A Study on the Variety of Prepositional Phrases and an Algorithm for Extracting them in English Texts

Morteza Abdolrahim Kashi and Abolfazal Keighobadi Lamjiri

Concordia University
Department of Computer Science
Montreal, Quebec, Canada
m_abdolr@cs.concordia.ca, a_keigho@cs.concordia.ca

Abstract— **There are different kinds of prepositional phrases (PPs) in English language. Extracting PPs is a very interesting topic in Computational Linguistics. The structure of each prepositional phrase is different. To extract each prepositional phrase (PP) in a text, we need a chunker. In this paper, we provide a grammar to chunk PPs. Prepositional phrases are divided into the following categories: Temporal PPs, Locative PPs, Directional PPs, Manner PPs, Purpose PPs, Beneficiary PPs Qualitative PPs, Quantitative PPs and Adjectival PPs[1], [2], [15], [16], [17]. In this paper, we explain the different kinds of Prepositional Phrases and we give a grammar to chunk PPs in a text. The result of the PP chunker will be given and we evaluate our PP chunker as it is concerned to its performance.**

I. Introduction

A prepositional phrase is a group of words including a preposition and a noun, pronoun, or group of words used as a noun phrase. In other words, a formula for prepositional phrases could be defined as:

Prepositional phrase = preposition + object of the preposition + modifiers [18]

Considering the above definition, we can classify prepositional phrases as following:

- adverbial prepositional phrases
- adjectival prepositional phrases
- multi-component prepositional phrases

Adverbial prepositional phrases are divided into 9 subcategories: locative, temporal, directional, manner, purpose, extent, and beneficiary, qualitative and quantitative prepositional phrases [1], [2], [15], [16], [17]. Locative prepositional phrases are divided into two subcategories by themselves. The first subcategory for locative prepositional phrases refers to PPs indicating a known or an explicit place. The other category is the category of locative prepositional phrases which refer to non-explicit or general locations. Similarly, temporal prepositional phrases are divided into two subcategories. The first category refers to the temporal prepositional phrases which refer to a specific and known time, while in the second category, an unknown time is pointed.

A directional prepositional phrase explains about the direction to which the verb has occurred [19]. A manner prepositional phrase is the one showing the manner by which a verb has occurred [16]. A purpose prepositional phrase explains about the purpose for which the verb occurs [19]. An Extent prepositional phrase refers to the extension in which a verb occurs. These PPs usually use prepositions such as for, along, across, during, etc [15]. In beneficiary prepositional phrase we can show that an action has been done for something or foe someone [17]. A qualitative prepositional phrase refers the quality of the action that has been done [1], [2]. A quantitative prepositional phrase refers to the quantity of an action [1], [2].

In the following sections, we explain the above classification by providing related examples. The grammar for a shallow parser is given for recognizing all kinds of prepositional phrases. A simple algorithm based on searching for keywords is also introduced and implemented for categorizing each PP detected by the PP chunker. Then, we will discuss the performance of this chunker, by evaluating our system with the precision and recall measures. Finally, we will bring the conclusion to this paper and the references we used in preparing this work.

II. Prepositional Phrase Categories

A. Adjectival Propositional Phrases

One category of prepositional phrases is the adjectival prepositional phrase class. An adjectival PP modifies a noun or pronoun. It always comes immediately

after the noun or pronoun it modifies. For example we can refer to the following sentence:

Mark is a journalist **with the highest rank.**

As we can see, this phrase modifies the noun phrase *"a journalist"*.

B. Adverbial Propositional Phrases

As we discussed earlier, one class of prepositional phrases is the adverbial category[1], [2]. An adverbial prepositional phrase modifies a verb, adjective, or adverb. It is used to tell *when, where, how, or to what extent* about the word it modifies. This category is divided into 9 subcategories (see figure 1): locative, temporal, directional, manner, purpose, extent, beneficiary qualitative and quantitative prepositional phrases.

B.1 Locative Propositional Phrases

Locative prepositional phrases are ones which explain the place where something has occurrence. For example we can refer to the following sentence:

Fred swam *in his swimming pool.*

The adverbial prepositional phrase is shown in bold. This prepositional phrase explains about where swimming has occurred. As another example see the following sentence:

John found his stolen car left **in Saint Luis Street.**

This prepositional phrase explains the place of *finding the stolen car.*

Locative PP category can be divided into two subcategories by itself. The first category is prepositional phrases which point to explicit locations such as Washington, Boston, or education ministry. Explicit locations are all the known locations such as cities, provinces, countries, organizations, and regions. For example, the following example shows an explicit locative PP:

He was born **in Canada.**

The other kind of locative prepositional phrase is a PP pointing to the non-specific locations. Those locations do not have a proper known names, but general names. For example school, university, coffee shop, and street are non-explicit locations. We can refer to the following example to show a locative prepositional phrase which point to a non-explicit location:

He usually eats **in school.**

B.2 Temporal Propositional Phrases

Temporal prepositional phrases are the ones which refer to time[4]. For example,

John ate his breakfast **after making a phone call.**

This category is divided into two subcategories. In the first subcategory, temporal prepositional phrases refer to a specific and known time. For example, in

John eats his lunch **at 1 pm.**

the prepositional phrase points to a specific time which is *1 Pm.* In the second subcategory of temporal prepositional phrases, the phrase points to an inexact time, but it still refer to a kind of time. For example in

John sleeps **during the day.**

The prepositional phrase provides an interval in which the sleeping has happened.

B.3 Directional Prepositional Phrases

A directional prepositional phrase explains about the direction to which a verb has occurred[19]. For example we can refer to the following example:

John and Jennifer walked **towards the woods**

As you can see the phrase *towards the woods* is a directional prepositional phrase since it refers to the direction to which John and Jennifer walked to.

B.4 Manner Prepositional Phrases

A manner prepositional phrase is the one explaining the manner in which a verb occurs[16]. For example we can refer to the following sentence:

Our teacher explains **in a good way**.

As you can see the phrase *in a good way* is a manner prepositional phrase which explains the manner in which the teaching process occurs.

B.5 Purpose Prepositional Phrase

A purpose prepositional phrase is a PP which explains about the purpose of occurrence of a verb[19]. For example we can refer to the following example:

Save money **for future confidence**

As you can see the prepositional phrase *for future confidence* explains the purpose of saving money in our sentence.

B.6 Extent Prepositional Phrases

An extent prepositional phrase refers to the extension of an action which has been done by a verb[15]. These PPs usually use prepositions such as for, along, across, during, etc. For example we can refer to the following sentence:

I drove **for 5 miles yesterday**

As you can see, the prepositional phrase *for 5 miles* shows the extension and duration of driving in our sentence.

B.7 Beneficiary prepositional phrases

A beneficiary prepositional phrase shows that an action has been done for something or someone[17]. These kinds of PPs usually use preposition *for* in their structure. For example we can refer to the following sentence:

Jack bough a present **for Susan**

As you can see, the prepositional phrase *for Susan* shows the action of buying has occurred for Susan.

B.8 Qualitative Propositional Phrases

Some propositional phrases modify the verb phrases in a sentence. They describe the quality in which the verb has occurred. We can refer to the following sentence:

John left the room **in a quiet manner**

Prepositional phrases like this are less frequent compared with temporal and locative PPs. Some of them can easily be converted to adverbs, or adjectival phrases that modify the subject or object. For example, the above situation can be expressed as *John left the room quietly* that replaces the phrase with an adverb. Here is an example of a more complex phrase.

He can assist me **in choosing the right service**

B.9 Quantitative Propositional Phrases

Another category of propositional phrases describe some quantity for the object in the sentence. Consider the following example:

The company paid **for $1,000 principal amount**

Here, paying is general. However, the attached PP explains about the quantity of the payment. This is another example from Wall Street Journal:

his four-stock portfolio is still exposed **to a great deal**

C. Multi-components Propositional Phrases

The third type of prepositional phrases is multi-component prepositional phrases[1], [2]. These kinds of prepositional phrases happen when two or more prepositional phrases follow each other. They may modify the same word, or one phrase may modify the object in the preceding phrase. For example we can refer to the following phrase:

I arrived at university **on time.**

This prepositional phrase is composed of two prepositional phrases: *at university* and *on time*. The first one is a locative prepositional phrase and the second is a temporal one. Both are adverbial prepositional phrases. The point here is that both prepositional phrases modify the verb *arrived*. This is not always the case however, sometimes in a multi-component prepositional phrase, one PP can modify one object and another component modifies another object. For example we can refer to the following sentence:

The house is **on the top of the mountain**

In this example, the phrase *on the top of the mountain* is a multi-component prepositional phrase. This phrase is made of the following components: *on the top* and *of the mountain*, and the phrase *on the top* modifies *is* while *of the mountain* modifies the *top*.

The following chart summarizes the above discussion:

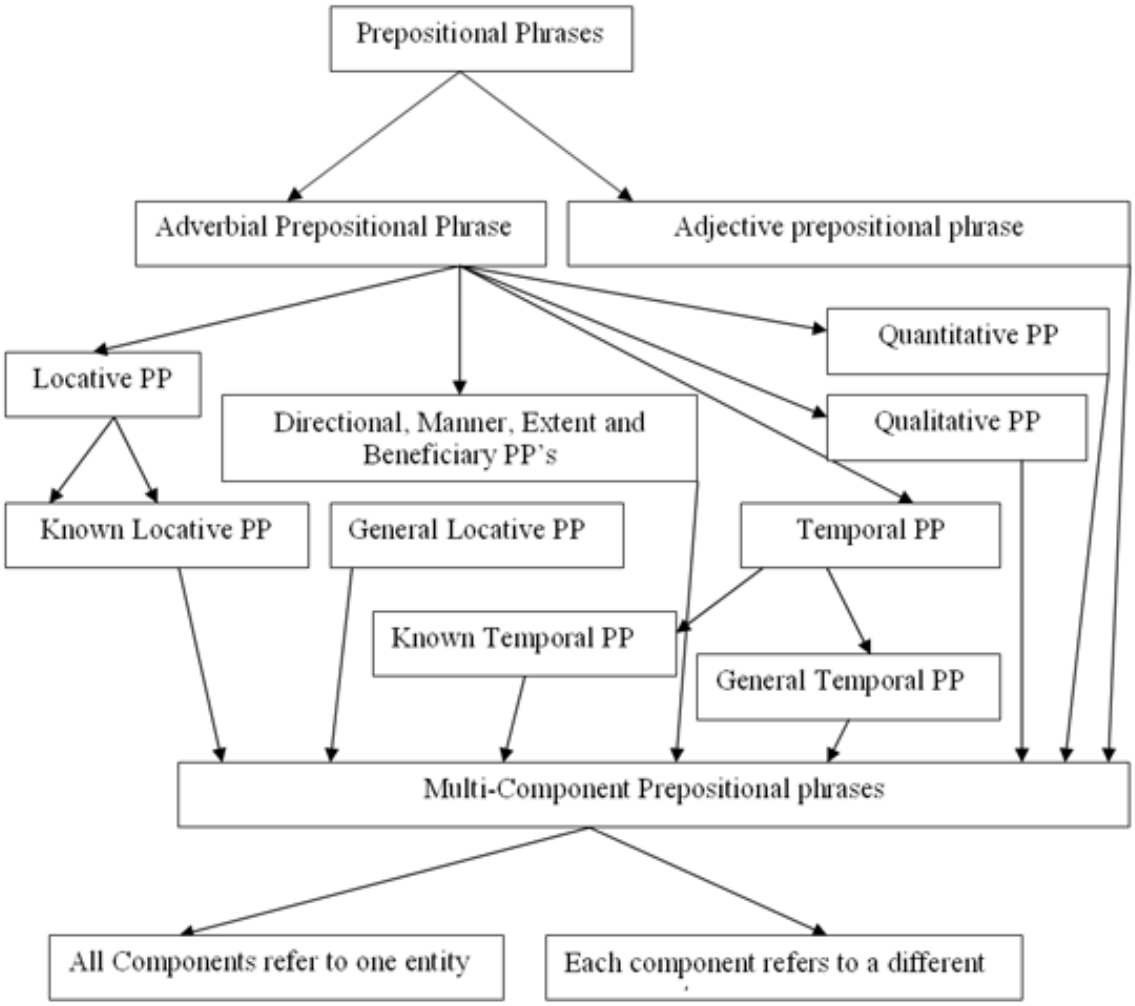

Fig. 1. Prepositional phrase categories.

III. Catching Prepositional Phrases in Our Grammar

The grammar for chunking prepositional phrases is enclosed in Appendix. The law of our grammar for catching prepositional phrases is as follow:

Any prepositional phrase is made of the following two structures: a- PP+NP b- PP+PP

That is: any prepositional phrase is made of a preposition followed by a noun phrase or a prepositional phrase followed by another prepositional phrase.

In the system we have available for parsing, prepositions fall into two categories. The first category has only one preposition which is *to*. This preposition is tagged as *TO* in our parser system. The other category of the prepositions is the rest of the prepositions rather than *to*. These prepositions are tagged as *IN*. Please find the grammar for chunking the prepositional phrases in the Appendix.

IV. Recognizing Different Kind of Prepositional Phrases

For classifying PPs, we use the head of the prepositional phrase. The head of each prepositional phrase is the last word in that PP. For example, if we have the phrase *in one minute* as our prepositional phrase, the head of our PP is the word *minute*. As we can see, this word has the sense of time. So we can conclude that our PP is a temporal PP. We have found most temporal heads from Gazetteer website and put them in a separate file to make a database to compare the head of our prepositional phrases with the entries in our database. We have done the same thing with all other kinds of prepositional phrases.

V. Analysis of the results

Our parser is designed to chunk prepositional phrases. A C++ source code is used (ppExtractor.cpp) to pick up prepositional phrases which are chunked by our parsing system(The parsing code is given in Appendix).Another C++ source code called matcher.cpp is used to specify all sorts of prepositional phrases extracted by our parser.

The above result has been taken from 10 different texts. A number of sample texts are given in Appendix A. All different kind of prepositional phrases are extracted by our system and the statistical results are given in above table.

Statistics	% Recall	% Precision
All PP's	42%	62%
Adjective PP	36%	50%
Locative	38%	71%
Temporal	52%	76%
Directional	35%	49%
Manner	21%	65%
Purpose	26%	81%
Extent	46%	80%
Beneficiary	48%	85%
Qualitative	25%	69%
Quantitative	33%	65%

TABLE I

The Recall and precision for extracting each kind of prepositional phrase and that of all prepositional phrases are given in the above table..

VI. Related Work

Noun-phrases should be categorized beforehand. *Name Entity Recognition* is done in a variety of ways. [22] exploits Decision Trees for learning NEs. [23] uses statistical learning methods enhanced with strong heuristics for catching some features of words in a category; *Mr.* for example is learned as a good predictor term for the PERSON category.

VII. Future Work

There are some situations where the PP chuker grammar does not perform well; for example, two parted verbs like *pick up* will provide a preposition which is always considered to begin a prepositional phrase (i.e. *up her third award*). Our grammar also is weak in parsing long, complex sentences, and usually fails to process them. If relaxed however, the number of parse trees produced will be very high, that is usually not acceptable because of high structural ambiguity it provides.

WordNet can be embedded in our program to help categorizing NPs in PPs; it would help in recognizing categories like Manner, Direction and Extent which are very difficult to be expressed in wordlists. WordNet however does not contain technical terms, which wordlists can provide easily.

Another extension to our algorithm would be to pick a better candidate word from a PP. Now decisions are always done on the last word in a PP, assuming it is the head of the included NP; there are contra-

dictory examples to this. For example, *at school daily* is categorized as a temporal PP, while it is actually a locative one.

In addition, Taking just one word as the representative of a propositional phrase leads to some mistakes in pp-categorization. For example, *Saint Luis* is a location while *Luis* is the name of a person.

Finally, using a lemmatizer is one of the extensions which would greatly affect the performance of our program. Usually, preparing every possible derivation of a word in wordlists is very difficult, if not impossible. It results in Huge word lists which degrades searching performance. The head word in a PP can be lemmatized as the first step, and the root of the word then can be used in matching against the wordlists available in the system.

References

[1] Prepositional phrases URL: http://www.english.uiuc.edu/cws/wworkshop/grammar/prepositional_phrases.htm
[2] Groups and Phrases URL: http://www.arts.usyd.edu.au/departs/english/grammar/GroupsandPhrases/GroupsandPhrases8.htm
[3] The Comparison of Machine Learning Algorithms for Prepositional Attachment URL: http://www.ai.mit.edu/people/jimmylin/papers/Mitchell02.pdf
[4] Categorial Grammar and the Semantics of indexical Prepositional Phrases URL:http://vivaldi.sfs.nphil.uni- tuebingen.de/ arnim10/Lehre/WS03.4/Perfektseminar/FrancezSteedman240303
[5] A syntactic Analysis of Instrumental prepositional Phrases URL: http://www.bib.uab.es/pub/linguistics/1132256Xv9p53.pdf
[6] Empirical Study of predictive powers of simple attachment schemes for post- modifier prepositional phrases URL: http://acl.ldc.upenn.edu/P/P90/P90-1004.pdf
[7] Realistic parsing: Practical solutions of difficult problems URL: http://www.uqtr.ca/ delisle/Recherche/conf-papers/PACLING-95.pdf
[8] A Rule-Based Approach to Prepositional Phrase Attachment Disambiguation URL: http://acl.ldc.upenn.edu/C/C94/C94-2195.pdf
[9] Classifying preposition semantic roles using class-based lexical association URL: http://www.cs.nmsu.edu/TechReports/2002/013.pdf
[10] Extensible shallow parsing for semantic nets URL: http://www.research.ibm.com/ecvg/pubs/jhc-nets.pdf
[11] Tense and space in natural language URL: http://www.sfu.ca/person/dearmond/801/Space.Tense.Paper.pdf
[12] Ferro L., Sundheim B., Gerber L., Instruction Manual for the Annotation of Temporal Expressions, Terqas Workshop, January 2002.
[13] The TIPSTER Gazetteer CLR: Computing Research Laboratory, clr.nmsu.edu ftp site. Accessed: Nov 2003. URL: http://clr.nmsu.edu/Tools/CLR/
[14] [preposition list] J. Essberger, English Prepositions Listed, English Club ESL Resources
[15] Spatio-Temporal URL:http://www.darmstadt.gmd.de/publish/komet/gen-um/node142.html
[16] Parts of Speech URL: http://eslus.com/LESSONS/GRAMMAR/POS/pos7.htm
[17] The Markers SA and KAY/KINA http://www.seasite.niu.edu/Tagalog/Grammar
[18] Phrase Formulas URL: http://leo.stcloudstate.edu/grammar/phraseformulas.html
[19] PREPOSITIONS URL: http://www.ccbc.cc.pa.us/lc/OWL/SentCon/Prep.html
[20] Adjectice Prepositional Phrases URL: http://www.faculty.uaf.edu/ffmee1/adjective_prep_phrases.html
[21] The Semantics of Locative Prepositional Phrases URL: http://cl.snu.ac.kr/nam/papers/abstract.PDF
[22] Sekine S., Grishman R., *A Decision Tree Method for Finding and Classifying Names*
[23] Palmer D.D., and Day S.D., "A Statistical Profile of the Named Entity Task", Proceedings of the Fifth Conference on Applied Natural Language Processing, Washington, D.C., March 31- April 3, 1997

Appendix

Five Test Texts Text1 Contrary to what might be expected based on the headline on John Dorfman's recent Money Matters article , I was able to stand proudly before my undergraduate finance students and proclaim that the findings of your yearlong experiment on stock picking is completely cons with what they have been taught in the classroom.

In particular I do not find the fact that your group of pros' monthly selections of four stocks outperforms the market in general to be inconsistent with market efficiency.

Mr. Dorfman states that an investor who invested $100,000 a year ago in the first four stocks selected by your pros and then sold those one month later purchasing the four new pro picks , and repeated this process for the year would have accumulated $166 537 excluding account dividends , taxes and commissions .

In contrast an investor holding the Dow Jones portfolio over the year would have accumulated only $127446.

Accepted theories of asset pricing offer a perfectly legitimate explanation.

Accepted theories state that investors require higher returns on riskier investments.

Thus, rather than seeing the excess returns to the pro-selected portfolio as being abnormal , I see those returns as simply compensations for taking on added risk .

I believe the risk for each individual stock selected by your pros is very large.

If you asked me to select a stock with the highest expected return, I would select a stock with the greatest amount of undiversifiable risk, as I am sure your pros do.

Your hypothetical investor is simply being compensated for taking on this added risk.

Moreover, your hypothetical investor has forsaken the gains to be had in reducing risk by diversifying his portfolio.

Four-stock portfolio is still exposed to a great deal of unnecessary risk.

This means the returns can vary a great deal.

Mr Dorfman provides confirming evidence of this phenomenon when he reports that your staff of dart throwers would have accumulated only $112383 by randomly selecting four new stocks to be held in a portfolio each month .

Your Investment Dartboard article misses the target.

The fact that stock pickers have bested a randomly selected portfolio in eight of 12 months has no bearing on the efficient-market theory .

What matters is that the stocks recommended by your pros tend to be substantially riskier than a diversified portfolio .

For example, your pickers' recommendations for the coming month are, on average, 22.5% riskier than holding the market portfolio according to Value Line's Beta estimates .

James Morgan's pick for October is a substantial 35% riskier than the market portfolio; his lauded Thermo Electron pick is 40

OUTPUT-Text 1

```
on John Dorfman 's recent Money Matters
before my undergraduate finance/QUAN
that the findings
of your yearlong experiment
on stock/QUAN
In particular
in general to be inconsistent with market efficiency
in the first four stocks/QUAN
In contrast
over the year/TEMPORAL
of asset
that investors
on riskier investments
for each individual stock/QUAN
by your pros
with the greatest amount of undiversifiable risk
to a great deal of unnecessary risk
of this phenomenon
on the efficient-market theory
that the stocks
to Value Line 's Beta estimates/QUAN
```

Text2

Use Word 2000 to create and format documents easily by using templates incorporate comments for collaborative learning or save documents as Web pages

Gather and manage data in Excel 2000 by using worksheets and the ability to publish on the Web Create and organize powerful presentations that include animation, and sounds with PowerPoint.

Communicate easily with colleagues and students manage your schedule by using the Calendar and much more by using the collaborative features of Outlook 2000. Use the structure and flexibility of Access 2000 to create databases to store and track important information.

Publish professional quality newsletters Web sites for your class or school easily by using the innovative features of Publisher 2000.

This comprehensive Table of Contents provides a list of the chapters in the Microsoft Office 2000 Professional guide.

Learn how to use the tools and features of Office 2000 Professional effectively in the classroom. This four-page chapter is a brief summary of the components of Office 2000 Professional.

This 12 page chapter contains a summary of the features of Office 2000 Professional including a description of Office tools and information about how to use Toolbars. The Clip Gallery to help you and your students

get the most out of Office 2000.

Learn how to use Word 2000 effectively with the information in this 24-page chapter. Use the many new features in Word 2000 to create Web, e-mail, and print documents quickly and easily. Take advantage of the new Web features that make it easy to create and post information to the Web without needing to know HTML.

This 22 page chapter provides the information you need to gather research data create worksheets and integrate your findings into reports. It charts effectively with Excel 2000. Use new features information sharing and drag and drop to collaborate with others and share data on the Web.

OUTPUT-Text 2

```
for collaborative learning
 with PowerPoint
 of Access 2000/QUAN
 of Contents
 in the Microsoft Office 2000 Professional guide
 of the components of Office 2000 Professional
 of the features
 of Office 2000/QUAN
 on the Web
```

Text3:

Sun Microsystems Inc., a computer maker, announced the effectiveness of its registration statement it is for $125 million of 6 3.8% convertible subordinated debentures due Oct 15, 1999. The company said the debentures are being issued at an issue price of $849 for each $1,000 principal amount. They are convertible at any time prior to maturity at a conversion price of $25 a share .The debentures are available through Goldman .

OUTPUT-Text 3

```
of its registration statement
 for \$125 million of 6 3.8\% convertible subordinated debentures due Oct/TEMPORAL
 at an issue price of \$849/QUAN
 for each \$1,000 principal amount/QUAN
 through Goldman/BENIFICIARY
```

Text4:

The BBC's disability sport website has picked up a prestigious award at the British Wheelchair Sports Awards. The website, which was launched in June, won the Media Coverage award at the ceremony in Reading on Saturday.

The judging panel paid tribute to the site's high level of commitment to the disability sports movement ahead of the 2004 Paralympics in Athens.

Other winners were Peter Norfolk and Emma Brown who took the Outstanding Male and Female Achievement awards.

Norfolk finished 2003 as world number one in wheelchair tennis' Quad division after winning a number of top tournaments including the French Open, British Open and US Open.

Brown, who is the powerlifting World, European and Paralympic champion, picked up her third award at the end of a year which saw her break her own world record on six occasions.

Great Britain wheelchair tennis number one Kay Forshaw won the Best Newcomer award.

The Team Event award was won by table tennis duo James Rawson and Neil Robinson, the Great Britain Class Three pair who won European Championship gold in 2003.

Coach of the Year was British women's powerlifting coach Jon Amos while the Team Sport award was won by the Great Britain wheelchair rugby team.

OUTPUT-Text 4

```
up a prestigious award
```

```
at the British Wheelchair Sports Awards
in June/TEMPORAL
at the ceremony in Reading on Saturday/TEMPORAL
up her third award at the end of a year/TEMPORAL
by table tennis
in 2003/TEMPORAL
by the Great Britain wheelchair rugby team
```

Text5:

HONG KONG (Dow Jones)–Pan Sino International Holdings Ltd. (8260.HK), the first cocoa bean trading company to be listed on the Hong Kong bourse, said its share placement was five times subscribed.

The company initially offered 240 million shares to institutional investors at 45 HK cents apiece.

ADVERTISEMENT

SingaporeAll the web

Shares of Pan Sino are scheduled to start trading Tuesday on the Growth Enterprise Market. The company is the fourth largest cocoa bean exporter in Indonesia, which is the third largest cocoa bean producer in the world.

Due to "favorable subscription rates", bookrunner and lead manager SBI E2-Capital Securities Ltd. allocated an additional 36 million shares, representing 15

As a result, the share placement was enlarged to 276 million shares from 240 million shares originally, Pan Sino said in a statement late Sunday. The public float will be equivalent to 47.5

Pan Sino is currently focused on the export market. Its clients are established cocoa bean product importers in Europe.

"Riding on the upward trend in the price and demand of cocoa beans, the company plans to diversify into cocoa the processing business which has a higher profit margin. We are therefore confident that we can bring better investment return to our shareholders and realize our (targeted) 30

Celestial Capital Ltd. is the sponsor of the listing.

OUTPUT-Text 5

```
on the Growth Enterprise Market
in Indonesia /LOCATIVE
in the world /LOCATIVE
of the shares
to 276 million shares /QUAN
in a statement late Sunday/TEMPORAL
in Europe/LOCATIVE
```

Appendix

PP Categorizer program in C++

```
#include "stdafx.h"
#include <fstream.h>
#include <string.h>
#include <strstrea.h>

const int MAX = 10000;
const int MAX_WLIST = 9;
enum ppTypes { locative, temporal, quan, qual, dir, manner, extent, beneficiary, purpose };

char *wlist[MAX_WLIST][MAX];
char *fileNames[MAX_WLIST] = {"LOCATIVE", "TEMPORAL", "QUAN", "QUAL", "DIR", "MANNER", "EXTENT",
int keySize [MAX_WLIST];
```

```
int loadKeys(ppTypes type) {
ifstream keyF(fileNames[type]);
keySize[type] = 0;
while (!keyF.eof() && keySize[type] < MAX) {
int top = keySize[type];
wlist[type][top] = new char [80];
keyF.getline(wlist[type][top],80);
strupr(wlist[type][top]);
if (wlist[type][top] != '\0')
keySize[type]++;
}
keyF.close();
return 1;
}
// Binary Search through any one of wordlists
int bs(ppTypes type, const char *word) {
bool found = false;
long l = 0, r = keySize[type]-1, m;
while (!found && (l<=r)) {
m = (l+r)/2;
if (strcmp(word,wlist[type][m]) < 0)
r = m-1;
else if (strcmp(word,wlist[type][m]) > 0)
l = m+1;
else
found = true;
}
if (found)
return 1;
return 0;
}

char* lastToken(const char *s) {
int idx = 0;
while (strstr(&s[idx]," ") != 0)
idx += strlen(&s[idx]) - strlen(strstr(&s[idx]," "))+1;
char * ans = new char [strlen(s)-idx+1];
strcpy(ans,&s[idx]);
return ans;
}

int main(int argc, char* argv[])
{
for (int i=0; i<MAX_WLIST; i++) {
keySize [i] = 0;
loadKeys((ppTypes)i);
}

            char pp[80], *np;
```

```
ifstream pps("resultver2/outputPPs5.txt");
ofstream taggedF("resultver2/ppsout5.txt");
while (!pps.eof()) {
pps.getline(pp,80);
taggedF << pp;
np = lastToken(strupr(pp));
for (i=0; i<MAX_WLIST; i++)
if (bs((ppTypes)i, np))
taggedF << "/" << fileNames[i];

taggedF << endl;

}
pps.close();
taggedF.close();

return 0;
}
```

Appendix

PP Chunker program

```
#include "stdafx.h"
#include <fstream.h>
#include <string.h>
#include <strstrea.h>

int process(char *s, char *pp) {
int n = 0;
int idx = strlen(pp);
int state = 0;
for (int i = 0; i<strlen(s); i++) {
if (s[i] == '(')
n++;
else if (s[i] == ')')
n--;

if ((s[i-1] == ' ') && (s[i] == '"')) {
state = 1;
pp[idx++] = ' ';
}
else if ((s[i] == '"') && (s[i+1] == ' '))
state = 0;
else if (state)
pp[idx++] = s[i];

}
pp[idx] = '\0';
return n;
}
```

```
int main(int argc, char* argv[])
{
char pp[80], s[80];
ifstream ppRaw("resultVer2/output9");
ofstream ppExtr("resultVer2/outputPPs9.txt");
while (!ppRaw.eof()) {
ppRaw.getline(s,80);
if (strstr(s,"PP")) {
int sum = 1;
pp[0] = '\0';
while ((sum > 0) && !ppRaw.eof()) {
ppRaw.getline(s,80);
sum += process(s,pp);
}
do {
ppRaw.getline(s,80);
if (ppRaw.eof())
break;
} while (strstr(s,"tree(s)") == NULL);
ppExtr << pp << endl;

}
}


ppRaw.close();
ppExtr.close();

return 0;
}
```

Appendix

PP Chunker Grammar

```
;;;;;;;;;;;;;;;;;; begin of grammar file
(define grammar '(       ;; obligatory line

(ISOLATED_NP
  (S *end-of-sentence*)
;;  (S *comma*)
)

(S
  (XSKIP)
  (NP)
)

  (XSKIP
(SKIP_SINGLE XSKIP)
(SKIP_SINGLE)
(NP XSKIP)
```

```
  )

  (SKIP_SINGLE
    (*comma*)
    (*quote* NP *quote*)
    (*quote* NP *end-of-sentence* *quote*)
    (*quote* S *quote*)
    (*quote* S *end-of-sentence* *quote*)
    (*quote* VP *quote*)
    (CC)
    (PP)
    (WRB)
    (VP)
    (WDT VP)
    (WDT NP VP)
  )

(AP ;; Adjectivial Phrase
(JJ)
(JJ PP)
;; (JJ VP)
(RB JJ)
(RB VBD)
(RB VBP)
)

(RBPh ;; Adverbial Phrase
(RBPh RB)
(RB)
(RBR)
(RBS)
)

(VP
(VP CC VP)
(MD VP)
(VP VP)
(TO VP)
(VP  NP)
(VP  PP)
(Verb AP)
(Verb RBPh)
(RBPh Verb)
(Verb)
(Verb VBP)
)

(Verb
(VB)
(VBD)
(VBP)
```

```
(VBG)
(VBZ)
(VBN)
)

(Prep
(IN)
(TO)
)

(PP
(Prep NP)
(PP PP)
)

  (NP
    (NP NP)
    (NP POS NP)
    (NP PP)
    (NP AP)
    (CD NP)
    (CD)
    (DT CD NP)
    (DT CD)
    (PDT NP)
    (DT NP)
    (AP NP)
    (JJR NP)
    (JJS NP)
    (RBR NP)
    (RBS NP)
    (DT NP)
    (PRP$ NP)
    (NN)
    (PRP)
    (WP)
    (WP$)
    (NNS)
    (NNP)
    (NNPS)
    (NP CC NP)
  )
))                          ;;end of grammar file
;;;;;;;;;;;;;;;;;;;;;;;;end of grammar file
```