ANS$\mathcal{F}$ORM: Answer Formulation For *Q*uestion-*A*nswering

Glenda Anaya

A Major Report

in

The Department

of

Computer Science

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Computer Science at
Concordia University
Montreal, Quebec, Canada

December 2002

ABSTRACT

Ans$\mathcal{F}$orm Answer formulation – for Question-Answering

Glenda Anaya

The goal of Question-Answering (QA) systems is to find short and correct answers to open-domain questions by searching a large collection of documents. The subject of this research is focused on finding patterns to formulate a "complete" and "natural" answer to questions, given the short answer. Finding such patterns is important as they can be used to enhance existing QA systems to find and provide answers to the user in a more "natural way" and providing a pattern to find the answer. Based on a number of patterns of type of answer formulation compiled from a survey carried out at the beginning of this project, the work of producing long and natural answers to specific type of questions is reduced to patterns matching, but additional research could be done to process and analyze the context of the question and the short answer given, in order to provide a more relevant, natural and correct answers.

The first chapter of this major report gives a general description of the nature and scope of our project, the second chapter introduces the fields of Natural Language Generation (NLG) and Question-Answering systems (QA) as a background to our system, which we expect to be used as a resource to increase the potential of QA system, to answer questions of a wide variety of topics in different grammatical and natural ways. The third chapter presents the process, analysis and results of the survey applied to find the answer formulation patterns. The fourth chapter describes the system, its requirements, scope, analysis and results, and finally chapter five shows the evaluation of the system and gives the conclusions and future research avenues.

Acknowledgements

This work would not be possible without the help of many people. First, I would like to express my largest gratitude to my supervisor, Prof. Leila Kosseim for her continuous support, encouragement and guidance throughout this investigation. She is acknowledged in the NLP community, in particular in Question-Answering and I am grateful for her insight and wisdom during our work. She taught me the way of proceeding research, and to be passionate about this field.

I also express my thanks to members of CLaC, especially to Frank Rudzic. He offered me his friendship, and his unconditional help in resolving a number of issues related to development of this application. His suggestions and insights have helped me tremendously.

A special thanks to all my friends, especially to Angel, Naybell, Martha and Endre; their advice, patience, and support helped me out in a lot of difficult moments during my studies.

Finally I am grateful to my family who encouraged me and gave me their emotional support.

Table of Contents

List of figures

List of tables

Chapter 1

1 Introduction

Unlike most Question-Answering (QA) systems that face the problem of how to find short and correct answers to open-domain questions by searching a large collection of documents, this project is focused on finding patterns to formulate a "complete" and "natural" answer to questions, given the short answer. Finding such patterns is important as it can be used to enhance existing QA systems to provide answers to the user in a more "natural way".

For example given this question, and the short answer between parenthesis

What two US biochemists won the Nobel Prize in medicine in 1992? (Edwin Krebs Edmond Fischer)

Our goal is to generate the following answers:

- Edwin Krebs and Edmond Fischer are the two US biochemists who won the Nobel Prize in medicine in 1992.
- The two US biochemists who won the Nobel Prize in medicine in 1992 are Edwin Krebs and Edmond Fischer.
- Edwin Krebs and Edmond Fischer won the Nobel Prize in Medicine in 1992
- Edwin Krebs and Edmond Fisher, two American biochemists, won the Nobel Prize in medicine in 1992
- In 1992, Edwin Krebs and Edmond Fischer were the two US biochemists that have won the Nobel Prize in medicine.
- The Nobel Prize in medicine in 1992 was won by Edwin Krebs and Edmond Fischer

Another interesting motivation for finding these patterns is to facilitate QA systems to retrieve exact answers with more probabilities of success. More specifically the QA system will get the answer formulation pattern for a specific question and try to find this complete sentence or phrase, either on the web or in a document collection, and retrieve the answer according the pattern which will indicate where the answer is located in the sentence structure and what type of answer must be; either a name, date, location, etc. This technique was used in the web component of *Quantum* [17], as well in the question answering system *QALC* [5], and the project *'Patterns of Potential Answer Expressions as Clues to the Right Answers'* [21].

In chapter two, we review the literature in the fields related to our project: specifically, Natural Language Generation and Question-Answering. Based on the review, we describe the requirements for an answer-formulation pattern system. Showing how external systems such as the Brill-Tagger, and ClaC's Noun-Phrase-Extractor (NPE) fulfil them.

In chapter three we describe in detail how the patterns for question and answers were established for any type of question, following these 2 steps:

- First we conducted a survey of English-speaking people to capture the variety of answer formulation.

- Second we analysed the formulation of the answers to extract their patterns. The patterns indicate the different grammatical and lexical structures to best answer specific types of questions. The formulations of the answers for each question are

displayed using a natural language structure. The analysis of the survey shows that there are different ways to answers a question.

In chapter four, we explain the design and implementation of the system Ans$\mathcal{F}$orm, which gets as input a text file with a series of types of question, extracts the pattern for each input question, matches them with the question-answer patterns extracted during the analysis of the survey (training-set), and generates a different answer formulation for each type of question. Also, we evaluate the results from the matched questions, to determine if the variation of answers produced for the system, according with the pre-established answers-patterns, were grammatically correct.

In chapter five, we show the evaluation of the system, and we discuss future work to be done according to the evaluation results. we feel that additional research to process and analyse the context of the question and the short answer given, would provide more relevant, natural and correct answers.

Chapter 2

2 Literature Review

We conducted a research about references of previous projects, covering the same objectives as ours, which is to get answer formulation patterns for a specific question, either to provide a long natural answer or to try to find this complete sentence or phrase on the web or in a document collection, to extract specific answers or information and we found that some systems have come up with the same idea but with a different approach. For example, to improve the probability of retrieving answers the system *"Specific Expressive Forms"* by Lawrence and Giles [12] attempt to transform queries (questions) into specific phrases (possible answer). For example, a question of the form "What is X?" can be used to generate phrases such as "X is" or "X refers to", and apply these sets of generated queries to each search engine. Another approach is the work of Eugene Agichtein, Steve Lawrence, and Luis Gravano [1] who present *"Tritus"*, a system that automatically learns to transform natural language questions into sets of effective search engine queries, optimized specifically for each search engine.

In contrast to previous research, we introduce a system that does not only look for specific phrases in the question that may be contained in the answer; we try to formulate all possible sentences (answer patterns) derived from question to provide complete and natural answers. These initial answer formulations were handcrafted for this project; for future work, we expect to find answer patterns through a learning process, asking the user to introduce possible answers to specific types of questions, to automatically generate the

answer pattern. Actually only the patterns for questions are generated automatically, this is fully explained in chapter 4.

For this project it was also useful to have some documentation on related fields. The work described in this report lies at the juncture of several fields: it incorporates Questions-Answering (QA) and Natural Language Generation (NLG).

2.1 Question-Answering

A question answering (QA) system provides direct answers to user questions by consulting its knowledge base. Since the early days of artificial intelligence in the 60's, researchers have been fascinated with answering natural language questions. However, the difficulty of natural language processing (NLP) has limited the scope of QA to domain-specific expert systems.

In recent years, the combination of web growth, improvements in information technology, and the explosive demand for better information access have increased the interest in QA systems. The availability of huge document collections (e.g., the web itself), combined with improvements in information retrieval (IR) and NLP techniques, has attracted the development of a special class of QA systems that answers natural language questions by consulting a repository of documents. A QA system utilizing this resource has the potential to answer questions on a wide variety of topics, and will

constantly be kept up-to-date with the web itself. Therefore, it makes sense to build QA systems that can scale up to the web.

The ultimate goal of Question-Answering is the creation of an integrated suite of analytic tools capable of providing answers to complex, multi-faceted questions involving judgment terms that analysts might wish to pose to multiple, very large, very heterogeneous data sources that may physically reside in multiple agencies and may include:

- Structured and unstructured language data of all media types, multiple languages, multiple styles, formats, etc.,
- Image data to include document images, still photographic images, and video; and
- Abstract/technical data.

Some QA Applications:

- Lunar [23]. High numbers of questions were answered correctly, but on a very restricted domain.
- MURAX [11]. Uses on-line encyclopedia as a source of answers for closed-class questions.
- Student [3]. Solved high school algebra problems.
- START [10]. Uses annotations to process questions from the web.
- Tritus [1]. Machine learning transformation of queries.
- AnswerBus [24]. An open-domain question answering system based on sentence level information retrieval.

2.1.1 QA systems and TREC

To support research within the information retrieval community, providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies, the *Text Retrieval Conference* (TREC) - co-sponsored by the National Institute of Standards and Technology (NIST) and the Defense Advanced Research Projects Agency (DARPA)- was created. In particular, the TREC workshop series is the occasion where most of the research on Question-Answering is evaluated and where researchers exchange ideas, techniques and methodologies in order to accelerate the transfer of technology from research labs into commercial products.

The TREC-QA workshop involves retrieving a short exact answer to a set of test questions. Participants run their own retrieval systems on the data, and return a list of the retrieved answers. NIST pools the individual results, judges the retrieved answers for correctness, and evaluates the results.

Since the TREC test data collections are available to the retrieval research community to evaluate their own retrieval systems at any time, we took advantage of it and we used some of the data (questions) available to build our training-set, which is detailed in Chapter 3.

2.2 Natural Language Generation

2.2.1 Introduction

Natural Language Generation (NLG) systems simulate the production of written or spoken language. Based on human language studies - carried out from different disciplines such as linguistics, computer science, and psychology - NLG investigates how to generate a high-level language output (text) from computer data encoded in a knowledge database.

Usually these systems are driven by systematically planning and transforming writer's and hearer's goals into natural and intelligible written or spoken expressions, to improve the communication, which is the aim of NLG.

Most NLG systems divide the generation process into two (2) main components. The first component "Strategy", which is delegated to a document-planning module, needs to decide *What* should be said, and the second component, "Tactics", delegated to a micro-planning and surface realization modules, needs to decide *How* to say it.

2.2.2 Architecture of a NLG system

Starting with the communicative goal of the user as the input to the NLG system, we describe each module in a pipeline where in general the output of the document planning is the input to the micro-planning, and the output of the micro-planning is the input to the realization module. The document planner determines the content and the structure of the information. The micro-planner deals with the lexicalization, referring expression generation, and aggregation. Finally, the surface realizer converts the representation into real text and adds the annotations required by the system. (See figure 2-1)

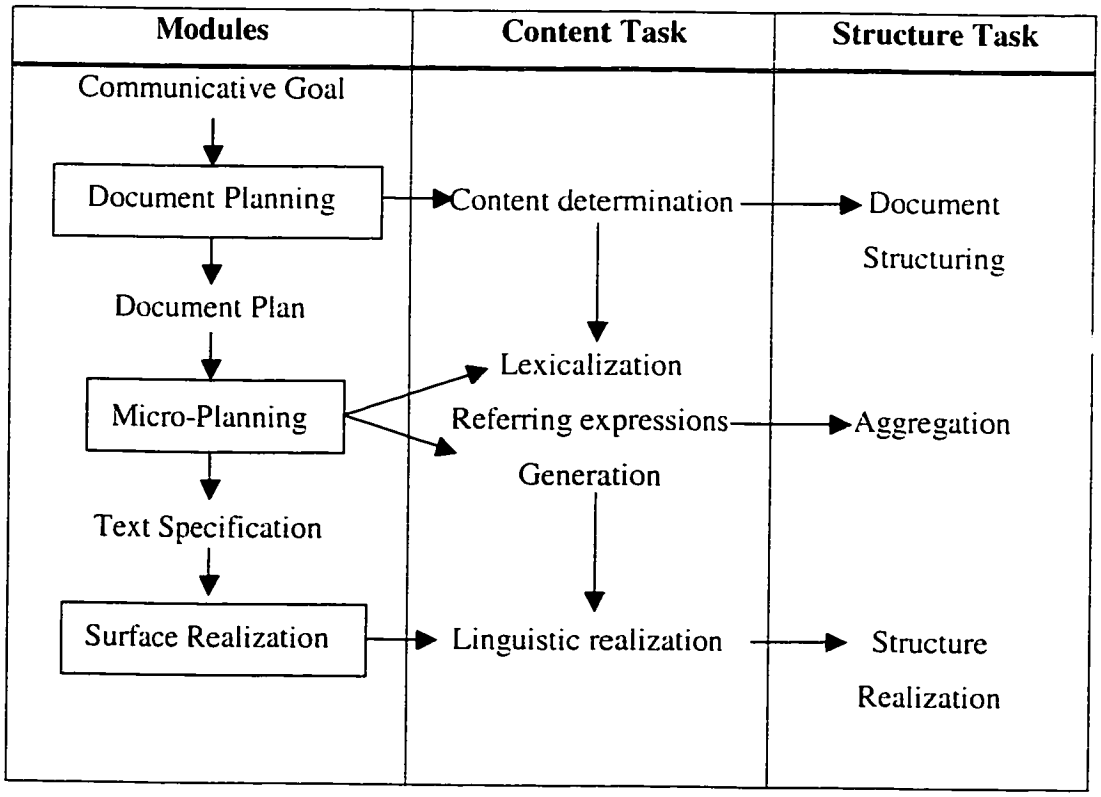


Figure 2-1 NLG system architecture (Modules and Tasks)

This architecture above defines an intermediate representation as a result of processing each module. In this architecture presented by Ehud Reiter and Robert Dale in the book

"Building Natural Language Generation Systems" [18] there are two representations: the *document plan*, a tree representation which specifies structural information and message forms in its internal node and leaf respectively. And as the second representation, we have *text specification*, a tree that specifies the structure of the text (internal node) and its sentences (leaf nodes).

2.2.2.1 Document Planning

The planning module produces a document plan representing the structure and the content of the text, using domain and application knowledge on what information is appropriate for the specified communicative goal, user model, and discourse history, taken as an input of the process. Specifically, to construct a document plan, this module creates messages from the original information, deciding which messages should be communicated, and carrying out the presentation of them in a consistent and fluent text to satisfy the goals of the user.

> *"Document planning is probably the most application-dependent aspect of natural language generation. While micro-planning and surface realization make use of what we can think of as more general linguistic knowledge, document planning is mostly based on application-specific knowledge governing what information is considered to be important, what conventions need to be obeyed in presenting this information..."[18].*

2.2.2.2 Micro-planning

The micro-planning module intents to refine the document plan produced in the previous module, determining how best to use knowledge about language and effective writing to package information into sentences (words, syntactic and grammatical structures in the language to be generated), and then produce an abstract specification of the text's content and structure, which are carrying out for the content and structure tasks specified in section 2.1.2.4. Also the micro-planner may specify the tense of a sentence by means of an abstract category such as past, present, or future, and state that a set of sentences should be grouped into a paragraph.

The micro-planning module might produce as a result of its process, a *text specification* that provides a full specification of the document to be generated. This specification must be mapped into a real text (sentences and paragraphs) that includes: pronoun specification, theme signaling, content aggregation to remove unnecessary redundancies, the ordering of prepositional phrases, adjectives, etc.

2.2.2.3 Surface Realization

The surface realization module converts abstract specifications (text specification) into real text. The text specification, a tree-like discourse structure of instructions, whose internal nodes specify paragraphs and leaf nodes are phrase specifications, will generate grammatically correct sentence based on its hierarchically input and its underlying text content. The surface realization module has also the choice to select the more suitable verb inflections in addition of any grammatically necessary auxiliary word and converts the paragraph specification into the mark-up symbols required by the NLG system, which is the target document presentation system.

2.2.2.4 Content and Structure Tasks

Two types of task are performed in the 3 modules: *Document Planning*, *Micro-Planning* and *Surface Realization*: content and structure tasks. (See figure 2-1)

- **Content Determination:** This task involves decisions about what information should be expressed to the user in the final text document. The decision to determine the content is based basically on: the communicative goals that may possibly require diverse information, the expertise of the reader/hearer on the domain application, the basic information available and some constraints on text formatting upon the output document.

- **Document Structuring:** This task decides how pieces of content should be ordered and structured in a document over the information to be communicated. This basic structure of a text goes further than the sequencing of sentences, instead sentences are viewed as part of a whole document analyzed in terms of a tree structure, whose relationships between the constituents are determined for the information to be presented and their discourse relations over the text document.

- **Lexicalization:** This task involves the selection of the most appropriate words to express the content, selected by the *content determination task;* this task includes linguistics resources that express meaning using syntactic structures. The selection between possible lexicalization depends on what has already been said, what is available from context, what are the goals; what effect the hearer/speaker wishes to produce/communicate.

- **Referring Expression Generation:** This task decides what expressions should be produced to refer to entities in a given context, to allow the reader/hearer to identify them. Basically this task needs to distinguish between an *initial reference* – the entity has not been mentioned before- and the *subsequent references* – the entity has already been mentioned in the discourse. When subsequent references are present, they are abbreviated or they are presented using pronouns.

- **Aggregation:** This task decides how the tree structures created by the *document structure task* should be mapped onto linguistics structures and textual element (sentences and paragraphs) to express the same information content in a variety of ways. To achieve this, *aggregation* considers grouping and ordering of sentences and paragraphs into the document to describe different pieces of information making use of the process of lexicalization.

- **Linguistic Realization:** This task translates abstract representations of sentences into syntactically and morphologically correct text, using the characterization of a set of grammar rules, which specify what is a well-formed sentence in the language.

- **Structure Realization:** This task converts abstract structures such as paragraphs and sections into the mark-up symbols understood by the NLG system.

2.2.3 NLG Input

In order to make these complex choices, language generators need various knowledge resources:

- **Discourse history** – information or previous explanations about what has been presented to avoid repetitions of present facts.

- **Domain knowledge** – categorization and knowledge of the domain.
- **User model** – specifications of the user's domain knowledge, goals, beliefs, plans, and interests.
- **Grammar** – a grammar of the target language, which is used to generate linguistically correct sentences.
- **Lexicon** – a lexicon entry for each word, containing typical information like part of speech, inflections class, etc.

Although our project generates complete answers from a question and its exact answer, it cannot be considered as an NLG system. We do not go through the typical phases of NLG, our input is not as rich and abstract as the one specified in section 2.2.3 and we do not produce discourse, but only single sentences. However our system does have strong common points with NLG systems as the goal is the same: To generate natural sentences and as such, we have been influenced by work in this area.

Chapter 3

3 Ans$\mathcal{F}$orm – Corpus Analysis

In our project, to accomplish the goal of generating "natural" and complete answers derived from a specific type of question, and their question-answer formulation patterns, we have followed four steps:

- Perform a survey to capture and analyze the variety of answer formulation for any type of question (Training-set). This chapter is devoted to explain in detail how this survey was conducted, its results, analysis and conclusions.
- Evaluate NLP tools available, using the training-set, to automatically generate patterns for questions and the answer to extract. Basically we looked for a part-of-speech Tagger, a Parser, and a NP-Chunker. Chapter 4 covers the evaluation of the tools during the implementation.
- Implement the Answer-Formulation System (Ans$\mathcal{F}$orm), and
- Evaluate the results of Ans$\mathcal{F}$orm; these two last steps are covered in detail in chapter 4.

3.1 The Corpus Analysis - Methodology

The main objectives of the survey were:

- To identify patterns on answer formulation for any type of open-domain question.
- To gather enough information to construct a corpus composed of grammatical and lexical patterns for the answers of each type of question.
- To develop an application that will automatically generate natural answers from the patterns, in one sentence having enough context to understand the answer without knowing the question.

The survey was conducted for a period of three weeks from May 13 to June 7, 2002. At the end of this period the questionnaires were compiled in one file to easily tabulate the answers for each question.

The questionnaire was designed for completion within 45 to 60 minutes. It was structured with a total of 150 open questions and answers, taken semi-randomly[1] from the [TREC-8, TREC-9 and TREC-2000] [2] question sets, to get an evaluation of each type of question (*Who, Where, Who, What, Name* and *Which* type questions), and their formulations. From the participants it was expected to get a complete answer in one sentence according to the formulation of the question.

The target population was English-speaking people, not necessarily from the NLP community; in order to complete the questionnaire it was only required to know the

[1] The question were not selected completely random because we wanted samples of each type of question, and also with different structures, to built a "complete" set of patterns.
[2] http://trec.nist.gov/data/qa.html.

English-grammar in order for us to get the most natural structured answers from different kind of users. The questionnaire was distributed to a total of 40 people; 21 people answered the survey. The identification of the target group for the survey was semi-random, the target could have been either people related with the NLP field (students and professors), or just colleagues or friends. However, the observations we have drawn from the results of the survey can apply to a large proportion of the QA information/service seekers.

Not a specific geographic area was covered; due to the advent of Internet, the questionnaires were distributed by email and the participants were asked to reply also by email to an email address clearly specified in the questionnaire. Figure 3-1 shows an example of the questions and short answer used in the questionnaire. See Appendix A for the complete questionnaire.

For the following questions, given the short-answer between (), please give a complete answer, in one sentence, for example question 1;

1. Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"? (Hugo Young)

   Ans.: Hugo Young is the author of the book "The Iron Lady: A Biography of Margaret Thatcher".

   Ans.: The author of the book"..." is Hugo Young

   Ans.: "The Iron Lady:.. "book was written by Hugo Young

   Ans.: ...

Figure 3-1 Excerpt of questions and answers from the survey

Two indicators based on the grammatical and the lexical structures of the responses were used, this is specified more clearly on the analysis of the survey covered in the next section.

The methodology used for the survey suffered some limitations in particular: the time to complete the survey and make an analysis was limited, due to the time to submit the major report and developing the application. This made the size of the sample and the size of the questionnaire rather small.

3.2 Processing the Responses

The collection of results was carried out on June 7, 2002. Three files were setup to be easily reused for the final set of results.

- **Collection file:** This file contains the collection of all the questionnaires with a total of 3150 answers (150 questions x 21 people); the answers were put together in the order the questionnaires were received. Figure 3-2 shows an example taken from the original document.

142. For how long is an elephant pregnant? (22-month)

1. *Ans.: For 22-month.*
2. *Ans.: Female elephants are pregnant for 22 months.*
3. *Ans.:* an elephant pregnant for 22 months
4. *Ans.: An elephant is pregnant for ( )*
5. *Ans.: An elephant is pregnant for a ( ) period.*
6. *Ans.:* The elephant has a 22-month pregnancy.
7. *Ans.: An elephant is pregnant for 22-months*
8. *Ans.:* An elephant pregnant for 22-month.
9. *Ans.:* An elephant pregnant for 22-months.
10. *Ans.: the pregnancy of an elephant last ( )*
11. *Ans.: That is 22-month.*
12. *Ans.: An elephant is pregnant for ( )*
13. *Ans.:* an elephant pregnant for 22-months.
14. Ans.: An elephant's pregnancy normally lasts for 22-month
15. *Ans.: Elephant pregnancy lasts for 22 months.*
16. *Ans.: An elephant is pregnant for ( )*
17. Ans.: An elephant is pregnant for 22-month.
18. *Ans.: an elephant is pregnant for ( )*
19. *Ans.: An elephant is ...( ).*
20. *Ans.: An Elephant is pregnant for ( ).*
21. *Ans.: The gestation period in elephants is ( ).*

Figure 3-2 Excerpt of the collection file (AllQuestionnaire.doc)

- **Tabulation file:** The second file gathers the type of answers for each question according to the frequency of the answers, and tabulates the results. Figure 3-3 shows that the answers were tabulated according to their grammatical construction; the variation on the lexical items used is shown on the answer indented to the right.

- **Analysis File:** The third file groups the tabulated-questions by type and formulation. Due to the size of the files only this file was included in this document (See Appendix B).

The analysis and ranking of each set of question was a time consuming process for the variety of formulation of each answer. On June 16, 2002, the final set of results was collected from the files and analyzed.

142. For how long is an elephant pregnant? (22-month)
(47.61%) 10/21 Ans.: An elephant is pregnant for 22-months.
1/14 Ans.: Female elephants are pregnant for 22 months.
(19.00%) 4/21 Ans.: An elephant pregnant for 22-months.
(9.52%) 2/21 Ans.: An elephant's pregnancy normally lasts for 22-month
1/2 Ans.: Elephant pregnancy lasts for 22 months.
(4.76%) 1/21 Ans.: For 22-month.
(4.76%) 1/21 Ans.: That is 22-month.
(4.76%) 1/21 Ans.: The elephant has a 22-month pregnancy.
(4.76%) 1/21 Ans.: the pregnancy of an elephant last ()
(4.76%) 1/21 Ans.: The gestation period in elephants is ().

Figure 3-3 Excerpt of the Tabulation file (TabulatedQ.doc)

3.3 Analysis of the Results

Of the 21 persons who answered the survey, 85% effectively answered the questions, as it was required: with a long, natural and complete sentence; the other 15 % considered that the short answer given was enough for the majority of the questions formulated in the questionnaire. This 15% of the answers were still taken into account because they are a good indicators of what usually people give as a natural answer, and a way to know when people considered a long answer - for some questions - completely unnecessary and not "natural".

The results can be divided into eight (8) main categories of types of question: *What, Which, When, Where, Who, Name, How* and *Why*. But within those categories there are variations of the same type of question and this is described in the next section.

3.3.1 Results – Training-Set

To build our training-set of question-answer patterns, we used the collection of 150 questions from the survey, which were tagged using part-of-speech tags (e.g., noun, adverb) to each word in the question. The tag notations used to express the patterns were done manually and were essentially based on Penn Treebank tag-set [13] (see table 1). Three extra tags were used: *NP, PP* and *AA*. *NP* is used to tag noun phrases, e.g. [The/DT red/JJ flower/NN]; *PP* indicates prepositional phrases, e.g. [in/IN 1969/CD]; and *AA* indicates where in the pattern, the exact answer should be located.

The collection was manually grouped by type of question and question structure, e.g. type of question *What* followed by verb "to be" in present or past, plural or singular: What "is"...?, What "were"...?. The reason to group questions is that they will possibly share the same grammatical structure on their answer formulations.

POS Tag	Description	Example
CC	coordinating conjunction	And
CD	cardinal number	1, third
DT	Determiner	The
EX	existential there	there is
FW	foreign word	d'oeuvre
IN	preposition/subordinating conjunction	in, of, like
JJ	Adjective	Green
JJR	Adjective, comparative	greener
JJS	Adjective, superlative	greenest
LS	list marker	1)
MD	Modal	could, will
NN	noun, singular or mass	table
NNS	noun plural	tables
NNP	proper noun, singular	John
NNPS	proper noun, plural	Vikings
PDT	predeterminer	both the boys
POS	possessive ending	friend's
PRP	personal pronoun	I, he, it
PRP$	possessive pronoun	my, his
RB	adverb	however, usually, naturally, here, good
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	give up
TO	to	to go, to him
UH	interjection	oh!
VB	verb, base form	take
VBD	verb, past tense	took
VBG	verb, gerund/present participle	taking
VBN	verb, past participle	taken
VBP	verb, sing. present, non-3d	take
VBZ	verb, 3rd person sing. present	takes
WDT	wh-determiner	which
WP	wh-pronoun	who, what
WP$	possessive wh-pronoun	whose
WRB	wh-abverb	where, when

Table 3-1 Part-Of-Speech Tag-set

We generated all possible answer patterns from the answers given in the questionnaire, and to generate a complete and correct answer formulation, we manually added some special notation and words to the answer according to the structure and the type of question. For example, *Who questions*, followed by a "noun": *Who invented the paper-*

clip? One answer pattern will take the same words provided in the question *"(AA) invented the paper-clip"*, where (AA) indicates the exact answer. But another answer pattern, for the same question, could use different words to build another complete and correct answer; e.g. "the paper-clip *(was)* invented *(by)* (AA)", in this case "was" and "by" were added to the answer pattern. Another example is question type *Why*, whose answers could include the word(s) "because" or "In order to" or "that is the reason why".

The following are examples showing type of questions grouped by structure, possible answer formulations, and a graph showing which given type of answer was predominant. See Appendix C to a complete analysis.

What + "NOUN" + "To Be" +...

Pattern 1: What (WP) + NP1+ VBD1(Z) + PP1

Example Pattern 11: What/WP1 soviet seaport/NP1 is/VBZ1 on the Black Sea/PP1? (Sevastopol)

Answer Patterns:
A. (AA) + VBZ1 + NP1+ PP1
B. NP1+ PP1 + VBZ1 + (AA)
C. Other
D. (AA) + VBZ1 + PP1

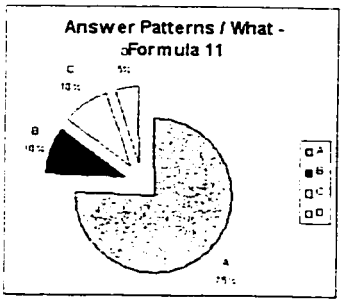


Figure 3-4 What-Pattern 11

The tag notation used to express the patterns has some slight variation from the one shown in table 1. We added a number to the tag (e.g. WP1), to indicate the number of times that tag appears in the question structure formulation. This was done for convenience at the moment of programming the application.

The special notations and words used to build the answer patterns were also surrounded by parenthesis (), indicating that for some question those words and notations are optional in the answer formulation, except the (AA) notation that represents the exact answer, which is not optional.

Here are some samples of patterns for type of question; *Why*, *Who*, and *Name*:

Why + AUX VBD(Z)+...

Pattern 1: WP + VBD1 + NP1 + VB + NP2
Example Pattern 1: Why/WP1 does/VBD1 the moon/NP1 turn/VB1 orange/NP2? (eclipse)
Answer Patterns:
A. NP1 + VB1 (tense) + NP2 + (because) + (of-an) + (AA)
B. NP1 + VB1 (tense) + NP2 + (during) + (of-an) + (AA)
C. (Because) + (of-an) + (the) + (AA)
D. NP1 + VB1 (tense) + NP2 + (when) + (there is) + (of-an) + (AA)
E. (An) + (AA) + (is) + (the reason) + WP1 + NP1 + VB (tense) + NP2
F. Other

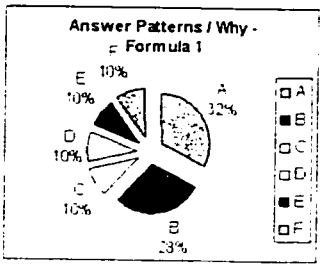


Figure 3-5 Why-Pattern 1

Who + To Be +...

Pattern 1: WP + VBD1(Z) + NP1 + IN1 + NP2 + VBD2 + IN2 + NP3 + PP1+ PP2
Example Pattern 1: Who/WP1 was/VBD1 the leader/NP1 of/IN1 the Branch Davidian Cult/NP2 confronted/VBD2 by/IN2 the FBI/NP3 in Waco, Texas/PP1 in 1993/PP2? (Mr. David Koresh)
Answer Patterns:
A. (AA) + VBD1 + NP1 + IN1 + NP2 + VBD2 + IN2 + NP3 + PP1+ PP2
B. NP1 + IN1 + NP2 + VBD2 + IN + NP3 + PP1+ PP2 + VBD1 + (AA)
C. (AA) + VBD1 + NP1 + IN1 + NP2
D. PP2 + NP1 + IN1 + NP2 + VBD1 + IN2 + NP3 + PP1+ VBD2 + (AA)
E. Other

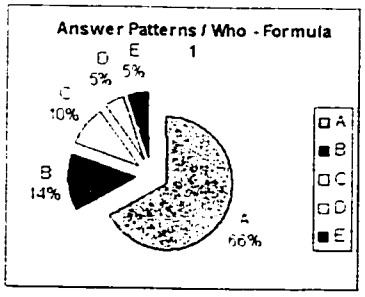


Figure 3-6 Who-Pattern 1

Name + NP +...

Pattern 1: NN + NP1 + IN1 + VBZ1 + VBN1 + NP2 + PP1

Example Pattern 1: Name/NN a film/NP1 that/IN1 has/VBZ1 won/VBN1 the Golden Bear/NP2 in the Berlin Film Festival/PP1? (In The Name Of The Father)

Answer Patterns:

A. (AA) + VBN1 + NP2 + PP1
B. (AA) + (IS) + NP1 + IN1 + VBZ1+ VBN1 + NP2 + PP1
C. NP1 + (AA) + IN1 + VBN1 + NP2 + PP1
D. other
E. (AA) + (IS) + (THE) + NN1 + (OF) + NP1 + IN1 + VBN1 + NP2 + PP1

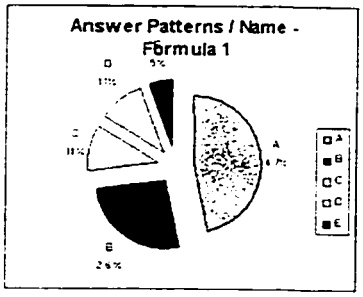


Figure 3-7 Name-Pattern 1

3.4 Survey - Conclusions

As it was required in the questionnaire, most people tried to give a complete answer as naturally as possible, but in general people used a grammatical structure and words inspired by the question. It also appears from the responses that people tend to use more varied grammar when the question is composed of more complex structure with one or more prepositional phrase(s). We found a great variety of formulation in the answers for the questions of type *What*, *Name* and *Why* (see figure 3-7). The other types of questions have fewer variations; the answers are more. For example, the following pattern (figure 3-8) for question *Where* has only two answer patterns:

Pattern 2: WP + VBZ1 + NP1 + VBD1
Example Pattern 2: Where/WP1 is/VBZ1 Belize/NP1 located/VBD1? (Central America)

Answer Patterns:
A. NP1+ VBZ1 + VBD1 + (IN) + (AA)
B. NP1+ VBZ1 + (IN) + (AA)

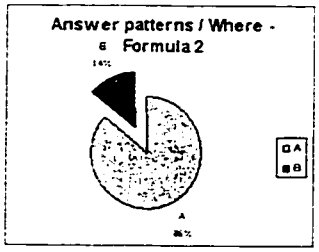


Figure 3-8 Pattern 2

Due the variety of questions and formulations, the numbers of patterns found for the questions are considerable –for 150 questions, 92 patterns were found-, which makes us consider that the evaluation on one type of question would have been preferable to get a better representative number of patterns for that type of question under any formulation. But we leave this consideration for future work, and focus on the next step of this project: the development of the application for the 8 types of question and their answer formulation, where the automation of the manual process of tagging questions and answers, extracting their noun phrases, and prepositional phrases was considered as an important requirement, using a POS Tagger, a Parser, and a NP-Chunker. This is fully explained in the next chapter.

Table 3-2 shows how the 92 patterns were distributed among the eight (8) types of questions. It is worth noting that the type of question *What* can have an large variety of question formulations and thus, ended-up with a large number of question-answer patterns; and the type of question *How* that has different subtypes, such as *How long*, *How many*, *How much*, *How large*, *How big*, *How old*, *How tall*, *How hot*, *How far*, *How did*, *How do*, *How does*, that showed almost one question pattern for each of its formulations.

# Type Questions	Type of question	No. Patterns by type of question
74	*What*	43
22	*How*	20
28	*Who*	10
9	*When*	5
3	*Which*	3
2	*Why*	2
6	*Where*	3
6	*Name*	6
150		**92 Patterns**

Table 3-2 Distribution of the patterns from the training set

Table 3-3 shows 9 questions formulations out the 43 patterns found for question *What* in the training set. The first column presents the number of questions that follow the pattern in the second column, and the third column presents an example of an answer formulation for that question. As we can observe the patterns presented here have slight differences between them. For example questions that were "*basically*" constructed by a verb "to be" following by a noun, or questions constructed with a noun, a verb, following by another noun, were frequently found, but what made the difference was the context of the question, which required additional information to be understandable and complete.

Type of question WHAT			
#	**Question Pattern**	**Example**	**Example Answer Formulation**
14	WP+[VBZ/VBD/VBP]+NP	What/WP is/VBZ an atom/NP?	An atom/NP is/VBZ (AA)
7	WP+[VBZ/VBD]+NP+IN+NP	What/WP1 is/VBZ1 the population /NP1 of/IN1 Mexico /NP2?	(AA) is/VBZ1 the population /NP1 of/IN1 Mexico /NP2
2	WP+[VBZ/VBD]+NP+PP+PP	What/WP1 is/VBZ1 the acronym /NP1 for the rating system /PP1 for air conditioner efficiency /PP2?	For the rating system /PP1 for air conditioner efficiency/PP2 the acronym/NP1 is/VBZ1 (AA)
6	WP+[VBZ/VBD]+NP+[VB/VBP]	What/WP1 does/VBZ1 Knight Ridder /NP1 publish/VB1?/	Knight Ridder /NP1 publish/VB1 (tense) (AA)
9	WP+NP+[VBZ/VBD/VBP]+NP+ [VB/VBN/VBP]	What/WP1 year /NP1 did/VBD1 WWII /NP2 begin/VB1?	WWII /NP2 begin/VB1 (tense) (in) (AA)
2	WP+NP+VBZ+NP	What/WP1 state /NP1 has/VBZ1 the most Indians /NP2 ?	(AA) (is) (the) state /NP1 (that) has/VBZ1 the most Indians /NP2
3	WP+NP+VBZ+NP+IN+NP	What/WP1 country /NP1 is/VBZ1 the biggest producer /NP2 of/IN1 tungsten /NP3 ?	The biggest producer /NP2 of/IN1 tungsten /NP3 is/VBZ1 (AA)
2	WP+NP+VBZ+PP	What/WP1 soviet seaport /NP1 is/VBZ1 on the Black Sea /PP1 ?	(AA) is/VBZ1 on the Black Sea /PP1
1	WP+NP+VBD+NP+PP+PP	What/WP1 two US biochemists/NP1 won/VBD1 the Nobel Prize/NP2 in medicine/PP1 in 1992/PP2 ?	In 1992/PP2 (AA) won/VBD1 the Nobel Prize/NP2 in medicine/PP1

Table 3-3 Example of formulations for question *What*

As in the previous table, the following tables (3.4 to 3.10) describe some formulation pattern for each type of questions. Questions of type *Who*, *When*, *Why* and *Where* show no significant variation on their formulations. Questions of type *Name*, *Which* and *How* have a great difference in their formulations, so the number of patterns, compared with the number of questions evaluated are not too representative.

Type of question HOW			
#	**Question Pattern**	**Example**	**Example Answer Formulation**
1	WRB+JJ+VBD+NP+JJ	How/WRB1 long/JJ1 did/VBD1 the Charles Manson murder trial /NP1 last/JJ2 ?/.	the Charles Manson murder trial /NP1 last/JJ2 (AA)
1	WRB+JJ+NP+MD+NP+VB+PP	How/WRB1 much/JJ1 fiber /NP1 should/MD1 you /NP2 have/VB1 per day/PP1 ?/.	Per day/PP1 you /NP2 should/MD1 have/VB1 (AA) fiber /NP1
1	WRB+JJ+NP+VBP+RB	How/WRB1 many/JJ1 Great Lakes /NP1 are/VBP1 there/RB1 ?/.	There/RB1 are/VBP1 (AA) Great Lakes /NP1
2	WRB+JJ+VBZ+NP	How/WRB1 big/JJ1 is/VBZ1 Australia /NP1 ?/.	Australia /NP1 is/VBZ1 (AA) big/JJ1
2	WRB+NP+VBD+NP+VB+PP+ PP2	How/WRB1 much /NP1 did/VBD1 Manchester United /NP2 spend/VB1 on players /PP1 in 1993/PP2.	In 1993/PP2 Manchester United /NP2 spend/VB1 (tense) on players /PP1 (AA)
1	WRB+RB+RB+VBZ+NP	How/WRB1 far/RB1 away/RB2 is/VBZ1 the moon /NP1 ?/.	the moon /NP1 is/VBZ1 (AA) far/RB1
1	WRB+VBD+NP1+VB1	How/WRB1 did/VBD1 Janice Joplin/NP1 die /VB1 ?/.	Janice Joplin/NP1 die /VB1 (tense) (cause of) (AA)

Table 3-4 Example of formulations for question *How*

Type of question WHO			
#	Question Pattern	Example	Example Answer Formulation
10	WP+VBD+NP	Who/WP1 found/VBD1 Hawaii /NP1 ?/	Hawaii /NP1 (was) found/VBD1 (tense) (by) (AA)
7	WP+VBZ+NP+IN+NP	Who/WP1 is/VBZ1 the president /NP1 of/IN1 Stanford University /NP2 ?/.	(AA) is/VBZ1 the president /NP1 of/IN1 Stanford University /NP2
4	WP+VBZ+NP+PP1	Who/WP1 is/VBZ1 the richest person /NP1 in the world /PP1 ?/.	the richest person /NP1 in the world /PP1 is/VBZ1 (AA)
1	WP+VBD+NP+WDT+MD+VB	Who/WP1 made/VBD1 the first airplane /NP1 that/WDT1 could/MD1 fly/VB1 ?/.	(AA) made/VBD1 the first airplane /NP1 that/WDT1 could/MD1 fly/VB1
1	WP+VBD+NP+NP	Who/WP1 wrote/VBD1 the song, /NP1 "Stardust" /NP2 ?/.	"Stardust" /NP2 (was) wrote/VBD1 (tense) (by) (AA)

Table 3-5 Example of formulations for question *Who*

Type of question WHEN			
#	Question Pattern	Example	Example Answer Formulation
1	WRB+VBD+NP	When/WRB1 was/VBD1 the San Francisco fire /NP1 ?/	The San Francisco fire /NP1 was/VBD1 (in) (AA)
2	WRB+VBD+NP+VB	When/WRB1 did/VBD1 Nixon /NP1 die/VB1 ?	Nixon /NP1 die/VB1 (tense) (in) (AA)
1	WRB+VBD+NP+VB+NP	When/WRB1 did/VBD1 Hawaii /NP1 become/VB1 a state /NP2 ?/	(In) (AA) Hawaii /NP1 become/VB1 (tense) a state /NP2
4	WRB+VBD+NP+VBD	When/WRB1 was/VBD1 Algeria/NP1 colonized/VBD2 ?/.	Algeria /NP1 was/VBD1 colonized/VBD2 (in) (AA)

Table 3-6 Example of formulations for question *When*

Type of question WHICH			
#	Question Pattern	Example	Example Answer Formulation
1	NP+IN+NP+RB+VBN+NP+PP +IN+WDT+NP	The U.S. Department /NP1 of/IN1 Treasury /NP2 first/RB1 issued/VBN1 paper currency /NP3 for the U.S. /PP1 during/IN2 which/WDT1 war /NP4?/.	The U.S. Department /NP1 of/IN1 Treasury /NP2 first/RB1 issued/VBN1 paper currency /NP3 for the U.S. /PP1 during/IN2 (AA) war/NP4
1	NP+VBD+NP+IN+WDT+NP	George Bush /NP1 purchased/VBD1 a small interest /NP2 in/IN1 which/WDT1 baseball team /NP3 ?/.	George Bush /NP1 purchased/VBD1 a small interest /NP2 in/IN1 (AA)
1	WDT+NP+PP+VBZ+NP+IN+N P	Which/WDT1 city /NP1 in China/PP1 has/VBZ1 the largest number /NP2 of/IN1 foreign financial companies /NP3 ?/.	The largest number /NP2 of/IN1 foreign financial companies /NP3 in China/PP1 (is) (AA)

Table 3-7 Example of formulations for question *Which*

Type of question WHY			
#	Question Pattern	Example	Example Answer Formulation
1	WRB+VBD+NP+VB+NP+PP	Why/WRB1 did/VBD1 David Koresh /NP1 ask/VB1 the FBI /NP2 for a word processor /PP1 ?/.	David Koresh /NP1 ask/VB1 (tense) the FBI /NP2 for a word processor /PP1 (because) (AA)
1	WRB+VBZ+NP+VB+NP	Why/WRB1 does/VBZ1 the moon/NP1 turn/VB1 orange/NP2 ?/.	(Because) (AA) the moon/NP1 turn/VB1 (tense) orange/NP2

Table 3-8 Example of formulations for question *Why*

Type of question WHERE			
#	Question Pattern	Example	Example Answer Formulation
2	WRB+VBZ+NP	Where/WRB1 is/VBZ1 Perth /NP1 ?/.	Perth /NP1 is/VBZ1 (in) (AA)
1	WRB+VBZ+NP+PP	Where/WRB1 is/VBZ1 the highest point /NP1 in Japan /PP1 ?/.	(AA) is/VBZ1 the highest point /NP1 in Japan /PP1
3	WRB+VBZ+NP+VBD	Where/WRB1 is/VBZ1 Belize /NP1 located/VBD1 ?	Belize /NP1 is/VBZ1 located/VBD1 (in) (AA)

Table 3-9 Example of formulations for question *Where*

Type of question NAME			
#	Question Pattern	Example	Example Answer Formulation
1	NN+CD+IN+PP	Name/NN1 one/CD1 of/IN1 the major gods of Hinduism /NP1 ?/.	(AA) (is) one/CD1 of/IN1 the major gods of Hinduism /NP1
1	NN+NP+NP+VBZ+VBN+IN	Name/NN1 a ballet company/NP1 Mikhail Baryshnikov/NP2 has/VBZ1 danced/VBN1 for/IN1 ?/.	A ballet company/NP1 Mikhail Baryshnikov/NP2 has/VBZ1 danced/VBN1 for/IN1 (is) (AA)
1	NN+NP+TO+VB+PP	Name/NN1 the first private citizen/NP1 to/TO1 fly/VB1 in space/PP1 ./.	The first private citizen/NP1 to/TO1 fly/VB1 in space/PP1 (is) (AA)

Table 3-10 Example of formulations for question *Name*

Chapter 4

4 Answer Formulation Tool (Ans$\mathcal{F}$orm) – Design and Implementation

4.1 Problem domain

During the last few years, Question-Answering (QA) designers have attempted to extract correct but short answers to questions. In our project the main goal is to formulate natural and complete answers, using a number of predefined patterns to formulate well-formed answers to questions according to their type (*What, Where, When, Who, Which, Why, How* and *Name*), and their grammatical formulation, and match them with open-domain questions and short answers given in a text-file as a result of an QA application.

Our first problem was to establish the patterns to formulate answers for each type of question. For that, we designed a survey composed of 150 questions and short answers, taken semi-randomly from the results of the TREC-8, TREC-9 and TREC-10 competition[3]. The methodology and the results of this analysis were fully explained in the previous chapter.

Once the analysis of the patterns was established, the next step was to implement a tool that would use these patterns to match any question from the input and provide a natural answer. For this project the analysis of the answers was based only on their grammatical

[3] http://trec.nist.gov/data/qa.html

structure, due to the difficulties of dealing with lexical choice, as it was explained in our literature review about *lexical generation* (see section 2.1.2.4).

4.2 The Goals

The goal of our implementation is improve QA systems by:

1. Allowing the construction of well-formed and natural phrases derived from the question structure (patterns), to provide answers to the user in a more "natural way".

2. Incorporating Answer Formulation as a QA method to facilitate the search and retrieve answers from a document collection, or from the web. In consequence QA systems will increase their level of accuracy.

4.3 Analysis Process

As we explained in the problem domain in section 4.1, the main strategy of the Ans$\mathcal{F}$orm Tool is to make use of pre-determined patterns to match any question form, and provide one or more answers, which will defer in their grammatical structure. Those pre-defined patterns (rules) for questions and answers were manually established from the training-set during the analysis of the survey (chapter 3), but to facilitate the tagging of the input questions, to be matched up with the patterns, we decided to make use of the NLP tools

available to automatically generate the question patterns. The tools considered were the following:

- CLAWS. A part-of-speech tagger, using its own tag-set [8].
- Brill's Tagger. A part-of-speech tagger that uses Penn Treebank tag-set [4].
- ClaC's NPE. A Noun Phrase Extractor that integrates Brill's Tagger and a context-free grammar [2].
- Link Parser. A syntactic parser, that uses its own tag-set and algorithm to generate a parse-tree [22].

4.3.1 Evaluation of NLP Tools

To select the most appropriated tool for our project, we ran some experiments with each tool, using the questions from the training-set, to judge the patterns automatically extracted against those found manually. The results of the evaluation are the following:

- From the two part-of-speech taggers available, Brill's and CLAWS, we found that for some questions one was more accurate than the other (and vice versa); in particular, to decide on what should be a noun or a verb in different context questions. But in general they produce the same result, and have the same limitations. In the following example CLAWS performed better for the first question, tagging "sink" as a *verb*, but in the second question both taggers gave a wrong result tagging "leave" as a *noun*.

- Brill's Tagger Problems

What/WP ocean/NN did/VBD the/DT Titanic/NNP **sink/NN** in/IN ?/.

What/WP debts/NNS did/VBD Qintex/NNP group/NN **leave/NN** ?/.

- CLAWS Tagger

- What_DTQ ocean_NN1 did_VDD the_AT0 Titanic_NP0 **sink_VVI** in_AVP ?_?

- What_DTQ debts_NN2 did_VDD Qintex_NP0 group_NN1 **leave_NN1** ?_?

- We evaluated NPE, and the Link Parser, and although the performance of the Link Parser was suitable for tagging and extracting noun phrases, we decided to use NPE because it was the one developed in our lab, it integrates Brill's tagger and has a satisfactory performance. Both NPE and the Link Parser were not created to parse questions. The following are the problem encountered during the evaluation of NPE:

CLaC's NPE problems due to Brill's Tagger: NPE will extract the wrong noun phrase if the tagger tags words incorrectly. Problems arise specifically when the tagger tags a word as a noun instead of a verb (and vice versa). The following are the NPE output for some of the same questions used to evaluate the taggers, getting as a result a list of "noun phrases" (between double-quotes), that are not completely right.

<S> What/WP <NP17> debts/NNS </NP> did/VBD <NP18> Qintex/NNP group/NN leave/NN </NP> ?/./*end-of-sentence*</S>

Right ⟶ "debts"

Wrong ⟶ "Qintex group leave"

<S> What/WP <NP163> ocean/NN </NP> did/VBD <NP164> the/DT Titanic/NNP sink/NN </NP> in/IN ?/./*end-of-sentence*</S>

Right ⟶ "ocean"

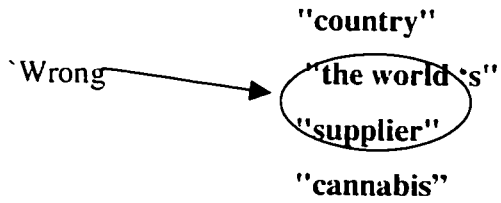
Wrong ⟶ "the Titanic sink"

NPE problems: Being a tool still under development that was not created to parse questions, NPE by itself presented some problems tagging incorrect noun phrases or extracting no noun phrase at all. For example NPE will not extract any noun phrase from questions of type "Name", even if that question might have some noun phrases to extract. The following example shows that the noun phrases "a country", and "a magnetic levitation railway system" were not parsed because of the presence of the isolate noun "Name" at the beginning of the question:

<S> **Name/NN a/DT country/NN** that/WDT is/VBZ developing/VBG **a/DT magnetic/JJ levitation/NN railway/NN system/NN** ?/./*end-of-sentence*</S>

Questions that contain compound noun phrases, having a verb as part of the phrase, were also not be parsed correctly by NPE[4]. In the following example, the complete noun phrase should be "**the world's** leading **supplier**".

<S> What/WP **<NP90> country/NN </NP>** is/VBZ **<NP91> the/DT world/NN 's/POS </NP> leading/VBG <NP92> supplier/NN </NP>** of/IN **<NP93> cannabis/NN </NP>** ?/./*end-of-sentence*</S>



Questions that contain a possessive tag (POS) as part of the phrase also were not parsed correctly by NPE for specific questions formulations.

<S> When/WRB was/VBD **<NP38> Dubai/NNP 's/POS </NP>** first/JJ concrete/JJ **<NP39> house/NN </NP>** built/VBN ?/./*end-of-sentence*</S>

"Dubai 's"
"house"

[4] This work was effectively an evaluation of NPE and some errors have been corrected since.

But for some questions NPE will extract the correct noun phrase composed with possessive tag:

<S> What/WP is/VBZ **<NP142> California/NNP 's/POS state/NN tree/NN** **</NP>** ?/./*end-of-sentence*</S>

"California 's state tree"

Most problems with the output received from NPE are due to:

- Missing grammar rules.
- Decisions to only parse minimal NP and not allow two NPs directly succeeding each other.
- The method of selecting alternatives chart parse trees generated by NPE.

Due to the problems with Brill's Tagger and NPE, a manual revision was made over the *tags* and the *noun phrases* extracted from the training-set (150 questions), as well from the input questions.

Also given the large number of patterns generated from the training set -over 150 questions 92 patterns were found-, we decided to take large constituents, as Noun Phrase, and Prepositional Phrase. This was done manually for the training-set and the input questions and is detailed in the next section 4.4.

4.4 Ans$\mathcal{F}$orm Flow Process

Figure 4-1 describes the process flow of Ans$\mathcal{F}$orm, explaining each process, its input and its output. The system is comprised of 2 main processes, "Extract-Tags" (generates patterns) and the "Matching Process", where 6 steps are covered to complete the expected results.

4.4.1 Extract Tag Process

This process is covered by two independent operations, such as the extraction of patterns from the training-set and the extraction of patterns from the input-questions. The extraction of pattern operation consists of 4 steps from the whole flow process of Ans$\mathcal{F}$orm.

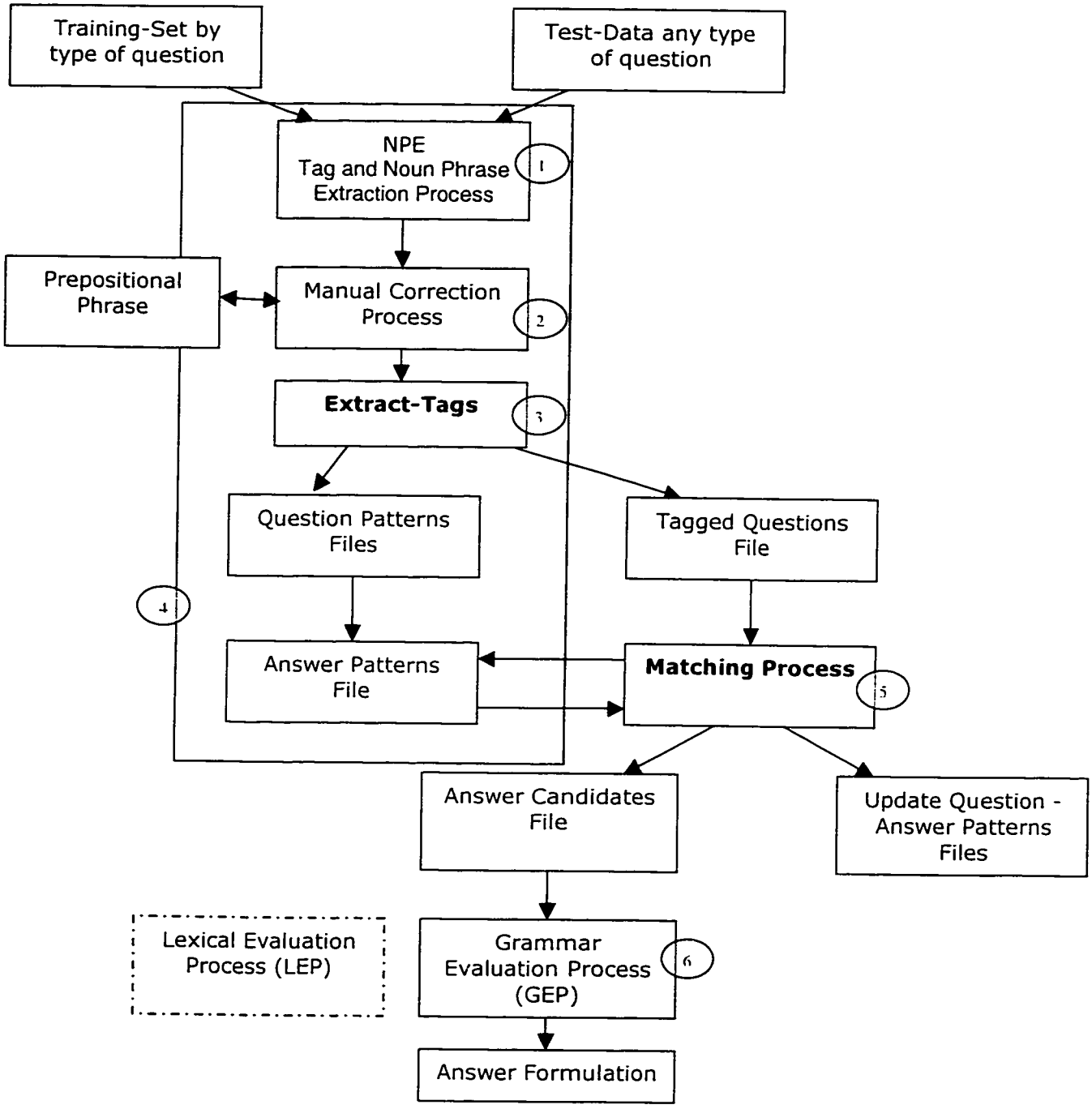


Figure 4-1 Process flow of Ans$\mathcal{F}$orm

Step 1 "NPE. Tag and Noun Phrase Extraction Process". Involves annotating sentences boundaries for the questions in each file to be processed by NPE (see Figure 4-2). Once the formatting is done, each file is sent to NPE. The application generates a tagged file for each text, with a list of questions marked with by word/tag pairs, where the former is an input token and the tags are in uppercase characters.

```
...
<s> How much did Manchester United spend on players in 1993 ?   </s>
<s> How much could you rent a Volkswagen bug for in 1966 ?   </s>
<s> What country is the biggest producer of tungsten ?   </s>
<s> When was London ' s Docklands Light Railway constructed ?   </s>
<s> How long did the Charles Manson murder trial last ?   </s>
<s> Who was the first Taiwanese President ?   </s>
...
```

Figure 4-2 Formatted Questions File

NPE also generates another file that integrates the tagged question and the noun phrases. This file inserts SMGL-type delimiters around the noun phrases and also displays a list of the noun phrases detected. (see Figure 4-3)

```
<S> Who/WP is/VBZ <NP1> the/DT author/NN </NP> of/IN <NP2> the/DT book/NN </NP>
,/,/*comma* "/" *quote* <NP3> The/DT Iron/NNP Lady/NNP </NP> : : <NP4> A/DT Biography/NNP
of/IN Margaret/NNP Thatcher/NNP </NP> "/"/*quote* ?/./*end of-sentence*</S>
"the author"
"the book"
"The Iron Lady"
"A Biography of Margaret Thatcher"

<S> What/WP was/VBD <NP5> the/DT monetary/JJ value/NN </NP> of/IN <NP6> the/DT Nobel/NNP
Peace/NNP Prize/NNP </NP> in/IN <NP7> 1989/CD </NP> ?/./*end of-sentence*</S>
"the monetary value"
"the Nobel Peace Prize"
"1989"

<S> What/WP does/VBZ <NP8> the/DT Peugeot/NNP Company/NNP </NP> manufacture/VB ?/./*end-
of-sentence*</S>
"the Peugeot Company"

<S> How/WRB <NP9> much/NN </NP> did/VBD <NP10> Mercury/NNP </NP> spend/VB on/IN <NP11>
advertising/NN </NP> in/IN <NP12> 1993/CD </NP> ?/./*end-of-sentence*</S>
"much"
"Mercury"
"advertising"
"1993"
```

Figure 4-3 Tagged-NP File

The output files go through an automatic preprocessing process done by Ans*Form*, where all SMGL-type delimiters around the noun phrases are substituted by an NP tag at the end of the noun phrase, and the list of noun phrases detected (below the question) are deleted. (see Figure 4-4)

```
Who/WP is/VBZ the author/NP of/IN the book/NP, "The Iron Lady/NP: A Biography of Margaret
Thatcher/NP "?

What/WP was/VBD the monetary value/NP of/IN the Nobel Peace Prize/NP in/IN 1989/NP ?

What/WP does/VBZ the Peugeot Company/NP manufacture/VB ?
```

Figure 4-4 Ans*Form* preprocessing process

Step 2 "Manual Evaluation". As mentioned in section 4.3.1, due to the problem encountered with Brill's tagger and NPE, a manual revision of the tags and the noun phrases is done on each file (training-set and input question). Also given the large number of patterns generated from the training set (see section 4.3), we decided to manually take large constituents, such as NP and PP with some assumptions for their extraction:

An NP can be:

- Anything between quotes "".

  Which/**WDT1** comedian' s signature line/**NP1** is/**VBZ1** **"Can we talk"/NP2** ?/.

- The nouns phrases + prepositional phrases before the question auxiliary (e.g. "did", "is").

  What/WP1 **time of day/NP1** did/VBD1 Emperor Hirohito/NP2 die/VB1 ?/

  What/WP1 **type of bridge/NP1** is/VBZ1 the Golden Gate Bridge /NP2 ?/

All prepositional phrases starting with the preposition *"of"* will not be annotated as a prepositional phrase using the tag PP because they may generate grammatically incorrect answer formulations if they are moved from the original position they held in the input question. Contrary prepositional phrases starting with *"on"*, *"in"*, *"for"* or *"between"* generate correct answer formulations whether we placed them at the beginning or in any position of the sentence, therefore they were annotated as prepositional phases using PP. Figure 4-5 shows examples of questions with prepositional phrases, between square brackets, however only those phrases starting with "for", "between", "on", or "in" were tagged as a PP.

```
What/WP1 is/VBZ1 the name/NP1 [of/IN1 the managing director/NP2]
[of/IN2 Apricot Computer/NP3]?

What/WP1 is/VBZ1 the fare cost/NP1 [for the round trip/PP1]
[between New York and London/PP2] [on Concorde/PP3]?

What/WP1 was/VBD1 the monetary value/NP1 [of/IN1 the Nobel Peace
Prize/NP2] [in 1989/PP1]?
```

Figure 4-5 Prepositional Phrase exceptions

Step 3. After each file is revised, we extract the words and tags from each file and write the tags (patterns) from both the training-set and the input-questions file out to separate files with patterns for later comparison. (see Figure. 4-6)

Step 4. This step entails the manual extraction of answer patterns for each type of question, according to the answers given in the questionnaire (see tables 3-3 to 3-10 for details).

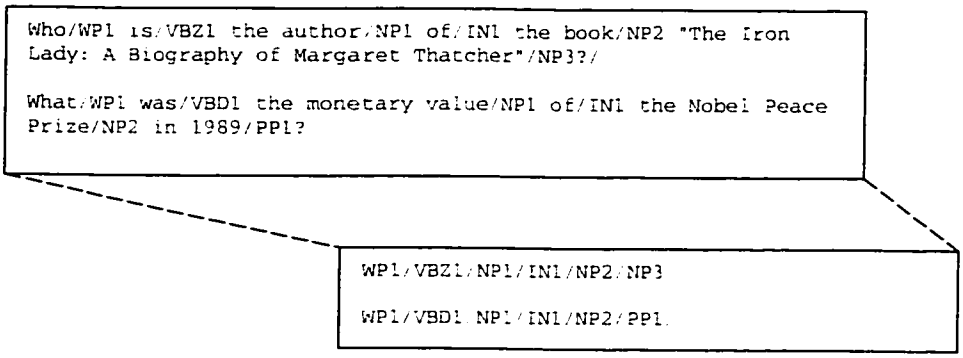


Figure 4-6. Extracting tags/word process

4.4.2 Matching Process

The matching process involves two final steps:

Step 5. Matching the pattern of the input question with the patterns computed from the training set. This process looks for an exact match of the tag sequence of each input question (see Step 3) against the patterns generated from the training set. Upon finding a match the system will produce the relevant answer sentence(s).

Step 6. A manual evaluation of the grammar on each question successfully matched is carried out to confirm that the answer formulations were grammatically correct. (see section 5.2).

4.5 Design and Implementation Process

After the analysis of the requirements and processes to follow for our application, and studying the advantages and disadvantages of the programming languages available, we decided to work with Perl for the design and implementation. The advantage of Perl over the others was that it has many built-in features for character and string manipulation. Mastering these features makes it possible to write string code quickly and concisely. In our case the extraction of tags and the pattern-tag matching make Perl the perfect selection. The other advantage is that Perl is implemented on a wide variety of platforms, so provides a useful tool for portable code.

4.5.1 Our Development process

The problem was broken down into two parts: data storage and data manipulation.
For storage, our options were rather open, with the restriction that we have a trustworthy correlation between a pattern (tag) in one file and the questions (words), which the pattern describes in another. Thus, for any data set such as

*<S> What/WP does/VBZ <NP8> the/DT Peugeot/NNP Company/NNP </NP> manufacture/VB ?/.*end-of-sentence*</S>*

The data structure used for doing our actual look up; is a hash of arrays, where the keys in the hash are the sequence of tags for a particular question, and the values are arrays of the corresponding question that have that structure. We used arrays as the values because hashes guarantee the uniqueness of individual keys and in the instance of two grammatically identical sentences (according to our tag schema) the newer entry will overwrite the older.

We will need the arrays to handle cases of multiple question/answer pair per rule since hashes eliminate duplicate keys. Initially, as we saw in the previous section of working process of AnsForm (see figure 4-1), we generated files in each step of the process because we can have as many duplicate entries as we want, but we ended up placing both files into arrays and then combining them into a hash table. Then, as we run through a list of rules for which we wish to find question/answer pairs, we only have to do the hash lookup.

To do a matching against another tag (rule) file, we use each line to do a key lookup in our hash. This returns a reference to an array of questions that have that structure (see figure 4-6).

```
while (my $rule = chomp(<in>) {
      my $questions = $pairs{$rule};
      # Process data returned
    }
```

Figure 4-6. Matching Perl code

4.5.2 Example of the internal running process of Ans$\mathcal{F}$orm

In section 4.4 we have described the process flow of Ans$\mathcal{F}$orm. Let us now go through a specific example. Figure 4-7 shows how the Ans$\mathcal{F}$orm system behaves: as input we have two files, the *training-set* and the *input-questions*, which both contain questions from the TREC collection. From the *training-set* we automatically extracted the patterns for the type of question and manually generated the variation of patterns of answers from each type of question. Once we produced the pattern files for both the questions and their corresponding answers, we proceeded to automatically tag and extract the patterns from each question of the *input-questions* file. The tagged questions are then matched with the questions patterns of the *training-set*. If a counterpart is found, the system checks the answer patterns associated with the matched question and automatically generates the candidate answer(s) as our output.

The candidate answers were produced matching each tag from the answer-pattern to each pair word/tag from the question - evaluated and matched against the questions in the patterns file, to finally reformulate the question (its words) following the order which the tags in the answer-pattern suggested.

The next chapter presents and evaluates the results of the system and discusses further work that we could perform to improve the Ans$\mathcal{F}$orm system.

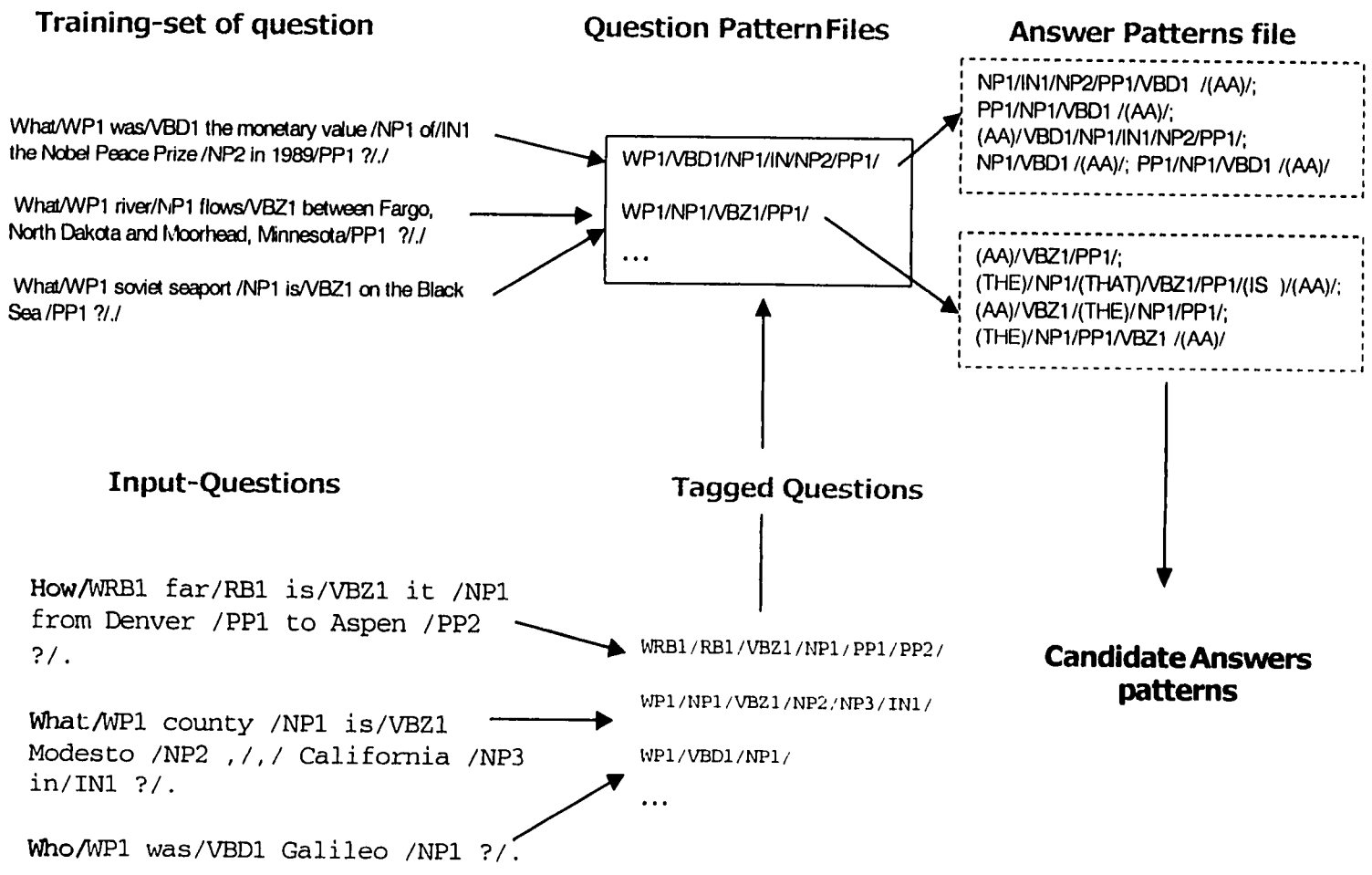


Figure 4-7 Example of the internal running process of AnsForm

Chapter 5

5 AnsForm Evaluation Results and Conclusions

5.1 Evaluation of the System Matching Process

To evaluate our approach, we performed one experiment with a test-set of 120 questions taken semi-randomly[5] from the TREC question collection. Table 5-1 shows the results of the experiment. The coverage of the 120 questions was only 47%, that is, about half of the questions of the test-set matched one of the question-patterns.

# Type Questions	Type of question	% Matched Questions		No. Patterns by type of question
64	What	31	48%	43
16	How	6	37%	20
15	Who	9	60%	10
8	When	7	88%	5
7	Which	0	0%	3
3	Why	0	0%	2
7	Where	3	49%	3
0	Name	-	-	6
120		**56/47%**		**92 Patterns**

Table 5-1 Results analysis of matching process for type of question.

The low performance of our system is due to the large variety of type of questions we used (eight types), the variety of formulations for each type, and the small size of the training set. As we explained in section 3.4, from these results, we arrived at the conclusion that we should consider the evaluation of only one (1) type of question to have a better representative number of patterns for that type of question under any formulation, therefore that could represent a better performance of our system.

[5] We made sure that none of the questions in the test-set were also in the training-set.

Table 5-1 shows how the 120 questions from the test-set were distributed among the eight (8) types of questions, what was the coverage (the percentage of questions that matched for each type), and the number of patterns that we actually have for each type of question. We showed the number of patterns for each type of questions in order to see if there is a direct relation between the number of pattern for type of questions and the percentage of the matched question obtained (the coverage). In our experiment, the system performed poorly on the types *What* and *Who* which for their particular formulations, need a larger number of patterns. For example, for question type *How* that has different subtypes, the 20 patterns we produced were almost distributed by one in each subtype, such as *How long*, *How many*, *How much*, *How large*, *How big*, *How old*, *How tall*, *How hot*, *How far*, *How did*, *How do*, *How does*, *How Could*, etc. that made us have not enough patterns for the different formulations we can get from each subtype, hence the low coverage.

The results for questions of type *Which*, *Why* and *Where* were not a surprise, considering the number of patterns we produced from the training-set for each of them. Besides, the question type *Where* got a high score even if we have only three (3) patterns for this type. As mentioned in section we believe that this is because *Where* questions have a more stereotypical structure.

5.2 Evaluation of the System Grammar Evaluation Process (GEP)

The evaluation of the *A*nswer *F*ormulations using the test-set of questions, resulted in only 88% of grammatically correct answer formulations. Table 5-2 shows in the last column the percentage of questions, which were grammatically correct formulated out of the number of matched questions for each type of question.

# Type Questions	Type of question	% Matched Questions		% Grammatically correct	
64	What	31	48%	24	77%
16	How	6	37%	6	100%
15	Who	9	60%	6	66%
8	When	7	88%	7	100%
7	Which	0	0%	-	-
3	Why	0	0%	-	-
7	Where	3	49%	3	100%
0	Name	-	-	-	-
120		**56/47%**		**46/88%**	

Table 5-2 Result of the grammar analysis for each type of question.

To describe the process of the grammar evaluation, we considered different cases:

- Different questions, sharing the same pattern will generate identical grammatically correct answer formulations.

Pattern in common: WP1/NP1/VBZ1/PP1/

What/WP1 person's head/NP1 is/VBZ1 on a dime /PP1 ?/.

(AA) is on a dime | (THE) person's head (THAT) is on a dime (IS) (AA) | (AA) is (THE) person's head on a dime | (THE) person's head on a dime is (AA)

What/WP1 soviet seaport/NP1 is/VBZ1 on the Black Sea/PP1 ?/./

(AA) is on the Black Sea | (THE) soviet seaport (THAT) is on the Black Sea (IS) (AA) | (AA) is (THE) soviet seaport on the Black Sea | (THE) soviet seaport on the Black Sea is (AA)

- But, with the same question and answer patterns, different questions will not generate grammatically correct answer formulations.

  Pattern in common: WP1/VBD1/NP1

  Who/WP1 was/VBD1 Galileo /NP1 ?/.

  Galileo (WAS) was (BY) (AA)

  Who/WP1 discovered/VBD1 radium /NP1 ?/.

  radium (WAS) discovered (BY) (AA)

In this particular case, the two questions share the same question patterns, but do not share the same set of answer patterns. While the second question can be answered by *radium (WAS) discovered (BY) (AA)*; the first cannot use the same answer pattern as the output formulation *Galileo (WAS) was (BY) (AA)* is grammatically incorrect. A solution to this problem would be to assign different tags to different types of verbs in order to produce distinct patterns for each case and generate correct answer formulations.

- Here is another example of a case when sharing the same question and answer patterns, different questions will not generate grammatically correct answer formulations

What/WP1 metal/NP1 has/VBZ1 the highest melting point/NP2 ?/.

the highest melting point has **(A)** (AA) metal | **(THE)** metal has (AA) | (AA) has the highest melting point | (AA) **(IS)** **(THE)** metal **(THAT)** has the highest melting point | (AA) **(IS)** **(THE)** metal (WITH) the highest melting point

What/WP1 type of bridge/NP1 is/VBZ1 the Golden Gate Bridge /NP2 ?/./

the Golden Gate Bridge is **(A)** (AA) type of bridge | **(THE)** type of bridge is (AA) | (AA) is the Golden Gate Bridge | (AA) **(IS)** **(THE)** type of bridge **(THAT)** is the Golden Gate Bridge | (AA) **(IS)** **(THE)** type of bridge **(WITH)** the Golden Gate Bridge

What/WP1 state /NP1 has/VBZ1 the most Indians /NP2 ?/.

(AA) **(IS)** **(THE)** state **(THAT)** has the most Indians

In this case the problem was the optional words (between parenthesis) we added to the answer patterns (see section 3.3.1), which for some question formulations work well but for others don't make any sense at all. For example the answer pattern with the following sequences of tags (AA) + (IS) + (THE) + NP1 + VBZ1 + NP2 works well with the question *What state **has** the most Indians?*, but it does not with the question *What type of*

bridge is the Golden Gate Bridge?, which gave a formulation as follows: *Suspension (AA) is the type of bridge that is the Golden Gate Bridge*. The phenomenon is due to the different type of verb used, which makes the answer ungrammatical.

After the analysis of the results from the test-set, we concluded that most of the problems were due to the tagging process; some of the verbs were tagged with the same category, so they ended-up having the same tag, which looking in the context of the question and the type of question, they should have different question pattern and therefore different answer formulations.

5.3 Conclusions

The answer formulation techniques presented here show that the system can effectively produce answers for most of the matched questions, but the average percentage of grammatically correct constructions of each type of questions, and the low coverage of matched questions, made us conclude that the techniques used were not enough. To increase the coverage, a larger training set is required, but to increase the percentage of grammatically correct answers, we need to do additional research to successfully produce high performance score in the formulation of answers grammatically correct.

"Answer formulations is more than a part-of speech pattern matching process"

This means that we should also concentrate on the context of the question to generate the patterns and produce the answer formulations. Thus this research could be considered as

a good start point for the research to be done, since the techniques used during the whole process of the Ans$\mathcal{F}$orm system took into account part-of-sppech tagging, noun phrase and prepositional phrase extraction, important steps to evaluate the context of a phrase.

5.4 Suggestions for future work

Based on the evaluation of the system (see sections 5.1 and 5.2), it appears that it would be useful to do additional work on the NLP tools used, to improve the input to our system. (Taking into account that those tools were not created to parse questions.) Another improvement will be in the techniques our system was based on, and the following are the main steps we should concentrate for further work:

- Concentrate on only one type of question to develop patterns of better quality.
- Include our own annotations on Brill's tagger and NPE to avoid problems in the extraction of Tags and NPs required, and generate them automatically.
- Include semantic analysis (e.g. dates).
- Build our own grammar for answer formulation.
- Incorporate our work into a working QA system.

Ideally, we want not only to formulate answers following a static structure, but rather use the knowledge of the context of the two structures, the question and the short answer to generate a natural and complete answer.

Appendix A: Survey

Concordia University
Department of Computer Science

Please find enclosed a Questionnaire with 150 questions. You are invited to take part in this survey of research by completing this questionnaire and returning it as soon as possible by email to the following e-address: anayaca@cs.concordia.ca

Purpose of this Survey:

The aim of the Questionnaire is to gather information about how people usually answer specific questions. We will use the information gathered to improve "Question-Answering" (QA) applications, which could provide - in one sentence- a complete answer according to the formulation of the question and the short answer given by the system. Specifically that information will be used to train a new module, which will make possible the improvement of the QA application. This survey and the module to be developed are part of my research, required to complete my major report and the master's program.

So your response is important to the success of this research.

The report of this research will be produced at the end of August 2002, and the result will be available in http://www.cs.concordia.ca/~grad/anayaca/

Contact: Glenda Anaya (anayaca@cs.concordia.ca)

Questionnaire

Name of the person completing this questionnaire:
Profession:
Organization:
E-mail:

For the following questions (1), given the short-answer between (), please give a complete one type of answer, in one sentence, for example question 1;

1. Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"? (Hugo Young)
   Ans.: Hugo Young is the author of the book "The Iron Lady: A Biography of Margaret Thatcher".
   Ans.: The author of the book... is ()
   Ans.: "The Iron Lady:.." book was written..()
   Ans.: ...

2. What was the monetary value of the Nobel Peace Prize in 1989? ($469,000)
Ans.:

3. What does the Peugeot Company manufacture? (cars)
Ans.:

4. How much did Mercury spend on advertising in 1993? (Pounds 12m)
Ans.:

5. What is the name of the managing director of Apricot Computer? (Peter Horne)
Ans.:

6. Why did David Koresh ask the FBI for a word processor? (To record his revelations)
Ans.:

7. What debts did Qintex group leave? (A Dollars 1.5bn)
Ans.:

8. What is the name of the rare neurological disease with symptoms such as: involuntary movements (tics), swearing, and incoherent vocalizations (grunts, shouts, etc.)? (Tourette's Syndrome)
Ans.:

9. How far is Yaroslavl from Moscow? (150 miles)
Ans.:

10. Name the designer of the shoe that spawned millions of plastic imitations, known as "jellies". (Andrea Pfister)
Ans.:

11. Who was President Cleveland's wife? (Frances Folsom)
Ans.:

12. How much did Manchester United spend on players in 1993? (Pounds 4m.)
Ans.:

13. How much could you rent a Volkswagen bug for in 1966? ($1 a day)
Ans.:

14. What country is the biggest producer of tungsten? (China)
Ans.:

15. When was London's Docklands Light Railway constructed? (1980s)
Ans.:

16. What two US biochemists won the Nobel Prize in medicine in 1992? (Edwin Krebs and Edmond Fischer)
Ans.:

17. How long did the Charles Manson murder trial last? (9 1/2-month)
Ans.:

18. Who was the first Taiwanese President? (Lee Teng-Hui)
Ans.:

19. Who was the leader of the Branch Davidian Cult confronted by the FBI in Waco, Texas in 1993? (Mr. David Koresh)
Ans.:

20. When was Dubai's first concrete house built? (1956)
Ans.:

21. Who is the president of Stanford University? (Donald Kennedy)
Ans.:

22. Who invented the road traffic cone? (David Morgan)
Ans.:

23. Who was the first doctor to successfully transplant a liver? (Dr. Thomas Starzl)
Ans.:

24. When did Nixon die? (April 22, 1994)
Ans.:

25. Where is Microsoft's corporate headquarters located? (Redmond, Wash)
Ans.:

26. How many calories are there in a Big Mac? (562)
Ans.:

27. What is the acronym for the rating system for air conditioner efficiency? (SEER)
Ans.:

28. Name a film that has won the Golden Bear in the Berlin Film Festival? (In The Name Of The Father)
Ans.:

29. Who was President of Costa Rica in 1994? (Rafael Angel Calderon)
Ans.:

30. What is the fare cost for the round trip between New York and London on Concorde? ($6,400)
Ans.:

31. What brand of white rum is still made in Cuba? (Havana Club)
Ans.:

32. What is the name of the chronic neurological autoimmune disease, which attacks the protein sheath that surrounds nerve cells causing a gradual loss of movement in the body? (Multiple Sclerosis)
Ans.:

33. What nuclear-powered Russian submarine sank in the Norwegian Sea on April 7, 1989? (Komsomlets)
Ans.:

34. Who is the voice of Miss Piggy? (Frank Oz)
Ans.:

35. Name a country that is developing a magnetic levitation railway system? (Japan/ Germany)
Ans.:

36. Name the first private citizen to fly in space. (Christa McAuliffe)
Ans.:

37. What is the longest river in the United States? (the Mississippi)
Ans.:

38. What does El Nino mean in Spanish? (boy child)
Ans.:

39. Who came up with the name, El Nino? (Peruvian fishermen)
Ans.:

40. Who is the founder of Scientology? (L. Ron Hubbard)
Ans.:

41. Which city in China has the largest number of foreign financial companies? (Shanghai)
Ans.:

42. Who released the Internet worm in the late 1980s? (Robert Morris)
Ans.:

43. Who first circumnavigated the globe? (Ferdinand Magellan)
Ans.:

44. Who wrote the song, "Stardust"? (Hoagy Carmichael)
Ans.:

45. What country is the worlds leading supplier of cannabis? (Ghana)
Ans.:

46. What time of day did Emperor Hirohito die? (6:33 a.m./ 1.33 p.m. Friday PST)
Ans.:

47. How large is the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast? (19 million acre)
Ans.:

48. Where is the highest point in Japan? (Mt. Fuji)
Ans.:

49. What is the term for the sum of all genetic material in a given organism? (genome)
Ans.:

50. What is considered the costliest disaster the insurance industry has ever faced? (Hurricane Andrew)
Ans.:

51. What was the name of the first Russian astronaut to do a spacewalk? (Aleksei A. Leonov)
Ans.:

52. Where is Belize located? (Central America)
Ans.:

53. How much folic acid should an expectant mother get daily? (400 micrograms)
Ans.:

54. What type of bridge is the Golden Gate Bridge? (suspension)
Ans.:

55. What is the population of the Bahamas? (250,000)
Ans.:

56. How far away is the moon? (quarter of a million miles)
Ans.:

57. What is Francis Scott Key best known for? (The Star Spangled Banner)
Ans.:

58. What state has the most Indians? (California)
Ans.:

59. Who invented the paper clip? (Norwegian Johan Varler)
Ans.:

60. How many dogs pull a sled in the Iditarod? (8-12)
Ans.:

61. Where did bocci originate? (Italian)
Ans.:

62. Who made the first airplane that could fly? (Orville and Wilbur Wright)
Ans.:

63. How many astronauts have been on the moon? (12 men)
Ans.:

64. Who is Coronado? (Francisco Vasquez de Coronado) (Explorer)
Ans.:

65. Name one of the major gods of Hinduism? (Brahma)
Ans.:

66. What does the abbreviation OAS stand for? (Organization of American States)
Ans.:

67. Who is Barbara Jordan? (Former U.S. Rep)
Ans.:

68. How big is Australia? (nearly 3 million square miles)
Ans.:

69. Who found Hawaii? (Captain Cook)
Ans.:

70. Who is the richest person in the world? (Sultan Hassanai Bolkiah)
Ans.:

71. How many films did Ingmar Bergman make? (more than 50 films)
Ans.:

72. What is the federal minimum wage? ($5.05)
Ans.:

73. What did brontosauruses eat? (plant-eaters)
Ans.:

74. What is California's state tree? (Redwood)
Ans.:

75. How many types of lemurs are there? (30 types)
Ans.:

76. What is leukemia? (blood cancer)
Ans.:

77. Who was the first coach of the Cleveland Browns? (Paul Brown)
Ans.:

78. How many people die from snakebite poisoning in the U.S. per year? (13)
Ans.:

79. What is the size of Argentina? (size of India)
Ans.:

80. What do manatees eat? (unwanted vegetation)
Ans.:

81. When was the San Francisco fire? (April 18, 1906)
Ans.:

82. What was the man's name who was killed in a duel with Aaron Burr? (Alexander Hamilton)
Ans.:

83. What is the population of Mexico? (85.8 million)
Ans.:

84. When was the slinky invented? (1943)
Ans.:

85. How hot is the core of the earth? (5,000 degrees Celsius)
Ans.:

86. What is porphyria? (rare metabolic disorder)
Ans.:

87. What ocean did the Titanic sink in? (Atlantic)
Ans.:

88. Who was the 33rd president of the United States? (Harry Truman)
Ans.:

89. At what speed does the Earth revolve around the sun? (30 km/sec.)
Ans.:

90. Who is the emperor of Japan? (Akhito)
Ans.:

91. What does Final Four refer to in the sports world? (NCAA basketball Final Four)
Ans.:

92. What does Knight Ridder publish? (29 Daily Newspapers)
Ans.:

93. What task does the Bouvier breed of dog perform? (herding cattle)
Ans.:

94. What sport do the Cleveland Cavaliers play? (basketball)
Ans.:

95. What year was the Avery Dennison company founded? (1990)
Ans.:

96. What's the population of Biloxi, Mississippi? (46,000)
Ans.:

97. Name a ballet company Mikhail Baryshnikov has danced for? (George Balanchine's New York City Ballet)
Ans.:

98. What was the name of the television show, staring Karl Malden, that had San Francisco in the title? (Streets of San Francisco)
Ans.:

99. Who was the founding member of the Pink Floyd band? (Roger Waters)
Ans.:

100. What did Delilah do to Samson's hair? (cut it off)
Ans.:

101. What is an atom? (the smallest units of matter.)
Ans.:

102. When did Hawaii become a state? (1959)
Ans.:

103. How tall is the Sears Building? (1454 feet.)
Ans.:

104. George Bush purchased a small interest in which baseball team? (Texas Rangers)
Ans.:

105. Why does the moon turn orange? (eclipse)
Ans.:

106. When did John F. Kennedy get elected as President? (1960)
Ans.:

107. How old was Elvis Presley when he died? (42)
Ans.:

108. Where is the Orinoco River? (Venezuela)
Ans.:

109. How much fiber should you have per day? (35 grams)
Ans.:

110. How many Great Lakes are there? (5)
Ans.:

111. Material called linen is made from what plant? (Flax)
Ans.:

112. What river flows between Fargo, North Dakota and Moorhead, Minnesota? (Red River)
Ans.:

113. What do bats eat? (Insects)
Ans.:

114. What state did the Battle of Bighorn take place in? (Montana)
Ans.:

115. Who was Abraham Lincoln? (president)
Ans.:

116. What do you call a newborn kangaroo? (Joey)
Ans.:

117. What are spider veins? (enlargement of the blood vessels)
Ans.:

118. What day and month did John Lennon die? (Dec. 8)
Ans.:

119. What strait separates North America from Asia? (Bering)
Ans.:

120. Who was the first female United States Representative? (Jeanette Rankin)
Ans.:

121. What are xerophytes? (Store-water plants)
Ans.:

122. What country did Ponce de Leon come from? (Spain)
Ans.:

123. The U.S. Department of Treasury first issued paper currency for the U.S. during which war? (Civil War)
Ans.:

124. What is desktop publishing? (computer application)
Ans.:

125. What is the temperature of the sun's surface? (above 10,800 degrees Fahrenheit)
Ans.:

126. What is the length of the coastline of the state of Alaska? (35,000 miles)
Ans.:

127. What is the name of Neil Armstrong's wife? (Jan)
Ans.:

128. What is Hawaii's state flower? (Yellow hibiscus)
Ans.:

129. Who won Ms. American in 1989? (Gretchen Elizabeth)
Ans.:

130. When did the Hindenberg crash? (1937)
Ans.:

131. What mineral helps prevent osteoporosis? (Calcium)
Ans.:

132. What was the last year that the Chicago Cubs won the World Series? (1908)
Ans.:

133. Where is Perth? (Australia)
Ans.:

134. What year did WWII begin? (1939)
Ans.:

135. What were Christopher Columbus' three ships? (Nina, Pinta, Santa Maria)
Ans.:

136. What does Phi Beta Kappa mean? (Academic honor society)
Ans.:

137. What is nicotine? (Alkaloid)
Ans.:

138. What is another name for vitamin B1? (Thiamine)
Ans.:

139. Who discovered radium? (Marie and Pierre Curie)
Ans.:

140. What are sunspots? (Cool spots)
Ans.:

141. When was Algeria colonized? (19th Century)
Ans.:

142. What is the capital of Ethiopia? (Addis Ababa)
Ans.:

143. For how long is an elephant pregnant? (22-month)
Ans.:

144. How did Janice Joplin die? (Overdose of heroin)
Ans.:

145. What is the primary language in Iceland? (Icelandic)
Ans.:

146. What is the Moulin Rouge? (cabaret)
Ans.:

147. What soviet seaport is on the Black Sea? (Sevastopol)
Ans.:

148. What is the atomic weight of silver? (110)
Ans.:

149. What currency do they use in Brazil? (Cruzeiros)
Ans.:

150. What are pathogens? (agents of disease)
Ans.:

Reference:
(1) http://trec.nist.gov/data/qa.html. TREC-8; QUESTIONS 2-19, 50-69, 90-100. TREC-9; QUESTIONS 201-211, 234-238, 292-302, 325-331, 374-78, 670-679. TREC-10; QUESTIONS: 897-902,928-934, 956-963, 981-986, 1006-14, 1041-47, 1161-64, 1262-66.

Appendix B: Analysis of each type of questions
Analysis of *WHAT* type of questions

1. **What was the monetary value of the Nobel Peace Prize in 1989? ($469,000)**
   (70%) 14/20 Ans.: The monetary value of the Nobel Peace Prize in 1989 was $469,000.
   (1/14) Ans.: The value of the Nobel Peace Prize in 1989 was $469,000
   (5%) 1/20 Ans.: The Nobel Peace Prize value in 1989 was $469,000.
   (5%) 1/20 Ans.: In 1989 the Nobel Peace Prize was worth $469,000
   (5%) 1/20 Ans.: $469,000 was the monetary value of the Nobel Peace Prize
   (5%) 1/20 Ans.: The monetary value of the 1989 Nobel Peace Prize was $469,000.
   (5%) 1/20 Ans.: The value was $469,000.
   (5%) 1/20 Ans.: In 1989, the monetary value was $469,000

2. **What is the name of the managing director of Apricot Computer? (Peter Horne)**
   (50%) 10/20 Ans.: Peter Horne is the managing director of Apricot Computer.
   (25%) 5/20 Ans.: The managing director of Apricot computer is Peter Home.
   1/5 Ans.: The manager of Apricot Computer is Peter Home
   (10%) 2/20 Ans.: The name of the managing director of Apricot Computer is Peter Home.
   (5%) 1/20 Ans. Peter Home manages Apricot Computer.
   (5%) 1/20 Ans.: His name is Peter Home
   (5%) 1/20 Ans.: Peter Horne is who directs Apricot Computer

3. **What is the name of the rare neurological disease with symptoms such as: involuntary movements (tics), swearing, and incoherent vocalizations (grunts, shouts, etc.)? (Tourette's Syndrome)**
   (50%) 10/20 Ans.: Tourette's Syndrome is the name of the rare neurological disease...
   1/10 Ans.: Tourette's Syndrome is the name of the disease...
   (15%) 3/20 Ans.: The name of that rare neurological disease is Tourette's Syndrome.
   1/3 Ans.: the name of the disease is ()
   (10%) 2/20 Ans.: Tourette's Syndrome is the rare neurological disease...
   1/2 Ans.: Tourette's Syndrome is the neurological disease...
   (10%) 2/20 Ans.: the "rare neurological" disease... is "called" Tourette's Syndrome.
   (5%) 1/20 Ans.: The name of the "rare neurological" disease... is Tourette's Syndrome.
   (5%) 1/20 Ans.: It is Tourette's Syndrome
   (5%) 1/20 Ans.: Tourette's Syndrome.

4. **What is the acronym for the rating system for air conditioner efficiency? (SEER)**
(55%) 11/20 Ans.: SEER is the acronym for the rating system for air conditioner efficiency
(30%) 6/20 Ans.: The acronym for the rating system for air conditioner efficiency is SEER.
(5%) 1/20 Ans.: SEER is the acronym used for rating air conditioner efficiency.
(5%) 1/20 Ans.: SEER is the acronym used as a rating system for air conditioner efficiency
(5%) 1/20 Ans.: The acronym used to rate air conditioner efficiency is SEER.

5. **What is the fare cost for the round trip between New York and London on Concorde? ($6,400)**
(55%) 11/20 Ans.: the fare cost for the round trip between New York and London on Concorde is $6,400
1/11 Ans.: The cost for a round trip between New York and London with Concorde is $6,400
(10%) 2/20 Ans.: It costs $6,400 for a round trip between New York and ...
(10%) 2/20 Ans.: $6,400 is the fare cost for the round trip between New York and London on Concorde.
(15%) 3/20 Ans.: The round trip fare between New York and London on a Concorde is $6,400
1/3 Ans.: the round trip between New York and London on Concorde costs $6,400.
1/3 Ans.: A round trip from New York to London on the Concorde costs $6,400.
(5%) 1/20 Ans.: A round trip fare for traveling between NY and London costs $6,400
(5%) 1/20 Ans.: The round trip fare cost on Concorde between New York and London is $6,400.

6. **What is the name of the chronic neurological autoimmune disease, which attacks the protein sheath that surrounds nerve cells causing a gradual loss of movement in the body? (Multiple Sclerosis)**
(55%) 11/20 Ans.: Multiple Sclerosis is the name of the chronic neurological autoimmune disease, which attacks the protein sheath that surrounds nerve cells causing a gradual loss of movement in the body+
(10%) 2/20 Ans.: Multiple Sclerosis is the chronic neurological autoimmune disease, which (that) attacks the protein sheath that surrounds nerve cells causing a gradual loss of movement in the body
(10%) 2/20 Ans.: The chronic neurological autoimmune disease, which attacks the protein sheath that surrounds nerve cells causing a gradual loss of movement in the body, is called Multiple Sclerosis.
(15%) 3/20 Ans.: The name of the... is Multiple Sclerosis
(10%) 2/20 Ans.: Multiple Sclerosis.

7. **What is the longest river in the United States? (the Mississippi)**
(80%) 16/20 Ans.: the Mississippi is the longest river in the United States
(20%) 4/20 Ans.: The longest river in the US is the Mississippi

8. **What is the term for the sum of all genetic material in a given organism? (genome)**
(50%) 10/20 Ans.: Genome is the term for the sum of all genetic material in a given organism
1/10 Ans.: The Genome is the term used for the sum of all genetic material in a given organism
(20%) 4/20 Ans.: the term for the sum of all genetic material in a given organism is Genome.

(10%) 2/20 Ans.: the sum of all genetic material in a given organism is called Genome.
1/20 Ans.: Genome is the term given the sum of all genetic material in a given organism
(5%) 1/20 Ans.: The term Genome is used to define the sum of all genetic material in a given organism.
(5%) 1/20 Ans.: Genome is the sum of all genetic material in a given organism.
(5%) 1/20 Ans.: The genome.

9. **What is considered the costliest disaster the insurance industry has ever faced? (Hurricane Andrew)**
(65%) 13/20 Ans.: Hurricane Andrew is considered the costliest disaster the insurance industry has ever faced
(5%) 1/20 Ans.: Hurricane Andrew is considered the costliest disaster in the Insurance Industry.
(5%) 1/20 Ans.: Hurricane Andrew is the costliest disaster the insurance industry ...
(5%) 1/20 Ans.: The costliest disaster the insurance industry has ever faced is considered to be Hurricane Andrew
(5%) 1/20 Ans.: the costliest disaster ... is Hurricane Andrew
(5%) 1/20 Ans.: The costliest disaster the insurance industry faced was Hurricane Andrew.
(5%) 1/20 Ans.: the costliest disaster of the insurance industry is Hurricane Andrew
(5%) 1/20 Ans.: Hurricane Andrews.

10. **What was the name of the first Russian astronaut to do a spacewalk? (Aleksei A. Leonov)**
(40%) 8/20 Ans.: Aleksei A. Leonov was the first Russian astronaut to do a spacewalk
(35%) 7/20 Ans.: The first Russian astronaut to do a spacewalk was Aleksei A. Leonov.
(10%) 2/20 Ans.: Aleksei A. Leonov was the name of the first Russian astronaut..
(10%) 2/20 Ans.: The name of the first Russian astronaut to do a spacewalk was Aleksei A. Leonov.
(5%) 1/20 Ans.: Aleksei A.Leonov

11. **What is the population of the Bahamas? (250,000)**
(55%) 11/20 Ans.: The population of the Bahamas is 250,000.
(15%) 3/20 Ans.: the Bahamas has a population of 250,000 habitants
(15%) 3/20 Ans.: Bahamas population is 250,000.
(10%) 2/20 Ans.: 250.000 is the population ...
(5%) 1/20 Ans.: 250.000

12. **What is Francis Scott Key best known for? (The Star Spangled Banner)**
(95%) 19/20 Ans.: Francis Scott Key is best known for "The Star Spangled Banner".
(5%) 1/20 Ans.: Francis Scott is the author of "The Star Spangled Banner"

13. **What is the federal minimum wage? ($5.05)**
(85%) 17/20 Ans.: the federal minimum wage $5.05
(5%) 1/20 Ans.: The minimum wage is $5.05.
(5%) 1/20 Ans.: $5.05 is the federal minimum wage.

(5%) 1/20 Ans.: $5.05

14. **What is California's state tree? (Redwood)**
(50%) 10/20 Ans.: Redwood is California's state tree.
(45%) 9/20 Ans.: California's state tree is the Redwood
(5%) 1/20 Ans.: Redwood is the state tree of California.

15. **What is leukemia? (blood cancer)**
(90%) 18/20 Ans.: Leukemia is a blood cancer.
1/18 Ans.: It is a blood cancer.
(5%) 1/20 Ans.: Leukemia is a type of blood cancer.
(5%) 1/20 Ans.: Leukemia is cancer of the blood.

16. **What is the size of Argentina? (size of India)**
(60%) 12/20 Ans.: Argentina is the size of India.
2/20 Ans.: Argentina is the same size as India.
(10%) 2/20 Ans.: The () is the Size of Argentina
(20%) 4/20 Ans.: the size of Argentina is comparable to the size of India
1/4 Ans.: The size of Argentina is equal to the size of India.
1/4 Ans.: The size ... is around the size of India
1/4 Ans.: It is about the size of India.

17. **What was the man's name who was killed in a duel with Aaron Burr? (Alexander Hamilton)**
(30%) 6/20 Ans.: Alexander Hamilton was killed in a duel with Aaron Burr.
(20%) 4/20 Ans.: Alexander Hamilton was the man who was killed in a duel with Aaron Burr
(15%) 3/20 Ans.: The man who was killed in a duel with Aaron Burr was Alexander Hamilton.
(10%) 2/20 Ans.: The name of the man who was killed in a duel with Aaron Burr is Alexander Hamilton.
(10%) 2/20 Ans.: Alexander Hamilton was the man's name who was killed in a duel with Aaron Burr
(5%) 1/20 Ans.: Alexander Hamilton is the name of the man killed in a duel with Aaron Burr.
(5%) 1/20 Ans.: Alexander Hamilton.
(5%) 1/20 Ans.: It was Alexander Hamilton.

18. **What is the population of Mexico? (85.8 million)**
(65%) 13/20 Ans.: The population of Mexico is 85.8 million.
(15%) 3/20 Ans.: Mexico has a population of 85.8 million
(10%) 2/20 Ans.: 85.8 milllion.
(5%) 1/20 Ans.: Mexico's population is 85.8 million.
(5%) 1/20 Ans.: Mexico has 85.8 mn of habitants

19. **What is porphyria? (rare medabolic disorder)**
(90%) 18/20 Ans.: porphyria is a rare medabolic disorder
(5%) 1/20 Ans.: It is a rare medabolic disorder.
(5%) 1/20 Ans.: Rare medabolic disorder.

20. **What's the population of Biloxi, Mississippi? (46,000)**
(80%) 16/20 Ans.: the population of Biloxi, Mississippi is 46,000 habitants
(10%) 2/20 Ans.: 46,000.
(5%) 1/20 Ans.: 46,000 is the population...
(5%) 1/20 Ans.: Biloxi, Mississippi has 46,000 inhabitants

21. **What was the name of the television show, staring Karl Malden, that had San Francisco in the title? (Streets of San Francisco)**
(40%) 8/20 Ans.: Streets of San Francisco was the name of the television show, staring Karl Malden, that had San 6/20 Ans.: The name of the television show, staring Karl Malden, that had San Francisco in the title is Streets of San Francisco.
(15%) 3/20 Ans.: () was the television show, staring by Karl Malden.
(5%) 1/20 Ans.:Karl Malden starred in the television show entitled The Streets of San Francisco.
(5%) 1/20 Ans.: Karl Malden stared in the television show Streets of San Francisco
(5%) 1/20 Ans.: It was Streets of San Francisco.
(5%) 1/20 Ans.: Streets of San Francisco.

22. **What is an atom? (the smallest units of matter.)**
(100%) 20/20 Ans.: an atom is the smallest units of matter.
(5%) 1/20 Ans.: It is the smallest unit of matter.
(5%) 1/20 Ans.: Atoms are the smallest units of matter.

23. **What are spider veins? (enlargement of the blood vessels)**
(100%) 20/20 Ans.:Spider veins are an enlargement of the blood vessels
(5%) 1/20 Ans.: They are enlargement of the blood vessels.

24. **What are xerophytes? (Store-water plants)**
(90%) 18/20 Ans.: xerophytes are Store-water plants
1/18 Ans.: Xerophytes are water storing plants.
1/18 Ans.: They are store-water plants.
(5%) 1/20 Ans.: Store-water plants are called Xerophytes
(5%) 1/20 Ans.: Store-water plants.

25. **What is desktop publishing? (computer application)**
(90%) 18/20 Ans.: desktop publishing is a computer application
(5%) 1/20 Ans.: It is a computer application.
(5%) 1/20 Ans.: Computer application is desktop publishing.
(5%) 1/20 Ans.: Computer application.

26. **What is the temperature of the sun's surface? (above 10,800 degrees Fahrenheit)**
(85%) 17/20 Ans.: the temperature of the sun's surface is above 10,800 degrees Fahrenheit
(10%) 2/20 Ans.: The sun's surface "temperature" is above 10,800 degrees Fahrenheit.
(5%) 1/20 Ans.: Above 10,800 degrees Fahrenhaeit.

27. **What is the length of the coastline of the state of Alaska? (35,000 miles)**
(55%) 11/20 Ans.: The length of the coastline of the state of Alaska is 35,000 miles
(25%) 5/20 Ans.: The coastline of the state of Alaska is 35,000 miles long.
(5%) 1/20 Ans.: Its coastline is 35,000 miles.
(5%) 1/20 Ans.: Alaska has a coastline of 35,000 miles.
(5%) 1/20 Ans.: 35,000 miles is the length of the coastline...
(5%) 1/20 Ans.: 35000

28. **What is the name of Neil Armstrong's wife? (Jan)**
(30%) 6/20 Ans.: Jan is the name of Neil Armstrong's wife.
(30%) 6/20 Ans.: the name of Neil Armstrong's wife is Jan
(15%) 3/20 Ans.: Neil Armstrong's wife is called Jan.
(10%) 2/20 Ans.: Neil Armstrong's wife name is Jan
(10%) 2/20 Ans.: Jan.
(5%) 1/20 Ans.: Neil Armstrong is married to Jan Armstrong.

29. **What is Hawaii's state flower? (Yellow hibiscus)**
(50%) 10/20 Ans.: Hawaii's state flower is the Yellow hibiscus
1/10 Ans.: It is the Yellow hibiscus.

(45%) 9/20 Ans.: The Yellow Hibiscus is Hawaii's state flower.
(5%) 1/20 Ans.: Yellow hibiscus

30. What was the last year that the Chicago Cubs won the World Series? (1908)
(45%) 9/20 Ans.: The last year that the Chicago Cubs won the World Series was 1908.
(25%) 5/20 Ans.: "The" Chicago Cubs "last" won the World Series in 1908.
(10%) 2/20 Ans.: The Cubs won the World series in 1908.
(5%) 1/20 Ans.: The Chicago Cubs won their last World Series in 1908.
(20%) 4/20 Ans.: 1908 was the last year when Chicago Cubs won the World Series
(10%) 2/20 Ans.: 1908

31. What were Christopher Columbus' three ships? (Nina, Pinta, Santa Maria)
(55%) 11/20 Ans.: Christopher Columbus' three ships were the Nina, Pinta and Santa Maria
1/1 Ans.: They were Niña, Pinta, and Santa Maria.
(25%) 5/20 Ans.: The Nina, Pinta, Santa Maria were Christopher Columbus' three ships
(5%) 1/20 Ans.: the three ships of Christopher Columbus were Nina, Pinta and Santa Maria
(5%) 1/20 Ans.: The names of Christopher Columbus' three ships were Niña, Pinta and Santa María
(5%) 1/20 Ans.: Nina, Pinta, Santa Maria

32. What is nicotine? (Alkaloid)
(90%) 18/20 Ans.: Nicotine is an alkaloid.
1/18 Ans.: It is an Alkaloid.
(5%) 1/20 Ans.: Alkaloid is nicotine.
(5%) 1/20 Ans.: Alkaloid

33. What is another name for vitamin B1? (Thiamine)
(65%) 13/20 Ans.: Thiamine is another name for vitamin B1
(20%) 4/20 Ans.: Vitamin B1 is also known as Thiamine.
2/4 Ans.: Vitamin B1 is also called Thiamine.
(5%) 1/20 Ans.: Vitamin B1 is named Thiamine
(5%) 1/20 Ans.: another name for vitamin B1 is Thiamine
(5%) 1/20 Ans.: Thiamine

34. What are sunspots? (Cool spots)
(75%) 15/20 Ans.: sunspots are Cool spots
(10%) 2/20 Ans.: Sunspots are Cool spots on the sun.
1/2 Ans.: They are cool spots on the sun.
(10%) 2/20 Ans.: Cool spots are sunspots.
(5%) 1/20 Ans.: Cool spots

35. **What is the capital of Ethiopia? (Addis Ababa)**
(65%) 13/20 Ans.: Addis Ababa is the capital of Ethiopia
(30%) 6/20 Ans.: The capital of Ethiopia is Addis Ababa.
(5%) 1/20 Ans.: Addis Ababa.

36. **What is the primary language in Iceland? (Icelandic)**
(55%) 11/20 Ans.: Icelandic is the primary language in Iceland
(35%) 7/20 Ans.: The primary language in Iceland is Icelandic
(5%) 1/20 Ans.: It is the Icelandic.
(5%) 1/20 Ans.: Icelandic

37. **What is the Moulin Rouge? (cabaret)**
(100%) 20/20 Ans.: the Moulin Rouge is a cabaret

38. **What is the atomic weight of silver? (110)**
(65%) 13/20 Ans.: the atomic weight of silver is 110
(25%) 5/20 Ans.: 110 is the atomic weight of silver.
(10%) 2/20 Ans.: 110.

39. **What are pathogens? (agents of disease)**
(90%) 18/20 Ans.: Pathogens are agents of disease.
1/18 Ans.: They are agents of disease.
(5%) 1/20 Ans.: () are called pathogens
(5%) 1/20 Ans.: pathogens mean agents of disease

40. **What soviet seaport is on the Black Sea? (Sevastopol)**
(75%) 15/20 Ans.: Sevastopol is the soviet seaport on the Black Sea.
1/15 Ans.: Sevastopol is the Soviet seaport located in the Black Sea
(10%) 2/20 Ans.: Sevastopol
(10%) 2/20 Ans.: the soviet seaport on the Black Sea is Sevastopol
(5%) 1/20 Ans.: Sevastopol is on the Black Sea.

41. **What country is the biggest producer of tungsten? (China)**
(60%) 12/20 Ans.: China is the biggest producer of tungsten
(20%) 4/20 Ans.: The biggest producer of tungsten is China
(5%) 1/20 Ans.: China.
(5%) 1/20 Ans.: China is the biggest producing country of Tungsten

(5%) 1/20 Ans.: (China) is the biggest country producing tungsten
(5%) 1/20 Ans. :It is china.

42. **What brand of white rum is still made in Cuba? (Havana Club)**
(30%) 6/20 Ans.: Havana Club is the brand of white rum still made in Cuba
2/6 Ans.: Havana Club is a brand of white rum still made in Cuba
(25%) 5/20 Ans.: Havana Club is a brand of white rum "that is" still made in Cuba
2/5 Ans.: Havana Club is the brand of white rum "that is" still made in Cuba
(15%) 3/20 Ans.: Havana Club is still made in Cuba.
1/3 Ans.: Havana Club rum is still made in Cuba.
(5%) 1/20 Ans.: Havana Club is still "being" made in Cuba.
(5%) 1/20 Ans.: () is the rum that is still made in Cuba
(5%) 1/20 Ans.: Havana Club is the thing.
(5%) 1/20 Ans.: Havana Club is a white rum still made in Cuba
(5%) 1/20 Ans.: It is Havana Club.
(5%) 1/20 Ans.: *Havana Club*

43. **What country is the worlds leading supplier of cannabis? (Ghana)**
(80%) 16/20 Ans.: Ghana is the worlds leading supplier of cannabis.
(10%) 2/20 Ans.: The world's leading supplier of cannabis is Ghana
(5%) 1/20 Ans.: The country of Ghana is the worlds leading supplier of cannabis.
(5%) 1/20 Ans.: Ghana is the country that is the worlds ...

44. **What state has the most Indians? (California)**
(35%) 7/20 Ans.: California has the most Indians.
(25%) 5/20 Ans. California is the state that has the most Indians
(15%) 3/20 Ans.: California is the state with the most Indians.
(5%) 1/20 Ans.: California is the state with the highest number of Indians
(5%) 1/20 Ans.: State of Californian is that with the most Indians.
(5%) 1/20 Ans.: The state of California has the most Indians
(5%) 1/20 Ans.: California state has the most Indians.
(5%) 1/20 Ans.: the state that has most Indians is California

45. **What year was the Avery Dennison Company founded? (1990)**
(95%) 19/20 Ans.: the Avery Dennison company was founded in 1990
(5%) 1/20 Ans.: 1990.

46. **Material called linen is made from what plant? (Flax)**
(55%) 11/20 Ans.: Linen is made from Flax.
3/11 Ans.: Linen is made of Flax.

2/11 Ans.: linen is made from Flax plant.
1/11 Ans.: It is made from Flax.
(25%) 5/20 Ans.: The Material called linen is made from Flax.
2/5 Ans.: the Material called linen is made from Flax Plant.
(15%) 3/20 Ans.: linen is made from a plant called Flax
(5%) 1/20 Ans.: Flax.

47. **What type of bridge is the Golden Gate Bridge? (suspension)**
(80%) 16/20 Ans.: the Golden Gate Bridge is a suspension bridge
(5%) 1/20 Ans.: The type of Golden Gate Bridge is suspension.
(5%) 1/20 Ans.: The Golden Gate bridge is of suspension type.
(5%) 1/20 Ans.: It is a suspension bridge.
(5%) 1/20 Ans.: suspension.

48. **What two US biochemists won the Nobel Prize in medicine in 1992? (Edwin Krebs Edmond Fischer)**
(35%) 7/20 Ans.: Edwin Krebs and Edmond Fischer are the two US biochemists who won the Nobel Prize in medicine in 1992.
2/7 Ans.: Edwin Krebs and Edmond Fischer were the two US biochemists who won the Nobel Prize in medicine in 1992.
1/7 Ans.: Edwin Krebs and Edmond Fischer are two US biochemists who won the Nobel Prize in medicine in 1992.
(15%) 3/20 Ans.: The two US biochemists who won the Nobel Prize in medicine in 1992 are Edwin Krebs and Edmond Fischer.
1/3 Ans.: Two biochemists that won ... were Edwin Krebs and Edmond Fischer
(15%) 3/20 Ans.: Edwin Krebs and Edmond Fischer won the Nobel Prize in Medicine in 1992
(10%) 2/20 Ans.: Edwin Krebs and Edmond Fisher, two American biochemists, won the Nobel Prize in medicine in 1992
(5%) 1/20 Ans.: Edwin Krebs and Edmond Fischer were two biochemists who won the 1992 Nobel Prize in medicine.
(5%) 1/20 Ans.: In 1992, Edwin Krebs and Edmond Fischer were the two US biochemists that have won the Nobel Prize in medicine.
(5%) 1/20 Ans.: The Nobel Prize in medicine in 1992 was won by Edwin Krebs and Edmond Fischer
(5%) 1/20 Ans.: US biochemists Edwin Krebs and Edmond Fischer won...
(5%) 1/20 Ans.: Edwin Krebs and Edmond Fischer.

49. **What nuclear-powered Russian submarine sank in the Norwegian Sea on April 7, 1989? (Komsomlets)**
(55%) 11/20 Ans.: The Komsomlets was the nuclear-powered Russian submarine that sank in the Norwegian Sea on April 7, 1989.
(10%) 2/20 Ans.: Komsomlets sank in the Norwegian Sea on April 7, 1989.
(10%) 2/20 Ans.: The Russian nuclear-"powered" submarine "Komsomlets" sank in the Norwegian Sea on April 7, 1989.
(5%) 1/20 Ans.: It was Komsomlets.
(5%) 1/20 Ans.: Komsomlets sank on that day.
(5%) 1/20 Ans.: () is the nuclear-powered Russian submarine, which sank.
(5%) 1/20 Ans.: The Komsomlets was the name of the nuclear-powered Russian submarine that sank in the Norwegian Sea on April 7, 1989.
(5%) 1/20 Ans.: the nuclear-powered Russian submarine that sank in the Norwegian Sea on...was Komsomlets

50. What river flows between Fargo, North Dakota and Moorhead, Minnesota? (Red River)

(80%) 16/20 Ans.: The Red River flows between Fargo, North Dakota and Moorhead, Minnesota.
(10%) 2/20 Ans.: The river which flows between Fargo, North Dakota, and Moorhead, Minnesota is the Red River.
(10%) 2/20 Ans.: The Red River.

51. What strait separates North America from Asia? (Bering)

(80%) 16/20 Ans.: The Bering Strait separates North America from Asia
5/16 Ans.: The strait of Bering separates North America from Asia.
2/16 Ans.: Bering separates North America from Asia.
(10%) 2/20 Ans.: North America is separated from Asia by the straits of Bering
(5%) 1/20 Ans.: Bering.
(5%) 1/20 Ans.: Bering is the Strait that separates North...

52. What mineral helps prevent osteoporosis? (Calcium)

(50%) 10/20 Ans.: Calcium helps prevent osteoporosis.
(25%) 5/20 Ans.: Calcium prevents osteoporosis.
(20%) 4/20 Ans.: Calcium is the mineral that helps to prevent osteoporosis
(5%) 1/20 Ans.: Calcium

53. What does the Peugeot Company manufacture? (cars)

(100%) 20/20 Ans.: Peugeot Company manufactures cars.
(10%) 2/20 Ans.: The Peugeot Company manufactures cars
(5%) 1/20 Ans.: The Peugeot Company produces cars.
(15%) 3/20 Ans.: Peugeot manufactures cars
(5%) 1/20 Ans.: The Peugeot manufactures cars
(5%) 1/20 Ans.: They manufacture cars.

54. What does Phi Beta Kappa mean? (Academic honor society)

(95%) 19/20 Ans.: Phi Beta Kappa means Academic Honor Society
7/19 Ans.: Phi Beta Kappa is an Academic honor society
1/19 Ans.: Phi Beta Kappa refers to Academic Honor Society.
1/19 Ans.: It is an Academic honor society.
(5%) 1/20 Ans.: Academic honor society

55. What does El Nino mean in Spanish? (boy child)

(60%) 12/20 Ans.: El Nino means boy child in Spanish.
(10%) 2/20 Ans.:"El Niño" means 'boy child'
(5%) 1/20 Ans.: El Nino is Spanish for Boy child.
(5%) 1/20 Ans.: In Spanish El Nino means Boy child
(5%) 1/20 Ans.: Boy child is the English translation of the Spanish phrase El Nino

(5%) 1/20 Ans.: Boy child.
(5%) 1/20 Ans.: El nino in Spanish means Boy child
(5%) 1/20 Ans.: The Spanish "El Nino" means boy child in English.

56. What do manatees eat? (unwanted vegetation)
100% (100%) 20/20 Ans.: Manatees eat unwanted vegetation.

57. What did brontosauruses eat? (plant-eaters)
(50%) 10/20 Ans.: brontosauruses ate plant-eaters.
3/10 Ans.: (the) brontosaures ate plants
(30%) 6/20 Ans.: The brontosauruses were plant-eaters.
(10%) 2/20 Ans.: They were plant-eaters.
(5%) 1/20 Ans.: Brontosauruses are herbivores.
(5%) 1/20 Ans.: Plant-eaters.

58. What does the abbreviation OAS stand for? (Organization of American States)
(55%) 11/20 Ans.: OAS stands for the Organization of American States.
(35%) 7/20 Ans.: The abbreviation OAS stands for Organization of American States.
(5%) 1/20 Ans.: OAS is the abbreviation for Organization of American States
(5%) 1/20 Ans.: OAS is the short... for Organization of American States

59. What does Final Four refer to in the sports world? (NCAA basketball Final Four)
(30%) 6/20 Ans.: In the sports world Final four refers to NCAA basketball Final Four
(25%) 5/20 Ans.: Final Four refers to NCAA basketball Final Four in the sports world.
(15%) 3/20 Ans.: Final Four in the sports world refers to the NCAA basketball Final Four.
(15%) 3/20 Ans.: Final Four refer to NCAA basketball Final Four.
(5%) 1/20 Ans.: Final Four is the common name of a basketball event of the NCAA
(5%) 1/20 Ans.: It refers to NCAA basketball Final Four.
(5%) 1/20 Ans.: NCAA basketball Final Four

60. What does Knight Ridder publish? (29 Daily Newspapers)
(95%) 19/20 Ans.: Knight Ridder publishes 29 Daily Newspapers
1/19 Ans.: It publishes 29 Daily Newspapers.
(5%) 1/20 Ans.: 29 Daily Newspapers

61. What did Delilah do to Samson's hair? (cut it off)
(55%) 11/20 Ans.: Delilah cut off Samson's hair.
(25%) 5/20 Ans.: Delilah cuts it off
(10%) 2/20 Ans.: She cut it off.
(15%) 3/20 Ans.: Delilah cut Samson's hair off.

(5%) 1/20 Ans.: Delilah did cut Samson's hair off

62. What do bats eat? (Insects)
(100%) 20/20 Ans.: Bats eat insects
(5%) 1/20 Ans.: They eat insects.

63. What do you call a newborn kangaroo? (Joey)
(75%) 15/20 Ans.: A newborn kangaroo is called Joey.
(10%) 2/20 Ans.: You call a newborn kangaroo Joey.
(5%) 1/20 Ans.: Joey is What you call a newborn Kangaroo.
(5%) 1/20 Ans.: We call a newborn kangaroo Joey.
(5%) 1/20 Ans.: Joey.

64. What year did WWII begin? (1939)
(95%) 19/20 Ans.: WWII began in 1939
1/19 Ans.: It began in 1939.
(5%) 1/20 Ans.: 1939

65. What currency do they use in Brazil? (Cruzeiros)
(30%) 6/20 Ans.: The Cruzeiros is the currency used in Brazil*
(20%) 4/20 Ans.: The currency used in Brazil is the Cruzeiros.
(10%) 2/20 Ans.: they use Cruzeiros in Brazil.
(10%) 2/20 Ans.: Cruzeiros.
(5%) 1/20 Ans.: Cruzeiros are used in Brazil
(5%) 1/20 Ans.: In Brazil they use Cruzeiros
(5%) 1/20 Ans.: In Brazil the currency is ()
1/18Ans.: Cruzeiros is the currency they use in Brazil
(5%) 1/20 Ans.: Cruzeiros is the currency of Brazil
(5%) 1/20 Ans.:Cruzeiro is the national currency of Brazil

66. What debts did Qintex group leave? (ADollars 1.5bn.)
(65%) 13/20 Ans.: Qintex group left (ADollars 1.5bn) in debts. ($1.5bn or 1.5bn Dollars)
1/13 Ans.: They left (1.5 billion dollars) in debt.
1/13 Qintex group left (ADollars 1.5bn) debts.
3/13 Qintex group left (ADollars 1.5bn) of debts.
6/20 Ans.: Qintex group left debts for ($1.5bn). (ADollars 1.5bn)
3/6 Ans.: The Qintex group left a debt of (1.5bn). (1.5b US. Dollars or ( ))
1/6 Ans.Qintex group left debts of A Dollars 1.5bn.
(5%) 1/20 Ans.: Qintex left ()

67. **What time of day did Emperor Hirohito die? (6:33 a.m./ 1.33 p.m. Friday PST)**
(90%) 18/20 Ans.: (the) Emperor Hirohito died at 6:33 a.m./ 1.33 p.m. Friday PST.
(5%) 1/20 Ans.: He died at 6:33am.
(5%) 1/20 Ans.: 6:33 a.m./ 1.33 p.m. Friday PST.

68. **What ocean did the Titanic sink in? (Atlantic)**
(95%) 19/20 Ans.: The Titanic sank in the Atlantic ocean.
1/19 Ans.: It sank in the Atlantic Ocean.
(5%) 1/20 Ans.: Atlantic

69. **At what speed does the Earth revolve around the sun? (30 km/sec.)**
(70%) 14/20 Ans.: The Earth revolves around the sun at 30 km/sec.
(15%) 3/20 Ans.: The earth revolves around the sun at a speed of 30 km/sec.
(10%) 2/20 Ans.: The earth revolves 30 km/sec around the sun.
(5%) 1/20 Ans.: 30 km/sec

70. **What task does the Bouvier breed of dog perform? (herding cattle)**
(30%) 6/20 Ans.: The Bouvier breed of dog performs herding cattle.
(25%) 5/20 Ans.: The Bouvier breed of dog herds cattle.
(15%) 3/20 Ans.: the Bouvier breed of dog can perform the task of herding cattle
(5%) 1/20 Ans.: The Bouvier breed of dog performs cattle herding tasks.
(5%) 1/20 Ans.: the Bouvier breed of dog is herding cattle
(5%) 1/20 Ans.: The Bouvier is a herding cattle breed of dog.
(5%) 1/20 Ans.: Bouvier breeds do herding cattle
(5%) 1/20 Ans.: Bovier are bred for herding cattle.
(5%) 1/20 Ans.: herding cattle

71. **What sport do the Cleveland Cavaliers play? (basketball)**
(85%) 17/20 Ans.: The Cleveland Cavaliers play basketball.
1/17 Ans.: They play basketball.
(5%) 1/20 Ans.: Basketball
(5%) 1/20 Ans.: The Cleveland Cavaliers are a basketball team
(5%) 1/20 Ans.: The Cleveland Cavaliers play the sport of basketball.

72. **What state did the Battle of Bighorn take place in? (Montana)**
(95%) 19/20 Ans.: the Battle of Bighorn took place in Montana
3/19 Ans.: The Battle of Bighorn took place in the state of Montana.
1/19 Ans.: The Battle of Bighorn occurred in Montana.
1/19 Ans.: It took place in Montana.
(5%) 1/20 Ans.: Montana.

73. **What day and month did John Lennon die? (Dec. 8)**
(85%) 17/20 Ans.: John Lennon died on Dec. 8.
(5%) 1/20 Ans.: He died on Dec. 8.
(5%) 1/20 Ans.: John Lennon died on the 8th day of Dec.
(5%) 1/20 Ans.: Dec.8.

74. **What country did Ponce de Leon come from? (Spain)**
(90%) 18/20 Ans.: Ponce de Leon came from Spain
(5%) 1/20 Ans.: He is from Spain.
(5%) 1/20 Ans.: Spain.

Analysis for *When*, *Why*, and *Where* type of questions

75. **When was London's Docklands Light Railway constructed? (1980s)**
(95%) 19/20 Ans.: London's Docklands Light Railways was constructed in the 1980s.
2/19 Ans.: London's Docklands Light Railways was constructed during the 1980s.
2/19 Ans.: It was constructed in the 1980s
(5%) 1/20 Ans.: It was in 1980s.

76. **When was Dubai's first concrete house built? (1956)**
(75%) 15/20 Ans.: Dubai's first concrete house was built in 1956
1/15 Ans.: It was built in 1956
(10%) 2/20 Ans.: The first concrete house in Dubai's was built in 1956.
(5%) 1/20 Ans.: The first concrete house built in Dubai was in 1956
(5%) 1/20 Ans.: In 1956.
(5%) 1/20 Ans.: In 1956 Dubai's first concrete house was built.

77. **When was the slinky invented? (1943)**
(95%) 19/20 Ans.: the slinky was invented in 1943
1/19 Ans.: It was invented in 1943.
(5%) 1/20 Ans.: 1943.

78. **When was Algeria colonized? (19th Century)**
(95%) 19/20 Ans.: Algeria was colonized in 19th Century
(5%) 1/20 Ans.: Algeria was colonized during the 19th century
(5%) 1/20 Ans.: Algeria colonized in 19th Century.
(5%) 1/20 Ans.: It was colonized in the 19th Century.

(5%) 1/20 Ans.: 19th Century

79. **When was the San Francisco fire? (April 18, 1906)**
(85%) 17/20 Ans.: the San Francisco fire was in April 20, 1906
(10%) 2/20 Ans.: April 20, 1906.
(5%) 1/20 Ans.: The San Francisco fire occurred April 20, 1906.

80. **When did Hawaii become a state? (1959)**
(100%) 20/20 Ans.: Hawaii became a state in 1959
1/20 Ans.: It became a state in 1959.

81. **When did John F. Kennedy get elected as President? (1960)**
(85%) 17/20 Ans.: John F. Kennedy got elected as President in 1960.
3/17 Ans.: John F. Kennedy was elected as President in 1960
(5%) 1/20 Ans.: In 1960.
(5%) 1/20 Ans.: He was elected in 1960.
(5%) 1/20 Ans.: JFK became president in 1960.

82. **When did the Hindenberg crash? (1937)**
(90%) 18/20 Ans.: the Hindenberg crashed in 1937
1/18 Ans.: It crashed in 1937.
(5%)1/20 Ans.: The Hindenberg crash was in 1937
(5%) 1/20 Ans.: 1937

83. **Why does the moon turn orange? (eclipse)**
(30%) 6/20 Ans.: The moon turns orange because of an eclipse.
(30%) 6/20 Ans.: The moon turns orange during an eclipse
(10%) 2/20 Ans.: Because of the eclipse.
(10%) 2/20 Ans.: the moon turns orange when there is an eclipse
(5%) 1/20 Ans.: The moon turns orange due to an eclipse.
(5%) 1/20 Ans.: An Eclipse is the reason why the moon can turn orange
(5%) 1/20 Ans.: An eclipse is the cause of the moon turning orange.
(5%) 1/20 Ans.: Because it is eclipsed.

84. **Why did David Koresh ask the FBI for a word processor? (To record his revelations)**
(55%) 11/20 Ans.: David Koresh asked the FBI for a word processor to record his revelations.
(10%) 2/20 Ans.: Because David Koresh wanted to record his revelations
(10%) 2/20 Ans.: David Koresh asked for a work processor to record his revelations
1/2 Ans.: He asked for a word processor to record his revelations
(10%) 2/20 Ans.: To record his revelations, David Koresh requested a word processor from the FBI.

1/2) Ans.: to record his revelations David Koresh asked the FBI for a word processor
(5%) 1/20 Ans.: David Koresh asked the FBI for a word processor for recording his revelations
(5%) 1/20 Ans.: Peter Koresh asked the FBI for a computer in order to record his revelations.
(5%) 1/20 Ans.: David Koresh asked the FBI for a word processor so that he could record his revelations

85. Where is the Orinoco River? (Venezuela)
(95%) 19/20 Ans.: the Orinoco River is in Venezuela
1/19 Ans.: It is in Venezuela.
(5%) 1/20 Ans.: Venezuela.

Where is Perth? (Australia)
(95%) 19/20 Ans.: Perth is in Australia.
(5%) 1/20 Ans.: Australia

86. Where is the highest point in Japan? (Mt. Fuji)
(75%) 15/20 Ans.: Mt. Fuji is the highest point in Japan.
(25%) 5/20 Ans.: the highest point in Japan is Mt. Fuji.

87. Where is Microsoft's corporate headquarters located? (Redmond, Wash)
(90%) 18/20 Ans.: Microsoft's corporate headquarters is located in Redmond, wash.
7/18 Ans.: Microsoft's corporate headquarters are located in Redmond, wash
1/18 Ans.: Microsoft's headquarters are located in Redmond, wash
(5%) 1/20 Ans.: In Redmond, wash.
(5%) 1/20 Ans.: Redmond, Wash is the location of Microsoft's corporate headquarters.

88. Where is Belize located? (Central America)
(85%) 17/20 Ans.: Belize is located in Central America
(15%) 3/20 Ans.: Belize is in Central America.

89. Where did bocci originate? (Italian)
(70%) 14/20 Ans.: bocci originated in Italy.
(25%) 5/20 Ans.: Bocci is Italian.
(5%) 1/20 Ans.: The origin of Bocci was Italian

Analysis for *How*, and *Which* type of questions

90. How much did Mercury spend on advertising in 1993? (Pounds 12m)
(55%) 11/20 Ans.: Mercury spent Pounds 12m on advertising in 1993
(15%) 3/20 Ans.: Mercury spent Pounds 12m.
(20%) 4/20 Ans.: In 1993 Mercury spent 12m pounds on advertising.
(5%) 1/20 Ans.: it spent 12m pounds
(5%) 1/20 Ans.: Mercury advertising expenditure for 1993 was Pounds 12m.

91. How much did Manchester United spend on players in 1993? (Pounds 4m.)
(80%) 16/20 Ans.:Manchester United spent 4 million Pounds on players in 1993.
(15%) 3/20 Ans.: In 1993, Manchester United spend 4m Pounds on players.
(5%) 1/20 Ans.: MU spent 4m Pounds on players that year.

92. How much folic acid should an expectant mother get daily? (400 micrograms)
(60%) 12/20 Ans.: an expectant mother should get 400 micrograms of folic acid daily.
2/12 Ans.: An expectant mother should get 400 micrograms of folic acid per day
1/12 Ans.: an expectant mother should get 400 micrograms of folic acid every day
(10%) 2/20 Ans.: an expectant mother should get daily 400 micrograms of folic acid
(5%) 1/20 Ans.: an expectant mother should get 400 micrograms daily of folic acid
(5%) 1/20 Ans.: An expectant mother should get daily 400 micrograms.
(5%) 1/20 Ans.: An expectant mother should get 400 micrograms daily.
(5%) 1/20 Ans.: An Expectant mother should get a daily dose of folic acid of 400 micrograms
(5%) 1/20 Ans.: 400 micrograms folic acid should an expectant mother get daily
(5%) 1/20 Ans.: 400 micrograms

93. How much fiber should you have per day? (35 grams)
(90%) 18/20 Ans.: You should have 35 grams of fiber per day.
(10%) 2/20 Ans.: 35 grams.

94. How much could you rent a Volkswagen bug for in 1966? ($1 a day)
(35%) 7/20 Ans.: In 1996 you could rent a Volkswagwen bug for $1 a day
(30%) 6/20 Ans.: You could rent a Volkswagen bug for $1 a day in 1966
(10%) 2/20 Ans.: You could rent it for $1 a day.
(10%) 2/20 Ans.: you could rent a Volkswagen bug for in 1966 for $1 a day.
(5%) 1/20 Ans.: I could do it for $1 a day.
(5%) 1/20 Ans.: To rent a Volkswagen bug for one day in 1966, you should pay $1.
(5%) 1/20 Ans.: A Volkswagen bug could be rented for $1 a day in 1966.

95. How far is Yaroslavl from Moscow? (150 miles)

(55%) 11/20 Ans.: Yaroslav is 150 miles from Moscow
(20%) 4/20 Ans.: Yaroslavl is 150 miles "far" from Moscow.
1/4 Ans.: Yaroslavl is 150 miles "away" from Moscow.
1/4 Ans.: Yaroslavl is 150 miles "far away" from Moscow.
(5%) 1/20 Ans.: Yaroslavl is situated 150 miles distant from Moscow.
(5%) 1/20 Ans.: Yaroslavi is located at () from Moscow
(5%) 1/20 Ans.: 150 miles.
(5%) 1/20 Ans.: It is about 150 miles.
(5%) 1/20 Ans.: The distance is 150 miles.

96. How far away is the moon? (quarter of a million miles)

(65%) 13/20 Ans.: the moon is quarter of a million miles away
3/13 Ans.: the moon is a quarter of a million miles far.
2/13 Ans.: the moon is a quarter of a million miles far away.
(20%) 4/20 Ans.: The moon is a quarter of a million miles away from earth.
(5%) 1/20 Ans.: The distance from the earth to the moon is quarter of a million miles
(5%) 1/20 Ans.: It is about quarter of a million miles.
(5%) 1/20 Ans.: The moon is a () away from you right now.

97. How long did the Charles Manson murder trial last? (9 1/2-month)

(95%) 19/20 Ans.: the Charles Manson murder trial lasted 9 1/2-month
2/19 Ans.: the Charles Manson trial lasted 9 1/2-month
1/19 Ans.: His trial lasted 9 ½ months.
(5%) 1/20 Ans.: The murder trial of Charles Manson lasted 9 ½ months.

98. For how long is an elephant pregnant? (22-month)

(70%) 14/20 Ans.: An elephant is pregnant for 22-months
1/14 Ans.: Female elephants are pregnant for 22 months.
(10%) 2/20 Ans.: An elephant's pregnancy normally lasts for 22-month
1/2 Ans.: Elephant pregnancy lasts for 22 months.
(5%) 1/20 Ans.: For 22-month.
(5%) 1/20 Ans.: That is 22-month.
(5%) 1/20 Ans.: The elephant has a 22-month pregnancy.
(5%) 1/20 Ans.: the pregnancy of an elephant last 22 months

99. How many dogs pull a sled in the Iditarod? (8-12)

(75%) 15/20 Ans.: 8-12 dogs pull a sled in the Iditarod
(5%) 1/20 Ans.: a Sled in the Iditarod pulls 8-12 dogs
(5%) 1/20 Ans.: The sled is pulled by 8-12 dogs in the Idatirod.

(5%) 1/20 Ans.: In the Iditarod, a sled is pulled by 8-12 dogs
(5%) 1/20 Ans.: It takes between 8 and 12 dogs to pull a sled in the Iditarod.
(5%) 1/20 Ans.: 8-12 dogs.

100. How many people die from snakebite poisoning in the U.S. per year? (13)
(65%) 13/20 Ans.: 13 people die from snakebite poisoning in the U.S. per year.
1/13 Ans.: Thirteen people die from snakebite poisoning in the U.S. each year.
(15%) 3/20 Ans.: In the US 13 die ...per year
(5%) 1/20 Ans.: 13 people die from snakebite per year.
(5%) 1/20 Ans.: 13 people die from snakebite poisoning per year in the U.S.
(5%) 1/20 Ans.: 13 people die in the US per year from snakebite poisoning
(5%) 1/20 Ans.: 13

101. How many astronauts have been on the moon? (12 men)
(75%) 15/20 Ans.: 12 men astronauts have been on the moon.
5/15 Ans.: 12 men have been on the moon
3/15 Ans.: 12 astronauts have been on the moon.
1/15 Ans.: Twelve male astronauts have been on the moon.
(10%) 2/20 Ans.: There have been 12 astronauts on the moon (had)
(5%) 1/20 Ans.: Up to now, 12 astronauts have been on the moon
(5%) 1/20 Ans.: A total of 12 men has been on the moon.
(5%) 1/20 Ans. 12 men is the amount of Astronauts that have been on the moon.

102. How many films did Ingmar Bergman make? (more than 50 films)
(95%) 19/20 Ans.: Ingmar Bergman made more than 50 films.
(5%) 1/20 Ans.: More that 50 films.

103. How many calories are there in a Big Mac? (562)
(65%) 13/20 Ans.: There are 562 calories in a Big Mac.
(35%) 7/20 Ans.: A Big Mac has 562 calories
2/7 Ans.: The Big Mac has 562 calories

104. How many types of lemurs are there? (30 types)
(100%) 20/20 Ans.: There are 30 types of lemurs.

105. How many Great Lakes are there? (5)
(100%) 18/18 Ans.: There are 5 Great Lakes.
(5%) 1/20 Ans.: There are 5
(5%) 1/20 Ans.: 5.

106. How big is Australia? (nearly 3 million square miles)

(75%) 15/20 Ans.: Australia is nearly 3 million square miles
(10%) 2/20 Ans.: Australia is nearly 3 million square miles in size.
(5%) 1/20 Ans.: Australia measures nearly 3 million square miles
(5%) 1/20 Ans.: Australia nearly has 3 mn square miles
(5%) 1/20 Ans.: The size of Australia is nearly 3 million square miles

107. How large is the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast? (19 million acre)

(45%) 9/20 Ans.: the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast "is" 19 million acre " ".
1/9 Ans.: the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast "is" 19 million acre large.
2/9 Ans.: the Arctic refuge to preserve unique wildlife and wilderness value on Alaska's north coast measure (has) 19 million acres

(15%) 3/20 Ans.: The "Alaska's" Artic refuge is 19 million acres.

(15%) 3/20 Ans.: The Artic refuge has 19 million acres to ...

(5%) 1/20 Ans.: A 19 million acre area on Alaska's north coast is the Arctic refuge to preserve unique wildlife and wilderness value.

(5%) 1/20 Ans.: the Arctic refuge that preserves unique wildlife and wilderness value on Alaska's north coast "is" 19 million acre in size

(5%) 1/20 Ans.: 19 million acres is the size of the artic refuge...
(5%) 1/20 Ans.: To preserve unique wildlife and wilderness value on Alaska's north coast, the Artic refuge has 19 million acres
(5%) 1/20 Ans.: 19 million acre.

108. How hot is the core of the earth? (5,000 degrees Celsius)

(60%) 12/20 Ans.: the core of the earth is 5,000 degrees Celsius
(15%) 3/20 Ans.: The core of the earth is 5,000 degrees Celsius hot
(10%) 2/20 Ans.: The temperature of the core of the earth is 5000 degrees centigrade.
(5%) 1/20 Ans.: The temperature of the earth's core is 5,000 degrees Celsius.
(5%) 1/20 Ans.: The Earth's core is 5,000 degrees Celsius
(5%) 1/20 Ans.: The core is 5,000 degrees Celsius.

109. How tall is the Sears Building? (1454 feet.)

(90%) 18/20 Ans.: The Sears building is 1454 feet tall.
(5%) 1/20 Ans.: The Sears Building's height is 1454 feet.
(5%) 1/20 Ans.: It is 1454 feet.

110. How old was Elvis Presley when he died? (42)

(65%) 13/20 Ans.: Elvis Presley was 42 when he died

4/20 Ans.: Elvis Presley was 42 years old when he died.
(10%) 2/20 Ans.: Elvis died at the age of 42.
1/2 Ans.: He died at age 42.
(5%) 1/20 Ans.: 42.

111. How did Janice Joplin die? (Overdose of heroin)
(90%) 18/20 Ans.: Janice Joplin died of an Overdose of heroin
4/18 Ans.: Janice Joplin died from an overdose of Heroin.
1/18 Ans.: She died of an overdose of heroin.
1/18 Ans.: Janice Joplin died because of an overdose of heroin
(5%) 1/20 Ans.: Janice Joplin died of heroin overdose.
(5%) 1/20 Ans.: Janice Joplin overdosed on heroin.

112. Which city in China has the largest number of foreign financial companies? (Shanghai)
(45%) 9/20 Ans.: Shanghai has the largest number of foreign financial companies in China
(20%) 4/20 Ans.: In China, Shanghai has the largest number of foreign financial companies.
(10%) 2/20 Ans.: Shanghai is the Chinese city with the largest number of foreign financial companies.
(10%) 2/20 Ans.: Shanghai is the city with the largest number of foreign financial companies in China
(5%) 1/20 Ans.: Shanghai.
(5%) 1/20 Ans.: The city in China with the largest number of foreign financial companies is Shanghai.
(5%) 1/20 Ans.: Shanghai is the city in Chine that has the largest...

113. George Bush purchased a small interest in which baseball team? (Texas Rangers)
(80%) 16/20 Ans.: George Bush purchased a small interest in the Texas Rangers baseball team.
(10%) 2/20 Ans.: The Texas Rangers is the baseball team where George Bush purchased a small interest
(10%) 2/20 Ans.: Texas Rangers.

114. The U.S. Department of Treasury first issued paper currency for the U.S. during which war? (1862)
(75%) 15/20 Ans.: The U.S Department of Treasury first issued paper currency for the U.S. during the war of 1862.
2/15 Ans.: The U.S. Department of Treasury first issued paper currency for the U.S. during the 1862 war.
4/15 Ans.: The U.S. Department of Treasury first issued paper currency for the U.S. in 1862.
(5%) 1/20 Ans.: The first paper currency issued for the U.S was during 1862 war
(5%) 1/20 Ans.: During the 1862 war.
(5%) 1/20 Ans.: 1862.
(5%) 1/20 Ans.: Paper currency was issued for the first time in the US during the 1862 war
(5%) 1/20 Ans.: During the war of 1862 the U.S. Department of Treasury first issued paper currency for the U.S.

Analysis of *WHO* and *NAME* type of questions

116. Who was President Cleveland's wife? (Frances Folsom)
(50%) 10/20 Ans.: Frances Folsom was President Cleveland's wife.
(45%) 9/20 Ans.: President Cleveland's wife was Frances Folsom
1/9 Ans.: His wife was Frances Folsom.
(5%) 1/20 Ans.: It is Frances Folsom.

117. Who was the first Taiwanese President? (Lee Teng-Hui)
(55%) 11/20 Ans.: The first Taiwanese President was Lee Teng-Hui.
(30%) 6/20 Ans.: Lee Teng-Hui was the first Taiwanese President.
(5%) 1/20 Ans.: Taiwan's first President was Lee Teng-Hui.
(5%) 1/20 Ans.: The first President of Taiwan was Lee Teng-Hui.
(5%) 1/20 Ans.: Lee Teng-Hui

118. Who was the leader of the Branch Davidian Cult confronted by the FBI in Waco, Texas in 1993? (Mr. David Koresh)
(70%) 14/20 Ans.: Mr. David Koresh was the leader of the Branch Davidian Cult confronted by the FBI in Waco, Texas in 1993
(15%) 3/20 Ans.: The leader of the Branch Davidian Cult confronted by the FBI in Waco, Texas in 1993 was Mr. David Koresh
(10%) 2/20 Ans.: David Koresh was the leader of the Branch Davidian Cult.
(5%) 1/20 Ans.: In 1993, the leader of the ... in Waco, Texas was Mr. David Koresh

119. Who is the president of Stanford University? (Donald Kennedy)
(70%) 14/20 Ans.: Donald Kennedy is the president of Stanford University.
(25%) 5/20 Ans.: The president of Stanford University is Donald Kennedy.
(5%) 1/20 Ans.:Donald Kennedy.

120. Who was the first doctor to successfully transplant a liver? (Dr. Thomas Starzl)
(75%) 15/18Ans.: Dr. Thomas Starzl was the first doctor to successfully transplant a liver.
(25%) 5/20 Ans.: The first doctor to successfully transplant a liver was Dr. Thomas Starzl

121. Who was President of Costa Rica in 1994? (Rafael Angel Calderon)
(50%) 10/20 Ans.: Rafael Angel Calderon was President of Costa Rica in 1994.
5/10 Ans.: Rafael Angel Calderon was the President of Costa Rica in 1994.
(20%) 4/20 Ans.: The President of Costa Rica in 1994 was Rafael Angel Calderon
(15%) 3/20 Ans.: In 1994, the president the Costa Rica was Rafael Angel Calderon
(5%) 1/20 Ans.: Rafael Angel Calderon
(5%) 1/20 Ans.: Costa Rica's president was Rafael Angel Calderon in 1994.
(5%) 1/20 Ans.: Rafael Angel Calderon was president.

122. Who is the voice of Miss Piggy? (Frank Oz)

(70%) 14/20 Ans.: Frank Oz is the voice of Miss Piggy
(20%) 4/20 Ans.: The voice of Miss Piggy is Frank Oz
(5%) 1/20 Ans.: Frank Oz does the voice for Miss Piggy.
(5%) 1/20 Ans.: Frank Oz.

123. Who is the founder of Scientology? (L. Ron Hubbard)

(75%) 15/20 Ans.: Ans.: L. Ron Hubbard is the founder of Scientology
(15%) 3/20 Ans.: The founder of Scientology is L. Ron Hubbard.
(10%) 2/20 Ans.: L. Ron Hubbard.

124. Who is Coronado? (Francisco Vásquez de Coronado) (Explorer)

(40%) 17/20 Ans.: Francisco Vasquez de Coronado is an explorer.
5/17 Ans.: Coronado is an explorer
1/17 Ans.: He was an explorer.
2/17 Ans.: Francisco Vasquez de Coronado is Coronado.
(5%) 1/20 Ans.:"Coronado" is the explorer Francisco Vasquez de Coronado
(5%) 1/20 Ans.: Francisco Vasquez de Coronado, the Explorer, is Coronado.
(5%) 1/20 Ans.: Francisco Vasquez de Coronado is known as Coronado
(5%) 1/20 Ans.: ?

125. Who is Barbara Jordan? (Former U.S. Rep)

(90%) 18/20 Ans.: Barbara Jordan is a former U.S. Rep.
(5%) 1/20 Ans.: She was a former U.S. Rep.
(5%) 1/20 Ans.: Former U.S. Rep is Barbara Jordan.

126. Who is the richest person in the world? (Sultan Hassanai Bolkiah)

(55%) 11/20 Ans.: Sultan Hassanai Bolkiah is the richest person in the world
(35%) 7/20 Ans.: The richest person in the world is Sultan Hassanai Bolkiah
(5%) 1/20 Ans.: Sultan Hassanai Bolkiah is attributed as the richest person in the world
(5%) 1/20 Ans.: Sultan Hassanai Bolkiah.

127. Who was the first coach of the Cleveland Browns? (Paul Brown)

(70%) 14/20 Ans.: Paul Brown was the first coach of the Cleveland Browns.
(20%) 4/20 Ans.: The first coach of the Cleveland Browns was Paul Brown
(5%) 1/20 Ans.: Paul Brown was the first coach.
(5%) 1/20 Ans.: Paul brown.

128. Who was the 33rd president of the United States? (Harry Truman)

(55%) 11/20 Ans.: The 33rd president of the United States was Harry Truman.

(40%) 8/20 Ans.: Harry Truman was the 33rd president of the United States.
(5%) 1/20 Ans.: Harry Truman.

129. Who is the emperor of Japan? (Akhito)
(65%) 13/20 Ans.: Akhito is the emperor of Japan.
(25%) 5/20 Ans.: The emperor of Japan is Akhito.
(5%) 1/20 Ans.: the name of the emperor of Japan is Akhito
(5%) 1/20 Ans.: Akhito

130. Who was the founding member of the Pink Floyd band? (Roger Waters)
(65%) 13/20 Ans.: Roger Waters was the founding member of the Pink Floyd band
1/20 Ans.: Roger Waters was one of the founding members of the Pink Floyd band
(15%) 3/20 Ans.: the founding member of the Pink Floyd band was Roger Waters
(10%) 2/20 Ans.: Roger Waters founded the Pink Floyd band.
(5%) 1/20 Ans.: Roger Waters.

131. Who was Abraham Lincoln? (president)
(90%) 18/20 Ans.: Abraham Lincoln was a president.
(5%) 1/20 Ans.: Abraham Lincoln was the first president of the US
(5%) 1/20 Ans.: He was one of the presidents of the U.S

132. Who was the first female United States Representative? (Jeanette Rankin)
(65%) 13/20 Ans.: Jeanette Rankin was the first female United States Representative
(25%) 5/20 Ans.: The first female United States Representative was Jeanette Rankin.
(10%) 2/20 Ans.: Jeanette Rankin.

133. Who invented the road traffic cone? (David Morgan)
(55%) 11/20 Ans.: David Morgan invented the road traffic cone.
(45%) 9/20 Ans.: The road traffic cone was invented by David Morgan

134. Who came up with the name, El Nino? (Peruvian fishermen)
(75%) 15/20 Ans.: Peruvian fishermen came up with the name, El Nino
(10%) 2/20 Ans.: Peruvian fishermen.
(5%) 1/20 Ans.: A Peruvian fishermen came up with the Spanish phrase called El Nino
(5%) 1/20 Ans.: It was Peruvian fishermen who came up with the name, El Nino
(5%) 1/20 Ans.: El Niño was so called by Peruvian fishermen

135. Who released the Internet worm in the late 1980s? (Robert Morris)
(65%) 13/20 Ans.: Robert Morris released the Internet worm in the late 1980s
(10%) 2/20 Ans.: The Internet worm was released in the late 1980s by Robert Morris

(5%) 1/20 Ans.: Robert Morris released in 1980s the Internet worm
(15%) 3/20 Ans.: In (the) late 1980s, Robert Morris released the Internet worm.
(5%) 1/20 Ans.: Robert Morris.

136. Who first circumnavigated the globe? (Ferdinand Magellan)
(80%) 16/20 Ans.: Ferdinand Magellan "was the" first "to" circumnavigate the globe.
5/16 Ans.: Ferdinand Magellan first circumnavigated the globe.
(10%) 2/20 Ans.: The first explorer to circumnavigate the globe was Ferdinand Magellan
(5%) 1/20 Ans.: Ferdinand Magellan.
(5%) 1/20 Ans.: () is the one who first circumnavigated the globe.

137. Who wrote the song, "Stardust"?(Hoagy Carmichael)
(75%) 16/20 Ans.: Hoagy Carmichael wrote the song, "Stardust"
1/16 Ans.: Hoagy Carmichael wrote it.
(20%) 4/20 Ans.: "The song", "Stardust" was written by Hoagy Carmichael.

138. Who invented the paper clip? (Norwegian Johan Varler)
(75%) 15/20 Ans.: The Norwegian johan Varler invented the paper clip
(20%) 4/20 Ans.:The paper clip was invented by a (the) Norwegian, Johan Varler
(5%) 1/20 Ans.: Norwegian Johan Varler.

139. Who made the first airplane that could fly? (Orville and Wilbur Wright)
(80%) 16/20 Ans.: Orville and Wilbur Wright made the first airplane that could fly
(15%) 3/20 Ans.: the first airplane that could fly was built by Orville and Wilbur Wright
(5%) 1/20 Ans.: The first airplane was made by Orville and Wilbur Wright.

140. Who found Hawaii? (Captain Cook)
(65%) 13/20 Ans.: Captain Cook founded Hawaii.
(35%) 7/20 Ans.: Hawaii was found by Captain Cook

141. Who won Ms. American in 1989? (Gretchen Elizabeth)
(80%) 16/20 Ans.: Gretchen Elizabeth won Ms. American in 1989
(10%) 2/20 Ans.: Gretchen Elizabeth.
(5%) 1/20 Ans.: Ms. American in 1989 was won by Gretchen Elizabeth
(5%) 1/20 Ans.: In 1989, Gretchen Elizabeth won Ms. American in 1989

142. Who discovered radium? (Marie and Pierre Curie)
(70%) 14/20 Ans.: Marie and Pierre Curie discovered radium
(15%) 3/20 Ans.: Radium was discovered by Marie and Pierre Curie

(10%) 2/20 Ans.: Marie and Pierre Curie.
(5%) 1/20 Ans.: () are the ones who discovered radium.

143. Name the designer of the shoe that spawned millions of plastic imitations, known as "jellies". (Andrea Pfister)
(40%) 8/20 Ans.: Andrea Pfister is the designer of the shoe that spawned millions of plastic imitations, known as "jellies"
(15%) 3/20 Ans.: Andrea Pfister is "the name of" the designer of the shoe that spawned millions of plastic imitations, known as "jellies" (was)
(5%) 1/20 Ans.: Andrea Pfister "is who" designed the shoe that spawned millions of plastic imitations
(5%) 1/20 Ans.: Andrea Pfister designed the shoe that spawned millions of plastic imitations
(5%) 1/20 Ans.: The designer "of the jellies" is Andrea Pfister
(5%) 1/20 Ans.: () is the designer of the shoe known as "jellies"
(5%) 1/20 Ans.: The designer is Andrea Pfister
(5%) 1/20 Ans.: The designer of the shoe that spawned millions of plastic imitations know as "jellies" is Andrea Pfister.
(5%) 1/20 Ans.: It is Andrea Pfister.
(5%) 1/20 Ans.:Andrea Pfister.
(5%) 1/20 Ans.: His name is Andrea Pfister

144. Name a film that has won the Golden Bear in the Berlin Film Festival? (In The Name Of The Father)
(40%) 8/20 Ans.: In The Name Of The Father won the Golden Bear in the Berlin Film Festival.
(25%) 5/20 Ans.: In The Name Of The Father is a film that has won the Golden Bear in the Berlin Film Festival.
1/5 Ans.: In The Name Of The Father is a film that who won the Golden Bear in the Berlin Film Festival.
(10%) 2/20 Ans.: In the Name Of The Father.
(5%) 1/20 Ans.: The film "In The Name Of The Father" won the Golden Bear in the Berlin Film Festival.
(5%) 1/20 Ans.: One of the films that won the Golden Bear in the Berlin Film Festival is The Name Of The Father
(5%) 1/20 Ans.: In The Name Of The Father" was the name of a film that won the Golden Bear in the Berlin Film Festival.
(5%) 1/20 Ans.: "In The Name Of The Father" has won the Golden Bear in the Berlin Film Festival.
(5%) 1/20 Ans.: A film that won the Golden Bear in the Berlin Film Festival is The Name Of The Father

145. Name a country that is developing a magnetic levitation railway system? (Japan/ Germany)
(45%) 9/20 Ans.: Japan is developing a magnetic levitation railway system. (Germany)
(40%) 8/20 Ans.: Japan is one of the countries "that is" developing a magnetic levitation railway system
(5%) 1/20 Ans.: Japan and Germany are two countries developing a magnetic...
(5%) 1/20 Ans.: One of the countries that is developing a magnetic levitation railway system is Japan.
(5%) 1/20 Ans.: The countries developing a magnetic levitation railway system are Japan

146. Name the first private citizen to fly in space. (Christa McAuliffe)
(50%) 10/20 Ans.: Christa McAuliffe was the first private citizen to fly in space
(45%) 9/20 Ans.: The first private citizen to fly in space was Christa McAuliffe.

(5%) 1/20 Ans.: Christa McAuliffe

147. Name one of the major gods of Hinduism? (Brahma)

(90%) 18/20 Ans.: Brahma is one of the major gods of Hinduism.

(10%) 2/20 Ans.: One of the major gods of Hinduism is Brahma.

148. Name a ballet company Mikhail Baryshnikov has danced for? (George Balanchine's New York City Ballet)

(65%) 13/20 Ans.: Mikhail Baryshnikov has danced for the George Balanchine's New York City Ballet

(20%) 4/20 Ans.: George Balanchine's New York City Ballet is a ballet company that Mikhail Baryshnikov danced for.

(10%) 2/20 Ans.: George Balanchine's New York City Ballet.

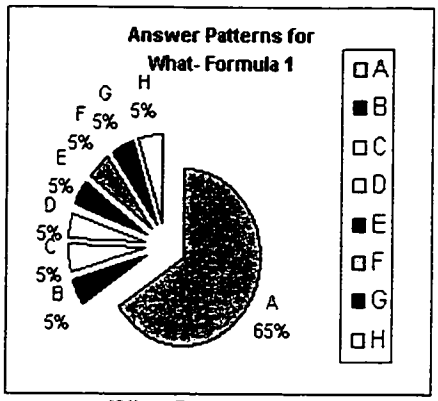
(5%) 1/20 Ans.: one ballet company Mikhail Baryshnikov has danced for is George Balanchine's New York City Ballet

Appendix C: Question and answer patterns

What + To Be +... - Type of Question

Pattern 1: What (WP) + Verb to-be (VBD) + NP1 + IN + NP2 + PP (IN+N)
Example Pattern 1: What/ was/ the monetary value/ of/ the Nobel Peace Prize/ in /1989? ($469,000)

Answer Patterns:
A. NP1 + IN + NP2 + PP + VBD + (AA)
B. NP2 + NP1 + PP + VBD + (AA)
C. PP + NP2 + VBD + (AA)
D. (AA) + VBD + NP1 + IN + NP2 + PP
E. NP1 + IN + N + NP2 + WBD + (AA)
F. NP1 + VBD + (AA)
G. PP + NP1 + VBD + (AA)
H. Other

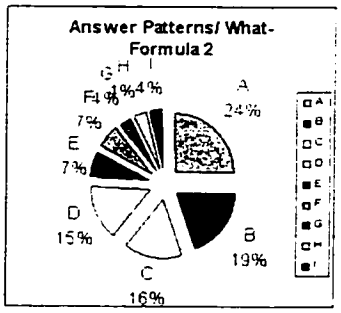


What-Pattern 1

1 out 150 questions follows this pattern

Pattern 2: What (WP) + Verb to-be (VBZ/VBD) + NP1 + IN + NP2 + IN(or to) + NP3 (or PP)
Example Pattern 2: What/ is /the /name /of /the managing director /of/ Apricot Computer? (Peter Horne)

Answer Patterns:
A. (AA) + ***NP2 (verb)*** + NP3
B. (AA) + VBD(Z) + NP2 + IN + NP3
C. Other
D. NP1 + IN + NP2 + IN + NP3 + VBD + (AA)
E. NP1 + (AA) + VBD(Z) + NP2 + IN + NP3
F. NP2 + IN + NP3 + VBD + (AA)
G. (AA) + VBD(Z) + ***WP + NP2 (inf. verb)*** + NP3
H. NP1 + IN + NP2 + VBD(Z) + (AA)
I. (AA) + VBD(Z) + NP1+ IN + NP2 + NP3

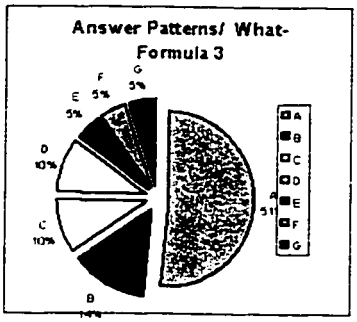


What-Pattern 2

6 out 150 questions follow this pattern

Pattern 3: What (WP) + Verb to-be (VBZ/VBD) + NP1 + IN + NP2 + IN + NP3 + PP

Example Pattern 3: What/ is/ the fare cost /for/ the round trip /between/ New York and London /on Concorde? ($6,400)



What-Pattern 3

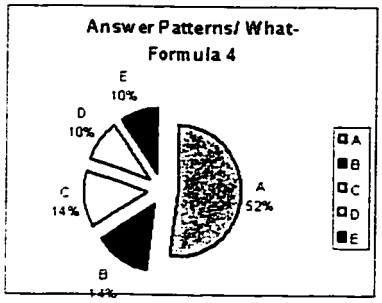
Answer Patterns:

A. NP1 + IN + NP2 + IN + NP3 + PP + VBD + (AA)
B. NP2 (noun NP1) + IN + NP3 + VBD + PP + (AA)
C. (AA) + VBD(Z) + NP1+ IN + NP2 + NP3 + PP
D. NP1(pro+3verb) + (AA) + IN + NP2 + IN + NP3 + PP
E. NP2 + NP1 (no det) + PP + IN + NP3 + VBD + (AA)
F. NP2 + NP1(noun) + IN + NP3 + VBD + (AA)
G. Other

1 out 150 questions follows this pattern

Pattern 4: What (WP) + Verb to-be (VBZ/VBD) + NP1 + IN + NP2 + WP + VBZ + PP

Example Pattern 4: What /is/ the name/ of/ the chronic neurological autoimmune disease/, which /attacks the protein sheath that surrounds nerve cells causing a gradual loss of movement in the body? (Multiple Sclerosis)



What-Pattern 4

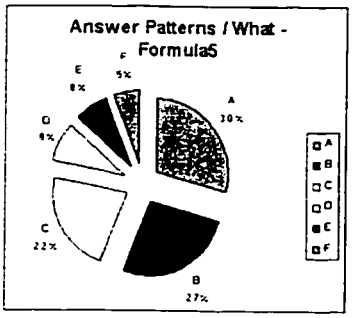
Answer Patterns:

A. NP2 + IN + WP + VBD(Z) + PP + VBD(Z) + (AA)
B. (AA) + VBD(Z) + NP1+ IN + NP2 + WP + VBD(Z) + PP
C. NP1 + IN + NP2 + WP + VBD(Z) + PP3 + VBD(Z) + (AA)
D. (AA) + VBD(Z) + NP2 + WP + VBD(Z) + PP
E. Other

1 out 150 questions follows this pattern

Pattern 5: What (WP) + Verb to-be (VBZ/VBD) + NP1 + IN + NP2

Example Pattern 5: What/ is/ the longest river/ in the United States? (the Mississippi)



What-Pattern 5

Answer Patterns:

A. NP2 + NP1 (no det) + VBD(Z) + (AA)
B. NP1 + IN + NP2 + VBD(Z) + (AA)
C. NP2 + VBZ (has) + NP1 (no det) + IN (of) + (AA)
D. NP2 + VBD(Z) + NP1 + IN + (/answer)
E. (AA) + VBD(Z) + NP1 + IN + NP2
F. Others

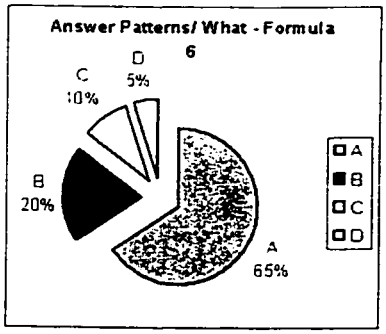
11 out 150 questions follow this pattern

Pattern 6: What (WP) + VP(VBZ+VBD) + NP1 + NP2 + VBZ + RB + VBN

Example Pattern 6: What/ is considered/ the costliest disaster/ the insurance industry/ has/ ever/ faced? (Hurricane Andrew)

Answer Patterns:

A. (AA) + VP + NP1 + NP2 + VBZ + RB + VBN
B. Other
C. (AA) + VP + NP1 + IN + NP2
D. NP1 + NP2 + VBZ + RB + VBN + VP + TOBE + (AA)



What-Pattern 6

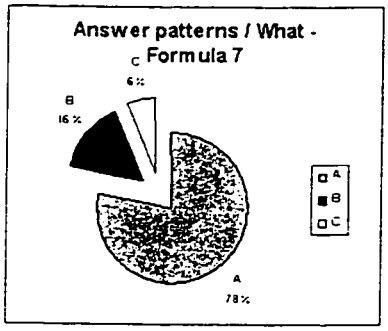
1 out 150 questions follows this pattern

Pattern 7: What (WP) + VBD(Z) + NP1

Example Pattern 7: What is an atom? (the smallest units of matter.)

Answer Patterns:

A. NP1+ VBZ + (AA)
B. (AA) + VBD(Z) + NP1
C. Other



What-Pattern 7

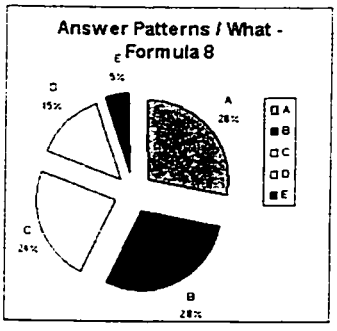
14 out 150 questions follow this pattern

Pattern 8: What (WP) + VBD1(Z) + NP1+ WP + VBD2(Z) + VBD3 + IN + PP

Example Pattern 8: What/ was/ the man's name/ who/ was /killed /in /a duel/ with /Aaron Burr? (Alexander Hamilton)

Answer Patterns:

A. (AA) + VBD2(Z) + VBD3 + IN + PP
B. (answer) + VBD1(Z) + NP1+ WP + VBD2(Z) + VBD3 + IN + PP
C. NP1+ WP + VBD2(Z) + VBD3 + IN + PP + VBD1 + (AA)
D. Other
E. (AA) + VBD1+ NP1+ VBD3 + IN + PP



What-Pattern 8

1 out 150 questions follows this pattern

Pattern 9: What (WP) + VBD1(Z) + NP1+ IN + NP2 + NP3 + IN + VBD2 + NP4 + PP

Example Pattern 9: What/ was/ the name/ of /the television show/, staring Karl Malden,/ that/ had /San Francisco/ in the title? (Streets of San Francisco)

Answer Patterns:

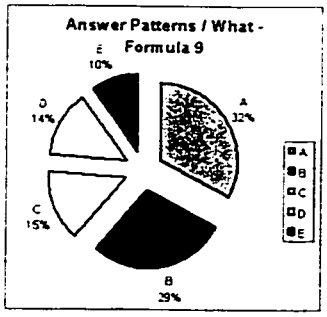
A. (AA) + VBD1(Z) + NP1+ IN + NP2 + NP3 + IN + VBD2 + NP4 + PP

B. NP1+ IN + NP2 + NP3 + IN + VBD + NP4 + PP + VBD1(Z) + (AA)

C. NP3 (inversed) + IN + NP2 + (AA)

D. (AA) + VBD1(Z) + NP2 + NP3

E. Other



What-Pattern 9

1 out 150 questions follows this pattern

Pattern 10: What (WP) + VBD1(Z) + NP1+ IN + NP2 + VBD2 + NP3

Answer Patterns:

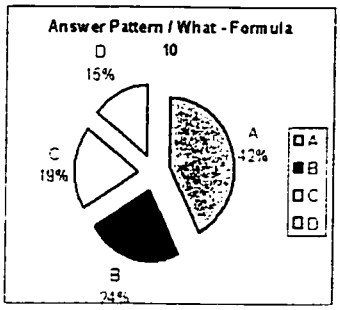
A. NP1+ IN + NP2 + VBD2 + NP3 + VBD1(Z) + (AA)

B. NP2 + VBD2 + NP3 + VBD1(Z) + IN + (AA)

C. (AA) + VBD1(Z) + NP1+ WDT + NP2 (no det) + VBD2 + NP3

D. NP3 (inversed) + IN + NP2 + (AA)

E. Other



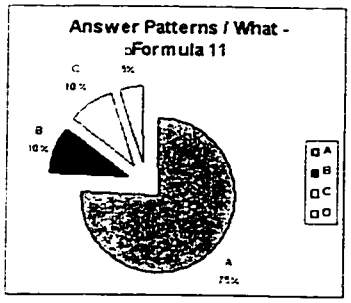
What-Pattern 10

1 out 150 questions follows this pattern

For the type of question *What* followed by a (noun phrase+ verb ToBe/ or verb (present/past)), (an auxiliary (present-past)+NP) or (NP+an auxiliary (present-past)), there are 19 formulas or variations of the same type of question, but just a sample of these formulas is showed in this report to make it easy to follow.

What + NOUN + ToBe +...

Pattern 11: What (WP) + NP1+ VBD1(Z) + IN + NP2
Example Pattern 11: What /soviet seaport/ is /on/ the Black Sea? (Sevastopol)
Answer Patterns:
A. (AA) + VBD1(Z) + NP1+ IN + NP2
B. NP1+ IN + NP2 + VBD1(Z) + (AA)
C. Other
D. (AA) + VBD1(Z) + IN + NP2

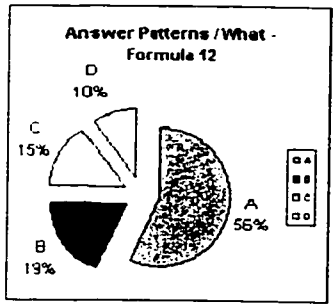


What-Pattern 11

1 out 150 questions follows this pattern

Pattern 12: What (WP) + NP1+ VBD1(Z) + NP2 + IN + NP3
Example Pattern 12: What /country/ is/ the biggest producer/ of tungsten? (China)

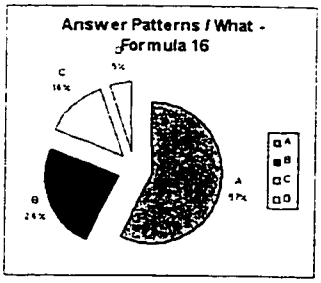
Answer Patterns:
A. (AA) + VBD1(Z) + NP2 + IN + NP3
B. NP2 + IN + NP3 + VBD1(Z) + (AA)
C. (AA) + VBD1(Z) + NP2 + NP1+ IN + NP3
D. Other



What-Pattern 12

1 out 150 questions follows this pattern

Pattern 16: NP1 + VBD1 + NP2 + VBD2(Z) + VBD3 + IN + What (WP) + NP3
Example Pattern 16: Material/ called/ linen/ is/ made/ from /What /plant? (Flax)
Answer Patterns:
A. NP2 + VBD2(Z) + VBD3 + IN + (AA)
B. (det) + NP1 + VBD1 + NP2 + VBD2(Z) + VBD3 + IN + (AA)
C. NP2 + VBD2(Z) + VBD3 + IN + NP3 + VBD1 + (AA)
D. Other



What-Pattern 16

1 out 150 questions follows this pattern

Pattern 18: What (WP) + NP1+ VBD1 + NP2 + IN + NP3 + PP

Example Pattern 18: What/ two US biochemists/ won/ the Nobel Prize/ in/ medicine/ in 1992? (Edwin Krebs Edmond Fischer)

Answer Patterns:

A. (AA) + **VBZ (ToBe)** + NP1+ **WP** +VBD1 + NP2 + IN + NP3 + PP

B. **Det** + NP1+ **WP** +VBD1 + NP2 + IN + NP3 + PP + **VBZ (ToBe)** + (AA)

C. (AA) + VBD1 + NP2 + IN1 + NP3 + PP

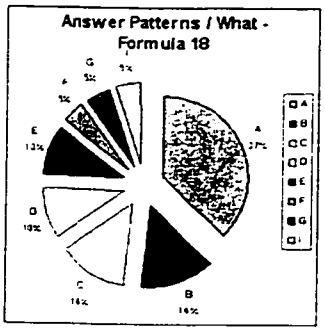
D. (AA) + NP1 + VBD1 + NP2 + IN1 + NP3 + PP

E. Other

F. NP2 + IN + NP3 + PP + **VBD (ToBe)** + VBD1 + **IN2** (by) + (AA)

G. NP1 + (AA) + VBD1 + NP2 + IN + NP3 + PP

H. PP + (AA) + **VBZ (ToBe)** + NP1 + **IN + (VBZ+VBN)** + NP2 + IN + NP3



What-Pattern 18

1 out 150 questions follows this pattern

What + AUX VBD(Z) + Noun... and What + Noun + AUX VBD(Z)

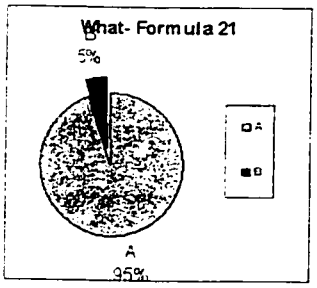
Pattern 21: What (WP) + AUX/VBD(Z) + NP1+ VB

Example Pattern 21: What/ does/ Phi Beta Kappa/ mean? (Academic honor society)

Answer Patterns:

A. NP1+ VB (tense) + (AA)

B. Other.



What-Pattern 21

1 out 150 questions follows this pattern

Pattern 24: What (WP) + AUX/VBD(Z) + NP1+ VB + TO+ IN + NP2

Example Pattern 24: What/ does/ Final Four/ refer /to/ in/ the sports world? (NCAA basketball Final Four)

Answer Patterns:

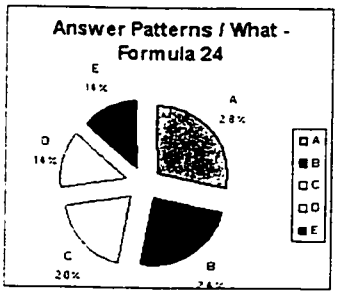
A. IN + NP2+ NP1 + VB (tense) + TO+ (AA)

B. NP1+ VB (tense) + TO+ (AA) + IN + NP2

C. Others

D. NP1+ VB (tense) + TO+ (AA)

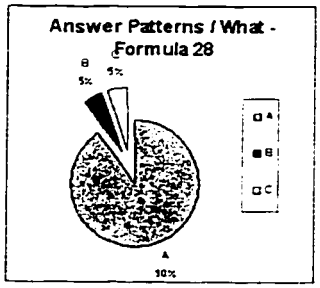
E. NP1+ IN + NP2 + VB (tense) + TO+ (AA)



What-Pattern 24

1 out 150 questions follows this pattern

Pattern 28: What (WP) + NP1 + AUX/VBD(Z) + NP2+ VB + IN
Example Pattern 28: What/ state/ did/ the Battle of Bighorn/ take/ place in? (Montana)
Answer Patterns:
A. NP2+ VB (tense) + IN + (AA)
B. NP2 + VBZ(D) (ToBe) + IN + (AA)
C. Other



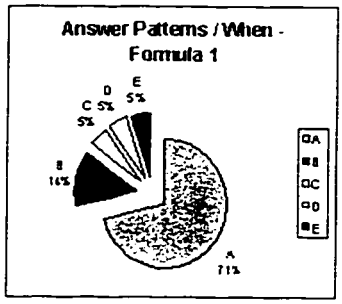
What-Pattern 28

1 out 150 questions follows this pattern

Samples of patterns for each type of question; *Who, Name, Which, How, Where* and *When* are the following:

When + To Be +... - Type of Question

Pattern 1: When (WP) + Verb to-be VBD1 + NP1 + NP2 + VB
Example Pattern 1: When /was/ Dubai's/ first concrete house/ built? (1956)
Answer Patterns:
A. NP1 + NP2 + VBD1 + VB + **IN** + (AA)
B. **det** + NP2 + **IN** + NP1 + VBD + VB + IN + (AA)
C. det + NP2 + VB + IN + NP1 + VBD + IN + (AA)
D. IN + (AA) + NP1 + NP2 + VBD1 + VB
E. Other.

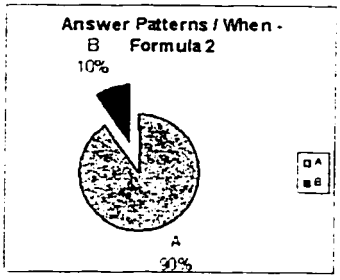


When-Pattern 1

1 out 150 questions follows this pattern

Pattern 2: When (WP) + Verb to-be VBD1 + NP1
Example Pattern 2: When / was/ the San Francisco fire/? (April 20, 1906)

Answer Patterns:
A. NP1 + VBD1 + **IN** + (AA)
B. Other



When-Pattern 2

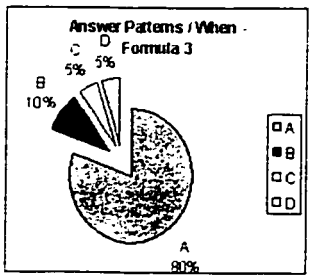
1 out 150 questions follows this pattern

When + AUX VBD+... - Type of Question

Pattern 3: When (WP) + AUX VBD1 + NP1 + VB + VBD2 + IN + NP2
Example Pattern 1: When/ did/ John F. Kennedy/ get/ elected/ as/ President? (1960)

Answer Patterns:
A. NP1 + VB (tense) + VBD2 + IN + NP2 + **IN** + (AA)
B. Other
C. NP1 + VBD + NP2 + IN + (AA)
D. **NP1 (pron) + VBD (ToBe)** + VBD2 + IN + NP2 + **IN** + (AA)

1 out 150 questions follows this pattern



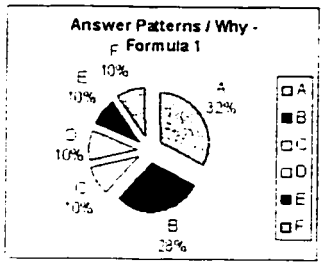
When-Pattern 3

Why + AUX +... - Type of Question

Pattern 1: Why (WP) + AUX VBD1 + NP1 + VB + NP2
Example Pattern 1: Why/ does/ the moon/ turn/ orange/? (eclipse)

Answer Patterns:
A. NP1 + VB (tense) + NP2 + "because" + **IN** + (AA)
B. NP1 + VB (tense) + NP2 + "during" + IN + (AA)
C. "because" + IN + det + (AA)
D. NP1 + VB (tense) + NP2 + **WDT (when)** +"there is" + IN + (AA)
E. IN + (AA) + **VBZ** (is) + **NP3** (the reason) + WDT (why) + NP1 + VB + NP2
F. Other

1 out 150 questions follows this pattern



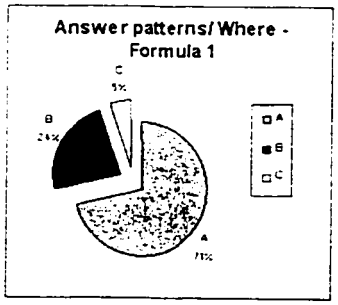
Why-Pattern 1

Where + Verb ToBe +... - Type of Question

Pattern 1: Where (WP) + VBD(Z) + NP1 + IN + NP2
Example Pattern 1: Where/ is/ the highest point/ in/ Japan? (Mt. Fuji)

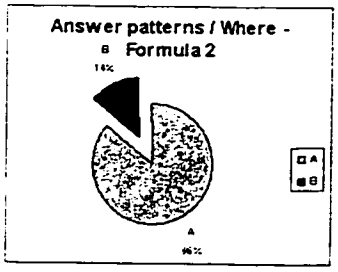
Answer Patterns:
A. (AA) + VBD(Z) + NP1 + IN + NP2
B. NP1 + IN + NP2 + VBD(Z) + (AA)
C. Other

1 out 150 questions follows this pattern



Where-Pattern 1

Pattern 2: Where (WP) + VBZ1(D) + NP1 + VBD2
Example Pattern 2: Where/ is/ Belize/ located? (Central America)
Answer Patterns:
A. NP1+ VBZ1(D) + VBD2 + IN + (AA)
B. NP1+ VBZ1(D) + IN + (AA)



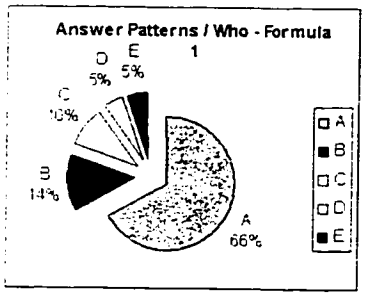
Where-Pattern 2

1 out 150 questions follows this pattern

Who + To Be +... - Type of Question

Pattern 1: Who (WP) + Verb to-be (VBD(Z)) + NP1 + IN + NP2 + VBD + IN + NP3 + PP1+ PP2
Example Pattern 1: Who/ was/ the leader/ of/ the Branch Davidian Cult/ confronted/ by/ the FBI/ in Waco, Texas/ in 1993? (Mr. David Koresh)

Answer Patterns:
A. (AA) + VBD + NP1 + IN + NP2 + VBD + IN + NP3 + PP1+ PP2
B. NP1 + IN + NP2 + VBD + IN + NP3 + PP1+ PP2 + VBD + (AA)
C. (AA) + VBD + NP1 + IN + NP2
D. PP2 + NP1 + IN + NP2 + VBD + IN + NP3 + PP1+ VBD + (AA)
E. Other
1 out 150 questions follows this pattern



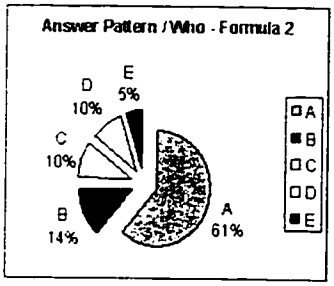
Who-Pattern 1

Who + VBD +... - Type of Question

Pattern 2: Who (WP) + VBD + NP1 + PP1
Example Pattern 2: Who/ released/ the Internet worm/ in the late 1980s? (Robert Morris)

Answer Patterns:
A. (AA) + VBD + NP1 + PP1
B. PP1 + (AA) + VBD + NP1
C. NP1 + **VBZ (ToBe)** + VBD + PP1 + **IN (by)** + (AA)
D. Other
E. (AA) + VBD + PP1 + NP1

1 out 150 questions follows this pattern



Who-Pattern 2

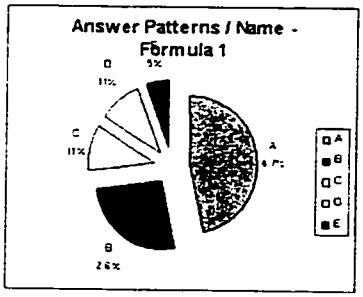
Name + NP +... - Type of Question

Pattern 1: Name + NP1 + IN + (VBZ+VBN) + NP2 + PP1

Example Pattern 1: Name/ a film/ that/ has won/ the Golden Bear/ in the Berlin Film Festival? (In The Name Of The Father)

Answer Patterns:

A. (AA) + VBN + NP2 + PP1
B. (AA) + VBZ (ToBe) + NP1 + IN + (VBZ+VBN) + NP2 + PP1
C. NP1 + (AA) + IN + VBN + NP2 + PP1
D. other
E. (AA) + VBD (ToBe) + det + Name + IN + NP1 + IN + VBN + NP2 + PP1



Name -Pattern 1

Bibliography

[1] Agichtein, E., Lawrence, S, and Gravano, L. Learning Search Engine Specific Query Transformations for Question Answering. *Tenth World Wide Web Conference*. Hong Kong, China. May 1-5, pp 1-10. 2001.

[2] Bergler, S. and Knoll, S. Coreference patterns in the Wall Street Journal. In: Percy, C.E., C. Meyer & I. Lancashire (eds.). Verlag. pp 85-96. 1996.

[3] Bobrow, D. STUDENT: A question-answering system for high school algebra word problems. *Proceedings of AFIPS conference*, 26, FJCC, pp 664. 1964.

[4] Brill, Eric. A Simple Rule-Based Part Of Speech Tagger. Proceedings of ANLP-92, *3rd Conference on Applied Natural Language Processing*. Trento, Italy. pp 152-155. 1992.

[5] de Chalendar, G. et al. The question answering system QALC at LIMSI: experiments in using Web and WordNet. In *Notebook Proceedings of TREC-11*. pp 457-465. Gaithersburg, Maryland, 2002.

[6] DFKI German Research Center for Artificial Intelligence. FLUIDS: Future Lines of User Interface Decision Support – Natural Language Generation. http://www.dfki.uni-sb.de/fluids/Natural_Language_Generation.html. Visited June 30th at 3:40 pm. 1999.

[7] EAGLES Lexicon Interest Group. Preliminary Recommendations on Semantic Encoding-Interim Report. 1998 http://www.ilc.pi.cnr.it/EAGLES96/rep2/node35.html Visited June 30th.

[8] Garside. R. The CLAWS word-tagging system. In Garside R., F. Leech, and G. Sampson, editors, *The Computational Analysis of English*. Longman, London and New York. 1987.

[9] Hovy, Eduard H.. Approaches to the planning of coherent text. In Cécile L. Paris, William R. Swartout, and William C. Mann, editors, *Natural language generation in artificial intelligence and computational linguistics*. Kluwer Academic Publishers, July 1991. Presented at the Fourth International Workshop on Natural Language Generation. Santa Catalina Island, California, July, 1988.

[10] Katz, B. From Sentence Processing to Information Access on the World Wide Web. *In Natural Language Processing for the World Wide Web: Papers from the 1997 AAAI Spring Symposium*, pp 77-94, 1997.

[11] Kupiec, J. MURAX: A robust linguistic approach for question answering using an on-line encyclopedia. *In 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp 181-190, Pittsburgh, 1993.

[12] Lawrence S. and Giles. C. L. Context and page analysis for improved web search. *IEEE Internet computing*, vol. 2, no. 4. pp 38–46. 1998.

[13] Marcus, M., Santorini, B. and Marcinkiewicz, M.A.: Building a large annotated corpus of English: The Penn Treebank. In *Computational Linguistics*. 1997. Volume 19, number 2, pp 313-330.

[14] McKeown, Kathleen R. Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text. *Studies In Natural Language Processing*. Cambridge University Press, 1985.

[15] Moore, J. and Swartout, W. A reactive approach to explanation: Taking the user's feedback into account. In C. Paris, W. Swartout, and W. Mann, editors, *Natural language generation in artificial intelligence and computational linguistics*. Kluwer Academic Publishers, July 1991. Presented at the Fourth International Workshop on Natural Language Generation. Santa Catalina Island, California, July, 1988.

[16] Paris, C. Generation and explanation: Building an explanation facility for the explainable expert systems framework. In C. Paris, W. Swartout, and W. Mann, editors, *Natural language generation in artificial intelligence and computational linguistics*. Kluwer Academic Publishers, July 1991. Presented at the Fourth International Workshop on Natural Language Generation. Santa Catalina Island, California, July, 1988.

[17] Plamondon L, Kosseim L, Lapalme G. The QUANTUM Question-Answering System at TREC-11. *In Notebook Proceedings of the Eleventh Text Retrieval Conference (TREC-11)*. pp 670-677. Gaithersburg, Maryland, 2002.

[18] Reiter Ehud and Dale Robert. *Building Natural Language Generation Systems*. 2000. Cambridge University Press.

[19] Reiter E., Mellish C., Levine J. Automatic generation of technical documentation. In: *Applied Artificial Intelligence*. 9, 1995. http://citeseer.nj.nec.com/article/reiter95automatic.html.

[20] SIGGEN. What is Text Generation?. ACL Special Interest Group on Generation. Web 2002 http://www.dynamicmultimedia.com.au/siggen/nlg.html. Visited August 9 at 12:30 pm. 2002

[21] Soubbotin M. and Soubbotin S. Patterns of Potential Answer Expressions as Clues to the Right Answers. *In Proceedings of the Tenth Text Retrieval Conference (TREC 10)*. pp 293-302. Gaithersburg, Maryland, 2002.

[22] Temperley, D., Sleator, D. and Lafferty, J. Grammatical trigrams: A probabilistic model of link grammar, [Abstract], *in Proceedings of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, Cambridge, MA, October 1992. Also issued as technical report CMU-CS-92-181, Department of Computer Science, Carnegie Mellon University, 1992.

[23] Woods, W. Semantics and quantification in natural language question answering. *In Advances in Computers*. Volume 17. Academic Press. pp 1-87. 1977.

[24] Zhiping Zheng. AnswerBus Question Answering System. *Proceeding of HLT Human Language Technology Conference (HLT 2002)*. San Diego, CA. March 24 - 27, 2002. pp 1-5.