in a text is the T-unit[2], it cannot be denied that "being a semantic unit, a text is REALIZED in the form of sentences, and this is how the relation of text to sentence can best be interpreted" (Halliday and Hasan, 1976, p.293). Hunt introduces the T-unit, or minimal terminal unit, to measure development of sentences cuts in the writing of grade-school children, for young children tend to frequently produce what are known in the classroom as run-on sentences. However, since the reading texts to be analyzed are written by adult editors, revised by native speakers, the run-on sentences are avoided.

Bardovi-Harlig (1992) also asserts that a sentence-based analysis is superior to a T-unit analysis for the description of syntactic complexity in written production of adult second language acquisition. A sentence analysis allows knowledge of coordination to be represented in quantitative descriptions, thereby more accurately reflecting learner knowledge. Besides, the addition of the coordination index captures the shift from the use of coordination by beginning learners to the use of embedding by more advanced learners.

Furthermore, Thompson (2004) confirms that the sentence is an idealization of the written language. Thus, it is agreed that "the term 'sentence' is therefore best reserved to label stretches of written text bounded by full stops or the equivalent" (Thompson, 2004, p. 22). Since clause constituents will be distinguished in the process of this study, and to avoid too complicated and trivial cuts within a single sentence, the sentence is adopted as the analysis unit. Thus, the sentence is defined as

---

[2] Hunt (1970) defines T-unit as "a main clause plus all subordinate clauses and noncausal structures attached to or embedded in it" (p.4). It is argued that the weakness of T-unit analysis lies in its treatment of syntax alone. Besides, the validity of the T-unit as a measure and the overall syntactic complexity as a variable have been questioned. Concerning the usefulness of T-unit analysis, Gaies (1980) points out two limitations. First, it does not seem to be appropriate for the analysis of data from subjects with relatively low proficiency. The second limitation involves the discrimination power of the T-unit. While the length of error-free T-unit will distinguish learners with low proficiency from those with a high degree of proficiency, it is not, in Gaies's opinion, as sensitive an indicator of second language development as might be desired.

the concept of clause complexes: two or more clauses linked by co-ordination and/or subordination in a larger structural unit, in which a full stop or the equivalent marks the boundary.

3.2 Steps for Analysis

For the convenience of analysis, sentences in each reading text are numbered in order of occurrence. For example, (1.1) stands for the first sentence in the first paragraph, (2.1) the first sentence in the second paragraph, and so on. And then each text length, i.e. number of sentences will be counted. Next, in accordance with Thompson's (2004) categorization, Theme constituents are to be distinguished from Rheme constituents. Then cohesive devices, proposed by Halliday and Hasan (1976), are to be identified in the process of determining the patterns of TP. Combined Daneš's (1974) four main progression types with Cloran's (1995) suggestion, the patterns to be applied in this analysis are listed as follows:

Type 1: Rheme>Theme, viz., Daneš's Simple Linear TP

Type 2: Theme>Theme, viz., Daneš's TP with a continuous (constant) theme

Type 3: Split Theme, viz., Daneš's TP with derived T's

Type 4: Split Rheme, viz., Daneš's Exposition of a Split Rheme

Type 5: Theme>Rheme

Type 6: Rheme>Rheme

3.2.1 Identifying Cohesive Ties in Texts

On the basis of Halliday and Hasan's (1976) categories, five cohesive ties (i.e. reference, substitution, ellipsis, conjunction, and lexical cohesion) are identified in this study. The following text is used to demonstrate these cohesive ties.

(3.1) (a) If *you*'d like to give someone a *phone* for Christmas, (b) there are *plenty* to *choose* from. (c) *Whichever you go for*, (d) if *it*'s to be used on the BT [British Telecom] network, (e) make sure *it*'s approved – (f) look for the *label* with a green circle to confirm *this*. (g) *Phones labeled* with

a red triangle are prohibited. (McCarthy, 1991, p. 25)

The italicized items all have something to do with previous clauses. In (b) *plenty* serves as ellipsis, meaning "plenty of phones"; in (c) *Whichever* is also ellipsis of "Whichever telephone"; in (c) *you* is same personal reference of you in (a); in (d) *it* refers back to the telephone, and in (e) *it* is understood as substitution of "using the telephone on the BT", and in (f) *this* is interpreted as demonstrative reference to "the fact that it is approved". These are features of grammatical cohesion, and there is lexical cohesion as well: in (c) *go for* is a synonym of *choose* in (b), and lexical repetition of *phone* and of *label* respectively in (g).

3.2.2 Discriminating Patterns of Thematic Progression (TP)

As mentioned above, six types of thematic progression are to be applied in the analysis. By tracing the cohesive connections between sentences, the six types can be found. Take the following text for example (Theme part is italicized and underlined words or phrases indicate the cohesive ties for progression across sentences):

(3.2) (a) *When <u>a human infant</u> is born into any community in any part of the world*, it has two things in common with any other infant, provided neither of them has been damaged in any way either before or during birth.

(b) *Firstly, and most obviously*, <u>new born children</u> are completely helpless.

(c) *Apart from a power capacity to draw attention to their helplessness by using sound*, there is nothing <u>the new born child</u> can do to ensure his own survival.

(d) *Without care from some other human being or beings, be it mother, grand mother, sister, nurse, or human group,* <u>a child</u> is very unlikely to survive.

(Text is taken from Lautamatti, 1987, p. 92)

In Theme (a) *a human infant* progresses to Rheme (b), using the cohesive device of synonym, and thus becomes *new born children* in Rheme (b), and then the nominal element *new born children* continues progressing to the following Rhemes by

applying the device of same word (lexical repetition), and thus becomes *the new born child* in Rheme (c) and *a child* in Rheme (d) respectively. In this way, the progression patterns should be Theme>Rheme>Rheme>Rheme.

To sum up, the criteria for this analysis are based on Thompson's (2004) theme categories in various syntactic structures, Halliday and Hasan's (1976) categories of five cohesive ties, and Daneš's (1974) four main thematic progression types with Cloran's (1995) two additional patterns.