Dependency Structures and Beyond: Assembling Drawings of Sentence Construction

Mosang Zhang*

Abstract

This paper aims at proposing a concise system of symbols to be used for assembling drawings in order to illustrate sentence constructions. The assembling drawings, purporting as a supplementary yet more linearized notation of dependency structures, reveal the dynamic process of how chunks are attached step by step according to their semantic dependency relations to form a partial structure which ultimately develops into the complete structure. It is identified by previous researchers that two movements exist within a sentence: the subjective movement and the objective movement, with the former being related to word order, while the latter is related to semantic dependency relations. Concepts of semantic stream and attachment are proposed and symbolised to represent semantic dependency relations and word order respectively, thus the assembling drawings can also represent both movements within a sentence. Sentence construction is ultimately accomplished when all the chunks in the sentence can be integrated, and the semantic streams merge together. The performance of the assembling drawings is investigated and exemplified via a selection of Mandarin Chinese and English sentences.

Key words: chunking, dependency, semantic stream, attachment, assembling drawing

* Peking University, Beijing Institute of Technology; zms118@pku.edu.cn

1 Introduction

In this paper, chunks occurring in a sequential order assemble together according to their semantic dependency relations to form partial structures. The partial structure is the processor's interpretation of the uttered language string, presenting the current state, while a transition of the state can be represented by the attachment of another chunk to the current chunk, causing change to this partial structure. A chunk can be a morpheme, word, phrase, or sentence segment.

In sentence analysis, both sentence producer and recipient rely on partial structures. Take the "tip-of-the-tongue", hesitation and regression situation, for instance, when the producer temporarily gets stuck, he/she tends to repeat the previous segment until the next occurs to him/her.

This paper attempts to analyse a number of Chinese sentence constructions, including those with flexible word order and special sentence patterning, as well as English sentences, including those with long sentence patterning.

Previously, much work has been done on dependency parsing. While most of this research is more focused on the relations between words, in a slightly different light, this paper delves into the inner workings of various inter-chunk relationships and creates a concise assembling drawing to illustrate this process.

2 Basic concepts

We first attempt to identify some key concepts brought forward in this study, including *attachment*, *semantic stream*, *subjective* and *objective movements* within sentences, etc. Their features are elaborated as a preparation for proposing a system of symbols and the assembling drawing of sentences.

2.1 Attachment of chunks

Attachment: *Attachment* in this paper refers to the action of attaching a chunk composed of word(s) or morpheme(s) to a partial structure. Since it causes a transition of states, the attachment of a chunk represents a transition.

The precondition of *attachment* is that there should be a semantic dependency relationship among the components within the two chunks. Furthermore, the posterior chunk could attach to a previous chunk even in such a situation where there are other chunks located in-between without participating in or being affected by their attachment.

Position of Attachment: When a chunk attaches at a certain position to the partial structure, this position is designated as the *position of attachment*. The *position of attachment* could span over a certain chunk or a combination of chunks.

Direction of Attachment: The direction of assembling the current chunk to the partial structure is deemed as the *direction of attachment*. Attachment itself is an action. The *direction of attachment* always points from the posterior chunk to a certain previous chunk in the partial structure. It indicates the direction of the movement of ideas which is related to the word order. In English and Mandarin Chinese, it is from the right to the left.

Target of Sentence: If we cut a sentence into two segments at a dividing point, the previous segment, represented by the partial structure, is the *target of sentence*. The posterior segment attaches to the previous part to form complete structure. The *target of sentence* is subjective in essence, which relates to the word order. It is always positioned before the dividing point, no matter whether the previous segment imports or exports semantic stream. The *target of sentence* is the partial structure which the processor pays the most attention to. Typically, within a sentence there is one *target of sentence*.

2.2 Semantic stream

A key concept proposed in this paper is the *semantic stream*, which indicates the dependency relations among the components within a sentence. *Attachment of chunks* occurs at the surface, while *semantic stream* reveals the underlying structure of meaning.

In a sentence, *semantic stream* flows from modifier to head; from direct object to predicate verb; from indirect object to the combination of predicate verb and direct object. Yet, differing from previous research on dependency relations, this paper has implemented some modifications: semantic stream is one which leads from patient to predicate verb; and from predicate verb to agent. Due to the need of distinguishing the notions of agent as opposed to patient in our system, this regulation differs from previous studies which regard both agent and patient as dependencies to the predicate verb. We hold that since the agent is the cause or initiator of an event, the relationship it bears to the predicate verb is different from that of the patient to the predicate verb. Therefore, distinctions are clarified.

Since the precondition of attachment between two chunks is that there is a semantic dependency relationship between them, to put it otherwise, the attachment chunk should either import or export semantic stream to a previous chunk.

2.3 Subjective movement and objective movement

According to Henri Weil in his influential thesis of 1844, syntactic movement is not equal to the movement of ideas. He argues that the movement of ideas is characterised in terms of word order: "the forward movement of our ideas is the natural basis of the order of words" (Weil 1844, 21-22). The movement of ideas is termed subjective movement and the syntactic movement is termed objective movement (Jiang, 2008).

Therefore, in sentence processing, the two simultaneously-occurring movements are in existence: the subjective movement and objective movement. Subjective movement is

the movement of ideas, which relates to the word order, while the objective movement is characterised in terms of semantic relations between sentence components, which relates to lexical dependency.

Representation of *subjective movement*: *Attachment*
Since the *subjective movement* is termed by the word order, it can be represented as the sequence of the *attachment* of chunks to the partial structure. The direction of subjective movement is from the posterior chunk to a certain preceding chunk, thus in English and Mandarin Chinese it is always pointing left.

Representation of *objective movement*: *Semantic stream*
Since *objective movement* is termed by dependency relations, the semantic relations between sentence components, it can be characterised in terms of *semantic stream* between chunks. Its direction is in accordance with the direction of *semantic stream*.

2.4 Recognition of chunks

Ever since the word "chunking" was proposed by George Miller in his 1956 work, a tremendous amount of research has been conducted on that topic. In this paper, chunking refers to organising items into familiar and manageable units. A chunk varies from as large as a clause, to as small as a morpheme. The similarity of this system to that of Abney (1995) is that it has been so mediated that no orphan words exist after parsing, and all the words in a sentence can be merged into their corresponding chunks. And after merging, all the chunks are non-overlapping and exhaustive.

The segmentation of a sentence into chunks hinges on the individual long-term memory patterns, as well as the specific task of analysing certain sentence patterning. Sentences can be parsed in varied ways. In fact, a trial parsing can take place during the processing. If the processor fails at a certain point, regression and reparsing take place, which are typical in garden-path sentences.

The success of parsing in this case is marked by the completion, such that the partial structure grows into a complete structure: the assembling drawing contains at least one main chunk with the predicate verb; all chunks can be integrated; and the semantic streams merge together.

2.5 Constructional elements of assembling drawings

This paper proposes a system of symbols for the graphic analysis of the relations between chunks within a sentence, namely via *assembling drawings*. In the drawings, arrows are used to represent the *direction of attachment*, with the arrow head always pointing to the *position of attachment*. The tag of *semantic stream* behind each arrow represents the direction of semantic stream. The tag of *semantic stream* together with the arrow can represent the *attachment* and the semantic dependency relations, or to

put it in another way, the *subjective movement* and the *objective movement*. Creating the *assembling drawing* firstly involves segmenting the sentence into chunks, with state points by each side of the chunk; then numbering each state point from 0 on in a sequential order; the number of a certain chunk is the same as the state point number on its right side. Then the chunk numbers are connected with arrows and other symbols in accordance with their semantic dependency relations, thus the *assembling drawing* takes shape, revealing the sentence structure as well as the assembling procedure.

3 System of Symbols

Having discussed the basic concepts, this section proceeds to propose a system of symbols which serves as a basic set of components for the *assembling drawing*. In the system of symbols, arrows and tags are applied to denote the previously-mentioned concepts.

3.1 Chunking and attachment symbols

Serial number of chunks: The serial number of chunks is defined as follows: the numbers in parentheses denotes the state point, and the numbers without parentheses denotes the chunk. The symbol (0) denotes the state point at the beginning of a sentence, the number of the state point to the right of the first chunk is (1) ... and the number of the state point to the right of the nth chunk is (n). The chunk on the left side of state point (1) is chunk 1; the chunk on the left side of state point (2) is chunk 2 ... and the chunk on the left side of state point (n) is chunk n. Take the sentence "I believe she has come" as an example. A pause is located between "believe" and "she". The sentence is segmented and numbered as follows:

(0) I believe (1) she has come (2).

In this case, chunk 1 is "I believe" and chunk 2 is "she has come".

For combined chunks, take as an example the combination of three chunks (chunk 1, chunk 2 and chunk 3), which is denoted as: 1∼3.

Main chunk: The chunk carrying the predicate verb is designated as the main chunk, which is signified by a single quotation mark ' appending to its right upper corner. In the previous example, if chunk 1 "I believe" contains the predicate verb, then the symbol 1' denotes the main chunk of the sentence.

Direction of attachment: For languages whose writing system is from left to right, such as English and Mandarin Chinese, the direction of an attachment points from the present chunk to a previous chunk in the partial structure. The *direction of attachment* is denoted by the left arrow ←.

Main attachment: The vast majority of sentence processing encounters the situation that there is one *main attachment* in a sentence, denoted by the thick arrow ← pointing at the *target of sentence*, which could be a chunk or a combination of chunks, and the segment following the thick arrow is the add-on. Therefore, the attachment in the sample sentence "I believe she has come" can be represented as 1'←2

Position of attachment: According to its definition, *position of attachment* refers to a certain point at which a chunk attaches to partial structure. The position of attachment could be a certain chunk or a combination of chunks to which the left arrow is pointing.

Hierarchy: Square brackets [ ] are introduced to indicate the boundaries of hierarchical relations among chunks.

3.2 Tag of semantic stream

The tag of semantic stream proposed in this system is placed to the right of the arrow.

Export: ←**e** is applied when the posterior chunk exports semantic stream to a previous chunk. The tag of exporting semantic stream is: **e**.

Import: ←**i** is applied when the posterior chunk imports semantic stream from a previous chunk. The tag of importing semantic stream is: **i**.

As for the aforementioned sample sentence, the object clause "she will come" is exporting semantic stream to the predicate "believe". Thus, the sentence can be further represented as: 1'←e2

3.3 Background chunk and Foreground chunk symbols

Background chunk reflects a situation where a particular chunk is inserted without participating in or being affected between two adjacent chunks, which directly import or export *semantic stream* from each other. Yet, this particular background chunk exports *semantic stream* to both of these two adjacent chunks. The symbolic representation of background chunk is: **b**. Specific sample sentences will be provided in the next section.

Suppose there are three chunks 3, 4, and 5 lined up in sequence, with chunk 4 acting as a background chunk. Then b4 shall export semantic stream to the combination of 3 and 5. Suppose chunk 5 exports the semantic stream to chunk 3 in this case.

The assembling drawing is: 3←b4←e5

Its equivalent drawing is: [3←e5]←e4

Conversely, *foreground chunk* reflects a situation where a particular chunk is inserted between two adjacent chunks which directly import or export semantic stream from each other. Yet, this particular foreground chunk imports semantic stream to the two adjacent chunks. The symbolic representation of foreground chunk is: **f**.

Suppose chunks 3, 4, and 5 are lined up in a sequence, with chunk4 being a foreground chunk. Then f4 shall import semantic stream to the combination of 3 and 5. Suppose chunk 5 exports the semantic stream to chunk 3 in this case.

The assembling drawing is: 3←f4←e5
Its equivalent drawing is: (3←e5)←i4

3.4 Parallel connection and serial connection symbols

Parallel connection: *Parallel connection* refers to the parallel relation of two chunks which import or export *semantic stream* equally to a third chunk. Parallel connection is denoted by the symbol: +. Therefore, the parallel connection of chunk 3 and chunk 4 is designated as: [3+4]

Take this sentence as an example:

Sue lent Joe a book, and Ben a paper. (cf. Milward 1991, 28)

This sentence can be segmented and numbered as follows:

(0) Sue (1) lent (2) Joe (3) a book (4), and Ben (5) a paper (6)

After attachment and tagging, it can be designated as:

1←e2' [←b3←e4+←b5←e6]

In this sample sentence, *background chunk* 3 and *background chunk* 5 are indirect objects in the sentence. The *semantic streams* gather to chunk 1 by the end. The items of *parallel connection* in the brackets obey distributive rules, as in Mathematics.

Serial connection: *Serial connection* refers to two chunks in a sequential order, exporting *semantic stream* to another preceding chunk, which functions as the cause of their events. *Serial connection* of chunks is denoted by the symbol ^. Serial connection is applicable in typical Mandarin Chinese sentence patterns, such as the serial predicate construction. For example,

他　　怕　　麻烦，　　马上　　　走了。
Ta　　pa　　mafan,　　mashang　　zuole
He fears trouble, immediately left.

This sentence can be segmented and numbered as:

(0) 他 (1) 怕 (2) 麻烦 (3), 马上走了 (4)

After attachment and tagging, the sentence is analysed as follows:

1←e2←e3^←e4'

This is a sentence with sequential serial verb construction; with the actions of "走了" following "怕", which is also the cause of the action "走了", "怕" and "走了" constitutes the serial predicate construction and the serial connection symbol ˆ is applied in-between.

3.5 Bracket lock, specific word location, attention cursor symbols

Bracket lock: The dot at the left side of a semantic tag is intended as a lock of the left bracket, ensuring the integration of the components within the brackets, which as a whole can be attached to other chunks, such as: [ ]←.e and [ ]←.i. Similarly, the dot at the right side of semantic tag locks up the right bracket, such as: ←e.[ ] and ←i.[ ]

Location of word in a chunk: In the *assembling drawing*, pointing out the location of a particular crucial word in a chunk is occasionally required, for which purpose the denotation of this location is exemplified as follows:

In the sample sentence "I believe she has come", the location of the word "believe", which is the second word in chunk 1, is designated as: 1/2.

Attention cursor: Attention cursor refers to the indicator showing the current position at which the processor arrives. When the attention cursor reaches a certain state point, the part of the sentence to its left becomes the partial structure representing the state at this point. In the *assembling drawing*, the symbol @ is used to indicate the current position of the attention cursor.

4 Application of a system of symbols in the assembling drawing

We are now in a position to illustrate the interpretation of sentence structures via the *assembling drawings* of typical examples discussed by previous researchers. The key to illustrate a sentence structure is to demonstrate the *attachment* of chunks as well as the flow of *semantic stream*.

Steps for the assembling drawing:

1. **Chunking and numbering**: the sentence is segmented into chunks, with state points tagged beside each chunk.
2. **Attachment of chunks**: each state point is assigned a serial number, and the chunk on its left side is assigned the same serial number.
3. **Assigning tags of semantic stream**: the serial numbers of chunks are connected through tags indicating the inter-chunk relations.
4. **Checking for the overall connectedness of the structure**: The outcome of this assembling drawing should be one in which all the chunks can be integrated into one, in other words, no isolated chunks exist, and all semantic streams merge together at certain chunk(s) in the end.

In this system, with the assembling of each chunk into the existing partial structure, the arrow points from right to the left, from the larger-numbered to the smaller-numbered chunk.

Consider the following sentence adapted from Abney (1994), a typical example of parsing by chunks in sentence processing:

1.)   When I read a sentence, I read it a chunk at a time.

a. Dependency graph:

When   I   read   a   sentence,   I   read   it   a   chunk   at   a   time.
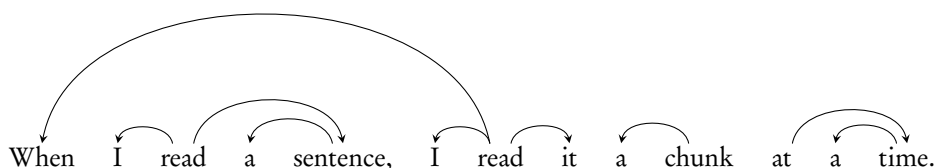
b. Assembling drawing:
Step one: chunking and numbering
According to Abney (1994), the sentence is segmented as follows:

(0) When I read (1) a sentence (2) I read it (3) a chunk (4) at a time (5)

The words in each pair of round brackets comprise one chunk.

Steps two and **three**: After attachment of chunks and tagging, the assembling drawing is as follows:

[1←e2]←i3'⇇e[4←i5]

This assembling drawing indicates that the main chunk 3 "I read it" imports semantic streams from both the combination [1←e2] and the combination [4←i5]. It reveals the process of parsing by chunks, facilitating sentence comprehension. The more detailed representation of this sentence is:

[1/3←e2]←i3'⇇e[4←i5]

"a sentence" in chunk2 is the object of the verb "read" in chunk 3, the precision of this drawing is specific to the word, with 1/3 denoting the 3rd word in chunk 1: "read".

Step four: Checking for the overall connectedness of the structure:
The outcome of this drawing is one in which all orphaned words are swept into adjacent chunks, all the chunks are integrated, and all semantic streams merge together in chunk 3 at the end. Thus, the analysis is completed.

Compared with the dependency graph, the *assembling drawing* presents a concise and linearized representation, which avoids putting arrows on specific words, thus providing more room for illustrating the complex, flexible sentence patterning in a dynamic way.

Semantic streams for resolving ambiguities:
The following typical examples are intended to present the applications of semantic stream to the resolution of ambiguities. The concept of semantic stream is proposed to enable the concise representation of the inter-chunk relations.

> 2.)   Flying planes can be dangerous. (cf. Chomsky 1965, 21)

Step one: chunking and numbering

(0) Flying (1) planes (2) can be dangerous (3)

Steps two and **three**: Attachment and tagging
There are two possible assembling drawings for this structurally ambiguous sentence:

> A. [1←e2]←e3
> B. [1←i2]←e3

As Chomsky has pointed out in *Aspects of the Theory of Syntax*, if the processor can detect the ambiguity of this sentence, his/her intuitive knowledge of the language is such that both of the interpretations (corresponding to "flying planes are dangerous" and "flying planes is dangerous") are assigned to the sentence by the grammar he/she has internalised in some form (Chomsky 1965, 21).

Assembling drawings reveal that the problem lies in the relations between "flying" and "planes": in 'assembling drawing A', the "planes" exports *semantic stream* to "flying", making "flying" the more weighted part, an action in this case; with "plane" being the patient. In 'assembling drawing B', the "planes" imports the *semantic stream* from "flying", thus the word "planes" is the more weighted part in this case, and "flying" becomes its modifier, gaining an adjective attribute.

> 3.)   Put the frog on the napkin into the box. (cf. Felser et al 2003)

This is a typical sample sentence in Psycholinguistics to test children's processing of ambiguous sentences. Two possible interpretations can be generated by the attachment of the second proposition phrase, the differences of which can be illustrated with the following assembling drawings:

Step one: chunking and numbering

(0) Put (1) the frog (2) on the napkin (3) into the box (4)

Steps two and **three**: Attachment and tagging
Two possible assembling drawings for this sentence are:

A. 1'←e[2←e3] 1←e4
B. 1'←e2 1'←e3 3/3←e4

In structure A, the verb "put" takes the big NP "the frog on the napkin", then takes CP "into the box" which indicates the destination of the action. The other interpretation in structure B indicates that both "on the napkin" and "into the box" act as destinations of the verb "put" in a sequential order.

Background chunk:
Advantageous over dependency graphs, our system gives a specific account of the inter-relations among three chunks. The concepts of background chunk and foreground chunk are proposed.

Background chunk is located between two adjacent chunks, which both import *semantic streams* from the background chunk. The following can be deemed as typical examples involving background chunks.

4).   病人是个近七十岁的老人，   老病加新病，   走路都困难。 (cf. Lü 1986)
Bingren shi ge jin qi shi sui de lao ren,   lao bing jia xin bing,   zoulu dou kunnan.
The patient is an about seventy-year-old person; new disease adding to old disease, even walking is difficult.
The patient is an old person about seventy years old, new disease plus old disease makes him in great difficulty even in terms of walking.

Step one: chunking and numbering

(0) 病人 (1) 是个近七十岁的老人 (2), 老病加新病(3), 走路都困难(4)

Steps two and **three**:
After attachment and tagging, the assembling drawing is as follows:

1←e2'←b3←e4

Without chunk 3, the sentence still has a complete structure. Chunk 3 is inserted in the sentence as a *background chunk*.

5).   不过　这个　过程　有人　长　有人　短　罢了。(cf. Lü 1986)
        Buguo zhege guocheng you ren chang you ren duan bale.
        Yet this process for someone long for someone short then.
        Yet for someone, this process is long; while for the others it is short.

Step one: chunking and numbering

(0) 不过 (1) 这个过程 (2) 有人 (3) 长 (4) 有人 (5) 短 (6) 罢了 (7)

Steps two and **three**: Attachment and tagging

[1←e [2←[b3←e4＋b5←e6]]]←e7

There is a *parallel connection* in this sentence structure: "有人长" and "有人短", in which "有人" is the *background chunk* in either of these fragments.

Foreground chunk:
Foreground chunk, however, is located between two chunks which both export *semantic stream* to it.

6).   我们　　路上　　　整整　　　走了　　　一个月。
        Women   lushang   zhengzheng zoule   yigeyue.
        We   on the way entirely   have walked for a month
        We have walked for an entire month on our way.

Step one: chunking and numbering

(0) 我们 (1) 路上 (2) 整整 (3) 走了 (4) 一个月 (5) 。

Steps two and **three**: Attachment and tagging

1←b2←e [3←f4'←i5]

This sentence has "整整" emphasised by moving it in front of "走了". Chunk 4 "走了", containing the main verb, is the *foreground chunk*. Chunk 4 imports *semantic stream* from the combination of both chunk 3 and chunk 5 "整整一个月".

7).   甭着急，　　　仗有你们打的。(Lü 1986)
        Beng zhaoji,   zhang you nimen da de.
        Do not worry, battles there will be for you to fight.
        Do not worry, there will be battles for you to fight.

Step one: chunking and numbering

(0) 仗 (1) 有 (2) 你们打的 (3)

Steps two and **three**: Attachment and tagging

1←f2'←e3

Chunk 3 exports semantic stream to chunk 1. Then they export *semantic stream* to *foreground chunk* 2, which contains the verb "有". "仗" is moved to the very beginning of the sentence in order to be emphasised and, in this inverted sentence, chunk 1 is the *target of sentence*.

Parallel connection:
Another advantage of *assembling drawings* is the clarity in representing the *parallel structures* and *serial connection*.

   Parallel connection refers to the relation between two chunks that import or export *semantic stream* equally to a third chunk. The following assembling drawing signifies a typical example for parallel connection:

   8). 大象   鼻子   长   耳朵   大。(Lü 1990)
        Daxiang bizi chang, erduo da.
        Elephant trunk long, ears big.
        Elephant has long trunk and big ears.

Step one: chunking and numbering

(0) 大象 (1) 鼻子 (2) 长 (3), 耳朵 (4) 大 (5)
Steps two and **three**: Attachment and tagging

1←e[2←e3 + 4←e5]

Chunks [2∼3] compose a subject-predicate phrase; similarly, chunks [4∼5] compose another subject-predicate phrase; [2∼3] and [4∼5] equally modify chunk 1, the target of this sentence. Therefore, they are parallel structures, connected by the symbol "+".

Serial connection:
Serial connection of two chunks refers to a kind of connection in which both chunks in a sequential order export semantic stream to another preceding chunk. The following is the sample sentence for serial connection:

9). 旧社会，穷人生了病，小病抗，大病躺，根本得不到医治。 (Lü 1986)

Jiu shehui, qiong ren sheng le bing, xiao bing kang, da bing tang, genben de bu dao yizhi.

In the old society, the poor fall ill, a minor illness (they) struggle against, a serious illness (they) lie down, cannot afford any medical treatment at all.

In the old days, when the poor fall ill, they have to struggle against minor illnesses, lie down when they are seriously ill, but they can never have access to medical treatment.

Step one: chunking and numbering

(0) 旧社会 (1), 穷人 (2) 生了病 (3), 小病 (4) 抗 (5), 大病 (6) 躺 (7), 根本得不到医治 (8)

Steps two and **three**: Attachment and tagging

1←i[2←e3'] ^←e[4←i5' + 6←i7'] ^←e8'

This is another sample sentence typical of Chinese flexible sentence patterning. Both *parallel and serial connections* arise in this sentence. [4∼5] and [6∼7] are parallel structures, connected by the symbol "+". As a whole, [4∼7] constitutes a serial connection with [2∼3], happening in a sequential order and being caused by the latter; similarly, chunk 8 also results from [2∼3], creating another *serial connection* relationship.

Bracket lock:

10). 王冕死了父亲。 (Shen 2006)
Wang Mian si le fuqing.
Wang Mian died father.
Wang Mian's father died.

Step one: chunking and numbering

(0) 王冕 (1) 死了 (2) 父亲 (3)

Steps two and **three**: Attachment and tagging
The assembling drawing is as follows:

1←.e [2'←i3]

In this sentence, the dot on the left side of the *semantic stream* tag "e" is a *bracket lock*, indicating that chunk 2 and chunk 3 jointly export *semantic stream* to chunk 1, since [2∼3] "*死了父亲*" as a whole explains what happens to the person named "*王冕*", Chunk 1, placed preceding the thick arrow, is the *target of sentence*.

Analysing long English sentences:

English long sentence patterning is another language phenomenon which the dependency graphs shall find it difficult to cope with, yet this problem can easily be solved with *assembling drawings*. Consider this example:

11).   It may, however, not be out of place to call to it the attention of the teachers of modern languages in secondary schools who, desiring to insure a good pronunciation for their pupils, will find Mr. Passy's book an invaluable handbook and guide with which to broaden and deepen out of their own knowledge the usually rather perfunctory phonetically introductions of the elementary modern language textbooks. Step one: Chunking and numbering

Step one: chunking and numbering

(0) It may (1), however, (2) not be out of place (3) to call (4) to it (5) the attention (6) of the teachers of modern languages in secondary schools (7) who, (8) desiring to insure a good pronunciation for their pupils, (9) will find Mr. Passy's book an invaluable handbook and guide (10) with which to broaden and deepen (11) out of their own knowledge (12) the usually rather perfunctory phonetically introductions (13) of the elementary modern language textbooks (14).

Steps two and **three**: Attachment and tagging

1←b2←e3' 1/1←e4←b5←e6←e7 7/3←e8←b9←e10 10/10←e11←b12←e13←e14

The assembling drawing of this long English sentence is composed of three interconnected parts, with the same number indicating the same chunk that is being overlapped. The inter-chunk relations are explicit. Specific words can be indicated in this drawing, "7/3" referring to "teachers", and "10/10" referring to the word "guide" are cases in point. A chunk can comprise only one single word, as is the case with chunk 8, for example. Therefore, chunking should comply with specific tasks.

Analysing garden paths sentence:

Garden paths sentences have received much attention for the fact that they lure the processor into a parse that turns out to be a dead end. When used in Psycho-Linguistics, garden path sentences illustrate the fact that when human beings read, they process language one word at a time (cf. Harley 2014). Previous studies on dependency graphs show that they generally provide a static representation. Due to the fact that the processing of the complexity of a garden paths sentence is incremental, the following samples are intended as an investigation into dynamic parsing of garden paths sentences by chunks with assembling drawings:

12).   The complex houses married and single students and their families. (Jurafsky 1991)

In their initial parsing, as Jurafsky (1991) has noted, most readers interpret "the complex houses" as a noun phrase that meets the strong frequency expectation.

Chunking and numbering the part of the sentence at this point provides this:

(0) The complex houses (1) @

The symbol @ denotes the present position of the *attention cursor*.

Here, the *assembling drawing* is: 1@

Next, as the cursor moves on, it presupposes the word "married" as a verb. At this point, houses cannot get married, a semantic incorrectness arises, and chunk 2 cannot directly attach to chunk 1, so the *attention cursor* leaves a question mark (?) between the words "houses" and "married".

At this point, chunking and numbering is as follows:

(0) The complex houses (1) married (2) @

The *assembling drawing* is: 1 (?) 2

Moving on, the cursor encounters the word "single", "married and single" can modify "students"; hence these three words are encoded into one chunk. Chunk 2 is expanded into "married and single students". Yet it still cannot directly attach to chunk 1, hence the question mark (?) is retained.

Chunking and numbering at this point provides this:

(0) The complex houses (1) married and single students (2) @

The *assembling drawing* is as follows: 1 (?) 2

Next, the cursor moves on and encounters "and their families".

Chunking and numbering at this point is as follows:

(0) The complex houses (1) married and single students (2) and their families (3) @

Here, the *assembling drawing* is as follows: 1 (?) 2+3

Now that the *attention cursor* has come to the end of the sentence, no following chunk could possibly rescue the initial partial structure, the question mark (?) is still retained. Chunks cannot be assimilated into one complete structure, signifying that an error exists in the initial parsing.

At this point, the cursor backtracks from the initial mis-parse to locate the problem and restarts parsing:

"Complex" takes its second meaning as a noun. The string "The complex" is re-grouped as a chunk.

The state of chunking and numbering is such:

(0) The complex (1) @

Next, the processor realises that the intended interpretation of "houses" is the third-person singular form of the verb.

Chunking and numbering thus are:

(0) The complex (1) houses (2) @

At this point, the assembling drawing is as follows:

1←e2'

Next, "married and single students" can be encoded it as chunk 3.

Chunking and numbering here becomes:

(0) The complex (1) houses (2) married and single students (3) @

The assembling drawing is:

1←e2←e3

Next, "and their families" is identified as chunk 4.

Chunking and numbering becomes:

(0) The complex (1) houses (2) married and single students (3) and their families. (4) @

The attention cursor has arrived at the end of sentence. The final assembling drawing is:

1←e2'←e [3+4]

This drawing indicates that all the chunks have been completely integrated, with the *semantic streams* merging together. Chunk 2 is the main chunk; and chunk 1 is the *target of sentence*. The interpretation of the sentence structure has been accomplished.

While dependency graphs mainly provide static representations, the *assembling drawings* reveal the step-by-step process of applying dynamic parsing by chunks to garden paths sentences.

5 Conclusion

This paper arises from an attempt to provide a concise *system of symbols* for *assembling drawings* to give an account of sentence construction in a more linearized fashion than is possible with dependency structures. The concept of *semantic stream* proposed in this paper is intended as a representation of the semantic dependency relations between chunks, which depicts how the meanings of sentence components are merged together.

Symbols serve as the basic elements for the *assembling drawings*: arrows and *semantic stream* tags together embody the *attachment* relation and the semantic dependency relation between chunks, which are also referred to as the objective movement and the subjective movement; chunk with apostrophes denote the *main chunk* with the predicate verb; thick arrows point to the *target of sentence*; square brackets are used to indicate the hierarchy of the sentence; *serial connection* and *parallel connection* are also denoted respectively; and the *attention cursor* symbol is intended to reflect the dynamic process of parsing by chunks. The following table summarises the basic symbols proposed in this study.

Symbol	Name	Explanation
(0), (1), ... (n)	Serial number of chunks	Parenthesis denotes the state point, and the number without parenthesis denotes the chunk
'	Main chunk	the chunk carrying the predicate verb
←	Direction of attachment	The direction of an attachment points from the present chunk to a previous chunk in the partial structure.
⬅	Main attachment	The thick arrow points at the target of sentence, the chunk or combination of chunks to which the processor pays most attention.
[ ]	Hierarchy	Indicator of the boundaries of hierarchical relations among chunks.
←e	Exporting semantic stream	←e is applied when the posterior chunk exports semantic stream to a previous chunk.
←i	Importing semantic stream	←i is applied when the posterior chunk imports semantic stream to a previous chunk.
b	Background chunk	a particular chunk inserted in-between, yet without being affected by its two adjacent chunks, it exports semantic stream to both of these two adjacent chunks
f	Foreground chunk	a particular chunk inserted in-between, yet without being affected by its two adjacent chunks, it imports semantic stream to both of these two adjacent chunks
+	Parallel connection	the parallel relation of two chunks which import or export semantic stream equally to a third chunk
^	Serial connection	two chunks, in a sequential order, exporting semantic stream to another preceding chunk, which functions as the cause of their events
.	Bracket lock	a lock of the left/right bracket that ensures the integration of the components within the brackets and, as a whole, can be attached to other chunks, such as: [ ]←.e and [ ]←.i or ←e.[ ] and ←i.[ ]
/	Location of word in a chunk	the fifth word in chunk 3 is designated as: 3/5
@	Attention cursor	the indicator showing the current position at which the processor arrives

A comparison between dependency graph and assembling drawings:

1. Assembling drawings are more concise and linearized representations than dependency graphs. In the assembling drawing, chunks are represented through numbers in accordance with the word order, in order to avoid the great complexity of having to tag and create arcs directly on the original sentence.
2. Dependency graphs are basically static analyses, while assembling drawings are dynamic.

3. Dependency graphs are decoder-oriented while assembling drawings are neutrally oriented between decoder and encoder.
4. Dependency graphs are more word-specific, while assembling drawings are more focused on chunks.
5. In the sample sentences, assembling drawings can more easily represent the semantic relationships among three chunks using labels, such as foreground chunk and background chunk.

Assembling drawings are relatively easy to produce, both in terms of manual charting as well as as computer-based images, via typing in Microsoft Word, for example. For instance, in order to create the arrow which occurs frequently as a representation of the *attachment* of chunks, it only takes a few steps with an English keyboard. Firstly, you type in the symbol "<", then follow this by typing "- -", which automatically creates the symbol "←"; alternatively, if one types in "= =" after the "<", the thick arrow "⇐" is created.

Building assembling drawings involves four steps: chunking and numbering; *attachment* of chunks; tagging the *semantic streams*; and checking for the overall connectedness of the structure. In the final drawing, at least one *main chunk* with a predicate verb, as well as one target of sentence is involved; all the chunks are finally integrated in one and all the *semantic streams* merge together at the end.

The *system of symbols* requires further expansion, defining chunks as specific types representing various special sentence patterns.

The application of *assembling drawings* provides a fertile ground for sentence analysis, which is left open to be explored in future studies.

References

Abney, S. 1994. "Partial Parsing." *ANLP-94*. Stuttgart.

Abney, S. 1995. "Chunks and Dependencies: Bringing Processing Evidence to Bear on Syntax." *Computational Linguistics and the Foundations of Linguistic Theory*. CSLI.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Felser, C., T. Marinis, and H. Clahsen, H. 2003. "Children's Processing of Ambiguous Sentences: A Study of Relative Clause Attachment." *Language Acquisition*, 11: 127-63.

Harley, T.A. 2014. *Psychology of Language, from Data to Theory*, 4th edition. Hove, UK: Psychology Press.

Jiang, W.Q. 2008. "Transmutations of the Concept of Theme." *Contemporary Linguistics II* (In Chinese). Beijing: Foreign Language Teaching and Research Press.

Jurafsky, D. 1991. "An On-line Model of Human Sentence Interpretation." *Proceedings of the 13th Annual Conference of the Cognitive Science Society (COGSCI-91)*: 449-454.

Kempson, R. 2001. *Dynamic Syntax: The Flow of Language Understanding*. Oxford: Blackwell.

Lü, S.X. 1986. "Flexibility of Chinese Sentence Patterns." *Studies of the Chinese Language* (In Chinese). Changsha: Hunan University Press.

Lü, S. X. 1990. "Formation and Elimination of Ambiguities." *Collected Essays by Lü ShuXiang* (in Chinese), Vol 3. Beijing: The Commercial Press.

Ma, X. Z. and F. Xia. 2014. "Unsupervised Dependency Parsing with Transferring Distribution via Parallel Guidance and Entropy Regularization." *ACL*: 1337-1348.

Miller. A. G. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." *The Psychological Review*, 63(2): 81-97.

Milward, D. 1991. *Axiomatic Grammar, Non-Constituent Coordination, and Incremental Interpretation.* PhD thesis, Cambridge: University of Cambridge.

Milward, D. 1994. "Dynamic Dependency Grammar." *Linguistics and Philosophy*, 17: 561-605.

Nivre, J. 2008. "Algorithms for Deterministic Incremental Dependency Parsing." *Computational Linguistics*, 34: 513–553.

Purver, M. & R. Kempson. 2004. "Incremental Parsing, or Incremental Grammar?" *Proceedings of the ACL Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*: 74-81.

Shen, J.X. 2006. "Generation of the Sentence 'Wangmian sile fuqing': On the Blending of New Patterns of Chinese Sentence Construction." *Studies of the Chinese Language* (In Chinese), 4: 291-300. Changsha: Hunan University Press.

Tiedemann, J. 2015. "Improving the Cross-lingual Projection of Syntactic Dependencies." *Nordic Conference of Computational Linguistics NODAL-IDA 2015*: 191.

Wang, L. 1956. "Definition of Subject and its Application in Mandarin Chinese." *Chinese Learning* (In Chinese): 271-285. Shanghai: Central China Book Company.

Weil, H. 1844. *De Vordre des Mots Dans les Langues Anciennes Comparées aux Langues Modernes. Thèse Principale Française*, Paris.