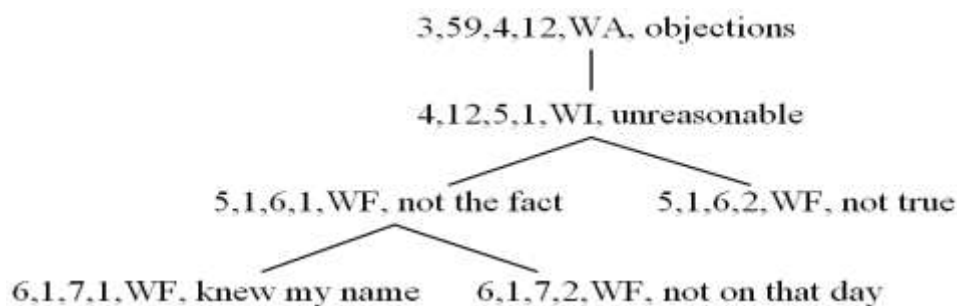
Defendant: <3,59,4,12,WA,b,  
objections>  
Yes, <5,1,6,1,WF,b, not the fact> what she  
said is not the fact, <4,12,5,1,WI,b,  
unreasonable> and not reasonable.  
<6,1,7,1,WF,b, knew my name> She knew  
my name. <6,1,7,2,WF,b, not on that day> It  
was not on January 20. <5,1,6,2,WF,b,  
unreal> And what she said about my mother  
and grandfather was not true.

Judge: <3,59,4,13,WA,b, any  
objections>  
Attorney, do you have any objections?

Extract 2 represents the questions and responses between the judge and the defendant in the phase of burden of proof and rebuttal. Due to the judge’s request for the opinion of evidence challenging, here the defendant’s answers are all <b> in terms of sharing categories, which means that the answers are the information only known to the defendant himself and reflect the information gap between the judge and the defendant, thus pushing forward the information flow and the courtroom communication (Du 2009). Since the defendant finishes his answers with six information units, information chunking of ‘Objections’ is realised through the

distribution of information levels and information knots (see Figure 4). Information flows from its information source, i.e. the state of information surplus, to its destination to fill in the judge's information vacancy. Thus the courtroom communicative objective has been achieved.

Figure 4 Information Chunking of 'Evidence Challenging'



In Figure 4, information chunking of defendant's 'Objections' is the process of information reasoning from the bottom information units to the top ones. Among the defendant's answers, the two lowest fact information units <WF> are at Level 7, which has concretised the detailed content of their superordinate information <5,1,6,1,WF,b, not the fact>. Together with an information unit <5,1,6,2,WF,b, unreal> at Level 6, the two fact information units provide the grounds for the inference information unit <4,12,5,1,WI> at Level 5, which proves unreasonableness of the evidence 'what she said' given by the prosecutor. The <unreasonable> inference information unit <WI> is just the prerequisite for <Objections>, so the logical reasoning from facts to inference and then to attitude (WF-WI-WA) has been established. Therefore, the five information units at lower levels serve for the head information unit <3,59,4,12,WA, objections>, and finally the information chunking of 'Objections' has been completed.

## 4 Features of discourse information chunks

### 4.1 Integrity and relative independence

Wu (1999) holds an opinion that chunks constitute thinking units, the integrity of chunks means that elements of a chunk are closely connected, and some chunks appear as a whole. Data analysis shows that DIC also boasts the feature of integrity, i.e. the information units inside a chunk, closely related to each other, are indispensable parts of the whole. For example:

Extract 3

审判长: <2,19,3,18,WT,A,  
原告举证>首先由原告举  
证。  
原告: <3,18,4,9,WB,A,证据

Judge: <2,19,3,18,WT,A, evidence submitted  
by Plaintiff> Plaintiff, submit the evidence first.  
Plaintiff: <3,18,4,9,WB,A, Evidence I> The  
evidence written by Yang himself and a working

一>杨某自己书写的证据、  
 \*\*区人民政府网站工作文  
 件，证明.....<3,18,4,10,WB,A,证  
 据二>第二份证据工资表，  
 证明.....<3,18,4,11,WB,A,  
 证据三>第三份证据，范某  
 的证言，证明.....  
 原告：<3,18,4,12,WB,A,证  
 据四>第四份证据，劳动合  
 同书，证明.....  
 原告：<3,18,4,13,WB,A,证  
 据五>第五份证据，调查笔  
 录，证明.....  
 审判长：<2,19,3,19,WT,A,  
 被告举证>下面由被告举  
 证。

document from the website of \*\*District  
 government, which prove  
 that.....<3,18,4,10,WB,A, Evidence II> The  
 second evidence is the pay sheets, which  
 prove..... <3,18,4,11,WB,A, Evidence III> The  
 third evidence, Fan’s testimony, proves...  
 Plaintiff: <3,18,4,12,WB,A, Evidence IV> The  
 fourth evidence, the labour contract, proves.....  
 Plaintiff: <3,18,4,13,WB,A, Evidence V> The  
 fifth evidence, the record of previous  
 investigation, proves.....  
 Judge: <2,19,3,19,WT,A, evidence submitted by  
 Defendant> Next, Defendant, submit your  
 evidence.

In Extract 3, what the plaintiff presents and states has realised the chunking of ‘Evidence Submitted by Plaintiff’. All the elements of the chunk include the head unit <2,19,3,18,WT,A, evidence submitted by Plaintiff>, five subordinate information units and the omitted parts. As the basis of those proofs, the five information knots <WB>, followed by some main points proved by the evidence respectively, have built the contents of ‘evidence submitted by Plaintiff’, which highlight the indispensability of each element in the chunk. Meanwhile, the five information units are parallel to each other and are developing in balance, of which all the sharing categories are information <A>, i.e. only known to the plaintiff. Therefore, the continuity and the integrity of information flow are reflected in the phase of evidence submitted by plaintiff in court.

Since it is an aggregation composed of some interconnected information units, a DIC is relatively independent (Yang et al 1999). In Extract 2, the information chunk ‘Objections’ is a relatively independent process of logical reasoning from facts to inference, and then to attitude (WF-WI-WA). The chunk is drawn forth by the judge’s question ‘被告人是否有意见? (Defendant, do you have any objections?)’ and is ended with the judge’s question ‘辩护人是否有意见? (Attorney, do you have any Objections?)’ The three information chunks, ‘Objections’ here, the preceding ‘Evidence Submitted by Plaintiff’ and the following ‘Evidence Submitted by Defendant’ are separate from each other. But the three are of the main contents of the phase ‘burden of proof and rebuttal’ in civil court. Therefore, such independence of chunks is relative. In Extract 3, a relatively independent chunk ‘Evidence Submitted by Plaintiff’ is composed of a serial of evidence and some key points proved by the evidence, which begins with the judge’s instructive words ‘首先由原告举证。(Plaintiff, submit the evidence first.)’ and ends with ‘下面由被告举证。(Next, Defendant, submit your evidence.)’. This chunk and the following chunks ‘Evidence Submitted by Defendant’ and the like are part contents of the phase ‘burden of proof



and rebuttal' in civil court. Thus the independence of the chunk is also relative. However, although different chunks are independent to each other, both the relatively independent information chunks and the relations among them will be taken into consideration when a certain discourse is being analysed so as to explore all kinds of inner relationships in the discourse.

## 4.2 Hierarchy and dynamism

Hierarchy of a DIC refers to the mutual relations between different chunks. A chunk at a higher level leads to those at lower levels and then to those at lowest levels, thus forming a tree structure with different levels that is similar to lexical chunks in a sentence (Abney 1991). Then a complete discourse has been constructed. For example:

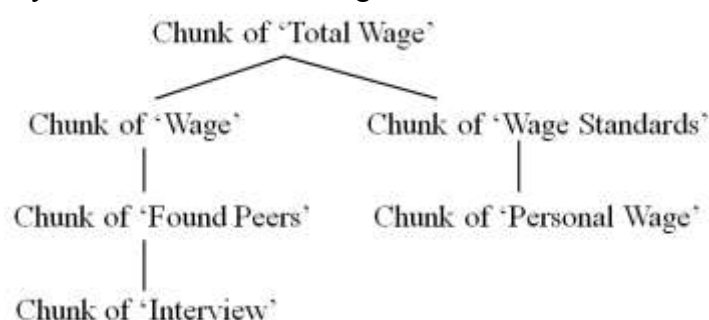
Extract 4

<p>原告：第三份证据..... &lt;3,18,4,12,WT,工资包干&gt;证明所谈工资的包干内容。.....</p> <p>审判长：&lt;3,20,4,21,WT,b&gt;你出庭要证实什么？</p> <p>证人：.....&lt;4,22,5,20,WF,b,工资待遇&gt;工资待遇是4个人工资大包干，1万元，&lt;5,20,6,6,WF,b&gt;每周休一天，&lt;5,20,6,7,WF,b&gt;有年假。&lt;5,20,6,8,WT,b,找同行&gt;我找了几个同行，&lt;5,20,6,9,WT,b&gt;根据这个待遇去找人.....&lt;6,9,7,5,WT,b,面试&gt;我对他们进行了初步面试，&lt;6,9,7,6,WT,b&gt;交代了工资待遇及简单的情况.....</p> <p>审判长：&lt;4,24,5,23,HW,b,工资标准&gt;工资标准是怎么约定的？</p> <p>证人：&lt;4,24,5,24,HW,b,工资标准&gt;酒店定的4个人工资共1万元，&lt;5,24,6,11,WF,b&gt;什么都包括。</p> <p>.....</p> <p>审判长：&lt;4,24,5,29,WF,b,个人工资&gt;你的月工资？</p> <p>证人：&lt;4,24,5,30,WF,b,个人工资&gt;4000元。.....</p>	<p>Defendant: The third evidence.....&lt;3,18,4,12, WT, total wage&gt;proves what I was responsible for the negotiated total wage.</p> <p>Judge: &lt;3,20,4,21,WT,b&gt;What are you going to prove in court?</p> <p>Witness: ..... &lt;4,22,5,20,WF,b, wage&gt;The total wage was RMB 10000 for four people who were responsible for what were specified in the contract. &lt;5,20,6,6 WF, b&gt; They had one day for holiday a week, &lt;5,20,6,7,WF,b&gt;and an annual holiday. &lt;5,20,6,8,WT,b, found peers&gt;I found several peers, &lt;5,20,6,9,WT,b&gt;according to the treatment just mentioned..... &lt;6,9,7,5,WT,b, interview&gt;I carried out a first interview on them, &lt;6,9,7,6,WT,b&gt; and introduced the wage and gave some basic information to them.....</p> <p>Judge: &lt;4,24,5,23,HW,b, wage standards&gt;What are the wage standards?</p> <p>Witness: &lt;4,24,5,24,HW,b, wage standards&gt; The four people's wage was RMB 10000 in total, &lt;5,24,6,11,WF,b&gt;with everything included.</p> <p>.....</p> <p>Judge: &lt;4,24,5,29,WF,b, personal wage&gt; What's your wage a month?</p> <p>Witness: &lt;4,24,5,30,WF,b, personal wage&gt; RMB 4000.....</p>
---	--

Extract 4 demonstrates that the third evidence has activated the information chunk 'Total Wage' in the phase of burden of proof in a civil trial. Among its subordinate information units, all the sharing categories are information <b>, which is known to the counterpart in the communication. There exists an information gap

between the witness' information source and the judge's information destination, making the information flow continuously. Through several rounds of question and response, information vacancy has been filled, meanwhile another five information chunks have also been activated. The head information units of the five chunks are <3,18,4,12, WT, total wage>, <4,22,5,20,WF, B, wage>, <5,20,6,8,WT,b, found peers>, <6,9,7,5,WT,b, interview>, <4,24,5,24,HW,b, wage standards> and <4,24,5,30,WF,b, personal wage>, which contains many subordinate information units respectively. The hierarchy of all the information chunks is clearly demonstrated, i.e. led by the head information chunk 'Total Wage', an information chunk cluster (see Figure 5) has been shaped like a chunk tree with four levels ---- the information units at Levels 3, 4, 5 and 6.

Figure 5 Hierarchy of the Chunk 'Total Wage'



In addition, as the arrows show in Extract 4, the information knots are constantly changing as information flow requested, due to which some new information chunks are continuously being formed. 'Total Wage' is the upper information chunk activated by the plaintiff's words, and the information knot of its head unit is <WT>, signifying the content being proved. This starts the judge's expectation to ask the witness with a question '你出庭要证实什么? (What are you going to prove in court?)', aiming to ask whether the witness is able to prove 'Total Wage'. So the information knot is also <WT>. Along with the information flow of witness' response, three other information chunks have been activated level by level by the information units <WF, wage>, <WT, found peers> and <WT, interview>. When these answers still cannot meet the needs of the chunk 'Total Wage' at the upper level and the information gap still exists, the judge activates the information chunk 'Wage Standards' with the information knot <HW> signifying the means, then asks further questions with the fact information <WF> to activate the information chunk 'Personal Wage' at the lower level. It is obvious that each independent information chunk, with the discourse information flow, will change its information levels and information knots constantly so as to activate new information chunks and to push forward the discourse. This shows the dynamism of information chunks. (Yang 1999; Wu 1999; Niu & Lü 2005).

In the light of the analysis above, it is known that information chunks at the discourse level, with the same features as lexical chunks in sentences and alphabetic chunks in words, possesses many features which become prominent in the discourse development. Certainly, these features, in turn, make information chunks exert some

influence on discourse information flow.

## 5 Influence of chunks on discourse information flow

### 5.1 Information chunks and information hyperplasia

In the discourse information flow, although the development of information hyperplasia is often restricted, speakers or authors sometimes will strive for opportunities to develop hyperplastic information (Du 2009) with certain purposes. When it develops to a relatively complete one, hyperplastic information will be chunking, exerting influence on the information flow. As a consequence, the information hyperplasia will be stopped, or be restricted for it is unacceptable by listeners, or even be truncated (Du 2009). For example:

#### Extract 5

审判长: .....<4,31,5,43,截车目的>当时你截车的目的是干什么?

被告人: <4,31,5,44,拿刀>因为我不是拿把刀从巷子出来嘛, <5,44,6,25,害怕>我就是害怕, <5,44,6,26,吸毒>吸毒, <5,44,6,27,感觉要被害>感觉有人要来害我。 <4,31,5,45,拿刀>然后就拿把刀, <5,45,6,28,不准备害人>不准备害人, <5,45,6,29,防备用>是拿来防备用的, 然后▲

审判长: ▼<4,31,5,46,截车目的>那又为什么要截车呢? .....

Judge: <4,31,5,43, purpose of carjacking>Then what was your purpose of carjacking the car?

Defendant: <4,31,5,44, took a knife>Because I came out of the alley with a knife. <5,44,6,25, very scared>I was just very scared. <5,44,6,26, took drugs>I took drugs, <5,44,6,27, felt to be killed>and I felt that someone would kill me. <4,31,5,45, took a knife> Then I took a knife, <5,45,6,28, not to kill someone> and I was not going to kill someone <5,45,6,29, protect myself> but to protect myself, then ▲

Judge: ▼<4,31,5,46, purpose of stopping a car> So what did you stop a car for? .....

In Extract 5, the defendant fails to give a straight answer to the judge's question <purpose of carjacking>. Instead, his answers produce several hyperplastic information units, like <took a knife>, <very scared>, <took drugs>, <felt to be killed>, <took a knife>, <not to kill someone>, and <protect myself>, which are at superordinate levels or subordinate ones. Among those hyperplastic information units, the two information units of <took a knife> are at the fifth level, i.e. <4,31,5,44> and <4,31,5,45>, the same level as that of the judge's two questions. The rest of the defendant's answers, as the subordinate information of those two information units, are all at the sixth level. Therefore, an information aggregation has been created, with the head information unit <took a knife> surrounded by its subordinate information units. The action 'Took a Knife' has been chunked. Although the word '因为 (because)', from the angle of language surface, seems establishes the surface connection between this information chunk and the judge's question, the judge

eventually interrupts the defendant's answer, showing that the judge regards the defendant's information as hyperplasia. According to the later interrogation in this trial, it is because of the defendant's fear of being arrest for taking drugs that contributes to his carjacking, and the information chunk 'Took a Knife' is the means of his carjacking. So the hyperplastic information for the local discourse may be transferred into non-hyperplastic information in the global discourse (Du 2009).

## 5.2 Information chunk and information loss

Information loss refers to some information which is valuable to a certain participant in the communication secedes from the information flow because the information has not been processed effectively by that participant (Du 2009). Since information chunks are integrated (Wu 1999), the phenomenon of information loss will appear if relevant information chunks cannot be activated because necessary information has not been produced in the development of discourse information. For example:

### Extract 6

审判员：<有无新意见>李\*\*，你有新的意见没有？

上诉人：有一些。他说，多么，我把多么看管，这，我不赞成。我没有打▲

审判员：▼<已说过>这个意见你刚才说过，<有无新意见>你还有新的意见没有？

上诉人：<晚上出去>新的意见是在那个我们家\*\*村，那天晚上出去，<打懵>有个人弄个棍把她打懵以后，<背出背回>背出去，背出去又把她背回来，.....

▲  
审判员：▼<已说过>这都叙述过了..... <发表新意见>辩护人发表新的意见。

Judge: <any new opinions>Li\*\*, do you have any new opinions?

Appellant: Yes. He said, how, how I kept watch on her. That, I don't agree with that. I didn't stun her▲

Judge: ▼<already said that>You have already said that. <any new opinions>Do you have any new opinions?

Appellant: <went out that night>My new opinion is that in our Village\*\*, she went out that night. <stunned her>Somebody stunned her with a stick, after that, <carried her out and then returned>he carried her out and then returned.....

▲  
Judge: ▼<already said that>You have already described that ..... <deliver new opinions>Attorney, deliver your new opinions.

Extract 6 is some of the questions and responses in the second half of courtroom debate. The judge asks the appellant whether he has any new opinions, but the appellant's answer is not new opinions but the content having been expressed previously in the first half of courtroom debate. After the judge interrupted the appellant with the information unit <already said that> and repeats the question, the information units transmitted by the appellant like <went out that night>, <stunned her> and <carried her out and then returned> are also old information. As a result, the information chunk 'New Opinions' cannot be activated by the appellant's words and the effective control of information flow in a trial (Pan & Du 2011) fails to be realised,

so the judge has to interrupt the appellant with another information unit <already said that> and gives up asking the appellant again. Instead, the judge directs the attorney to deliver new opinions. All these make the judge's interrogation information <any new opinions> secede from the information flow just because 'New Opinions' has not been chunked eventually.

### 5.3 Information chunk and information slow information

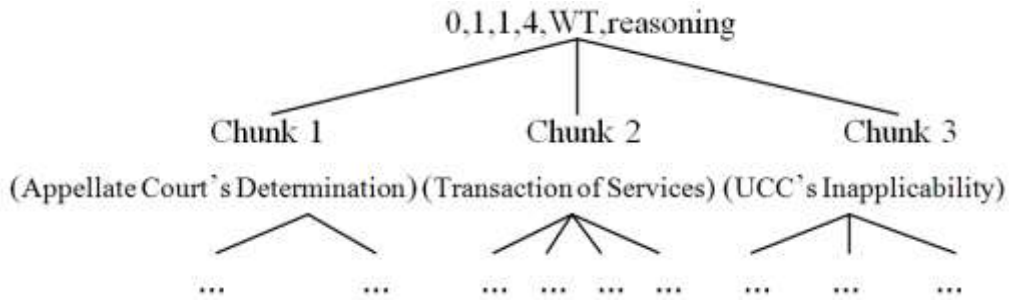
Du (2009) argues that sometimes the information flows smoothly with little backflow, few elements and simple information exchanges in the development of a discourse. That's the phenomenon of slow information. In the process of information flow, information groups may be chunked, with relatively simple inner structures and few subsidiary information units. But many parallel chunks may be activated constantly. The dynamism of chunks pushes discourse information flowing forward slowly and smoothly. For example:

#### Extract 7

<0,1,1,4,WT, reasoning>Reasoning: <1,4,2,9,WF, appellate court's determination>First, the state appellate court determined mixed transactions should be analyzed in terms of the transaction's dominant thrust, ... <2,9,3,6,WI>all of the contract should fall within Article 2 of the UCC.... <1,4,2,10,WF,clear-cut>Second, in this case, the transaction here is clear-cut.<2,10,3,8,WF> DPS was retained to design, develop and implement an electronic data processing system to meet Smith's specific needs not selling hardware to Smith.... <2,11,3,12,WF>Third, in this transaction, it is the skill and knowledge of the programmer which is being purchased in the main, ... <1,4,2,11,WJ,UCC not applicable>Thus, the provisions of the UCC do not apply.

Extract 7 is the reasoning part in a legal case brief. From the main information flow of the whole discourse, legal reasoning represents one of the main contents in a legal case brief, with fluency and simplicity as the key factors in evaluating the case analysis (Li 2008). During the process of reasoning here, three information chunks 'Appellate Court's Determination', 'Transaction of Services' and 'UCC's Inapplicability' have been formed in succession, serving their superordinate information unit <0,1,1,4,WT, reasoning> in details. These chunks separate from each other and develop in parallel and in balance. And the subsidiary information units of the three chunks are mainly at Level 3 without subordinate information. Since the reasoning is simple and discourse information flows smoothly (see Figure 6), the quality of case analysis has been guaranteed.

Figure 6 A Series of Information Chunks in Legal Reasoning



### 5.4 Information chunk and information vortex

In discourse information development and transmission, sometimes information will be processed intensively and information flows lingering around a certain target and the focus concentrated by the relevant communicative parties, resulting in information vortex (Du 2009), which is closely related to information chunking and chunk levels. In Extract 4, with the judge's further interrogation on Evidence III, the defendant enriched the new information continuously. Multi-level information chunks have gradually formed an information vortex there. Some chunks at lower levels have continuously been formed more and more deeply, and then the chunks at different levels have formed nested structures, with the core information of each chunk flowing to the information chunk 'Total Wage' at higher level. Then the function of Evidence III has been highlighted. Another example is as follows:

#### Extract 8

<p>审判员: &lt;2,6,3,32&gt;其他还有啥(上诉理由)没有了?</p> <p>上诉人: .....</p> <p>审判员: &lt;3,33,4,53&gt;这不是严格看管嘛, &lt;3,33,4,54&gt;刚才给你总结过了嘛。</p> <p>上诉人: &lt;2,6,3,34&gt;还有, &lt;3,34,4,55&gt;她晚上经常出去, 经常出去.....&lt;4,55,6,30&gt;她每天晚上都是一点多到二、两点多, 四点多她过来, &lt;4,55,6,31&gt;她出去干啥, 问她, &lt;4,55,6,32&gt;她说“你不要管我干啥!”.....&lt;4,55,6,36&gt;每天晚上出去, 都是一两点出去。</p> <p>审判员: &lt;2,6,3,35&gt;还有没有了?</p> <p>上诉人: &lt;2,6,3,36&gt;还有, &lt;4,56,5,37&gt;我们在新疆摘那个棉花, &lt;3,36,4,56&gt;也经常出去, &lt;4,56,5,38&gt;晚上都是半夜出去, &lt;4,56,5,39&gt;不知干啥。</p> <p>审判员: &lt;2,6,3,37&gt;有没有了?</p> <p>上诉人: &lt;2,6,3,38&gt;还有, &lt;3,38,4,57&gt;我们那晚上, 问她呢, &lt;3,36,4,58&gt;她不</p>	<p>Judge: &lt;2,6,3,32&gt;Do you have any other grounds for appeal?</p> <p>Appellant: .....</p> <p>Judge: &lt;3,33,4,53&gt;<u>That is to have kept watch on her?</u> &lt;3,33,4,54&gt;and I have just summarised for you.</p> <p>Appellant: &lt;2,6,3,34&gt;And, &lt;3,34,4,55&gt; <u>She often went out at night</u>, very often..... &lt;4,55,6,30&gt;<u>She came here at about one o'clock, two o'clock, four o'clock every night.</u> &lt;4,55,6,31&gt;<u>I asked what she went out for,</u> &lt;4,55,6,32&gt;<u>but she said, "it's none of your business!"</u>..... &lt;4,55,6,36&gt;<u>She went out every night,</u> always at one or two o'clock.</p> <p>Judge: &lt;2,6,3,35&gt;Any other thing?</p> <p>Appellant: &lt;2,6,3,36&gt;Yes. &lt;4,56,5,37&gt;When we picked cotton in Xinjiang, &lt;3,36,4,56&gt;<u>she also often went out,</u> &lt;4,56,5,38&gt; <u>always went out at midnight.</u> &lt;4,56,5,39&gt;I don't know what she had done.</p> <p>Judge: &lt;2,6,3,37&gt;Any other things?</p> <p>Appellant: &lt;2,6,3,38&gt;Yeah. &lt;3,38,4,57&gt;On that night, I asked her, &lt;3,36,4,58&gt;<u>but she didn't tell me what she had done.</u> &lt;3,36,4,59&gt;<u>She said, "it's none of your</u></p>
---	---

跟我说去干啥， <3,36,4,59>她说：“你 business!” <3,36,4,60>I asked her, but she did not say  
 不要管我！” <3,36,4,60>我问她呢，她 anything. <3,36,4,61>She said: “it’s none of your  
 不说， <3,36,4,61>她说：“你管我干啥 business!” .....  
 不干啥呢！”.....

In Extract 8, a typical information vortex has been exemplified by the hierarchical information chunk structures at two levels. At the macro level, the chunk ‘The Victim Not Kept Watch On Strictly’ at the higher level with three subordinate chunks and some subordinate information units (see Figure 7) has been activated by the judge’s questions and the appellant’s responses. At the micro level, the core contents of each chunks point to the upper chunk ‘The Victim Not Kept Watch On Strictly’ from the lower information levels (see the arrows in Extract 8). In other words, the information flows from Levels 5 and 6 to the head information unit <3,33,4,53> at Level 4. Although the appellant answers the judge’s question for three times, he always says that he did not keep watch on the victim strictly but the victim herself frequently went out at night without telling the appellant about that. Actually there exist subtle differences between the surface meanings of the appellant’s answers, but the contents all belong to the same chunk ‘Went Out At Night’, repeating virtually the grounds of appeal ‘not keep watch on her strictly’ which has been summarised by the judge. Thus in the discourse information flow, all subordinate information units are developing and transmitted around the core of the information vortex ‘the appellant denied that he kept watched on the victim strictly’.

Figure 7 Hierarchy of the Chunk ‘The Victim Not Kept Watch On Strictly’



## 6 Discussion and implications

Given discourse information studies attach great importance to the analysis at both macro and micro levels, the characteristics of legal discourse and its influence on information flow have been interpreted effectively. The research on chunks has been changed from the aspect of words, phrases and sentences into the discourse studies at both macro and micro levels.

The bottom-up analysis shows that discourse information chunking is a dynamic process, in which surplus information flows to the information vacancy and necessary information chunks are activated. In the Chunk ‘Total Wage’ in Extract 4, the

information sharing category has always been kept as <b>, which is known to the appellant and unknown to the judge. So the judge's information need results in the information changes, in which the information knots <WT> or <HW> have been converted to the knot <WF>. Then a variety of different chunks are activated by different information combinations, promoting the information flow and the development of discourse.

In the light of the top-down analysis, a head information unit serves as the core in each chunk, with subsidiary units developing level by level. Meanwhile some information chunks contain more subsidiary units than others do, and some chunks are surrounded by more subordinate chunks. In Extract 1, the head information unit in the chunk 'implied warranties' is developing constantly, forming the top-down hierarchy with three levels. And in Extracts 4 and 8, the levels of information chunks have taken on a top-down trend with the information becoming more and more complicated and the content deeper and deeper.

Although the studies on information chunk can be made to explore legal language as above, it will also be practically extended to the relations between chunks and language teaching in that information processing is always involved in the classroom context (Du 2015). The exploration and application of discourse information chunks will promote the expansion of chunks in language teaching for the enhancement of the efficiency of language teaching and language learning. For example, to teach the legal English text in Extract 1, a model for legal reading can be constructed based on chunk studies. Due to integrity and relative independence of information chunks, learners' ability will be improved in such fast reading processes as skimming and scanning; according to hierarchy and dynamism, learners' ability will be enhanced in their logical analysis and micro information access.

In addition, in terms of the features of information chunks, other pedagogical implications for teaching listening, speaking, writing and translation can also be ramified. For example, in interpreting classroom, the efficiency of interpreting training will be developed in that the information content through short-term memory or shorthand information capacity based on integrity and relative independence of information chunks. In teaching writing, learners' ability can be cultivated for their hierarchical, coherent and logical paper design based on hierarchy and dynamism of information chunks.

## **7 Conclusion**

This paper demonstrates that information at the discourse level flows in the same way as that at the inner-sentence level, with the information chunks pushing the discourse forward. A DIC, gradually being formed in discourse information flow, is an aggregation of some information units with a head unit, serving as the core surrounded by its subsidiary units. Through the analysis of information levels, key words, information knots and sharing categories of information units, it is found that information chunks have such features as integrity, relative independence, hierarchy and dynamism, which have an effect on information hyperplasia, information loss,



slow information and information vortex in the discourse information development. This paper has further explored Discourse Information Theory and discourse information flow. However, for the effective processing of discourse information, it is still necessary to study constantly the application of information chunks in various kinds of discourse and to compare discourse information chunks between different languages.

## Acknowledgments

I would like to thank the anonymous reviewers for their inspiring and helpful comments. This paper is substantially supported by a grant from “College English ESP Teaching Team” in Guangdong University of Foreign Studies.

## References

- Abney, S. 1991. Parsing by Chunks. In R. Berwick, S. Abney and C. Tenny (eds.) *Principle-Based Parsing*, 1-18. Kluwer Academic Publishers.
- Cowan, N. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences* (1), 87-114.
- Cowan, N. 2011. The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia* 49(6), 1401-1406.
- Chen, J. 2011. *The Construction of Prosecution-Defense-Judge Relationship: A Frame Analysis of Judges' Courtroom Discourse Processing*. Beijing: Science Press.
- Chen, J. 2015. Self-Assessment for Legal English Writing: From the Perspective of Information Output. *International Journal of Legal English* 3(2), 57-69.
- Du, J. (2007). A study of the tree information structure of legal discourse. *Modern Foreign Languages* (Quarterly). 30(1), 40-50.
- Du, J. 2009. Information Flow in Discourse. *Journal of Foreign Languages*. 32(3), 36-43.
- Du, J. 2015. Application of multimodal information corpus techniques in legal English teaching. *International Journal of Law, Language and Discourse* 2(4): 19-38.
- Du, J. 2014. *On Legal Discourse Information*. Beijing: People's Publishing House.
- Du, J. (eds.). 2013. *The Course of Discourse Analysis*. Wuhan: Wuhan University Press.
- Ellis, N. C. 2003. Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In C. J. Doughty & M. H. Long (Eds) *The Handbook of Second Language Acquisition* (pp. 63-103). Malden, MA: Blackwell Publishing Ltd.
- Ge, Y. 2014. Resolution of conflict of interest in Chinese civil court hearings: a perspective of discourse information theory. *The International Journal of Speech, Language and the Law* 21(1): 63-68.
- Gilchrist, A. L., Cowan, N. & Naveh-Benjamin, M. 2009. Investigating the childhood development of working memory using sentences: New evidence for the growth of chunk capacity. *Journal Of Experimental Child Psychology* 4(2), 252-265.
- Gobet, F., Lane, P.C., Croker, S., Cheng, P.C., Jones, G., Oliver, I. & Pine, J.M. 2001. Chunking Mechanisms in Human Learning. *Trends in Cognitive Sciences* (6), 236-243.
- Guan, X. 2015. Potential Speaker-Discriminating Power of Speaking Style: Application of Discourse

- Information Analysis to Forensic Speaker Recognition. *International Journal of Law, Language and Discourse* 5(1): 38-65.
- Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold & Co.
- Huang, Y. & Wang, H. 2011. A survey of the researches on L2 chunks in China. *Foreign Language World* (3), 74-81.
- Li, K. 2008. Composing and Translating Conditional Clauses in Legal Texts. *Chinese Translators Journal* (4), 71-77.
- Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* (63), 81-97.
- Niu, S. & Lü J. 2005. The Chunking Theory in Human Learning. *Journal of Chongqing University* (Social Science Edition) 11(1), 96-99.
- Pan, X. & Du, J. 2011. Information Flow of Process Control in Courtroom Question and Response. *Journal of Foreign Languages* (2), 56-63.
- Rubensson, E. H. & Rudberg, E. 2014. Chunks and Tasks: A programming model for parallelization of dynamic algorithms. *Parallel Computing* (40), 328-343.
- Song, D. 2002. Chunking Effect and Its Implications for Foreign Language Teaching. *Foreign Languages and Their Teaching* (9), 23-25.
- Wu, H. 1999. A Model for Chunk Theory of Creation and Discovery. *Studies in Dialectics of Nature* 15(6), 20-24.
- Yang, Z. et al. 1999. *Psychology of Memory*. Shanghai: East China Normal University Press.
- Zhao, J. & Huang, C. 1999. A Transformation-based System for the Identification of Base-NP in Chinese. *Journal of Chinese Information Processing* (2), 1-7.
- Zhao, J. 2011. *On the Information Structure and Its Linguistic Realization in Legal Discourse: A Chinese-English Contrastive Discourse Analysis*. Beijing: Science Press.
- Zhou, Q, Sun, M & Huang, C. 1999. Chunk Parsing Scheme for Chinese Sentences. *Chinese Journal of Computers* (11), 1-10.

### Appendix: transcription and annotation style

.....	ellipsis		
▲ ▼	interrupt		
( )	annotation		
<1,2,2,5,WT,A,宣读起诉书>	1,2,2,5	information level	
	WT	information knot	
	A	sharing category	
	宣读起诉书	key words	