Arab World English Journal
مجلة اللغة الانكليزية في العالم العربي
A QUARTERLY INTERNATIONAL PEER REVIEWED JOURNAL
ISSN: 2229-9327

AWEJ Vol.3 No.1 March 2012 pp.50-76

A Predictive Model on Level of Difficulty in English-Chinese Translation

Guey, Ching-chung
Department of Applied English
I-Shou University, Kaohsiung, Taiwan

Abstract

The paper seeks to propose a predictive model on level of difficulty in English (L2)-Mandarin (L1) translation. Among the three non-Verbial functions (i.e., nominal, adjectival, and adverbial) and five different structures (i.e., infinitives, gerunds, participles, clauses, and phrases) in English, there exists a general contingent relationship in the course of translating each of the structures into Mandarin in coordination of the functions each structure serves. This predictive model can be most useful when dealing with recursive and mixed structures, and thus offers a general guideline for English-Mandarin translators. In this paper a brief deduction of sentential structures of L2 will be introduced first with mathematic interpretations, followed by presenting the cognitive steps involved in English-Mandarin translation (i.e., chunking the structures, identifying their functions, and arranging the structure order), and drawing the predictions of levels of difficulty on each structure in terms of the cognitive steps involved. The validity of such a tentative prediction model requires further examination and elaboration.

Keywords: English-Chinese translation, cognitive steps, syntactic differences, recursive structures

A Predictive Model on Level of Difficulty in English-Chinese Translation

"I have not hesitated to propose a general principle of linguistic structure on the basis of observations of a single language" (Chomsky, 1980: 48). Indeed, as the mathematician and linguist Noam Chomsky put it, there is a general principle of linguistic structure that can be elaborated or interpreted mathematically. Since the general principle of linguistic structure may exist in both English (L2) and Mandarin (L1), thus there must also be a general principle that serves as a bridge (translation) between L2 and L1, and such a bridge can be expressed concisely from mathematic perspectives. Such a translation principle requires the understanding of both languages (target and source), especially of their syntactic structures as well as their cultural connotations (Liu, 1997).

As Wingfield and Titone (1998) indicates, to understand a sentence, one must determine its syntactic structure. For this, Noam Chomsky' transformational Syntax does provide a good framework in dealing with the assignment of words in a sentence to their relevant linguistic categories (i.e., sentence parsing), but still when one relates it to translating sentences of complex syntactic structures, especially those of recursive structures, difficulties will inevitably ensue for such cognitive model (simple cognitive steps) that focuses specifically on syntactic structures is yet to evolve even though there are a whole lot of discussions about how difficult sentences are produced, as in Noam Chomsky's universal grammar (1965), but few (if there is any) about how they are translated in another language (Kimble, 1973).

This paper proposes a predictive model of level of difficulty in translating complex English sentences (L2) into Mandarin (L1) by, firstly reviewing the general concepts about differences between L1 and L2, then working on the development from "eight parts of speech (8ps)" to "four parts of speech (4ps)", and further into "two parts of speech (2ps)",

then introducing the cognitive steps involved in English-Mandarin translation, next identifying the functions of different structures and determining levels of relative difficulty, and finally making predictions and giving suggestions based on the model for further investigation.

The number of Verbs in a sentence between L2 and L1

For Mandarin-English bilinguals (where Mandarin is L1), learning English (L2) is basically applying L1 system to that of L2 (Obler & Albert, 1977). In this way, there is no need to encode a completely new set of linguistic rules, but rather what is already available in L1. In the course of translation, if the input information of L2 corresponds to the constructs available in L1, then it can be integrated with the first language and form the shared conceptual and language stores (i.e., schema for L1-L2 compound). The interconnectedness between similar forms can, according to "cognitive load theory" (Sweller, 1994) reduces the load in the "working memory" for translation. By the same token, if there is a sharp contrast between the input data of L1 and L2 on both semantic and syntactic levels, then there should be certain transformation between L2 and L1 to accomplish the translation task (Chan, Chau & Hoosain, 1983). Thus, comprehending each language input as well as translating it into the other language involves quite a few cognitive steps. To specify, when dealing with sentences with only one VERB (L2), or simply put, the five basic sentence patterns in the course of English-Mandarin translation, translators can easily translate L2 into L1; however, when more than two VERBs appear in L2 sentences, then translating them into L1 will become naturally complex for the essential difference between L2 and L1 lies in the number of VERBs in a legitimate sentence of each language.

In specific, to understand the novel L2 input information, one must first monitor its

surface forms and encode its relevant semantic nodes to infer the plausible L2 semantic meaning, and then translate it into L1. In such a case, a direct pathway between L2 and L1 is not possible because there exist gaps between L2 and L1 (i.e., there is no schema for L2 and L1). That is, translating L2 means establishing links (as schemas) with cognitive steps between L2 and L1. As long as the links are established, the load of L2 processing in our working memory may be reduced, thus speeding up the translating processing. In this regard, the links are the necessary mediations for different syntactic structures between L2 and L1 (Brooks & Dansereau, 1983). See the Mandarin-English paired sentences as an example below:

Case 1: Wh-movement (L2-English, L1-Mandarin, L1 in terms of L2 order)

L2→ (What they do) Depends on (where they come from)

L1→ (他們是做什麼的) 依 (他們從那兒來) 而定

L1 order→ (they do what) depends on (they come from where)

To comprehend the above L2 sentence: *"What they do depends on where they come from,"* the translator needs to know the "Wh- Movement" (Radford, 1981) and how it is translated in L1. As is shown, the difference lies mainly in "word order" between these two languages. Translating such a structure would be relatively easier with the knowledge of WH-movement as a mediator. That is, WH-movement here may serve as the schema to bridging the gap between these two languages (Brooks & Dansereau, 1983). With the help of cognitive rules such as WH-movement, translation may be more efficient in the course of encoding and decoding of the novel structures of the sentence. What then, if such cognitive rules concerning differences between L1 and L2 are so numerous and trivial (in most cases they are) that they might cause difficulties in translation process?

In fact, there exist similarities (compound) or differences (coordinate) of surface forms between L1 and L2 (Weinreich, 1953; Kohlers, 1963), which may shed a new light on translation. The levels of compound and coordinate relationships of language structures themselves may have much to do with the nature of translation process. That is, more transformations in translation are required when language structures of L1 and L2 are more coordinate in nature. For example, the WH-movement as Noun Phrases are coordinate in one way, while compound in another, under which circumstance fewer transformations are required in the process of translation than the completely coordinates such as adjectival clauses (Wang, 1990). Besides WH-movement, similarities (compound) between L2 and L1, as mentioned earlier, lies in sentences with only one VERB (basic sentence patterns), and differences (coordinate) of surface forms between L1 and L2 are often seen in sentences with ADJECTIVAL structures (e.g., relative clauses or adjectival phrases), which may indeed cause cognitive load for translators especially when dealing with such structures in their embedded forms (e.g., adjectival structures embedded in sentences of adjectival structures.) Since there is much evidence for different amounts of compound and coordinate storage across bilinguals of different languages due to the differences in language structures (Chan et al., 1983; Fang, Tzeng, & Alva,1981; Poplack, 1980; Timm, 1975), a model which specifies different levels of compound or coordinate between L1 and L2 in relation to the levels of difficulty in translation is required.

To propose a prediction model of level of difficulty in L2-L1 translation, one must first resort to the existing syntactic system of L2 in the eyes of L1 users, then to examine such a system to see if it is cognitively convenient or efficient in our working memory when doing translation, to work out a mathematical expression in which level of difficulty can be

operational and quantified, and thus prediction can be made and further exploration through empirical validation with practical examples can be suggested. The mathematic model adopted in this study can be manifold. First, English sentences, as defined by the present model, are made up of two parts: verbs and non-verbs. This can be indicated as:

Sentence = Verb + Non-verb

Or,

$$S = V + \sim V \quad (1)$$

And,

$$\sim V = N \text{ (nominal)} + Adj \text{ (Adjectival)} + Adv \text{ (adverbial)} \quad (2)$$

Assumptions behind L2 sentence compositions

There are all kinds of assumptions and theories in any given field of discipline, including the field of English syntax. From macro perspective, there exist simple underlying assumptions about sentences with complex structures:

1) Only one main VERB is allowed in a sentence with the other verbs, if any, being transformed into infinitives, gerunds, participles, clauses or phrases.
2) Any unit other than VERB in a given sentence can be roughly classified into the functions of nouns, adjectives, and adverbs.
3) Sentences with symmetric structures can be created around the main VERB.
4) Structures of identical function are interchangeable and can be recursively created.

The Five non-VERB structures

First, suppose there is more than one VERB in a giving sentence (L2), how to deal with the extra-VERBs? Here are the variations of the second verb: (i.e., there are ways to make a VERB "non-VERB")

(1) Infinitives (Inf.)→ (to+V~)

→ They plan ***to leave.***

(2) Participles (Par.) → (Present –Ving~) / Past- Vpp-)

→ They are ***leaving***. / → They are ***left alone***.

(3) Gerunds (G.) → (Ving~)

→ They enjoy ***doing chores***.

(4) Clauses (Cl) → (Conj. +s+v~)

→ We learned ***that they would leave.***

(5) Phrases (Phr.) → (wh-to+V~, or Prep + N~)

→ They know ***where to go***.

→ They are on the point ***of leaving***.

These five Verbal variations can be coined as five structures, and they are "~V" (non-VERB). As mentioned, there are four parts of speech (N, V, Adj., Adv.) in a given sentence, these five ~V cannot be a VERB any longer, but they can be, according to assumption two, Nominal, Adjectival, and Adverbial. Thus, if we combine the five structures with the three functions, then logically we can have:

~V = N (Inf, Par, G, Cl, Phr) + Adj (Inf, Par, G, Cl, Phr) + Adv (Inf, Par, G, Cla, Phr) (3)

Or, similarly,

~V = Inf (N, Adj, Adv) + Par (N, Adj, Adv) + G (N, Adj, Adv) + Cla (N, Adj, Adv) +

Phr (N, Adj, Adv)

(4)

Note that formula (3) = formula (4).

Three functions and symmetric structures

As the second assumption indicates, sentences with symmetric structures can be created around the main VERB. Such assumption has to do with the permutation of structures in a legitimate sentence (L2). Four parts of speech (functions) in a given sentence actually suggest their individual relative positions in a given sentence. Principally, nominal structures are to be placed around the main VERB, adjectival structures around the modified nouns, whereas adverbial structures in every possible position (position free, but mostly around the main structure- N V N). With such a principle, we may start inducing the prototype of sentences as below: (note that permutation is basically symmetrically around the VERB, and the symmetrical structure only serves as a framework for structure identification, and one can also create a sentence without symmetrical structures.) Below are the steps of structure development:

Step 1: Create the most important Verb; namely,

V

Step 2: Create Nouns around the main Verb; namely,

N V N

Step 3: Create Adjectives around the Nouns; namely,

Adj **N** Adj **V** Adj **N** Adj

Step 4: Create Adverbs around the main structure (N V N) VERB; namely,

Adv, (Adj)+N+ (Adj) V (Adj)+N+(Adj) (,) Adv

(1) (2) (3) (4) (5) (6) (7) (8)

Step 5: Put all the ~V structures in the non-VERB positions, and then the global patterns of each structure can be created. To take infinitives as an example, we have:

[TO+V~], **(N)+**[TO+V~] **V** **(N)+** [TO+V~] (,) [TO+V~]
[TO+V~] [TO+V]
(1) (3) (4) (6) (7) (8)

Note that positions (2) and (5) are missing; they do not exist because by convention adjectival structures more than 1 unit must be placed behind the nominal they modify, and [to+V~] is obviously more than 1 unit. Note also that the other four structures (participles, gerunds, clauses, and phrases) can also be theoretically assigned to the positions like those of infinitives, with differences on certain positions (e.g., gerunds as adjectives must be on either (2) or (5), whereas adjectival clauses on either (4) and (7), etc.)

Interchangeability

Structures of identical function can be interchangeable and made recursive, as indicated in Equations (3) or (4), "~V = N (Inf, Par, G, Cl, Phr) + Adj (Inf, Par, G, Cl, Phr) + Adv (Inf, Par, G, Cla, Phr)." See the elaborations below:

1) To+V~ (Infinitive) -(Nominal) (a)/ (Adjectival) (b) / (Adverbial) (c)

2) Ving/Vpp~ (Participles) -(Nominal) (d)/ (Adjectival) (e)/ (Adverbial) (f)

3) Ving~ (Gerund) -(Nominal) (g)/ (Adjectival) (h)/ (Adverbial) (i)

4) Cj+S+V~ (Clause) -(Nominal) (j)/ (Adjectival) (k)/ (Adverbial) (l)

5) (wh-to+V~/Prep + N~) Phrases-(Nominal) (m)/ (Adjectival)(n)/ (Adverbial)(o)

In short, we can find that, functionally, (a) = (d) = (g) = (j)= (m); (b) = (e) = (h) = (k)= (n); (c) = (f) = (i) = (l) = (o). Note also that the interchangeability mentioned above is basically on syntactic aspect, rather than on the semantic one, and it is theoretically valid, but scarce in language data in some; thus, adjectival gerunds and nominal participles may not be elaborated here forth.

Recursive structures

What makes English-Mandarin translation most intriguing is the translation of sentences embedded with recursive structures, especially those embedded with recursive adjectival clauses. Recursive structures can be both homogeneously recursive (e.g., repetition of identical structures) and heterogeneously recursive (e.g., repetition of mixed structures). Sentences such as *He went home to ask her mother to purchase reference book* are homogeneously recursive on infinitives), whereas Sentences such as *He went visiting a professor that lives in the village located on the hill* are heterogeneously recursive on participles, clauses, and phrases. On the other hand, sentence embedded with recursive structures can also appear in nominal, adjectival, and adverbial positions. The recursive structures can be specified as:

$$\sum_{i=0}^{n} Random\ (\text{inf}\ initive_{\text{i}} + gerund_{\text{i}} + participle_{\text{i}} + clause_{\text{i}} + phrase_{\text{i}}) \qquad (5)$$

If "i=0," then no such structure exists in a given sentence, but "i" can also be any positive number. If we combine functions with structures, then the complexity can be specified as:

$$\sum_{j=1}^{3}\sum_{i=0}^{n}\ (\text{inf}\ \ initive_{\text{i}} + gerund_{\text{i}} + participle_{\text{i}} + clause_{\text{i}} + phrase_{\text{i}}) \times (function)_{\text{j}} \quad (6)$$

Note also that "j" is from 1 to 3. (e.g., 1=adverbial, 2=nominal, 3=adjectival, though arbitrarily assigned. To further specify, suppose we have three different clauses embedded in the sentence without other structures, then:

$$\sum_{j=1}^{3}\sum_{i=0}^{n}\ (clause_{\text{i}}) \times (function)_{\text{j}} \qquad (7)$$

With the mathematical expression of the sentence patterns on the bases of the four assumptions, we still cannot decide the level of complexity or difficulty until the

introduction of English-Mandarin translation principles in which cognitive load is taken into account.

General L2-L1 translation principles

It is generally accepted that nominal structures in L2 and those in L1 are compound by nature, so when dealing with nominal structures, one can always follow the Mandarin word order, thus little problems exist in translation. As to adjectival structures in L2, especially those with more than one word unit, there is a reverse order in the correspondence of L1, so one must deal with the adjectival structures first prior to the nominal structures they modify. And such adjectival structures may easily cause problems, and thus considered more difficult. Lastly, for the adverbial structures, there is consistency between L2 and L1 sentences, the latter of which allows either following or reversing word order in a given sentence.

Word Order based on different functions of structure

As we have learned that any constituent of a given sentence can be categorized in terms of 4ps (four parts of speech), the next task is to make a comparison between L2 and L1 in terms of nominal, adjectival, and adverbial structures. We wonder whether there is any rule behind the differences between L2 and L1. Let's start from clausal structures (Cases 2, 3, and 4). The bracketed unit in Case 2 is nominal clause, in Case 3 adjectival clause, and in Case 4 adverbial clause. See the attached L1 translation of each sentence below:

Case 2: nominal clause

L2 → We know (that the girl is gracious).

L1 → 我們知道 (那女孩是典雅的)。

L1 order→ We know the girl is gracious.

Case 3: Adjectival clause

L2 → The girl (that you met) is gracious.

L1 → (你遇到的)女孩是典雅的。

L1 order→ the (you met) girl is gracious.

Case 4: Adverbial clause

L2 → The girl is gracious (that we all admire her.)

L1 → 這女孩是典雅的 (所以我們都愛慕她)。

L1 order→ the girl is gracious (that we all admire her).

As is shown, there are some rules behind the order of translations between L1 and L2. In Case 2 (where nominal clause lies), both L1 and L2 follow the same word order, while in Case 3 (adjectival clause), the adjectival clause is translated prior to its modified noun, and in Case 4 (adverbial Clause), like in Case 2, both L1 and L2 remain the same in word order. Thus we may draw tentative principles regarding translating L2 into L1 below:

Principle 1:

For nominal and adverbial clauses in translation, L1 and L2 follow similar word order.

Principle 2:

For adjectival clauses in translation, L1 and L2 are mutually reversed in word order.

As indicated above, it is then plausible to obtain a general rule regarding differences between L1 and L2 on different clausal structures. In reality, for adverbial clauses, one can either follow or reverse the structure order of each language (Guey, 2000b). Interestingly, such logic also applies to other structures (infinitives, gerunds, participles, and phrases).

Identifying the function of each structure unit

In most cases, the function of each structure unit can be identified through its semantic meanings (Chomsky, 1975). To specify, the structure unit can be nominal if it serves as a *Subject*, an *Object* of a VERB or followed by a preposition; it can be adjectival if it serves as a modifier of a noun; it can be adverbial if it provides additional message concerning time, space, state, cause, effect, result, condition, and etc. to the main structure. Confusions might arise if, for example, there are clausal structure units with conjunctions (such as "when", "that", "as" …etc.), because structures with the same conjunction may serve different functions (i.e., nominal, adjectival, and adverbial). In such a case, syntactic structure analysis may be supplementary. As mentioned earlier, the bracketed unit of the sentence in Case 2 is nominal clause in that it serves as the *Object* of the VERB. The sentence in Case 3 is adjectival clause because it modifies its antecedent (the girl). Finally, the sentence in Case 4 is adverbial clause for it is the result of the main structure. All the descriptions made above are mainly based on semantic connotations of each sentence.

To conduct syntactic analysis, we may focus attention on the differences of structures before or after the conjunction (the case in point is 'that') to see whether there is any rule regarding the differences. So we have (ignore 'that' for each sentence, and 'C*' stands for complement):

Case (2). We know (that) The girl is gracious.
S V S V C*
→ In the main structure, (S V) is incomplete.
In the nominal clause, "*The girl is gracious* (S V C*)" is complete.

Case (3). The girl (you met) is gracious.
S S V V C*

→ In the main structure, the girl is gracious (S V C*), it is complete.
In the adjectival clause, '__ *you met* (__S V)', it is incomplete and a subject is missing.

Case (4). The girl is gracious (we all admire her).
S V C* S V O

→ In both the main structure and the adverbial clause, the structures are both complete.

Again, such logic also applies to other structures (infinitives, gerunds, participles, and phrases).

So, we can summarize the above mentioned phenomena below:

Principle 3:

3.1 For nominal clauses, the main clause is incomplete, while the nominal clause complete.

3.2. For adjectival clauses, the main clause is complete, while the adjectival clause incomplete.

3.3 For adverbial clauses, both clauses are complete.

Though it is yet to be confirmed that such a principle may fit in all the possible structure units, it can still serve as a general guideline for identifying functions of each structure unit in a given sentence. In most cases, when we do the translation, information from other levels of analysis (semantic or syntactic) could freely interact with the message reflected at one level allowed to facilitate processing at other levels (Tyler & Malslen-Wilson, 1977). In this regard, it is suggested that one may start from syntactic analysis based on the principles proposed in the present paper, and shift to semantic analysis when there is ambiguity in the course of identifying function of each structure unit.

Determining levels of relative difficulty

Now comes one of the most important issues: how can we specify the levels of relative difficulty in translating any given structure from L2 to L1? As mentioned earlier, to conduct a quality translation, one needs to understand the semantic meaning of each single word, phrase and the syntactic structure on sentential level along with having a good command of relevant cultural connotations (Liu, 1997). In this paper we are going to confine the issue solely to the syntactic aspect of the text in translation. As Sweller put it (1994), the number of cognitive processes involved in a given task may substantially have to do with its level of difficulty. That is, the more cognitive processes involved in a given task, the more difficult it is. And cognitive processes can be operationally defined in terms of cognitive steps. We may start by specifying what possible steps are involved in translating a given sentence as below:

Step 1: Chunk each structure in terms of the five main structures (infinitives, participles, gerunds, clauses, phrasal expressions, and the main structure "S + V + O/C").

Step 2: Specify the function of each chunk in terms of its nominal, adjectival, and adverbial functions on the basis of theorem 3 (i.e., principles 3.1, 3.2, and 3.3).

Step 3: Decide the proper chunk order (when translating L2 into L1) on the basis of principle 1.

Step 4: Work on inserted structures (i.e., a certain chunk may be inserted by another chunk, as in recursive or mixed structures).

Still these cognitive steps involved in translation are not comprehensive, and some of the steps are of trivial difference, say, the steps involved in the recursive adjectival structures of *Subject* and of *Object*. The factors determining levels of difficulty may include: 1) the length of hunks, 2) the number of the structures, 3) the function of each chunk, 4) recursive

chunks, and 5) mixed chunks.

Length of chunks

Apparently, the longer the string of the structure in a sentence is, the more difficult it is. According to cognitive load theory (Sweller, 1988), if the number of chunks is beyond the capacity of working memory (e.g., more than 7), then it will cause cognitive load to our working memory, and thus makes it difficult. But the length of a sentence itself is not independent unless other factors are kept constant. To take the sentences below as examples:

Case 5

L2 → (a) The girl studies the book [*about psychology*].

(b) The girl studies the book [*which is about psychology*].

L1 →	(a)	這 女孩	讀	關於	心理學的	書
		the girl	studies	about	psychology	book
	(b)	這 女孩	讀	有關	心理學的	書

It can be predicted that the sentence (a) in Case 5 is easier because the length of the chunk is smaller than that in the sentence (b). This naturally leads to our first prediction stated below:

Prediction 1:

The longer the chunk in a given sentence, the more difficult it is to be translated, with other factors being equal.

Such a prediction can be expressed by the equation:

$$D = \mathrm{f}(S_L) \qquad (8)$$

That is, difficulty level (D) can be the function of structure length (S_L).

The number of chunks

The number of chunks and the length of chunks may not mean the same thing because some sentences are long but with few chunks. And also there are short sentences with many chunks. The number of chunks mentioned here may not have much to do with the length of the sentence. When considering cognitive load, the larger the number of the chunk in a given sentence, the more difficult it is. See the sentences below:

Case 6 (Radford, 1981, p. 49)

L2 → (b) Fred knows (that) John knows Joe.

(c) Jim knows (that) Fred knows (that) John knows Joe.

(d) Pete knows (that) Jim knows (that) Fred knows (that) John knows Joe.

L1 → (b) Fred 知道 John 認識 Joe

Fred knows John knows Joe.

(c) Jim 知道 Fred 知道 John 認識 Joe.

Jim knows Fred knows John knows Joe.

(d) Peter 知道 Jim 知道 Fred 知道 John 認識 Joe.

Peter knows Jim knows Fred knows John knows Joe.

From the sentence in Case 6 (b) to (d), the number of chunks is getting larger, we can easily predict that sentence (b) is the easiest, while sentence (d) the most difficult because of the different amount of cognitive load they have caused. Again other factors involved should be kept constant. Thus, we have:

Prediction 2:

The larger the number of the chunk in a given sentence, the more difficult it is to be translated, other factors being equal.

In the same vein, the prediction can be expressed by the equation:

$$D = \mathrm{f}(S_N) \qquad (9)$$

That is, difficulty level (D) can be the function of structure number (S_N).

The function of the chunk

As discussed earlier, L1 and L2 can be compound in nominal, coordinate in adjectival, and mixed in adverbial structures. Thus, the relative difficulty levels in translation among these three functional structures can be, nominal easiest, adjectival most difficult, while adverbial in between (e.g., the sentences (1), (2) and (3) on 'that' clauses mentioned previously.) Our prediction for this category is:

Prediction 3:

In translating a given sentence, the nominal chunks are the easiest, adjectival chunks the most difficult, whereas adverbial chunks in between, other factors being equal.

This prediction can be also elaborated as:

$$D = \sum_{i=1}^{3} F_i(S_L \times S_N) \qquad (10)$$

As is shown, the equation is a combination of predictions 1, 2, and 3, where 'F' denotes "functions" of structure (S); namely, three functions: nominal-1, adverbial-2, and adjectival-3 the weighed scores for each function 1, 2, 3 also indicate their relative difficulty, though arbitrary.)

Recursive chunks

Understanding recursive chunks can be less difficult than translating them into another language, because ordering the word sequence may cause cognitive load in the

working memory. This factor may also interact with length and the function of each structure in a 0given sentence. Hence, the more recursive structures in a given sentence, the more difficult in their translation especially when dealing with adjectival structures. See the sentences in Case 7 below:

Case 7. (Wingfield & Titone, 1998, p. 238)

L2 → (a) The mouse ate the cheese.

(b) The mouse (that) the cat bit ate the cheese. (n=1)

(c) The mouse (that) the cat (that) the dog chased bit ate the cheese. (n=2)

L1 →	(a)	這	老鼠		吃了	乳酪	
		the	mouse		ate	cheese.	
	(b)	這	(貓咬了的)		老鼠	吃了	乳酪
		the	(cat bit)		mouse	ate	Cheese.
	(c)	這	(狗追了的)	(貓咬了的)	老鼠	吃了	乳酪
		the	(dog chased)	(cat bit)	mouse	ate	cheese.

As is known, the “that” in the adjectival clauses embedded in the (a), (b), and (C) of Case 7 can be deleted because it serves *Object* case. Again, in relation to principle 2 (*For Adjectival Clauses in translation, L1 and L2 are mutually reversed in word order,)* we need to translate each of the recursive structure one by one. That is, we need to translate adjectival clause prior to its modified nouns in a sense that the L1 translation wording seems to be recursive accordingly. Apparently, sentence 7 (c) is more difficult than sentence 7 (b) when translated. Hence this naturally leads to:

Prediction 4:

In translating a given sentence, the more recursive chunks, the more difficult they are, other factors being equal.

The corresponding mathematical expression of this prediction can thus be:

$$D = \mathrm{f}(R_N) \qquad (11)$$

That is, difficulty level can be the function of the number of recursive structure (R_N).

$$D = \sum_{i=1}^{3} F_i (S_L \times S_N \times R_N) \qquad (12)$$

Again, the equation is a combination of Predictions 1, 2, 3, and 4. "R" denotes "recursive structures".

Mixed chunks

It is not uncommon to see sentences with chunks of various structures such as infinitive, participle, gerund, phrase, and clause combined (Rose & Carroll, 1974). Though the difficulties mainly lie in the identification of the function of each chunk, interactions with the length and the number of each chunk may aggravate cognitive load in the working memory when doing the translation. Considering the sentences below:

Case 8

L2 → (a) The girl is a champion.

(b) The girl seated by the river is a champion in a spelling bee contest.

(c) The girl seated by the river lying behind the hill is a champion in a spelling bee contest that is held once a year.

(d) The girl seated by the river lying behind the hill to be removed is a champion in a spelling bee contest that is held once a year.

L1 → (a) 這 女孩 是 冠軍

the girl is champion.

L1 → (b) 這坐在 河畔 的女孩 是 拼字 比賽的 冠軍

the seated by river girl is spelling contest Champion.

L1→ (c) 這坐在 山丘 後 河畔 的 女孩 是

the seated hill behind by river girl is

L1→ 每年 舉辦 一次 拼字 比賽 的 冠軍

every year hold once spelling contest champion.

L1→ (d) 這 坐在 將移除 的 山丘 後 的 河畔 的 女孩

the seated will remove hills behind by river girl

L1→ 是 每年 舉辦 一次 拼字 比賽 的 冠軍

is every year hold once spelling contest champion.

Note that English words under each L1 in the example above are re-ordered according to word order of Mandarin. From (a) to (d), the number of mixed structures is increasing, and it can be expected that the level of difficulty in translating them is increasing accordingly. Again, we can have:

Prediction 5:

In translating a given sentence, the more mixed chunks, the more difficult they are, other factors being equal.

Difficulty on sentences with mixed chunks is at least twofold in terms of cognitive load, first on chunking, and second on identification. The corresponding mathematical expression can be: (where m_i denotes the varieties of mixed structures, "*e*" –exponential function)

$$D = \sum_{i=1} e^{-m_i} \qquad (13)$$

Here we have extra exponential element "*e*," because exponential function arises whenever a quantity grows or decays at a rate proportional to its current value. It is assumed that additional 'recursive, mixed' structures (especially with adjectival clauses) will overwhelmingly aggravate the cognitive load in the course of sentence processing and translation. This can also be elaborated through the curve below:

For every extra- mixed structure (in numbers), the difficulty of processing them will be higher out of proportion (as in Y, denoting difficulty level). Such an assumption is not groundless, as suggested by the study on the segmentation of Chinese words during reading (Li, Rayner, & Cave, 2009). In their study, '$eccentricity_i = \ell^{-\gamma \times i}$' is used to indicate the level of eccentricity, where 'γ' refers to a parameter to be fit during the simulation, "*i*" refers to "location" of the target character. Though processing of words is basically different from that of sentences, it will be interesting to further examine such a mathematical interpretation on sentence processing level.

Then in combination of equations (12) and (13), thus, we have,

$$D = \sum_{j=1}^{3} \sum_{i=1}^{5} \mathrm{e}^{-\mathrm{m_i}} \times F_{\mathrm{j}}(S_{\mathrm{L}} \times S_{\mathrm{N}} \times R_{\mathrm{N}}) \qquad (14)$$

It is important to note from the above predictions that variables involved are interwoven, and thus to further clarify the relative influence of each variable requires carefully designed experimental studies. Indeed, it is quite difficult to precisely quantify or define each level of difficulty with regard to translating different structures in different situations with the knowledge at the present stage. The above-mentioned predictions along with their corresponding equations only serve as a tentative guideline when considering the possible

variables involved in the whole process.

Implications for further research

The present paper starts from introducing the assumptions about English syntax, then followed by the demonstration of sentence development on the basis of these assumptions, with the focus on the structural transformation laws (i.e., symmetry, interchangeability, recursion), which are essential for English-Mandarin translation on sentential level. Examples were given under each category to specify the laws, followed by the introduction of translation principles as the bridge between English and Mandarin syntactic structures. The relationships between functions (nominal, adjectival, and adverbial) and structures (infinitives, gerunds, participles, clauses, and phrases) were elaborated and specified to offer the framework for prediction of level of difficulty in English –Mandarin translation. Each of the five predictions was made on the basis of the amount of cognitive load involved in the course of translation under the guideline of translation principles, and the factors determining levels of difficulty (*the length of hunks, the number of the structures, the function of each chunk, recursive chunks, and mixed chunks)* were indicated through mathematical equations. Inevitably, there are problems that require further clarification and elaboration.

First, the assumptions proposed may not fit all the real life language data. For example, in the first assumption, **"**Only one main VERB is allowed in a sentence with the other verbs, if any, being transformed into infinitives, gerunds, participles, clauses or phrases," it is not uncommon to find sentences with more than one main VERB (e.g., *He came to the door, took out the key, plugged it in the keyhole, and turned*…). It is convenient to argue that these VERBs are essentially symmetrically aligned to fulfill description of an episode, and it is hard to imagine if the episode is described otherwise. To our conviction, the catch phrase

"better rules than without," is especially of much value when dealing with translation between two drastically different languages (e.g., English and Mandarin).

Second, the legitimacy of emphasis on syntax in building translation framework will beyond doubt invite criticism because translation is generally recognized as a cognitive process during which lexicons, collocations, syntax, cultural contexts, writers' intention are all interwoven, and none of which can be isolated and treated independently. Truly, all these are interdependent, but the syntactic differences do, in most cases, cause relatively more confusions than others in translation, especially when managing recursive adjectival clauses embedded in a sentence. As indicated, translating L2 into L1 is a complicated and exhaustive task, and dealing with syntactic aspects on two languages is certainly not enough, but it is well worth of more discussion since monitoring surface forms, as Wingfield and Titone (1998) put it, is the first step for understanding and translating L2 input information. In most cases, with the knowledge regarding how to chunk and specify each structure along with its individual function, one may then be more capable of assigning appropriate word orders in translations. In short, miscomprehension of structures can be fatal, whereas confusions derived from words can always be adjusted by accurate comprehension of syntax.

Next, the introduction of mathematical equations in the present study is, beyond doubt, an audacious, if not risky endeavor. Firstly, the mathematician and linguist, Noam Chomsky (1975) in his *The Logical Structure of Linguistic Theory* has suggested the underlying mathematical concepts, which did create an overwhelmingly fresh standpoint in looking at the nature of languages as well as their translation (e.g., from English to Mandarin.) Such a picture can be enlightening, thus contributive, given that the accuracy of such a modeling is at issue. Second, with the mathematical equations, studies of languages, specifically of

translation, can be made more operational and objective. For one thing, the embedded structures (especially those with adjectival units) are inherently difficult as compared with other structural units, and thus can be better elaborated and clarified through mathematical ideas. Besides, the mathematical concepts involved in the study are open to discussions, and therefore serve to trigger interdisciplinary inquiries. Third, like mathematical equations in other fields, translation of two languages is always faced with multi-dimensional complexity of various factors, and mathematical equations can be tools to facilitate scientific endeavor.

Interested researchers in the future may further elaborate 1) the validity of the mathematic equations proposed by designing experiments to test the predictions made in the present study (e.g., manipulating the number, the length, the recursive structures as independent variables, whereas time spent on translating each target sentence and level of accuracy as dependent variables), 2) proposition of translation instruction techniques or strategies to facilitate student translators' coping with sentences with high complexity.

References

Brooks, L. W., & Dansereau, D. F.(1983). Effects of structural schema training and text organization on expository prose processing. *Journal of Educational Psychology, 75*, 811-820.

Chan, M., Chau, H. L., & Hoosain, R. (1983). Input/Output switch in bilingual code switching. *Journal of Psycholinguistic Research, 12*, 407-416.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass : MIT Press.

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory (Logical Structure)*. New York: Plenum Press.

Chomsky, N. (1980). On cognitive structures and their development : A reply to Piaget. In: M. Piattelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 35 - 52). Cambridge, Massachusetts: Harvard University Press.

Fang, S. P. Tzeng, O. J. L., & Alva L. (1981). Intra-language vs. inter-language Stroop effects in two types of writing systems. *Memory and Cognition, 9*, 609-617.

Guey, C.C. (2000). *Macro scope of English Syntax (b)*. Taiwan: Language Learning Press.

Kohlers, P. A. (1963). Inter-lingual word associations. *Journal of Verbal Learning and Verbal Behavior, 2*, 291-300.

Kimble, J.P. (1973). Seven principles of surface structure parsing in natural languages. *Cognition, 2*, 15-47.

Li, X. S., Rayner, K., & Cave, K.R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology, 58*, 525-552.

Liu, M.(1997). *A Workbook for English-Chinese translation*. Taipei: Bookman..

Obler, L. & Albert, M. (1977). A monitor system for bilingual language processing. In M. Paradis (Ed.), *The fourth LACUS forum* (pp. 103-113). Columbia, SC: Hornbeam Press.

Poplack, S. (1980). Sometimes I will start a sentence in English termino en espanol: Towards typology of code-switching. *Linguistics, 18*, 581-618

Radford, A. (1981). *Transformational Syntax*. Cambridge university press, Cambridge.

Rose, R.G. & Carroll, J.F. (1974). Free recall of a mixed language list. *Bulletin of the Psychosomatic Society, 3*, 267-268.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312.

Timm, L. (1975). Spanich-English code-switching: el porque and how-not-to. *Romance Philology, 28*, 473-483.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 683-692.

Wang, Z. (1990). A *contrastive approach to language transfer in Chinese and English subject structures*. FL: Florida University Press.

Weinreich, U. (1953). *Languages in contact*. London: Mouton

Wingfield A., & Titone, D. (1998). Sentence Processing. In Gleason, J. B., & Ratner, N. B.(Eds.), *Psycholinguistics* (pp. 227-274). New York: Harcourt Brace College.