Increase The Efficiency of English-Chinese Sentence Alignment: Target Range Restriction and Empirical Selection of Stop Words

Wing-Kwong Wong[1], Hsi-Hsun Yang[2], Wei-Lung Shen[3],
Sheng-Kai Yin[2], and Sheng-Cheng Hsu[4,*]

[1] Department of Electronic Engineering
[2] Graduate School of Engineering Science & Technology
[3] Institute of Computer Science and Information Engineering
National Yunlin University of Science & Technology
Douliou, Yunlin, Taiwan, R.O.C.
{wongwk , g9417708, g9310811}@yuntech.edu.tw
[4] Department of Media and Design, Asia University
Wufeng, Taichung, Taiwan, R.O.C.
*Corresponding author: g9010802@yuntech.edu.tw (S.-C. Hsu)

Abstract: - In this paper, we use a lexical method to do sentence alignment for an English-Chinese corpus. Past research shows that alignment using a dictionary involves a lot of word matching and dictionary look ups. To address these two issues, we first restrict the range of candidate target sentences, based on the location of the source sentence relative to the beginning of the text. Moreover, careful empirical selection of stop words, based on word frequencies in the source text, helps to reduce the number of dictionary look ups. Experimental results show that the amount of word matching can be cut down by 75% and that of dictionary look ups by as much as 43% without sacrificing precision and recall. Another experiment was also done with twenty New York Times articles with 598 sentences and 18395 words. The resulted precision is 95.6% and the recall is 93.8%. Among all predicted alignment, 86% of the alignment is 1:1 (one source sentence to one target sentence), 8% is 1:2, and 6% is 2:1. Further analysis shows that most errors occur in alignments of types 1:2 and 2:1. Future work should focus on problems with these two alignment types.

Key-Words: - Sentence alignment, lexical method, statistical method, English-Chinese corpus, stop words, target range

1 Introduction

Machine translation is an important research topic in natural language processing and has been applied to different languages, such as English-Chinese [1, 2], Chinese-Japanese [3, 4], English-Japanese [5], English-France [6], English-German [7]. In general, a machine translation system needs to be trained with a lot of aligned data from a parallel corpus. One way to obtain aligned data is by doing automatic sentence alignment when given a source text and a target text. Thus, it is crucial to be able to do sentence alignment with minimum errors. Past research indicates that sentence alignment can be done with statistical, lexical and hybrid methods. Purely statistical methods, such as those based on sentence length, do not work well for languages with very different cultural origins, such as English and Chinese. Lexical methods require bilingual dictionaries and involve a large number of dictionary look ups and word matchings. We propose one technique to reduce the number of word matching and another technique to reduce the number of dictionary look ups.

Sentence alignment research has been active since early 1990s. Brown et al. [6] used an English-France parallel corpus to do alignment. They concluded that long source sentences were indeed aligned to long target sentences. Gale and Church [7], using English-France and English-German bilingual corpus, did alignment based on sentence length. Their method took advantage of the lengths of sentences to do alignment too. Differing from Brown's method in which prior probabilities of alignment types 1:1, 2:1, 1:2, etc. must be obtained, Gale and Church proposed an EM algorithm to compute the relevant variables. In Wu [8] and Xu and Tan [9], they used a Chinese-English bilingual corpus to do alignment. They combined the methods provided by Brown et al. and Gale and Church, with techniques for matching date, time and number. Their method got 96% precision.

McEnery and Oakes [10] showed that they used Gale and Church's method to align sentences from a Polish-English parallel corpus. Because the articles were from different domains, the precision was between 64.4% and 100%. When the method was used in a Chinese-English parallel corpus, the precision was less than 55%. The statistical method of sentence length does not require lexical knowledge and works fine when the source language and target language share similar a cultural origin, e.g., European languages such as English, French and German. When both languages differ in substantial ways, e.g, English and Chinese, one source sentence can be aligned to more than one target sentences and one target sentence can be aligned to more than one source sentences, poor precision and recall might result.

Haruno and Yamazaki [5] did sentence alignment of an English-Japanese parallel corpus. They compared statistical, lexical and hybrid methods and found that the hybrid method was better than the others, with 91.6% precision and 97.1% recall. Wu et al. [11], using the Chinese-English proceedings of the Hong Kong Legislative Council, made use of punctuation, the length of sentence and the number of bilingual sentences.

The above researches showed that alignment using a dictionary involves a lot of word matching and dictionary look ups. To address these two issues, we first restrict the range of candidate target sentences, based on the location of the source sentence relative to the beginning of the text. Moreover, careful empirical selection of stop words, based on word frequencies in the source text, helps to reduce the number of dictionary look ups

The rest of this paper is organized as follows. Our system architecture is introduced in Section 2. How to restrict the local region of target sentences when matching a source sentence is described in Section 3. Section 4 discusses how to empirically determine stop words in order to reduce dictionary look ups. Section 5 discusses some experiments and their results. Finally, discussion and conclusion are presented in Section 6.

2 System architecture

The architecture of our alignment system is shown in Figure 1. The inputs are one English text and its translation in Chinese. Each text is segmented into sentences. Each English sentence is broken into words and so is each Chinese sentence. Then the root form of each English word, which is derived from a WordNet library, is looked up in the Hownet [12] and the Sinica BOW WordNet [13] for its Chinese translations. The list is matched against the words of the target Chinese sentences, producing a match score. A good match between a pair of English-Chinese sentences produces a high score. For each Chinese sentence, its best match is the English sentence that produces the highest match score. Text processing is described in Section 1. Sentence matching is described in Section 2.

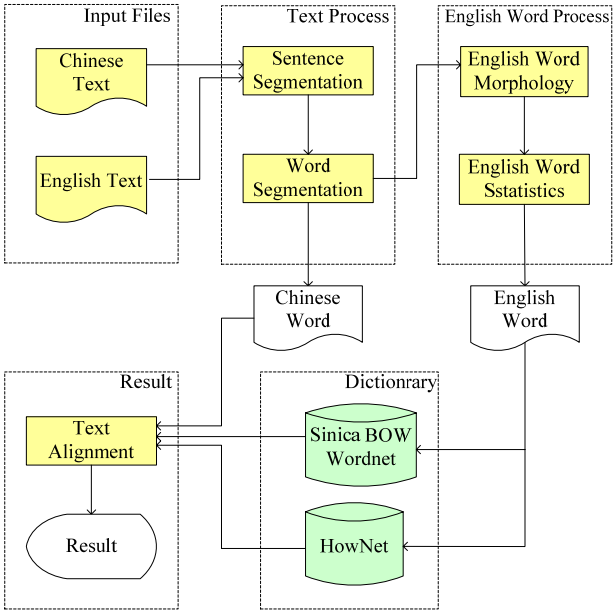

Fig. 1 System architecture

2.1 Segmentation of sentences and words

Each input English text is segmented into sentences with a Maximum Entropy method implemented by the OpenNLP library [14]. Such method would not mistake the period in "Mr." as the end of a sentence. Chinese sentence segmentation is based on punctuations like "。", "？", "。" each of which indicates the end of a sentence.

For each English sentence, white space and punctuations are used to separate words. For each Chinese sentence, we use Academia Sinica's CKIP [15] to extract the words, each of which might consist of more than one character.

After the words in an English sentence are obtained, we derive the root form of each word and count its frequency of occurrences in the given text. For each English word, we use the morphology library of WordNet.Net [16] to get the root form of the word. Without this morphology processing, we will not find the Chinese translations of tensed verbs and plural nouns in our bilingual dictionaries. If the word is a stop word, then it is ignored. Otherwise, its Chinese translations are found by looking up our

bilingual dictionaries. Stop words will be discussed in detail in a later section.

2.2 Bilingual dictionaries

Bilingual dictionaries are used to retrieve the Chinese translations of English words. These translations are used to match the words of each Chinese sentence. The Chinese sentence that produces the highest match score will be the target aligned sentence. The more dictionaries we use, the more translations an English word will have, thus increasing the chance of getting a match against a Chinese sentence. However, more dictionaries also increase the look-up time. Hence, there should be a balance between the total size of dictionaries and the computation time used for checking the dictionaries and matching the words of a Chinese sentence. To keep a good balance, we decide to use two electronic dictionaries HowNet and Sinica BOW WordNet, which is based on WordNet 1.6 (Table 1).

Table 1 Two electronic dictionaries for this project

Dictionary	**HowNet 2003**	**Sinica Bilingual WordNet**
#words	110,000	99,642

3 Restricted region of candidate target sentences

For the sake of simplicity, each English sentence can be matched against every sentence in the corresponding Chinese text. However, it is very unlikely for the first English sentence to match against a sentence near the end of the Chinese text. More generally, a region of English sentences should be aligned with a region of Chinese sentences at similar locations relative to the beginning of the English text and to that of the Chinese text. Based on this idea, we propose to restrict the matching region.

Let E_id be the index of the English sentence to be aligned, E_count, C_count be the number of English sentences and that of Chinese sentences respectively. Then the index of the first Chinese sentence in the region to be matched is:

$$Start\_c\_id = \begin{cases} E\_id, & if\ E\_id < \frac{1}{2}E\_count \\ \frac{1}{2}C\_count + E\_count - E\_id, & Otherwise \end{cases} \quad (1)$$

And the index of the last Chinese sentence in the region to be matched is:

$$End\_c\_id = \begin{cases} Start\_c\_id + \frac{1}{4}C\_count, \\ \quad if\ Start\_c\_id < \frac{3}{4}C\_Count \\ Ch\_Count, \quad Otherwise \end{cases} \quad (2)$$

For each word in an English sentence (say id = i) matching against the words of a Chinese sentence (say id = j), the Chinese translations of the English word are retrieved from the dictionaries and matched against the words of the Chinese sentence, working from long Chinese words to short words. Once a match is found, the Chinese translation is recorded and the same process repeats with the next English word of the sentence. Let M be the total number of matched pairs of English-Chinese words. Then the match score between English sentence i and Chinese sentence j is:

$$S(i,j) = \frac{2 * M}{Length(En(i)) + Length(Ch(j))} \quad (3)$$

After the match score S(i,j) of each pair (i,j) of English-Chinese sentences is obtained, we need to derive the aligned sentences with maximum likelihood. English sentence i is aligned with Chinese sentence m if S(i,m) is the maximum among S(i,Start_c_id),s(i, Start_c_id+1),….s(i,End_c_id). After we process all English sentences, three results are possible:

(1) Each English sentence is aligned with some Chinese sentence and each sentence Chinese is aligned with some English sentence. If one English sentence is aligned with a number of Chinese sentences, then the Chinese sentences are grouped together and aligned with the English sentence. Similarly, if one Chinese sentence is aligned with a number of English sentences, then the English sentences are grouped together and aligned with the Chinese sentence.
(2) If some English sentences remain unaligned. Then each remaining English sentence is matched against all Chinese sentences within the restricted range and the same process repeats as above.
(3) If some Chinese sentences will remain unaligned. Then each English sentence is matched against the remaining Chinese sentences within the restricted range and the same process repeats as above.

After this process, there can be several types of aligned sentences: 1:1, 1:n, n:1, n:m. Type 1:n means one English sentence is aligned with more than one Chinese sentences. Type n:1 means more

than one English sentences are aligned with one Chinese sentence. Type n:m means n English sentences are aligned with m Chinese sentences.

4 Empirical selection of stop words

Some common English words, e.g. "a", "be", are frequently used. They occur in many sentences in the source text and they have many translations that are likely to match words in many target sentences. Such words are generally not useful in identifying their corresponding Chinese words. These words are commonly known as "stop words". Most researchers simply provide a list of stop words in their sentence alignment projects. Instead, we propose to use an empirical method to identify stop words. For each word w, let Sentence_count be the number of English sentences in a text to be aligned against a Chinese text, and Sentence_hit_count(w) be the frequency of occurrences of w in the English text. Then R(w) is the average frequency of occurrences of w in one sentence. The equation (4) shows a typical list of words in a sample text and their corresponding r scores.

$$R(w) = \frac{Sentence\_hit\_count(w)}{Sentence\_count} \quad (4)$$

Table 2 The R scores of some words in an article

Word	R	Word	R	Word	R
the	2.25	Write	0.25	stone	0.15
of	1.75	900	0.2	than	0.15
and	1.25	date	0.2	think	0.15
in	1.25	other	0.2	university	0.15
be	1.2	san	0.2	wa	0.15
a	0.95	300	0.15	with	0.15
to	0.6	An	0.15	age	0.1
maya	0.55	archaeolog-ist	0.15	already	0.1
as	0.5	bartolo	0.15	around	0.1
at	0.5	blanca	0.15	art	0.1
b.c.	0.45	classic	0.15	before	0.1
preclassic	0.4	creation	0.15	carve	0.1
it	0.35	discovery	0.15	center	0.1
early	0.3	guatemala	0.15	century	0.1
from	0.3	have	0.15	ceremoni al	0.1
new	0.3	know	0.15	column	0.1
that	0.3	La	0.15	concept	0.1
time	0.3	Or	0.15	culture	0.1
by	0.25	paint	0.15	dark	0.1
find	0.25	period	0.15	emerge	0.1
mural	0.25	quatrefoil	0.15	epoch	0.1
on	0.25	say	0.15	example	0.1

Common words are those with high R scores so they are good candidates for stop words. The key question is how to set the threshold R value---all words with R value greater than the threshold should be considered stop words. If a big value is used, say 1, then there would be few stop words (e.g., the, of, and, in, be) and there would be little increase of efficiency. If a small value is used, say 0.1, there would be many stop words that might include those (e.g., university, stone, mural) that are informative for matching the target words. Below we will describe one experiment that indicates how to set the threshold value.

5 Experiment results

It is common to use the precision rate and the recall rate to evaluate the performance of an alignment method. The expert's alignment, which is assumed correct, is called a target. The alignment produced by the alignment method is a positive prediction. A positive alignment that is indeed a target is called a true positive, while one that is not a target is called a false positive. A negative prediction that is a non-target is called a true negative, while one that is a target is called a false negative (Table 3). The precision rate is the ratio of the number of true positives to the number of predicted positives (Equation 5). The recall rate is the ratio of true positives to the number of targets (Equation 6).

Table 3 Target and system prediction in a contingency table

Target (T) System prediction (S)	Target	¬Target
Positive	TP	FP
Negative	FN	TN

$$precision = \frac{|S \cap T|}{|S|} = \frac{TP}{TP + FP} \quad (5)$$

$$recall = \frac{|S \cap T|}{|T|} = \frac{TP}{TP + FN} \quad (6)$$

Table 4 Target range restriction methodlogy using in an article

	Precision	Recall	#matchings
Using target range restriction	90%	87%	40771
None	82%	70%	167178

We did a small experiment using one article to check which threshold R value should be used in classifying stop words. Figure 2 shows the accumulated number of look ups against the lexicon for words with R scores from 0.05 and up, if the translations of these words are used to match against the words in the Chinese sentence. The accumulated number of look-ups increases with the R score. If no stop words are used, then a total of 700 dictionary look ups are needed. A word with a high R score is used in many sentences and is a good candidate as a stop word. If a small threshold, say 0.5, is used, then many look ups are saved. Unfortunately, informative words are also considered stop words and ignored, resulting in poor precision of 71% and recall of 62% (Table 5). When the threshold is set to be 0.15, only about 400 dictionary look ups are needed, meaning a saving of 43% (Figure 2). This saving is achieved with no sacrifice in precision (90%) and recall (87%). When a higher threshold is set, less stop words are used and more look ups are done but precision and recall do not improve. Experiments with other articles show similar results. Thus, words with R scores greater than 0.15 are considered stop words.

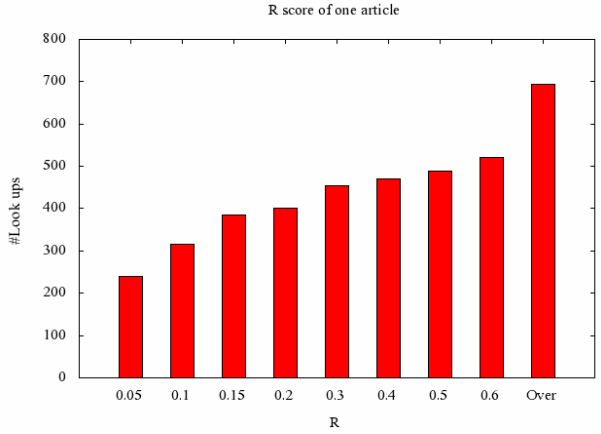

Fig. 2 R score of one article

Table 5 Precisions and recalls with different threshold R scores

Threshold R	Precision	Recall
0.05	71%	62%
0.10	79%	65%
0.15	90%	87%
0.20	90%	87%
0.30	90%	87%
0.40	90%	87%
0.50	90%	87%
0.60	90%	87%
Over	90%	87%

We tested our alignment method with 20 articles from the New York Times from January to December, 2006, with Chinese translations by the United Daily News of Taiwan. There were 598 sentences and 18395 words in these English articles. The predicted alignment types included only 1:1, 1:2 and 2:1 alignments (Table 6). Further analysis shows that most of the alignment errors occur in alignments of 1:2 and 2:1. The precision is 95.6% and the recall is 93.8 % (Table 7).

Table 6 Proportion of four alignment types in system prediction

Type	Frequency	Percentage
1-1	531	86%
1-2	51	8%
2-1	37	6%
Others	0	0%
Total	619	100%

Table 7 Precision and recall of our alignment method

System prediction	Target	¬Target	Total	Precision
Positive	598	20	618	95.6%
Negative	39	0	39	
Total	637	20	657	
Recall	93.8%			

6 Discussion and conclusion

Sentence alignment is important for automatically producing aligned sentences in a parallel corpus, which can be used to train machine translation systems. Past research shows that aligment using a dictionary can produce good precision and recall but can be time consuming in doing a lot of look ups of source words against the dictionary and a lot of matchings against the target words. We propose two techniques to address these issues. First, a local region of English sentences should be aligned with a similar region of Chinese sentences. By restricting the range of target sentences to be aligned, we reduce the number of word matchings by 75% while increasing the number of correct alignments. Moreover, we successfully reduce the number of look ups by as much as 43% in ignoring words with R scores above 0.15, with almost no sacrifice in precision and recall.

In an experiment involving twenty New York Times articles, which contain 598 sentences and 18395 words, the resulted alignment types include 86% of 1:1, 8% of 1:2, and 6% of 2:1. Further

analysis shows that most of the alignment errors occur in alignments of types 1:2 and 2:1. The precision rate is 95.6% and the recall rate is 93.8%. Being similar to the best results of English-Chinese sentence alignment published before (Table 8), these results are achieved with less computation. In future, we need to focus on the anlysis of the incorrect alignment pairs, find out what goes wrong and propose solutions to fix the problems, and extend to other language alignment like Chinese-Japanese.

Table 8 Precision rates of different English-Chinese alignment methods

Paper	Languages	Corpus	Method	Precision
Wu (1994)	English, Chinese	Hong Kong Hansard	Length based, time, date and number	96%
Xu and Tan (1996)	English, Chinese	News	Length based	96%
Wu et al. (2004)	English, Chinese	Hong Kong Hansard	Punctuation	98%
The present study	English, Chinese	News	Target range restriction, empirical selection of stop word	95.6%

Acknowledgments

This project is supported by the National Science Council, Taiwan (NSC 95-2520-S-224-001-MY3). The Sinica BOW WordNet is provided by Institute of Linguistics, Academia Sinica and Global View Co., Taiwan.

References:

[1] Z.-X. Chen and Q. Gao, English-Chinese machine translation system IMT/EC. *COLING* 1988, pp.117-122

[2] H.-H. Chen, C.-C. Lin, and W.-C. Lin, Building a Chinese-English WordNet for Translingual Applications. *ACM Transactions on Asian Language Information Processing*, Vol.1, No.2, 2002, pp.103-122.

[3] L. Mi, X. Luo, F. Ren, and S. Kuroiwa, How to Improve the Accuracy of Super-Function Based Chinese-Japanese Causative Sentence Machine Translation, *Proceedings of the 10th WSEAS International Conference on COMPUTERS*, 2006.

[4] D. Yin, M. Shao, P. Jiang, F. Ren, and S. Kuroiwa, Rule-based Translation of Quantifiers for Chinese-Japanese Machine Translation, *Proceedings of the 10th WSEAS International Conference on COMPUTERS*, 2006.

[5] M. Haruno and T. Yamazaki, High-Precision Bilingual Text Alignment Using Statistical and Dictionary Information. *Proceedings of the Annual Conference of the Association for Computational Linguistics*, 1996, pp.131-138.

[6] P. Brown, J. Lai, and R. Mercer, Aligning Sentences in Parallel Corpora. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 1991, pp.169-176.

[7] W. Gale and K. Church, A Program for Aligning Sentences in Bilingual Corpora. *Proceedings of the Annual Conference of the Association for Computational Linguistics*, 1991, pp.177-184.

[8] D. Wu, Aligning a parallel English-Chinese corpus statistically with lexical criteria. *Proceedings of 32nd Annual Meeting of ACL*, 1994, pp.80-87.

[9] D. Xu and C. L. Tan, Automatic alignment of English - Chinese texts of CNS news. *Proceedings of International Conference on Chinese Computing*, 1996.

[10] O. McEnery and M. Oakes, Sentence and Word Alignment in the CRATER Project. *In Thomas and Short (eds.) Using Corpora for Language Research*, 1996, pp.211-231. New York: Longman

[11] J.-C. Wu, T. C. Chuang, W.-C. Shei, and J. S. Chang, Subsentential translation memory for computer assisted writing and translation. *In The Companion Volume to the Proceeding of 42st Annual Meeting of the Association for Computational Linguistics*, 2004, pp.106–109

[12] HowNet, http://www.keenage.com/

[13] The Academia Sinica Bilingual OntoLogical Wordnet , http://bow.sinca.edu.tw.

[14] OpenNLP, http://opennlp.sourceforge.net/

[15] CKIP, A Chinese Word Segmentation System with Unknown Word Extraction and Pos Tagging, http://ckipsvr.iis.sinica.edu.tw/

[16] WordNet.Net, http://opensource.ebswift.com/WordNet.Net/

Appendix:- System's predicted alignment of an article from the New York Times

Type	En	From Murals and Glyphs, New Maya Epoch Emerges	Ch	壁畫和象形文字揭開馬雅文明新紀元
1-1	1	On the sacred (聖) walls (牆) and (和) inside (裡) the dark (黑暗) passageways (通道) of ancient ruins in Guatemala (瓜地馬拉), archaeologists (考古學家) are making (形成) discoveries that open (開) expanded vistas of the vibrant Maya civilization (文明) in its formative period (時期), a time (時) reaching back more than 1000 (千) years (年) before its celebrated Classic epoch (期) .	1	在瓜地馬拉(Guatemala)古代廢墟的聖(sacred)牆(walls)上和(and)黑暗(dark)的通道(passageways)裡(inside),考古學家(archaeologists)的發,使馬雅文明(civilization)形成(making) 期(epoch)生氣蓬勃的景象豁然開 (open)朗,那個時 (time)期(period)比著名的古典時期還早一千(1000)多年(years)。
1-1	2	The intriguing finds (發現) , including (包括) art (藝術) masterpieces (傑作) and (和) the earliest (早) known (知) Maya writing , are overturning (推翻) old ideas of the Preclassic period (時期) .	2	這些引人入勝的發現(finds),包括(including)藝術 (art)傑作(masterpieces)和 (and)已知(known)最早(earliest)的馬雅文字,已經推翻(overturning)過去對前古典時期 (period)的認知。
2-1	3	It was not (不是) a kind (種) of dark (黑暗) age (時代), as once (一) thought (想), of a culture that (那) emerged and bloomed in Classic times, at places like the spectacular royal ruin at Palenque beginning about A.D.250 and extending to its mysterious collapse around 900.	3	那 (that)不是 (not)過去所想(thought)的一(once)種(kind)黑暗 (dark) 時代(age)。
2-1	3	It was not a kind (等) of dark age (大) , as once thought , of a culture (文化) that (那個) emerged and (並) bloomed in Classic times (時) , at places (地方) like (像) the spectacular (壯觀的) royal ruin (廢墟) at Palenque beginning (發生) about (約) A.D.250 and extending (延) to its mysterious (神秘) collapse (崩潰) around 900 .	4	那個(that)在古典時(times)期興起並(and)盛綻的文化(culture),發生(beginning)在像(like)壯觀的(spectacular)帕倫克王室廢墟 (ruin)等 (kind)地方(places),始於西元 250 年左右,延(extending)續到大(age)約 (about)西元 900 年神秘(mysterious)崩潰(collapse)。
1-2	4	At the derelict ceremonial center (中心) of pyramids (金字塔) and (和) wide (寬闊) plazas (廣場) , a site (址) in remote (偏遠) northeastern Guatemala (瓜地馬拉) known as San Bartolo , archaeologists (考古學家) have uncovered the unexpected remains of murals (壁畫) in vivid (活) colors depicting (描繪) the Maya mythology (神話) of creation and (和) kingship (王權) .	5	在瓜地馬拉(Guatemala)東北偏遠 (remote)地區一個叫做聖巴特羅的遺址(site),考古學家 (archaeologists)在荒廢的金字塔(pyramids)祭典中心 (center)和 (and)寬闊(wide)廣場(plazas),挖掘到出人意表的鮮活(vivid)壁畫(murals)遺跡,壁畫描繪 (depicting)開天闢地和 (and)馬雅王權 (kingship)的神話(mythology),繪於西元前 100 年,旁邊一行象形文字見時間更早一、兩個世紀,證明是已發展成熟的書寫系統。

1-2	5	The murals (壁畫) date to 100 (世紀) B.C. , and (和) nearby , a column (塔) of hieroglyphs (象形文字) , a century or two (兩) oider , attests (證明) to an already (已) well-developed writing (書寫) system (系統) .	5	在瓜地馬拉東北偏遠地區一個叫做聖巴特羅的遺址,考古學家在荒廢的金字塔(column)祭典中心和 (and)寬闊廣場,挖掘到出人意表的鮮活壁畫(murals)遺跡,壁畫描繪開天闢地和馬雅王權的神話,繪於西元前 100 年,旁邊一行象形文字 (hieroglyphs)見時間更早一、兩 (two)個世紀(100),證明(attests)是已(already)發展成熟的書寫(writing)系統(system)。
1-1	6	News (新) of the discoveries (發現) , announced (布) in the last (去) six (六) months (月) by an American-Guatemalan team (團隊) led (引起) by William A. Saturno of the University (大學) of New Hampshire, is reverberating through the small (小的) community of Mayanists .	6	新(News)罕布(announced)夏大學 (University)的威廉沙特諾帶領的美國和瓜地馬拉團隊(team),過去 (last)六 (six)個月(months)發布的這些考古發現(discoveries),在小小的 (small)馬雅學界引起 (led)回響。
1-1	7	They (他們) see these (這些) and (和) other (其他) recent finds (發現) as strong (有力) evidence (證據) for (爲) the early (早期) origin (起源) and (和) remarkable continuity of the culture (文化) 's concepts (觀) of cosmology and possibly (可能) governance (支配) over more than a Preclassic millennium (一千年) .	7	他們 (They)認爲(for),這些 (these)和(and)近期的其他(other)發現(finds),是馬雅文化(culture)宇宙觀(concepts)的早期(early)起源(origin)和 (and)驚人延續的有力(strong)證據(evidence),可能(possibly)支配(governance)不只前古典時期那一千年(millennium)。
1-1	8	Coming (來) away (去) from a visit to San Bartolo , Michael D.Coe , a retired (退休) Yale Mayanist who was not (沒有) involved in the work (工作) , called the murals (壁畫) " one (一) of the greatest Maya discoveries (發現) of all (一) time (次) . "	8	耶魯退休 (retired)馬雅學者麥可．柯伊沒有(not)參與這次(time)考古工作(work),他去(away)了聖巴特羅一(one)趟,形容這些壁畫 (murals)是「有史以來(Coming)最偉大的馬雅發現(discoveries)之一(all)」。