

# Evaluation of the New Paraphrase and Question Formulation Capability in AURA

Working Note 37

Peter Clark

Boeing Research and Technology

February 1st, 2010

## 1. Introduction

In 2009 we introduced the use of paraphrases and some initial use of underspecification (KB guidance) into AURA, wrapped around a simple deductive question-answering facility. We also experimented with using full (rather than controlled) language interpretation for question understanding. We here refer to these developments together as "NewQF". The goal of NewQF was to allow a wider variety of question formulations to be correctly answered by AURA, in particular handle question wordings either identical or close to the original AP English. This document gives a brief summary of NewQF, and then reports three evaluations of its performance and the degree to which it meets its goal.

The main evaluation was performed on the 22 "difficult questions" identified in early 2009 from the Final Evaluation, and through ablation studies we evaluated the degree to which paraphrases, underspecification, and full language interpretation can help interpret the 22 questions using close to the original language in the AP exams.

In addition, we tested NewQF's performance on the Minimal Test Suite and on the Final Evaluation CPL questions (from the Q1 expert-expert quadrant), to verify that it was performing the intended deductive reasoning and identify any gaps in its performance. The results of those tests are also summarized here.

## 2. Overview of NewQF

### 2.1 New Developments

The new question formulation facility (NewQF) in AURA constitutes a significant redesign of how question formulation is performed in AURA. It includes three major advances:

#### 2.1.1 Use of Paraphrases

In language, many phrases can mean the same thing. To make sense of unanticipated phrasings in a question, the NewQF makes use of the DIRT paraphrase database to explore alternative phrasings when a phrase in a question cannot be directly understood. As a result, AURA should be able to accommodate a greater number of ways of stating a question.

#### 2.1.2 Initial Work on Underspecification (KB Guidance)

AURA's old QF facility used a traditional pipelined architecture for language processing:

Text → Parse → Logical Form → Word Sense Disambiguation (WSD) →  
Semantic Role Labeling (SRL) → Question-Answering

While often adequate, such a pipeline can also be brittle: a mistake at any stage in the pipeline can result in downstream question-answering failure. To address this brittleness, the pipeline has been revised to accommodate a degree of "underspecification", i.e., preservation of ambiguity, for subsequent downstream modules. Specifically, in 2009 we have implemented underspecification handling for (only) semantic role labeling (SRL), one of the weaker aspects of the old QF facility.

To illustrate this, consider asking the question:

Does a nucleus have chromosomes? [expected answer: yes]

In the old QF, AURA eagerly interpreted "have" as meaning has-part, and subsequently (and here undesirably) answer "No", as (according to the KB) the nucleus *contains* chromosomes, rather than has a *part* which is a chromosome. In the NewQF, however, "have" is left as an underspecified meronymic/containment relation<sup>1</sup>. Then during question-answering, AURA searches for a (non-empty) answer within that scope of meanings, the underlying assumption being that the question-asker's intent is that question ambiguity be resolved in the way that makes the most sense with respect the knowledge-base. In this case, NewQF finds that (according to the KB) all nuclei enclose chromosomes, and thus will answer "Yes", also showing that "have" was interpreted as "encloses". This mechanism thus adds a degree of robustness to question-answering, which we evaluate later in this document.

In the old QF the semantics of the original question was a single, fully specified logical formula. In the new QF, it is an underspecified logical formula, or equivalently a *set* of alternative, fully specified formulae that are then explored to see if they yield a (non-empty) answer. If an answer is found, then the fully specified formula which yields an answer is reported to the user. One can thus view the KB as guiding the completion of the interpretation as part of the question-answering process.

### 2.1.3 Replacing "Matching" with Search for Deductive Answers

A third, closely related development in the NewQF is replacing the notion of "heuristic question-answering" with search for a deductive answer. Rather than eagerly finding a single interpretation of the question and then performing some kind of heuristic matching to find an answer, question-answering is recast as exploring multiple interpretations, searching for one which *deductively* provides a positive answer. Again, the underlying assumption here is that the question-asker's intent is that question ambiguity be resolved in the way that makes the most sense with respect the knowledge-base. The importance of this is that it replaces heuristics with well-defined search plus deductive reasoning, thus providing a formal semantics and specification of correct behavior for the question-answering module in AURA. These issues and the semantics of questions are discussed in more depth [1].

In addition, we also explored removing the controlled language restrictions of CPL and using the full Boeing Language Understanding Engine (BLUE), to identify its potential utility in improving AURA's question interpretation abilities.

---

<sup>1</sup> AURA's graphical display of the interpretation is not currently able to display underspecified relations, and so the graph instead shows a fully specified relation although internally it is treated as underspecified.

## 2.2 Description, Coverage, and Current Limitations

We now provide a brief summary of the current NewQF facility and its behavior.

### 2.2.1 Coverage of Question Types

The NewQF facility covers four major question types:

- Find a value
- Is it true that
- What is the X (object identification)
- Is X a Y (class membership)

These cover many of the questions in the AP exams. It currently does not cover the following question types (which can be answered using existing mechanisms in BPS):

- What is a X (definition)
- What is the similarity/difference between X and Y (comparison)
- What is the relationship between X and Y
- What is an example of X
- How many X (counting)

It also does not handle the following types, also not covered by existing AURA:

- What are the superclasses/subclasses of X (taxonomy)
- Is it possible that ...

### 2.2.2 How Questions are Interpreted

#### a. Find-a-value and Is-it-true Questions

The NewQF currently treats the entire question as a query, i.e., does not do any "problem setup". (This clearly needs to be extended to also allow problem setup in 2010, but turns out to be surprisingly adequate for biology). Any declarative sentences in the question are treated as part of the query, for example, NewQF considers the following three questions as equivalent:

- (1) What does the nucleus of a eukaryotic cell contain?
- (2) A eukaryotic cell has a nucleus. What does the nucleus contain?
- (3) A eukaryotic cell has a nucleus. The nucleus contains some things. What are the things?

In biology at least, this equivalence is often what the author intended; for example, an original AP question like (1) is often rewritten by the user in CPL like (2), the intension being that the CPL reflects the same query as the original. In the biology final evaluation, which we describe later (Section 5), none of the 263 questions required an assertional component; similarly for the 60 questions in the biology minimal test suite, only 3 have assertional content (see Section 4). For physics, however, this would clearly be limiting as many questions include explicit assertions as part of problem setup.

For find-a-value questions, NewQF returns a set of *classes*<sup>2</sup> rather than instances, namely those classes which make a quantified form of the question formula true, where the quantified form

---

<sup>2</sup> with the exception of property values, where NewQF returns the angified value(s) (e.g. "10 millimeters").

tests whether *all* members of one class have the properties stated in the question. For example, the interpretation of questions (1)-(3) above can be written:

;;; Answer = the set of (most general) class(es) for which the question formula is true:  
{ class | is-true[ $\forall x \text{ isa}(x, \text{Euk-Cell}) \rightarrow \exists y, z \text{ isa}(y, \text{Nucleus}), \text{isa}(z, \text{class}), \text{have}(x, y), \text{contain}(y, z)$ ] }

where have() and contain() are underspecified relations.

For is it true questions, the interpretation is similarly to find the truth of the quantified form of the question formula.

## b. What-is-the Questions

Consider the question:

(4) A cell has a nucleus. What is the cell?

Again this is interpreted as a request for classes which make a quantified form of the question formula true, but this time where the classes are in the universally quantified part of the formula. In this case, (4) is interpreted as asking for the (most general) subclass(es) of cell for which the formula is true:

{ class | is-true[subclass-of(class, Cell) &  $\forall x \text{ isa}(x, \text{class}) \rightarrow \exists y \text{ isa}(y, \text{Nucleus}), \text{have}(x, y)$ ] }

Implementationally, this (here) requires a search through the subclasses of Cell to identify those for which the formula is true.

### 2.2.3 Implementational Limitations

We evaluate the performance of the NewQF facility shortly, but first summarize its known limitations and features (bear in mind this is still a work in progress):

- NewQF does not do any “problem setup” (as discussed above)
- The CPL/BLUE interpreter does not capture quantification in the question. For example, the following two questions produce the same CPL/BLUE interpretation, i.e., are currently indistinguishable to QA (both in the old QF and NewQF):

(5) Do eukaryotic cells have a nucleus?

(6) Are nuclei part of a eukaryotic cell?

NewQF proceeds by exploring both quantification patterns when answering the question, and returns any non-null answers found.

- Transitive closure (and other “grain size”) reasoning is performed in KM. This is currently depth-limited to a depth of 5, for reasoning times to be acceptable. (Strictly this is a self-imposed limitation of the KM knowledge base, not NewQF)
- Search for what-is-the answers (see b above) has a built-in timeout of 20 seconds, as it potentially involves searching a large number of classes. (For these, NewQF first does a quick non-inferencing search, followed by a full inferencing search). It is thus possible that NewQF will not find a (deducible) answer to a complex what-is-the question, although no such questions were encountered during this evaluation.
- Multiple queries within a single question are not currently handled.

- NewQF does not handle mathematical comparison operations (greater than, less than)
- In the evaluations described in this document, reasoning times are typically one or two seconds, but can be up to 30 seconds (running on a somewhat slow machine, a 2006 Dell Latitude D610 with 2GB RAM). See Appendices B, C, and D for samples of the reasoning time taken during the trials.
- The possible meanings of underspecified relations are currently stored in the NewQF code, rather than in the component library.
- The explanation presentation is still somewhat crude.

Despite these limitations, its ability to interpret questions is significantly improved compared with the previous CPL implementation, as we now describe.

### 3. Evaluation I: The "Difficult Questions"

#### 3.1 Creating the Test Suite (46 questions)

The primary evaluation of NewQF is based on the “difficult questions” identified in the biology Final Evaluation, and which we now describe. At the start of 2009, we identified 22 "difficult" biology AP questions in the biology Final Evaluation, where non-expert SMEs had difficulty formulating them in CPL and obtaining an answer. A detailed discussion of these questions and the various difficulties that they posed is presented in [2], and a list of them is also given in Appendix A. Our interest is in whether/how much paraphrases, underspecification, and use of the full BLUE language engine can improve AURA’s question interpretation capabilities.

To evaluate the NewQF facility, we proceeded as follows. First, the 22 questions were minimally rewritten as complete sentences, where the rewrites were intended to break up multiple questions into individual questions, and convert multiple choice options into sentence form while preserving fidelity to the original English. The phrasing and vocabulary was left unchanged, although parallelism<sup>3</sup> was removed and partial sentences were expanded into full sentences. The original questions and their rewrites are shown in Appendix A. All 22 questions had 5 multiple choice parts, and typically each multiple choice part became a separate question. As a result, the 22 AP questions became a test suite of 142 questions. (All the questions are true/false questions).

There was only one question which we subsequently modified, namely:

74c. [Is it true that] In DNA replication, the DNA polymerase builds a new strand from 5' to 3'?

AURA's parser does not properly tolerate "5'" and "3'", and so these were replaced with "the five prime end" and "the three prime end":

74c. [Is it true that] In DNA replication, the DNA polymerase builds a new strand from the five prime end to the three prime end?

Of the 142 questions, 47 have the official answer of "yes" (i.e., true). In ideal circumstances, AURA should also answer "yes" to these 47 (and "no" to the remaining). However, there was one question in the 47 whose official "yes" answer seemed somewhat questionable, namely:

---

<sup>3</sup> An example of parallelism is “Are A,B,C an X,Y,Z respectively?”, which would be rewritten as three questions “Is A an X?” “Is B a Y?” “Is C a Z?”

45d. [Is it true that] The function of the nucleus is to replicate DNA?

Although the official answer is "yes", biologically speaking it seems somewhat questionable (a more reasonable statement would be that the function of the nucleus is to regulate the cell's activity, or control the cell). As a result, we did not include this question in the set of questions with positive answers.

This left a test set of 46 questions where a positive answer should be expected, and which we used to evaluate the NewQF facility.

### **3.2 Out of Scope Questions**

Of the 46 questions, it turns out that 6 are out of scope of the existing framework for AURA. These six are:

68a [Is it true that] A DNA adenine can bond to a DNA thymine?

68c [...etc...] A DNA guanine can bond to a RNA cytosine?

68d A DNA adenine can bond to a RNA uracil?

68e A DNA guanine can bond to a DNA cytosine?

77c-1 During mitosis of plant cells, a cell plate begins to form at telophase?

77c-2 During mitosis of animal cells, a cleavage furrow is initiated at telophase?

The first four questions are "Is it possible that...?" rather than "Is it true that...?" questions, a question form not currently handled in AURA. One could imagine a weak method for answering such questions whereby AURA searches for an *example* of the assertions in the question, and if an example is found then the assertion is indeed possible. However, even this would not answer question 68c (A DNA guanine bonding to an RNA cytosine) as there *are* no examples in the KB of a DNA base bonding with an RNA base! Similarly for 68d.

77c-1 and 77c-2 are out of scope for different, but equally interesting reasons. Although the questions seem to involve a simple subevent relationship between mitosis and telophase, telophase is not strictly a subevent of mitosis, but is instead an event which temporally overlaps with the end of mitosis. Thus, to answer this question requires that temporal knowledge to be encoded and reasoned about in AURA, something outside AURA's current ontological scope.

### **3.3 Extending the KB**

Of the remaining 40 "in scope" questions, we found that the Biology Reference KB had the knowledge required to (in principle) answer 25 of them, but was missing the knowledge for the remaining 15. As we want to factor out "knowledge missing in the KB" as a reason for QA failure, the Biology SME added the missing knowledge for 14 of the 15 questions (Details of the additions are given in Table 1 at the end of this document). The remaining question (103b) appears to require general commonsense knowledge:

103b: [Is it true that] The cell membrane of a cell is in contact with the environment?

As it did not seem feasible to add the required knowledge without "cheating" (i.e., hardwiring the KB to give the answer we wanted), we did not extend the KB to support an answer to this question.

### 3.4 Evaluation

We ran the NewQF facility on the 46 questions, 39 of which are in principle answerable given the KB. *It is important to note that both the old QF and NewQF can answer all 39 questions given a favorable wording of the questions.* However, our interest is not in answering some formulation per se, but how many of the questions can be answered *without any rewording*, i.e., using the wording in the original AP questions, minimally rewritten as sentences. In several cases very simple rewordings would allow the question to be answered correctly, however this was not done as it was not the purpose of the evaluation.

For each question, we performed ablation tests to identify which questions required use of:

- the full BLUE language interpreter (i.e., were outside the scope of CPL)
- underspecification
- paraphrases

in order to find the answer. For questions which still did not produce an answer, we also produced a diagnosis of the cause of failure.

### 3.5 Summary of the Results

The full results are shown in **Table 1** at the end of this document, which provides full details for each question. For our purposes here, we provide an overall summary. There are two useful ways of summarizing the results:

- Module view:** How many of the 46 questions require the use of BLUE, underspecification, and paraphrases
- Cumulative view:** How the overall score on the 46 improves as the 3 new modules are added

Note that a and b differ, as many questions require more than one new module to obtain a correct answer.

#### 3.5.1 Module View: Summary of the Role of each Module

We can summarize the results on a per-module basis as follows: Of the 46 questions:

- 15 required extending the KB (i.e., the original KB did not contain the required knowledge)
- 18 require use of BLUE (i.e., 28 are within the scope of CPL)
- 12 require use of underspecification
- 16 require use of paraphrases
- 20 require some additional (unmade) extension to QA.

Note that the total is not 46, as these categories are not mutually exclusive. The figures above are the same as in the column totals in Table 1.

#### 3.5.2 Cumulative View: Cumulative Accuracy

Given the extended KB, we can also summarize the results in a cumulative way. The following number of questions are answerable as we add modules:

- 5 are answerable out of the box using AURA and CPL
- + 3 if we upgrade to BLUE
- + 5 if we allow underspecification

- + 5 if we allow paraphrases
- + 7 if we allow underspecification and paraphrases together

Total: 25

Thus the accuracy score rises from 5 to 25 (out of 46) with BLUE, underspecification and paraphrases together. For the remaining 21 questions which are currently not answerable:

- 6 are out of scope (discussed earlier)
- 1 requires general knowledge
- 14 require extensions to CPL/BLUE/QA reasoning

Specific details of these extensions to CPL/BLUE/QA are given in the keys to Table 1 at the end of this document. As noted earlier, an often trivial rewording of the question, as a SME might perform, would allow these 14 questions to be answered, but this was not the point of the study.

#### 4. Validation on the Minimal Test Suite (MTS)

As a validation exercise, we also ran the NewQF on the Minimal Test Suite containing 60 simple questions, considered to be straightforwardly answerable. The MTS was developed in 2009 as a basic fidelity check for AURA's question-answering behavior. Note that BLUE, paraphrases, and underspecification are *not* required to answer these questions. Thus this investigation constitutes a basic validation exercise on the core NewQF facility.

The trace of NewQF running on the MTS is given in Appendix C, and is summarized here. Of the 60 questions in the MTS:

- 22 are out of scope of the NewQF, with question types: What is a (2), Find an example (2), Similarities/differences (8), What is the relationship (6), How many (counting) (4). Note that these question types can be passed on to existing AURA BPS mechanism to answer, and thus are still covered in the overall AURA system.
- 30 NewQF gets the right answer

This leaves a residue of 8 questions where NewQF gets the wrong answer. Of these 8:

- 3 have assertional content. As described earlier, NewQF does not currently do any "problem setup", and thus these are currently out of scope. These three are:
  - MTS-31: There is an increase. The from-value of the increase is 10 celsius. Is it true that the increase is a heating?
  - MTS-32: There is an increase. The from-value of the increase is 10 celsius. Is it true that the increase is a cooling?
  - MTS-41: The length of a eukaryotic cell is 20 micrometers. Is it true that the length of the cell is less than 0.05 millimeters?
- 1 requires use of comparison operators, similarly not covered by NewQF at the moment:
  - MTS-39: Is it true that 10 m is less than 35 feet?
- 1 requires use of undefined relation (purpose). This question is also not answerable with BPS or PDQA for the same reason.
  - MTS-19: What is the purpose of cellular respiration?



- 3 questions seem to have bad answers in the test suite, where the reference answer given is not universally true:

MTS-7: What is the complement of adenine? [gold answer: thymine]

MTS-8: What is the complement of adenine? [gold answer: uracil]

MTS-25: A nucleus is a part of a cell. Is it true that the cell is a eukaryotic cell? [gold answer: yes]

For questions 7 and 8 (which are identical), neither given answer is universally true for all adenines. Similarly, there is not sufficient knowledge in the KB to prove that all cells with nuclei are eukaryotic cells. Rather, these 3 gold answers can be viewed as "partial answers" - answers which are true for some restricted subclass of the queried class. 2010 work for AURA includes enabling such questions to be answered, by finding the restricted class where a deductive answer is then obtainable.

Our conclusion from this exercise is that NewQF is performing as per its design.

## 5 Validation on the Final Evaluation Questions

As a final validation for the NewQF facility, we re-ran a set of CPL questions from the Final Evaluation using the NewQF. Specifically we used the Q1 (expert KF, expert QF) KB and SME-authored CPL questions dated 1/6/09. The data includes a semi-formal spreadsheet analysis of the CPL questions, answers, and correctness, thus providing a basis for comparing results. Our goal in this exercise is to validate that NewQF and BPS produce the same answers for the subset of questions within NewQF's scope, or provide an explanation for the difference.

The spreadsheet associated with this Q1 data contains 263 CPL questions. Of these:

- 23 are question types outside NewQF's scope (e.g., "What is a X?")
- 9 contain multiple questions within a single query (outside NewQF's scope)
- 1 is an assertion, not a question
- 48 are not clearly scored in the spreadsheet
- 73 are scored as BPS incorrect

The remaining 109 are in scope of NewQF and marked as BPS correct, and thus these 109 are of interest. They were re-run using NewQF and the NewQF results compared with those scored on the spreadsheet. The full trace of the NewQF run is shown in Appendix D.

Of these 109:

- 97 BPS and NewQF both got the correct answer (62 yes, 35 no)
- 1 BPS was scored correct solely from the documentation string it recited (59-1&2)
- 5 BPS were scored correct in the spreadsheet, but produced the incorrect answer during the re-run (correct behavior could not be reproduced), and NewQF was correct.
- 1 BPS apparently found an abductive answer
- 5 BPS produced a somewhat dubious answer

The discrepancies in the answers to the last 11 (5+1+5) questions are of interest, and so we now describe them in more detail.

The five marked as BPS correct in the spreadsheet but the BPS correct behavior could not be reproduced were 77c1, 80c, 82c, 83c, and 92b. In all these five cases the correct answer is "No".

During the re-run for this evaluation BPS incorrectly answered "Yes" with meaningless explanations for these, while NewQF correctly answered "No". It is possible that some minor versioning difference accounts for this changed behavior in BPS.

The question requiring some kind of "partial" or "abductive" reasoning with the knowledge in the KB was 36e:

36e. Is it true that a lysosome is inside a membrane? [correct answer is yes]

For this question, the KB does not contain the knowledge that all lysosomes are inside a membrane, only that all lysosomal-membranes enclose a lysosome, and so from a deductive point of view 36e cannot be proven true. Nevertheless BPS answered "Yes" to 36e, presumably performing some kind of abductive reasoning. NewQF's deductive "No" answer can be considered correct behavior here according to its specification.

The remaining five with "dubious" BPS answers were 55a1, 55a2, 90-1, 90-2b, and 90-3a, all of a similar form. The correct answer to all these five is "Yes", BPS answered "Yes", but NewQF answered "No":

Q55.a.1: Guanine is a part of a DNA.

Cytosine is part of a RNA.

Is it true that the guanine is the complement of the cytosine?

Q55.a.2: Cytosine is a part of a DNA.

Guanine is part of a RNA.

Is it true that the cytosine is the complement of the guanine?

Q90.1: Uracil is part of RNA.

Adenine is part of RNA.

Is it true that the uracil is the complement of the adenine?

Q90.2.b: Adenine is part of RNA.

Uracil is part of RNA.

Is it true that the adenine is the complement of the uracil?

Q90.3a: Cytosine is part of RNA.

Guanine is part of RNA.

Is it true that the cytosine is the complement of the guanine?

The first two questions are similar to the question considered out of scope during the earlier NewQF evaluation, namely "Is it possible that a DNA guanine can bond with a RNA cytosine?" (question 68, Section 3.2). Deductively, AURA cannot answer this as it has no knowledge about parts of DNA and RNA connecting. It is not clear how BPS derived a "Yes" answer, though most likely it ignored the setup information in these questions.

The last three questions appear straightforward to answer deductively. However inspection of the CPL shows that the question creates *two* (not one) RNA molecules, as there is no explicit coreference between the RNA in the first and second sentences. As a result, question 90.1 is interpreted as "Is the uracil of one RNA the complement of the adenine in another RNA?". As the KB only contains knowledge of complementarity within a single strand, the "Yes" answer is not deductively derivable from the KB. Again, it is not clear how BPS answered "Yes", again it perhaps ignored the setup information. If we replace "Adenine is part of RNA" with "Adenine is

part of *the* RNA" in question 90.1, thus establishing the coreference, then a "Yes" answer becomes deductively derivable (and NewQF does indeed derive it).

Our conclusion from this exercise is two-fold: First, that NewQF is performing as per its design. Second, that simple deductive reasoning is sufficient to obtain the same performance in biology as obtained during the 2009 final evaluation.

## **6. Summary**

The upgraded question formulation facility in AURA allows use of paraphrases, some initial use of underspecification, and some exploratory use of full (rather than controlled) language interpretation. Our evaluation suggests that these features significantly help in interpreting questions stated in a form close to the original English, and also points to a number of additional ways it can be further improved. Question answering itself consists of simple, straightforward calls to a deductive reasoning engine (KM). Our tests with the Minimal Test Suite indicate it is behaving correctly, and our tests with the Final Evaluation Q1 CPL data suggest that a simple deductive engine is sufficient to obtain the same performance in biology obtained during the 2009 final evaluation.

## **References**

- [1] Clark, P. The Semantics of Questions. Working Note 36, Boeing Research and Technology, 2010.
- [2] Clark, P. A Study of some "Hard to Formulate" Biology questions. Working Note 33, Boeing Research and Technology, 2010.

### Table 1: Results on the “Difficult Questions”

= correct answer with old QF  
 = correct answer with new QF

Qn No	Need kb addition	need BLUE	need under-specification	need para-phrases	other	
2a	kb1	y	y	y		A prokaryotic cell has a semi-fluid region consisting of cytosol.
2c				y	cpl1	A prokaryotic cell has a plasma membrane consisting of lipids and proteins.
2d		y				A prokaryotic cell has ribosomes that synthesize proteins.
2e		y				A prokaryotic cell has a rigid cell wall.
5abc-1	kb2		y	y		Peroxisomes can catalyze reactions involving hydrogen peroxide.
5abc-2	kb5	y		y		Mitochondria can provide cellular energy.
5a						Ribosomes can make proteins.
9a-1			y	y		Adenine is found in RNA molecules.
9a-2			y	y		Adenine is found in DNA molecules.
9c			y	y		Thymine is found in DNA molecules.
9d			y	y		Deoxyribose is found in DNA molecules.
9e			y	y		Ribose is found in RNA molecules.
37a		y				A eukaryotic cell has a membrane-bound nucleus.
37c-1						A prokaryotic cell has ribosomes.
37c-2						A eukaryotic cell has ribosomes.
37e						A prokaryotic cell has a nucleoid.
43a			y			The function of ribosomes is to synthesize proteins.
43b				y		The function of a nucleolus is to produce ribosomes.
43e			y			The function of chloroplasts is to photosynthesize.
53d		y			cpl2	During synapsis, segments of chromatids are exchanged during crossing over.
56d		y		y	cpl1	The cell membrane is composed mainly of lipids and proteins.
68a					out1	A DNA adenine can bond to a DNA thymine.
68c					out1	A DNA guanine can bond to a RNA cytosine.
68d					out1	A DNA adenine can bond to a RNA uracil.
68e					out1	A DNA guanine can bond to a DNA cytosine.
73c	kb12	y			cpl5,qa2	A DNA molecule consists of two antiparallel strands of nucleotides held together by hydrogen bonding.
74c	kb13,lex3	y	y			In DNA replication, the DNA polymerase builds a new strand from the 5 prime end to the 3 prime end.
77c-1	kb3	y			cpl8,out2	During mitosis of plant cells, a cell plate begins to form at telophase.
77c-2	kb3	y			out2	During mitosis of animal cells, a cleavage furrow is initiated at telophase.
79a				y		Ribosomes are associated with protein synthesis.
79b	kb6			y	qa1	Golgi bodies are associated with protein synthesis.
79c				y		The nucleus is associated with protein synthesis.
79d	kb7			y		The rough endoplasmic reticulum is associated with protein synthesis.
80b	kb4,lex1	y			cpl7	Lysosomes break down cellular debris.
81c	kb8	y			cpl6	A cell wall is a semi-rigid structure that supports a cell.
85a	lex2					Nucleoli reappear during the telophase of mitosis.
89b		y		y	cpl3	The nucleus is closely associated with the transcription activity of RNA.
91a		y			cpl4,qa2	During meiosis there are two rounds of prophase.
91b		y			cpl4,qa2	During meiosis there are two rounds of metaphase.
91c		y			cpl4,qa2	During meiosis there are two rounds of anaphase.
91d		y			cpl4,qa2	During meiosis there are two rounds of telophase.
103b					gen1	The cell membrane of a cell is in contact with the environment.
104a	kb9		y			Chloroplasts contain photosynthetic pigments.
106a	kb10		y			Ribosomes synthesize proteins in the cytoplasm.
121e	kb11				cpl8	The synaptonemal complex forms during prophase I of meiosis.
122d					cpl2	Homologous chromosomes separate during anaphase I.

TOTALS                      15    18    12    16                      21

## Keys from Table 1:

### Out of Scope Questions:

- out1** - This question is out of scope, as it is an "Is it possible..." rather than "Is it true...". We could search the KB for examples (as the "partial answer" method does) of the answer options, but even then this specific question is hard, as there are no examples of a DNA base bonded with a RNA base in the KB!
- out2** - Answering 77c requires temporal reasoning, as cell plate / cleavage furrow formation occur in an event [namely, cytokinesis] that is (partially) cotemporal with, rather than part of, mitosis and telophase. AURA doesn't support representation and reasoning with such temporal info.

### General Knowledge Requirements:

**gen1** - requires general world knowledge about space and the environment

### KB additions made to the Reference KB:

- kb1** - fixed cytoplasm encloses cytosol (was wrong way round in KB)
- kb2** - added peroxisomes do catalysis on hydrogen peroxide producing oxygen and water.  
In Peroxisome, change  
Catalyze-object→Hydrogen-Peroxide  
to Catalyze-enables→Synthesis, Synthesis (for the 2 synthesis events)  
Create concept of a Reaction (<| Action).  
In Synthesis, add Reaction as a 2nd superclass.
- kb3** - added plant-cell-mitosis and animal-cell-mitosis to kb.
- kb4** - In Lysosome, Insert concept Debris (a subclass of Physical-Object) and then unify it with (drag it on top of) Subcellular-Entity. This makes it a 2nd class for the Subcellular-Entity that the Lysosome takes apart.
- kb5** - added knowledge that ATP possesses cellular energy
- kb6** - need to represent the process of "protein synthesis and secretion", involving both protein synthesis and transport to Golgi bodies.
- kb7** - In Rough-endoplasmic-reticulum, specialize Synthesis (of proteins) to Protein-synthesis
- kb8** - Added cell-wall supports cell
- kb9** - Added photosynthetic pigment as a part of a chloroplast
- kb10** - Added that ribosomal synthesis of proteins is \*in the cytoplasm\*
- kb11** - Added Synaptonemal-complex is the result of Synapsis (in Synaptonemal-complex)  
In Synapsis, add superconcept Create (a subclass of Action) as a 2nd superclass of Synapsis. Then in Synapsis: Close then re-expand Synapsis, then specialize the result→Entity to Tetrad. add a second result→Synaptonemal-complex (<| Organic-Structure <| Physical-Object). This is to be consistent with the use of Synapsis in the concept Synaptonemal-Complex.
- kb12** - Add parallel/antiparallel spatial relations to AURA.
- kb13** - In DNA replication, KB incorrectly has DNA polymerase as the "by means of" DNA elongation. Delete the by-means-of link and replace it with an agent link.  
Similarly add a link between DNA-Elongation and Replication-Fork (e.g., instrument)  
In DNA replication, Add Second-DNA as the object of DNA elongation.  
In Polynucleotide, add Strand as an additional superclass.

In CLib, add Create as a superclass of Construction. [need to capture the Create ("build") aspect of Construction (DNA elongation)].

**lex1** - in CLib, added "break down" as synonym of Take-Apart

**lex2** - in CLib, added "reappear" as synonym of Create

**lex3** - in CLib, add "build" as synonym for Construction [not needed with kb13 CLib change, but should be done anyway]

### **Question-Answering Failures due to CPL/BLUE limitations:**

**cpl1** - unable to properly handle conjunction "and".

**cpl2** - "during" taken as subevent, but should be "equal"

**cpl3** - "activity" here needs to map to Event, not Activity

**cpl4** - Won't process non-initial "there are" properly. (But "There are two rounds of prophase during meiosis." is correctly handled)

**cpl5** - CPL won't process "together" constructions ("two strands are held together by hydrogen bonding" Should produce something like  
[Strand] ←object- Hydrogen-Bond -base→ [Strand])

Also see kb12.

**cpl6** - "wall supports a cell" → (\_CellWall0 supports \_Cell0) rather than a Support event in the KB. (CPL has misinterpreted "supports" as |supports| (an event-event causal relation). Note that NewQF will answer the variant "Is it true that a cell wall is a support of a cell?".

**cpl7** - Parser needs to prefer declared particle verbs (eg "break down") over normal verbs (eg "break"). See lex1 and q80b.

**cpl8** - "form" mapped to Organizing, but should map to Create

### **Question-Answering Failures due to reasoning limitations:**

**qa1** - need to look for a partial answer ("sometimes"), namely in the context of protein-synthesis-and-secretion (see kb6).

**qa2** - NewQF won't deal with set cardinality or count objects ("Are there two prophase events?")