LogicSolver - Solving Logic Grid Puzzles with Part-of-Speech Tagging and First-Order Logic

Ross Nordstrom
University of Colorado - Colorado Springs
1420 Austin Bluffs Pkwy,
Colorado Springs, CO 80918
rnordstr@uccs.edu

Abstract

This paper leverages natural language processing and first order predicate logic to programmatically solve logic grid puzzles, a constraint-based class of word problems. Solving logic grid puzzles has two main challenges: parsing natural language clues for semantic meaning and applying constraints against a model of the puzzle to solve it. We present *LogicSolver*, a tool advancing existing research in the area by introducing a more scrutinous evaluation process and automatically solving logic grid puzzles with minimal puzzle-specific semantics.

1 Introduction

This proposal seeks to implement a tool that can accurately and consistently solve logic grid puzzles, such as those provided by Puzzle Baron [3]. Logic puzzles are constraint-based word problems which present the player with a number of fact statements towards identifying how several entities relate to each other.

The driving intuition behind our selection of logic grid puzzles is that the language used in the puzzle clues is highly structured and task-oriented, and should therefore simplify parsing it for use in programatically solving the puzzle.

Looking towards broader applications, we find that the language and purpose of logic puzzles is representative of the growing field of language-controlled devices. This trend is clearly evident in the current trends of commercial technology. From smartphone voice-control such as Apple's Siri, Google's Assistant, and Amazon's Echo to home automation such as controlling your lights, music, and thermostat.

A general-purpose tool for solving logic puzzles should be extensible for use in a growing set of everyday and commercial tasks.

1.1 Logic Grid Puzzles

The logic grid puzzle class of word problems likely finds its roots in the Einstein puzzle (also known as the Zebra puzzle), first published by the Life International Magazine in 1962 [1]. The puzzle is a series of facts summarized with two questions, whose interpretation was one of the larger challenges Milicevic, et al. faced[10]. A key insight is that answering those questions is simply an extension work on solving the puzzle's constraint space – once you know all the relationships between entities in the puzzle, it takes only incremental work to answer questions about those relationships. For this reason, our work focuses on logic grid puzzles as can be found in printed and online form, such as those produced by Puzzle Baron's Logic Puzzles [3]. In contrast to the Einstein puzzle, logic grid puzzles have a more explicit puzzle space and assume a wider variety of forms including a larger variation in entities, sentence structure and tense.

Logic grid puzzles are presented to the user with three main components: Preamble providing context about the puzzle, Entities – in the form of the grid visualization – describing the subjects of the clues whose relationships we need to find, and Clues providing definitions of constraints on the entities' relationships. In general, the preamble is not necessary or helpful in solving the puzzle, but helps to explain how logic grid puzzles work to new players. We have observed that entities in logic grid puzzles have some implicit contraints, such as mutual-uniqueness in their names to avoid ambiguity, and the tendancy to include a quantifiable entity type such as time, size, or length.

Conventionally, logic grid puzzles are visualized with intersecting grids representing the possible relationships between entities of the different types. A typical approach for human puzzle solvers is to work through the clues, iteratively applying constraints. As they identify impossible relationships, they mark it with a "✓"; as they identify known relationships, they mark it with an "O" or a "✓" and "✗" out the remaining relationships in the row and column of the known one. For humans, it is relatively easy to parse the clues for semantic meaning, but more difficult to use that meaning to solve the puzzle; for computers, it is the reverse.

In our experience solving a variety of puzzles, we have noticed a handful of core clue structures. Below is a unseen-by-our-tool "moderate" logic grid puzzle, which we both parse and solve with 100% accuracy with *LogicSolver*. This example puzzle was selected because is representative of a wide variety of the clue structures we have come across.

Preamble

Howie, a professional nature guide, is taking a group of campers on a one-week backpacking trip through the Gros Ventre Wilderness in Wyoming. Using only the clues below, help Howie match each camper to their bag by determining the size (in liters), manufacturer and color of each.

Entities

- **Names:** Gene, Jeffrey, Leroy, Olga
- **Brands:** Bistric, Grennel, Pinkster, Travelore
- **Pack Size (liters):** 25, 30, 35, 40

Clues as Provided

1. The Pinkster pack is 5 liters smaller than the Grennel pack.
2. Olga's pack is 25 liters.
3. The 30 liter pack, the Bistric pack and the Grennel pack are all different packs.
4. Of Leroy's pack and the Travelore pack, one is 25 liters and the other is 30 liters.
5. Gene's pack is 35 liters.

Clues after FOP Parse

1. smaller(Pinkster, Grennel, 5 liters)
2. is(Olga, 25)
3. not(30, Bistric, Grennel)
4. is(xor(Travelore, Leroy), xor(25, 30))
5. is(Gene, 35)

Solution

Name	Brand	Pack Size
Olga	Travelore	25 liter
Leroy	Pinkster	30 liter
Gene	Grennel	35 liter
Jeffrey	Bistric	40 liter

Table 1: The solution to the example logic grid puzzle

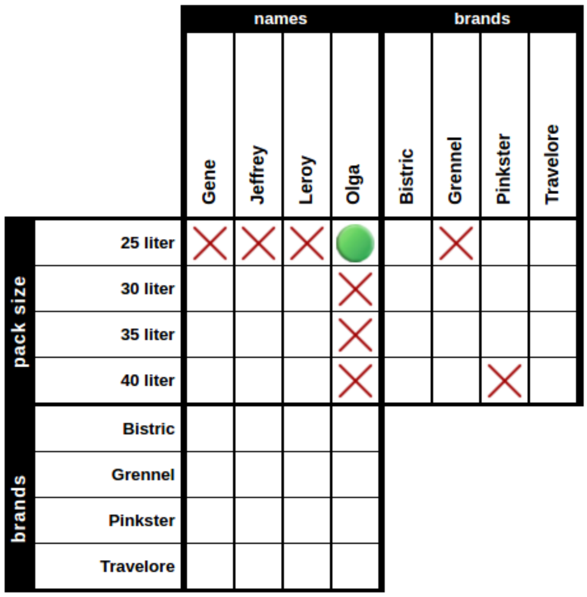

Figure 1: The example puzzle after applying some initial constraints from Clues 1 and 2. Clue 1 informs us that Pinkster cannot be the largest pack and Grennel cannot be the smallest. Clue 2 informs us of a known relationship: Olga's pack == 25 liters.

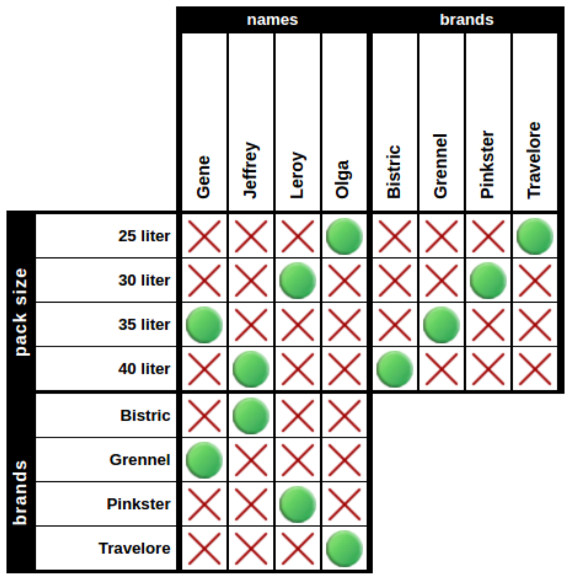

Figure 2: The example logic grid puzzle after solving. "✓"s mark mappings which cannot be, while "O"s mark matches.

In our example puzzle figures, we show a puzzle containing simple relationships (Clues 2 and 5), comparisons (Clue 1), mutual exclusivity (Clue 3), and conditional exclusivity (Clue 4). For motivating purposes, we've also highlighted the first-order-predicate statements our tool produces after parsing the clues. In our observations, these clues collectively represent all the clue structures we have observed. In practice, the other main variation we have seen in puzzle clues is the use of varying sentence tenses. While different clues may share the same structure, the use of – for example – past passive or future tense can produce very different part-of-speech parses, and was one of the main hurdles we faced in our efforts to produce a general-purpose parser.

We will discuss the variations in clue structure and tense, and how we handle them in our Design 4 and Implementation 5 sections. Additionally, we will explain and justify the first-order-predicate syntax we presented in our example puzzle. However, before touching on either of those topics, it is necessary to differentiate our efforts from previous works.

2 Related Work

There are a number of other works that focus on parsing meaning – and sometimes solving – word problems. In the area of logic puzzles, there has been some highly relevant work by Lev, et al. [7] and Milicevi, et al. [10].

In this section we will discuss the more general works done on the area of solving word problems and analyze more closely-related logic puzzle work. While the logic puzzle works are relevant and accomplished in their own right, we intend to convince the audience of some shortcomings in their approaches and evaluation. Improving upon these shortcomings is one of the main goals of our work in *LogicSolver*.

2.1 Word Problems

Before researching works in the logic puzzles space, it is useful to start more broadly in the word problems space. Word Problems are arguably a superset of logic puzzles, and have therefore been more extensively researched.

In the space of word problems, we highlight two works which focus on using natural language processing to generically solve word problems. Their approaches reflect the approaches we've seen and use for logic puzzles in that the core approach is to (1) symbolically parse natural language, (2) apply pattern matching and transformations while parsing to achieve semantic meaning, and (3) solve the problem with general solutions and inferences.

2.1.1 Number Word Problems - Shi, et al.

Shi, et al. [12] used a Context Free Grammar (CFG) parser to automatically solve number-based word problems. Their work discusses two main categories of approaches: symbolic and statistical learning. They cite two major shortcoming of symbolic approaches as a tendency to rely on pattern matching and ad-hoc transformations, as well as ambiguous evaluation criteria. The dependence on pattern matching and ad-hoc transformations hurts the applicability of the solutions to general problems. The lack of clear evaluation criteria in existing works makes it difficult to determine whether or not their results were authentically accurate, or just superficially accurate through overfitting against their test dataset.

While the solution and approach that Shi, et al. took is not entirely relevant to our work, their notes about the shortcomings of existing efforts echoes our main criticisms of Milicevic, et al.'s work [10].

2.1.2 Arithmetic Word Problems - Roy, et al.

Similarly, Roy and Roth's work to solve general aithmetic word problems [11], does not overwhelm with relevance to our solution; however, their work cites very similar motivations and general approach.

The appraoch Roy and Roth took consists of a few main elements: (1) break down the problem into smaller, more manageable tasks, (2) compose these smaller pieces together in order to understand the complex sentence in whole, and (3) apply constrained inferences to partition the text into applicable and inapplicable information.

Some constrained inferences they used include encoding knowledge about special words like "amount" and "how many," and which arithmetic operations or spaces to use for each.

2.2 Logic Puzzles

2.2.1 Solving Logic Puzzles - Lev, et al.

Lev, et al. [7] provide a strong basis upon which we designed our approach to solving logic puzzles. While their work focuses on solving "question-and-answer" problems like those found in the Law School Admission Test (LSAT) and the Graduate Record Exam (GRE), the structure of their problems is similar to the logic puzzles we solve in our work.

As is the theme in solving word and logic puzzles, Lev et al. used a processing pipeline containing parsing modules, followed by solving modules. Like our work, they were unable to avoid encoding puzzle-specific knowledge in their parsing and solving modules.

Much like Shi, et al. observed in the word problem space, we find the evaluation of this work to be somewhat ambiguous and narrow. They cite a 94% accuracy training dataset of 259 sentences and testing dataset of 46 unseen sentences. However, they follow that result with an undermining statement implying we can't fully rely on the results of the aforementioned 46 sentences.

> "A preliminary evaluation of another unseen puzzle shows that on 60% of the sentences, the parser's output is accurate enough to support the subsequent compputation of the semantics, and we expect this to be better after it is trained on puzzle texts."

To their credit, they refer to their work as a prototype, and highlight potential extensions of it in their Future Work section.

2.2.2 Puzzler - Milicevic, et al.

The work done by Milicevic, et al. in the tool they call *Puzzler* is the main reference of our work. Their approach was the typical pipeline of multi-step parsing and then solving the puzzle. We think it is worth reiterating their approach here since it is so closely relevant to what we present with *LogicSolver*. This section will conclude with the main points on which we differ from *Puzzler*, and will highlight our main criticisms with *Puzzler*, which we aim to improve upon.

The design approach Milicevic, et al. took was to: (1) parse the clue statements via a structured part-of-speech parser, Sleator et al.'s Link Grammar [13] – described in further detail in 3.1 – (2) apply pattern matching to identify the key components of each clue, (3) disambiguate the parsed components using wordnet for consistency in order to (4) semantically parse the clues into a set of constraints, and (5) find the full solution to the problem space by running those constraints through a constraints-based modeling language, Alloy [5].

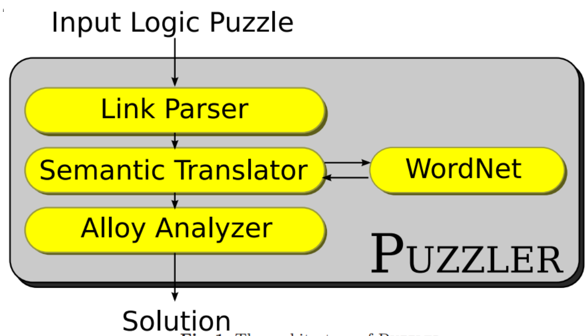

Fig. 1. The architecture of PUZZLER

Figure 3: *Puzzler*'s architecture, as shown in Fig. 1 of their paper [10]

Since we agree with the general approach *Puzzler* took – and since they had promising results – we have followed a similar approach, with the main exception of developing our own solution-finding program instead of using Alloy, which is a language Milicevic built. This choice was made primarily for the sake of personal development, but also under the assumption that it would be easier to apply the implicit logic grid puzzle constraints.

Another point of departure in our works is that *Puzzler* focuses on solving the Einstein puzzle as representative of other logic puzzles. The Einstein puzzle is arguably the foundation of modern logic puzzles, including the logic grid puzzles we present. Structurally, the puzzle is similar to the puzzles in our dataset; however, as a single puzzle, it lacks the variety and quantity we have collected. Unlike logic grid puzzles, the Einstein puzzle only implicitly presents the problem space, requiring more inferences than we need for logic grid puzzles. Additionally, the Einstein puzzle doesn't technically require a full solution, as it merely asks for the answers to two questions. Implicitly, one is to understand that the answer to the questions requires the full solution to the problem space. The full puzzle, as published by Life Magazine in 1962, is shown below. [1]

```
1. There are five houses.
2. The Englishman lives in the red house.
3. The Spaniard owns the dog.
4. Coffee is drunk in the green house.
5. The Ukrainian drinks tea.
6. The green house is immediately to the right
     of the ivory house.
7. The Old Gold smoker owns snails.
8. Kools are smoked in the yellow house.
9. Milk is drunk in the middle house.
10. The Norwegian lives in the first house.
11. The man who smokes Chesterfields lives in
    the house next to the man with the fox.
12. Kools are smoked in the house next to the
    house where the horse is kept.
13. The Lucky Strike smoker drinks orange
    juice.
14. The Japanese smokes Parliaments.
15. The Norwegian lives next to the blue house
    .

Now, who drinks water? Who owns the zebra?

In the interest of clarity, it must be added
    that each of the five houses is painted a
    different color, and their inhabitants are
     of different national extractions, own
    different pets, drink different beverages
    and smoke different brands of American
    cigarets [sic]. One other thing: in
    statement 6, right means your right.

- Life International, December 17, 1962
```

Interestingly, you can see the author of the puzzle explicitly calls out the problem space's constraints. This is something that is commonly left implicit in modern logic grid puzzles, and, in fact, is something we have encoded into our solution. Another notable attribute of the puzzle is that the questions' subjects "water" and "zebra" are never actually referenced in the puzzle's clues. Again, this commonly appears in logic grid puzzles as entities which are absent from the puzzle's clues, but present in the puzzle's grid.

The main point in which we have differed and improved upon the work in *Puzzler* is in our evaluation of *LogicSolver*. Milicevic et al. evaluated *Puzzler* against three puzzles: (1) the original Eintsein puzzle, (2) a modified (less ambiguous) Einstein puzzle, and (3) a randomly generated puzzle in the same puzzle. In our brief analysis of the puz-

Description	Penn type	Example
adjective phrase	ADJP	
adverb phrase	ADVP	
sentence clause	S	(**S** (NP (NP The commie) , (SBAR (WHNP who) (**S** (VP was (ADJP red)))) ,) (VP faded (PRT away)) .)
inverted subject-verb clause	SINV	
noun phrase	NP	
particle	PRT	(S (NP (NP The commie) , (SBAR (WHNP who) (S (VP was (ADJP red)))) ,) (VP faded (**PRT** away)) .)
prepositional phrase	PP	S (NP There) (VP were (NP (NP three) (**PP** of (NP them)))) .)
quantifier phrase	QP	
root clause	S	
subordinate clause	SBAR	(S (NP (NP The commie) , (**SBAR** (WHNP who) (S (VP was (ADJP red)))) ,) (VP faded (PRT away)) .)
verb phrase	VP	
wh-adverb	WHADVP	(S (NP The bridge) (VP was (ADJP unfinished)) (SBAR (**WHADVP** when) (S (NP it) (VP collapsed))) .)
wh-noun	WHNP	(S (NP (NP The commie) , (SBAR (**WHNP** who) (S (VP was (ADJP red)))) ,) (VP faded (PRT away)) .)
wh-prep	WHPP	(S (NP The commie (SBAR (**WHPP** to (WHNP whom)) (S (NP I) (VP was))) speaking) (VP was (ADJP red)) .)

Table 2: POS Tagging: Descriptions of the types of parse tags produced by the Logic Puzzle statement parser. (Taken from OpenCog's wiki [2])

zle generator they used, mensus.net [9], it struck us that the generated puzzles are very similar to the Einstein puzzle, and arguably represent a subset of the clue-types we find in the original Einstein puzzle. Again, we find Shi, et al.'s observation about the shortcomings of the evaluations of word problems are applicable to *Puzzler*.

With this in mind, the major emphasis of our work in *LogicSolver* is to robustly evaluate our tool, while following the established best practices for parsing and solving word problems, and ultimately striving towards a general solution while acknowledging the need for some puzzle-specific semantics and inferences to be encoded into the tool to successfully and consistently solve logic grid puzzles.

3 Background

Before presenting our solution, it is worth providing some background information about the driving concepts behind our work: part of speech tagging and first order predicate logic.

3.1 Part of Speech Tagging

Part of speech (POS) tagging is a common method for understanding human language in the fields of natural language processing and linguistics. Jurafsky et al. describe the value of POS tagging succinctly: [6]

> "The significance of parts-of-speech (also known as POS, word classes, morphological classes, or lexical tags) for language processing is the large amount of information they give about a word and its neighbors."

Part of speech tagging is a well understood space, with a number of resources available for tagging arbitrary sentences, trained on large datasets. A popular tagset is the Penn Treebank library [8], which contains 48 POS tags.

Our work utilizes the Link Grammar Parser [14], which inspects relationships (links) between words to more insightfully parse sentences. Link Grammar's approach to parsing is visually summarized in figures 5 6. The implementation we use is a continually-updated version of the original Link Grammar, maintained by OpenCog [15]. OpenCog's current version uses a dictionary of roughly 600,000 words with 117 link types. Since our task relies more on the word tags than the linkages, we utilize an output of the Link Grammar parser which assigns Treebank-like tags to each word in the sentence. We show a list of tags the Link Grammar Parser produces in table 2.

For our goal of identifying the semantic words in logic grid puzzle clues, POS tagging is a powerful technique allowing us to partition important from unimportant words in puzzle clues. The strategy we use for interpretting POS tag parses will be discussed in the Design section 4.2.2.

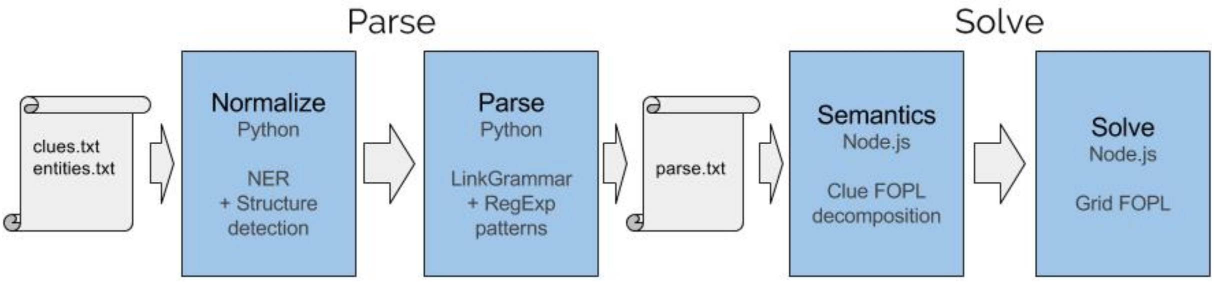

Figure 4: *LogicSolvers*'s architecture

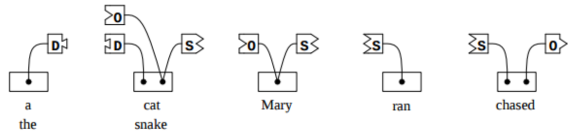

Figure 5: Link Grammar's approach to parsing assigns allowed linkages to each word. (Taken from Sleator et al.'s paper [14])

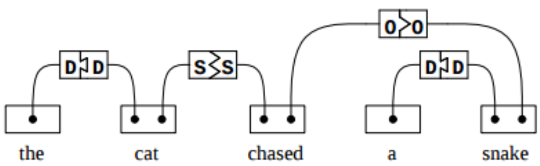

Figure 6: Link Grammar then connects as many words as possible *without* allowing links to cross. There are often several valid linkages, and thus many sentences can be interpretted in multiple ways. (Taken from Sleator et al.'s paper [14])

3.2 First-Order Logic

First-order logic (FOL), or First-order predicate logic (FOP/FOPL), is a method of modelling and reasoning about statements. It can be used to represent constraints and relationships in a consistent and simple form, allowing for unambiguous evaluation. Jurafsky et al. describe the value of First-order logic: [6]

> "[A]n attractive feature of FOL is that it makes very few specific commitments as to how things ought to be represented. ...[T]he specific commitments it does make are ones that are fairly easy to live with and that are shared by many [other schemes of meaning representation]; the represented world consists of objects, properties of objects, and relations among objects."

FOL, as used in the context of our work, can be summarized has containing three critical components: (1) *constants* defining known entities, (2) *functions* defining constraints and relationships, and (3) *quantifiers* augmenting the relationships. Some examples of FOL statements (clauses) are shown in figure 7.

```
is(Olga, 25)
smaller(Pinkster, Grennel, 5 liters)
is(xor(Travelore, Leroy), xor(25, 30))
```

Figure 7: FOL: Example predicate statements. The second statement demonstrates a function (*smaller*) relating two entity constants with a quantifier (*5 liters*). Note that the third statement is actually a second-order predicate, since it composes two first-order predicates. We will still refer to such statements as FOL statements for simplicity.

We describe the specific FOL language we use in *LogicSolver* in the Design section 4.3.

4 Design

Similar to previous works, we approached the problem with a general-purpose solution in mind, but made some concessions to special-purpose designs to accomplish the task of parsing and solving logic grid puzzles. As such, our overarching design – shown in figure 4 – resembles the conventional pipeline we have observed in previous works:
(1) transcribe the puzzle as input to the system,
(2) parse the clue statements into POS tag strings,
(3) translate the parsed statements into a specialized FOPL language model using puzzle-specific logis, and (4) solve the puzzle using a custom FOPL processor with built-in puzzle semantics.

4.1 Transcribing the Puzzle

To enable the program to work, we require a manual step to transcribe the puzzle into something consumable by the program. The transcribed data represents the knowledge a human would easily gather when presented with a puzzle. The transcriptions currently stored in text files with newline and comma separation between

Type	Predicate Structure	Description
comparison relationship	COMP(E1, E2 [,Q])	E1 is [Q] COMP than E2. Q is optional
equality	is(E1, E2, ...)	E1 and E2 are related
inequality	not(E1, E2, ...)	E1, E2, ... are unrelated
mutually-exlusive	is(E1, xor(E2, E3))	E1 is either E2 or E3
double mutex	is(xor(E1, E2), xor(E3, E4))	Of E1, E2; one is E3, other is E4

Table 3: Descriptions of the types of parse statements produced by the Logic Puzzle statement parser.

inputs. The core inputs are the puzzle's clue statements (`clues.txt`) and a definition of the entities (`entities.txt`), as presented by the grid visualization. During training, we include two extra files describing the expected FOPL statements (`parseExpected.txt`) and the solution to the puzzle (`answers.txt`). Examples of these files are shown in the appendix **??**.

The entities are stored in two sections: the first line defines the entity types, comma-separated. Subsequently, there is a line per entity type, listing each entity within that type, listed in the same order as is shown in the Logic Grid. The ordering is used later on for infer the meaning of comparisons.

The justification for this manual step is that it simply provides the program the same contextual information a human has when solving a logic grid puzzle. Automatically scraping the contents of a logic grid puzzle application, such as Puzzle Baron's [3], is outside the scope of our work. Additionally, it is always implied in these puzzles that each entity of a given entity type can only be matched to one entity of each other entity type. This implied logic is built into the program as a necessary feature to solving the puzzles.

4.2 Parsing the Statements

The intent of the parsing step is to convert a clue statement into a first-order predicate structure representing the relationships and constraints defined in the statement. It will then be up the semantics module 4.3 to apply the meaning of the predicates. An example of a clue and its associated FOPL parse is shown here.

"The kayaking trip was 2 years after the camping vacation."

$\rightarrow after(kayaking, camping, 2years)$

Before delving into more detail about parsing logic puzzle clues, it is important to highlight some intuitions and observations about the structure of logic grid puzzle clues. First, all clues contain at least two entities. The basic structre of clues is to define known relationships between two entities

"Olga's pack is 25 liters."

or comparisons between two entities,

"The Pinkster pack is 5 liters smaller than the Grennel pack."

We have observed that most clues with only two entities follow one of those two structures. Clues with more than two entities are typically "either or" statements,

"Of Leroy's pack and the Travelore pack, one is 25 liters and the other is 30 liters,"

or describe anti-relationships between all of the entities,

"The 30 liter pack, the Bistric pack and the Grennel pack are all different packs."

Comparison clues with quantifiers ("*35 liters* smaller") often contextualize the quanitifier with an entity type or a unit for the entity type. E.g. "liters" is a unit of the entity type "Pack Size".

These rules help us understand how to parse the clue statements, using a combination of named-entity recognition and POS tag pattern matching. By applying a first pass with NER, we can decide which structure the clue most likely is, and then apply the appropriate ruleset on the parsed POS tags for the clue to identify any comparisons or quantifiers.

4.2.1 FOP Structure in *LogicSolver*

While the parsing step doesn't provide any means for programmatically interpretting or understanding the FOPL statements, we can describe the intention of the syntax. The "function" defines the relationship, and the arguments describe the subjects of the relationship. We currently define four classes of operators: comparison relationships, statements of equality, statements of inequality, and statements of mutually-exclusive (mutex) relationships. These types are described in table 3.

If the relationship is a comparison between two entities, the first argument describes the actor of the comparison, the second argument describes the actee of the comparison, and the optional third argument describes a comparison qualifier. For example, if the predicate was `before(Independence Day, July 6, 2 days)`, then we know "Independence Day" is equal to "July 6" minus "2 days," or in other words is "July 4." If the qualifier were omitted, such as `before(my birthday,` Christmas), all we can infer is that "my birthday" has a date less than "Christmas" ()December 25).

Equality statements are the simplest ones, simply stating "X is Y," or "The thing with X and the thing with Y are the same." An example is the clue statement *"Gene's pack is 35 liters."*

Similarly, inequality statements simply describe anti-relationships. They behave in a manner opposite to equality statements, stating "X is not Y," or "The thing with X and the thing with Y are different." An example is the clue statement *"The 30 liter pack, the Bistric pack and the Grennel pack are all different packs."*

Conditional-exclusive relationships typically replace a definitive entity in an equality statement, saying "X is either Y or Z." An example of this is clue statement

> "The vacation with Dustin is either the 2004 holiday or the hang gliding holiday."

When parsed, this statement can be repesented with the predicate `is(Dustin, xor(hang gliding, 2004))`. It is important to note that some complex clue statements can include two conditional-exclusive relationships in a single equality, of the form "Either A or B is C or D," which we represent with the predicate `is(xor(A, B), xor(C, D))`.

4.2.2 Parsing Structure with Link Grammar

Sleator and Temperley's Link Grammar [13] is used for the first pass of parsing the structure of the clue statements. Since it focuses on the relationship between words in a sentences (represented by links between), it aligns well with our goal of parsing out the relationships between known entities in our clue statements. Additionally, the Python implementation of Link Grammar used in this project is pre-trained on the English language, helping to avoid the need for a substantial training suite of games.

Upon parsing a sentence, the Python Link Grammar library can produce a linkage diagram, as shown in Figure 8, as well as a breakdown of the constituent parts of the sentence, including their parts of speech, as shown in Figure 9. The constituent parts are offered both in a structure hierarchy, as well as a flat structure. In this project, the flat constituent parts are used to extract out the parts of speech of the clue statement. They are then converted into a string, and compared against several Regexes to identify which type of relationship is described by the statement.

4.2.3 Parsing Syntax with Assisted-NER

Given the type of the clue statement has been identified, we can parse the syntax of it using custom logic, driven by observations on the structures of

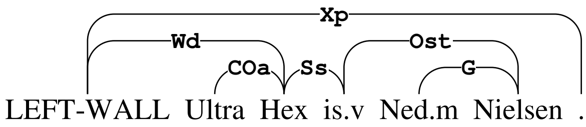

Figure 8: An example of a Link Grammar linkage diagram describing an example Logic Puzzle clue statement.

```
<S: .>
<NP: Ultra, Hex>
<VP: is>
<NP: Ned, Nielsen>
```

Figure 9: An example of Link Grammar constituent parts describing an example Logic Puzzle clue statement.

each type of statement. The custom syntax parsing used in this project combines knowledge about the parts of speech from Link Grammar to identify the relationships and qualifiers, and custom-made named entity recognition (NER) logic to parse the entities being operated on.

Equality and inequality statements are the simplest to parse. For these statements , we already know the statement type, and thus the operator (`is(...)` and `not(...)`, respectively); therefore, we can simply search for all known entities in the statement to find the arguments for the operator.

Relationship statements are more challenging to parse. The parsing logic must adapt to a number of variables, such as the tense of the statement and the part of speech of the entities (e.g. proper nouns like "Ultra Hex" versus adjectives like "orange" in "the orange shirt"). This project relies on the custom logic to identify the relationship operator and qualifier based on their part of speech, and then looks for a single entity on each side of the operator using the custom NER mentioned previously.

Either or (or "exclusive or") statements are parsed by locating "either" in the statement, then using custom NER to find the subject entities.

4.3 Semantics via First-Order Predicate Logic

Once parsed, we can use semantics to understand the relationship between the Entities. The parsed predicate statements act as constraints. *Puzzler* [10] used a modelling language and tool, Alloy [5], to solve the Logic Puzzles. For variation, semantics are implemented in this work through First-Order Predicate logic combined with applied set theory enforce the grid and statement constraints.

This section describes each category of First-Order Predicates (FOP) found in Logic Grid Puz-

$$\forall a \text{ in } A, \text{ discard if } \nexists b \text{ in } B | a > b$$
$$\forall b \text{ in } B, \text{ discard if } \nexists a \text{ in } A | a > b$$

Figure 10: General comparison for "less(A, B)"

$$\forall a \text{ in } A, \text{ discard if } \nexists b \text{ in } B | b = a + x$$
$$\forall b \text{ in } B, \text{ discard if } \nexists a \text{ in } A | a = b + x$$

Figure 11: Offset comparison for "less(A, B, x)"

zles, along with the approach used to apply the predicates. In general, the approach taken is to store potential relationships on each entity, and then modify those sets as each FOP constraint is enforced.

4.3.1 FOP: IS

The "IS" constraint simply means each entity present in the predicate's arguments are equivalent to each other. In terms of their sets of potential relationships, we simply remove all relationships except for the other arguments.

4.3.2 FOP: NOT

The "NOT" constraint simply means each entity in the predicate's arguments must be removed from each other argument's list of potential relationships.

4.3.3 FOP: XOR

The "XOR" constraint can arrive in three forms: "`A is XOR(B,C)`", "`XOR(B,C) is A`", and "`XOR(A,B) is XOR(C,D)`". The first two get standardized to `XOR(A, B,C)`, while the second form can be decomposed to `NOT(A,B), XOR(A, C,D), XOR(B, C,D)`.'

The application of the XOR contraint relies on conditionally introducing new predicates: if `is(A,B)` then apply `not(A,C)`, else if `not(A,B)` then apply `is(A,C)`. If neither condition is satisfied yet, then nothing can be done until more constraints have been applied.

4.3.4 FOP: Comparisons

The rest of the FOP statements found in logic puzzles can be considered "comparison" constraints, which describe a difference between two entities. The comparison can exist in both a general form and specific form with an offset. Below, the formula is described for each form. Comparisons occur as either a "less than" or a "greater than" comparison. Switching between the two is as simple as swapping the order of their arguments.

While comparisons can have many terms (shorter, smaller, less, ...), they all can be interpreted as "lesser comparisons," and can be paired to a complement (taller, bigger, more, ...).

An interesting challenge with comparisons is to understand which entity type they refer to. For example, consider a puzzle with the entity types Person, House, and Cost. If a person saw the statement "Bob paid more than John," they would intuitively understand that the context of "more" is the Cost entity type. However, a machine would need to be taught how to make inferences like this, or use a different means to learn the context of comparisons.

It turns out most logic puzzles will have at least one comparison between disparate types. If our puzzle has only three entity types, a comparison between two types tells us the context of the comparison is the third type.

An additional technique to identify a comparison's context is to simply look for numeric entities, and assume those are the context for comparisons.

5 Implementation

Our work was done in two parts: a parser module written in *Python*, and a solver module written in *Node.js*. The two modules share a text-file based interface, driven from the command line for simplicity. The modules were developed separately to maintain a better separation of concerns. Each should be able to operate independently of the other. This abstraction helps ensure each module could individually be extended or combined with other tools.

For extensibility, the parser and solver modules were both designed to take in directories of puzzles as input, making it easy to test the tool on many puzzles at once. The files used throughout the pipeline are defined here 5.1.1.

`entities.txt`: The entity types and entities. Entity types are comma-separated on the first line, followed by an empty line, and then a comma-separated line containing the entities for each type.

`clues.txt`: The clue statements driving the logic grid puzzle, separated by `newline`'s.

`parseActual.txt`: The actual parse result, modelling the clues as FOPL statements. This acts as the interface between the parser and solver.

`parseExpected.txt`: (Testing) The expected parse result of the clue statements, used to evaluate the parser.

`answers.txt`: (Testing) The solution to the puzzle, used to evalute the solver.

5.1 Parser

The parser module follows up to three steps to process each puzzle clue: (1) use named-entity recognition (NER) to decide if it is a special structure, and if not (2) use Link Grammar to identify POS tags, and (3) use pattern matching on the POS tags to identify the comparison, entities, and quantifier. In certain cases, we reprocess clues we fail to parse using NER to replace tricky entity names with simpler-to-parse words.

5.1.1 Special Cases with NER

As was discussed in the design section 4.2, when there are more than two entities in the clue, we have observed that the clue structure can only be a mutually-exclusive or inequality statement. For this reason, the first task the parser does for each clue is to scan and identify all referenced entities in the statement. A simple count tells us if we're dealing with an XOR or NOT statement. If this is the case, we have found that the simplest and most consistent means of parsing the statement is to look for keywords. Our "decision tree" is straight forward:

If the clue starts with "Of" and has four entities, it must be a double XOR statement

If the clue contains "either" and "or", it must be a simple XOR statement

Otherwise, if the clue has more than two entities, it must be a NOT statement.

Based on these simple conditions, we can easily construct the desired FOPL statement using the scanned entities. Here, we show an example of each condition:

"Natasha was either the person who danced second or the dancer who performed first."

$\rightarrow is(Natasha, xor(second, first))$

"Of Leroy's pack and the Travelore pack, one is 25 liters and the other is 30 liters."

$\rightarrow is(xor(Travelore, Leroy), xor(25, 30))$

"The 30 liter pack, the Bistric pack and the Grennel pack are all different packs."

$\rightarrow not(30, Bistric, Grennel)$

5.1.2 Pattern matching with POS tags

The remaining majority segment of clue statements contain two entities, and can either be equality statements, general comparisons without quantifiers, or specific comparisons with quantifiers. The biggest challenge in this portion of the parsing is handling the various tenses the clue sentences can be phrased.

In our training dataset, we identified four main categories of sentense tenses: simple statements, present tense, past passive tense, and future tense. For each category, we defined a couple Regular Expressions (regex) for identifying where the entities, comparisons, and quantifiers exist in the clue. These regexes are applied to each sentense's POS tagset. For each match, we do some validation to ensure it's a reasonable match. In practice, we use a simple searching technique, applying all regexes to each statement, taking the most detailed match. E.g. if there is a match describing a comparison and a quantifier, take that; otherwise take the first match describing a comparison; if there are not matches, assume it's an equality ("IS") statement.

One large concession we made to achieve better accuracy, was to train the parser with a list of known comparator strings. This is effectively a hard-coded list of strings which the parser uses when validating the regex matches it finds. If a potential comparison isn't defined in this list, we assume the statement is an equality statement. This is the main area future work should focus on improving.

5.2 Solver

The solver module, written in *Node.js*, consists of two main steps to model the puzzle state and iteratively apply constraints to the puzzle state, converging on a single solution. Our approach relies on modelling entities as objects with a set of potential relationships with each entity of the other entity types. As constraints are applied, we remove entities from the list of potential relationships, until only one valid entity-relationship exists between a given entity and each of the other types.

Four core predicates are used in the solver to represent all possible FOPL functions: is, not, xor, and lesserCompare. As discussed in the design section 4.3, `is` and `not` are simple predicates to apply. The `xor` predicate is nearly as simple to apply, and is a conditional application of the `is` or `not` predicates. Comparison predicates are more challenging to apply.

5.2.1 Applying Comparison Constraints

The first challenge in applying comparison constraints is identifying if the comparison (e.g. "smaller" or "after") is a "lesser" or "greater" comparison. We resolve this by defining known comparisons as one or the other. When applying greater comparisons, we simply reuse the lesser-Comparison predicate, but swap the order of the entities. E.g. `larger(Grennel, Pinkster)` is equivalent to `smaller(Pinkster, Grennel)`.

The larger difficulty in applying comparison constraints is inferring the "comparison context," or in other words, the entity type the comparison applies to. The comparison context tells the solver which relationships to compare between the two entity arguments. E.g. the predicate `smaller(Pinkster, Grennel)` doesn't directly tell you it refers to the

"Pack Size" entity type. A human can easily infer that from the types "Names," "Brands," and "Pack Size;" but a computer cannot make such a leap of intuition. To resolve this in our solver, we rely on two mechanisms: comparison context inferences and known contexts.

Context inferences are applied when a comparison refers to all but one of the entity types. In our observations, comparisons never apply to an entity type of one of the compared entities. This let's us infer that the comparison's context applies to the remaining entity type. An example of this would be "Jeffrey's pack is larger than the Grennel pack." (`larger(Jeffrey, Grennel)`): in this case, the only entity type not referenced is "Pack Size," and therefore is the comparison context.

Known contexts are a mapping of known comparisons to entity types. For example, we trained the solver that the comparison "taller" must apply to an entity type of "height" or "distance." A comparison "older" can apply to entity types "age", "month," "date," "time," or "year."

Our hard-coding of lesser vs greater comparisons and comparison contexts is the second main opportunity for improvement in future work, along with eliminating our hard-coded list of known comparisons in the parser.

6 Results

With the permission of Puzzle Baron [3], we used a dataset of over 60 generated puzzles with roughly 300 clues to evaluate *LogicSolver*. The puzzles were evenly split across Easy, Moderate, and Challenging 3x4 logic grid puzzles. Since the parser and solver modules operate independently, our evaluation considers each individually with the following criteria:

parsingAccuracy: `(# of statements correctly parsed) / (# total statements)`

solutionAccuracy: `(# of correctly identified relationships) / (# total relationships)`

To populate our dataset, we simply started new puzzles of the appropriate difficulty on Puzzle Baron's web site and transcribed the puzzle entities and clues into text files. Additionally, we then solved the puzzles (abusing hints for speed) for the solution and manually recorded the expected FOPL parses for each clue. The tediousness of the transcription process prevented us from collecting more puzzles for evaluation.

Overall, we accomplished decent results of 74% (221/297) parser accuracy and 78% (213/272) solver accuracy. Table 4 breaks down our results by puzzle difficulty.

	Accuracy	
Difficulty	**Parser**	**Solver**
Easy	77% (96/124)	83% (93/112)
Moderate	71% (62/ 87)	90% (72/ 80)
Challenging	73% (63/ 86)	75% (60/ 80)
All	74% (221/297)	83% (225/272)

Table 4: Accuracy/our parser and solver by puzzle difficulty

In general, the weakest area of the parser was its ability to identify and correctly parse comparison clues. Across all the puzzles, 81% of incorrect parses involved comparison clues which was parsed incorrectly as equality clues. We present a breakdown in table 5

Difficulty	Incorrect	Accuracy
Easy	82% (23/28)	77% (96/124)
Moderate	84% (21/25)	71% (62/ 87)
Challenging	75% (18/24)	73% (63/ 86)
All	81% (62/77)	74% (221/297)

Table 5: Portion of incorrect parses from incorrectly parsing comparison clues as equality clues, compared with the overall parsing accuracy.

We believe our success in correctly parsing non-comparison clues is because of the consistent structure found in "xor" and "not" statements. Using named-entity recognition to find known entities is a trivial task, and allowed for overwhelming success when parsing clues where we only had to identify entities. Our lower success rate on parsing comparison clues stems from how varied those sentences were, using a number of different sentence structures to mean the same thing.

Our failures in solving the puzzles were due to problems converging on a solution, rather than issues of selecting incorrect relationships. This is likely due to our puzzle heuristics not being as rich as they could be.

7 Future Work

We presently describe some areas where the work could be applied, and use these applications to focus the efforts of future work.

7.1 Applications

With a more generalized pipeline and parsing, our work could be applied broadly to any product or task which needs to process human language for intended actions or constraints. Parsing statements into an FOPL structure works excellently as an input for computer processing. Some examples of such processes include search engines and question-answering ("Ok Google, who's taller between Michael Jordan and an ostrich?"), voice control for devices ("Turns the lights on, but dimmer than last time"), and constraint-based logistics like planning seats at a wedding ("Couples must sit next to each, but don't seat John next to Uncle Bob").

7.2 Parser

The largest weakness of our parser is its inability to identify comparisons (e.g. "larger," "younger," "after") without hard-coded knowledge of them. One possibility for improving our parser is the use of the Stanford Parser [4], either replacing or augmenting our use of *LinkGrammar*. Their parser has explicit "comparative" and "superlative" POS tags, which could simplify the task of parsing comparison statemetns. Additionally, for most general-purpose uses we envisoin, our parser would need to be modified so it doesn't depend on a static list of known entities that could be involved in the parsed statements.

7.3 Solver

If we consider the "solver" to more generally be a "processor" of language statements to execute a task, there is less value in a general-purpose solution, as different applications will have different needs. Therefore, there is less usefulness in improving or generalizing our solver module. The main value in our work is the concept of parsing human language into the easy-to-consume format of FOPL statements. Applications will be more effective with specialized processors, which can consume FOPL statements like those used in our work.

8 Conclusion

In this paper we have presented *LogicSolver*, a system built upon a parser and solver, capable of automatically solving logic grid puzzles with reasonable accuracy. The emphasis of our work was establishing an example for more robust evaluation, with repeatable experimentation, while maintaining a general-purpose approach where possible. The intent of our work is to further existing works in the area of solving word problems and logic puzzles, with eventual commercial applications.

References

[1] Life international: Who owns the zebra? December 1962.

[2] Constituent tree - opencog. `http://wiki.opencog.org/w/Constituent_tree`, 2008. [Online; accessed 26-November-2016].

[3] Puzzle Baron. Logic puzzles, Sep 2015.

[4] Marie-Catherine de Marneffe and Christopher D. Manning. The stanford typed dependencies representation. In *Coling 2008: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation*, CrossParser '08, pages 1–8, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[5] Daniel Jackson. *Software Abstractions: logic, language, and analysis.* MIT press, 2012.

[6] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 2009.

[7] Iddo Lev, Bill MacCartney, Christopher D. Manning, and Roger Levy. Solving logic puzzles: From robust processing to precise semantics. In *Proceedings of the 2Nd Workshop on Text Meaning and Interpretation*, TextMean '04, pages 9–16, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

[8] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of english: The penn treebank. *COMPUTATIONAL LINGUISTICS*, 19(2):313–330, 1993.

[9] mensus.net. logic puzzles, Sep 2016.

[10] Aleksandar Milicevic, Joseph P Near, and Rishabh Singh. Puzzler: An automated logic puzzle solver.

[11] Subhro Roy and Dan Roth. Solving general arithmetic word problems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1743–1752, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[12] Shuming Shi, Yuehui Wang, Chin-Yew Lin, Xiaojiang Liu, and Yong Rui. Automatically solving number word problems by semantic parsing and reasoning. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1132–1142, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[13] Daniel DK Sleator and Davy Temperley. Parsing english with a link grammar. *arXiv preprint cmp-lg/9508004*, 1995.

[14] Daniel DK Sleator and Davy Temperley. Parsing english with a link grammar. *arXiv preprint cmp-lg/9508004*, 1995.

[15] Daniel DK Sleator and Davy Temperley. Link grammar parser. `http://www.abisource.`

com/projects/link-grammar/, 2014. [Online; accessed 26-November-2016].

9 Appendix

In this appendix, we demonstrate the actual results from using *LogicSolver* against our dataset, separated between Parsing and Solving, as well as by puzzle difficulty.

9.1 Parser: Successful Example

Here, the files and command-line output are shown for successfully parsing an example logic grid puzzle.

9.1.1 Entities file

The first line describes the entity *types*. The subsequent lines describe the entities for each of the entity types. Each entity row is ordered to match the entity types, and the entities are ordered as shown in the puzzle, which provides comparison indexes. E.g. we don't need to understand the semantics of entities "22 ft", "25 ft", "28 ft", "31 ft" since they are sorted in the entity definition.

```
names, owners, lengths

Adagio III, Irish Pride, Restless Wave, White Squid
Ann Alvarado, Betsy Becker, Ernesto Ellis, Hetta Hensley
22 ft, 25 ft, 28 ft, 31 ft
```

9.1.2 Clue Statements file

```
The Adagio III is 3 feet shorter than the Irish Pride.
The 25 ft vessel is owned by Betsy Becker.
The White Squid is either the 25 ft boat or the 31 ft boat.
Ann Alvarado's sailboat is somewhat shorter than Betsy Becker's vessel.
Of the Irish Pride and the 28 ft vessel, one is owned by Ernesto Ellis and the other is owned by Bet
```

9.1.3 Expected Parse file

```
shorter(Adagio III, Irish Pride, 3)
is(25 ft, Betsy Becker)
is(White Squid, xor(25 ft, 31 ft))
shorter(Ann Alvarado, Betsy Becker)
is(xor(Irish Pride, 28 ft), xor(Ernesto Ellis, Betsy Becker))
```

9.1.4 Actual Parse file

This is the output our parser actually produces. On successful parses, it should be identical to the Expected Parse file.

```
shorter(Adagio III, Irish Pride, 3)
is(25 ft, Betsy Becker)
is(White Squid, xor(25 ft, 31 ft))
shorter(Ann Alvarado, Betsy Becker)
is(xor(Irish Pride, 28 ft), xor(Ernesto Ellis, Betsy Becker))
```

9.1.5 Answers file

```
Adagio III, Ann Alvarado, 22 ft
Irish Pride, Betsy Becker, 25 ft
Restless Wave, Ernesto Ellis, 28 ft
White Squid, Hetta Hensley, 31 ft
```

9.1.6 Result output

When run, the parsing portion of *LogicSolver* prints the following on the example puzzle.

```
1. The Adagio III is 3 feet shorter than the Irish Pride.
2. The 25 ft vessel is owned by Betsy Becker.
3. The White Squid is either the 25 ft boat or the 31 ft boat.
4. Ann Alvarado's sailboat is somewhat shorter than Betsy Becker's vessel.
5. Of the Irish Pride and the 28 ft vessel, one is owned by Ernesto Ellis and the other is owned
     by Betsy Becker.

## SUCCESS
100% success - 5 of 5 total statements
```

9.2 Parser: Failure Example

Here, the files and command-line output are shown for unsuccessfully parsing an example logic grid puzzle.

9.2.1 Entities file

```
planets, stars, distances

Cheelia, Ereph, Llyrak, Xidat
BDF 198, HV 491, MKP 427, VJD 913
41 light years, 43 light years, 45 light years, 47 light years
```

9.2.2 Clue Statements file

```
The exoplanet orbiting star BDF 198 is 4 light years closer to us than Xidat.
Ereph is 2 light years closer to us than the planet orbiting star HV 491.
The planet orbiting star BDF 198 is 2 light years closer to us than Llyrak.
Of Xidat and Cheelia, one orbits star MKP 427 and the other is 45 light years from earth.
```

9.2.3 Expected Parse file

```
closer(BDF 198, Xidat, 4)
closer(Ereph, HV 491, 2)
closer(BDF 168, Llyrak, 2)
not(Xidat, Cheelia, MKP 427, 45 light years)
```

9.2.4 Answers file

```
Ereph, BDF 198, 41 light years
Llyrak, HV 491, 43 light years
Xidat, VJD 913, 45 light years
Cheelia, MKP 427, 47 light years
```

9.2.5 Result output

When run, the parsing portion of *LogicSolver* prints the following:

```
1. The exoplanet orbiting star BDF 198 is 4 light years closer to us than Xidat.
2. Ereph is 2 light years closer to us than the planet orbiting star HV 491.
x  is(Ereph, HV 491)   (Actual)
   closer(Ereph, HV 491, 2)   (Expected)
3. The planet orbiting star BDF 198 is 2 light years closer to us than Llyrak.
x  closer(BDF 198, Llyrak, 2) (Actual)
   closer(BDF 168, Llyrak, 2) (Expected)
4. Of Xidat and Cheelia, one orbits star MKP 427 and the other is 45 light years from earth.
x  is(xor(Xidat, Cheelia), xor(MKP 427, 45 light years))   (Actual)
   not(Xidat, Cheelia, MKP 427, 45 light years) (Expected)

## FAILURE
25% success - 1 of 4 total statements
```

9.3 Parser: All Results

Here we display the results of parsing all 28 Easy, 20 Moderate, and 20 Challenging puzzles in our dataset. Clues which were incorrectly parsed are marked with an "x," and show both the Actual (incorrect) parse and the Expected (correct) parse.

9.3.1 Easy Puzzles

```
../data/puzzlesEasy/puzzle01
1. The album that came out in June sold 200000 more copies than Sodium Green's record.
x  is(June, Sodium Green)   (Actual)
   more(June, Sodium Green, 200000)  (Expected)
2. Jagged Knee's album was released in September.
3. Jagged Knee's release sold somewhat more copies than Sodium Green's release.
4. The release that came out in August sold 100000 fewer copies than the album that came out in Janu
x  is(August, January)   (Actual)
   fewer(August, January, 100000)    (Expected)
```

5. Audio Array's album is either the record that came out in September or the release with 1.5 milli

../data/puzzlesEasy/puzzle02
1. Stanley will leave sometime before the person departing from Leland.
2. Chris will depart from Leland.
3. The person departing from Daly City will leave 2 days after Kyle.
4. The person departing from Jackman will leave 1 day before Kyle.

../data/puzzlesEasy/puzzle03
1. Vera was picked up at the Moore Mansion.
2. The customer who got picked up at the Space Needle paid 6 dollars less than the client who got pi
x is(Space Needle, Arctic Building)   (Actual)
   less(Space Needle, Arctic Building, 6 dollars)   (Expected)
3. The person who got picked up at the Space Needle paid 3 dollars more than Kate.
x is(Space Needle, Kate)   (Actual)
   more(Space Needle, Kate, 3 dollars) (Expected)
4. Max paid $15.50.

../data/puzzlesEasy/puzzle04
1. The Tela G5 costs 25 dollars less than the model made by Nectron.
2. The camera made by Lirios is either the $625 camera or the DM-5000.
3. The camera made by Lirios costs 25 dollars less than the model made by Torvia.
4. The four cameras are the G-290, the $625 camera, the $550 camera and the model made by Lirios.

../data/puzzlesEasy/puzzle05
1. The four videos are the viral video with 5 million views, "Beach Brouhaha", Rodney's viral video
2. Tonya's viral video is either the viral video with 4 million views or "Happy Hermits".
3. The video with 7 million views is "Happy Hermits".
4. Daryl's viral video is "You Otter Know".
5. Daryl's viral video has 1 million more views than "Beach Brouhaha".
x is(Daryl, Beach Brouhaha)   (Actual)
   more(Daryl, Beach Brouhaha, 1 million) (Expected)

../data/puzzlesEasy/puzzle06
1. The Z-Free XL has the 20000 BTUs cooling capacity.
2. The unit with the 25000 BTUs cooling capacity is either the Chiller Z or the $375 item.
3. The $570 unit, the air conditioner with the 20000 BTUs cooling capacity and the Chiller Z are all
4. The Binson C40 costs $440.
5. The air conditioner with the 10000 BTUs cooling capacity costs somewhat less than the air conditi
x is(10000 BTUs, 25000 BTUs) (Actual)
   less(10000 BTUs, 25000 BTUs)   (Expected)

../data/puzzlesEasy/puzzle07
1. The contestant who went second dove 6 meters deeper than the diver who went seventh.
x is(second, seventh)   (Actual)
   deeper(second, seventh, 6 meters)   (Expected)
2. Brit Bradley went to a maximum depth of 107 meters.
3. Hill Harvey went to a maximum depth of 104 meters.
4. The four divers were Gumbo Giles, the diver who went seventh, the person who reached a depth of 1

../data/puzzlesEasy/puzzle08
1. Ed Ewing finished 1000 votes ahead of the teacher.
2. The person who received 9000 votes was either the writer or Jed Jarvis.
3. Jed Jarvis is the teacher.
4. The person who received 9500 votes was either the doctor or Jed Jarvis.
5. Kelly Kirby finished 500 votes behind the doctor.

../data/puzzlesEasy/puzzle09

```
1. The dog Teresa trained had a training period that was somewhat longer than that of Jake.
x  is(Teresa, Jake)   (Actual)
   longer(Teresa, Jake)  (Expected)
2. Jake trained 1 week more than Gizmo.
3. Buster was in classes for 8 weeks.
4. The dog that was in school for 9 weeks was trained by Martha.
5. The one that was in school for 6 weeks is either the canine Martha trained or the one Odette trai

../data/puzzlesEasy/puzzle10
1. Preston is either the Texas native or the centenarian who is 109 years old.
2. The four people are the centenarian who is 112 years old, the Alaska native, Ernesto and Preston.
3. Ernesto is a native of Texas.
4. Olivia is 1 year older than the Virginia native.

../data/puzzlesEasy/puzzle11
1. Mt. Dawson is 260 feet shorter than the mountain first summited by Kermit Kramer.
x  is(Mt. Dawson, Kermit Kramer)  (Actual)
   shorter(Mt. Dawon, Kermit Kramer, 260 feet)   (Expected)
2. Mt. Norman is 14470 feet tall.
3. Mt. Ingram was first climbed by Gerry Gamble.
4. The peak that is 14210 feet tall is either Mt. Larsen or the peak first summited by Irving Igor.
5. The peak first summited by Kermit Kramer is shorter than Mt. Ingram.
x  is(Kermit Kramer, Mt. Ingram)  (Actual)
   shorter(Kermit Kramer, Mt. Ingram)   (Expected)

../data/puzzlesEasy/puzzle12
1. The person whose assignment begins at 10:00am will be working with Orlando.
2. Charlie will start sometime after the person who will be working with Rafael.
x  is(Charlie, Rafael)   (Actual)
   after(Charlie, Rafael)   (Expected)
3. Joann will start 2 hours before Annie.
4. Joann will be working with Karla.

../data/puzzlesEasy/puzzle13
1. First Crescent is run by Kip Kerr.
2. Citizen Trust is either the one with a return of 7% or the fund run by Kip Kerr.
3. The fund run by Ule Underwood has a 2 percent larger return than Hudson Brook.
x  larger(Ule Underwood, Hudson Brook, 2 return)    (Actual)
   larger(Ule Underwood, Hudson Brook, 2 percent)   (Expected)
4. The fund with a return of 7% is run by Walter Wade.

../data/puzzlesEasy/puzzle14
1. Istryn, the exoplanet orbiting star HV 491 and the planet orbiting star PLC 120 are all different
2. Vestor is either the exoplanet 47 light years from earth or the planet orbiting star BDF 198.
3. Istryn is closer to us than the exoplanet orbiting star MKP 427.
x  is(Istryn, MKP 427)   (Actual)
   closer(Istryn, MKP 427)  (Expected)
4. Vestor orbits star HV 491.
5. The exoplanet orbiting star PLC 120 is 2 light years closer to us than Rodul.
5 [NER]. The exoplanet orbiting star PLC 120 is 2 light years closer to us than Rodul.
   closer(PLC 120, Rodul, 2 light years)   (Expected)

../data/puzzlesEasy/puzzle15
1. Penny Perry's reservation is for 3 days.
2. Vicky Vaughan's reservation is either the 4 day rental or the Kia reservation.
3. Penny Perry's reservation is for a period 1 day longer than the Jeep rental.
x  is(Penny Perry, Jeep)    (Actual)
   longer(Penny Perry, Jeep, 1 day))    (Expected)
```

```
4. Rose Rowland's rental is for the Kia.
5. Penny Perry's reservation is either the 4 day rental or the Buick rental.

../data/puzzlesEasy/puzzle16
1. The 1225 sq ft rental rents for $1525/month.
1 [NER]. The 1225 sq ft rental rents for $1525/month.
   is(1225 sq ft, $1525/month)    (Expected)
2. The home in La Palma offers 1100 sq ft of living space.
3. The four houses are the home that rents for $1100/month, the rental in La Palma, the home in Mont
4. The home that rents for $825/month is 250 sq ft larger than the house in Eagle Grove.
4 [NER]. The home that rents for $825/month is 250 sq ft larger than the house in Eagle Grove.
   larger($825/month, Eagle Grove, 250 sq ft)    (Expected)

../data/puzzlesEasy/puzzle17
1. The Foltron is either the rocket developed by Omnipax or the rocket developed by Vexatech.
2. The rocket that will launch in February is either the Exatris or the rocket developed by Omnipax.
3. The four rockets are the rocket that will launch in January, the Foltron, the rocket developed by
4. The rocket that will launch in March is made by Vexatech.
5. The Athios is made by SpaceZen.

../data/puzzlesEasy/puzzle18
1. The ostrich that finished second was #128.
2. The ostrich that finished first was either #118 or #126.
3. The runner that finished third was either #126 or Bridget.
4. Ophelia finished second.
5. Stretch finished 2 places after Kermit.

../data/puzzlesEasy/puzzle19
1. The silver coin sold for somewhat more than the item found at Colbert Run.
x  is(silver coin, Colbert Run)   (Actual)
   more(silver coin, Colbert Run)    (Expected)
2. The cannonball sold for 75 dollars more than the piece found at Heffen Lane.
x  is(Cannonball, Heffen Lane)    (Actual)
   more(cannonball, Heffen Lane, 75 dollars)  (Expected)
3. The silver coin sold for 150 dollars less than the diamond ring.
4. The object that sold for $400 was found at Burr Woods.

../data/puzzlesEasy/puzzle20
1. Lonnie's purchase was either the whip spider or the arachnid that sold for $60.
2. Olga's purchase was either the raft spider or the arachnid that sold for $60.
3. Anthony's purchase cost 15 dollars less than the wolf spider.
4. The raft spider cost 30 dollars less than Lonnie's purchase.

../data/puzzlesEasy/puzzle21
1. The alfalfa honey is produced by Midge Mintz.
2. The lemon blossom product costs 2 dollars less than Linda Lynn's honey.
3. Midge Mintz's honey costs 1 dollar less than the lemon blossom product.
4. The sage product is produced by Keith Koch.

../data/puzzlesEasy/puzzle22
1. Bladescape was played in March.
2. The four games were the board game played in April, the game hosted at Ira's house, Planets Align
3. Vendetta was either the game hosted at Ira's house or the game played in February.
4. The game played in January was hosted by Lynette.

../data/puzzlesEasy/puzzle23
1. Merlin is somewhat older than Bubba.
2. Merlin is 10.2 feet long.
```

```
3. Bubba is 10 years older than the gator that is 14.0 feet long.
x  not(Bubba, 10, 14.0 feet)   (Actual)
   older(Bubba, 14.0 feet, 10)    (Expected)
4. Barnabas is 12.0 feet long.

../data/puzzlesEasy/puzzle24
1. Luther scored 49 points.
2. Bertha is either the person who scored 56 points or the person who made the smoked pork.
3. Hattie scored 7 fewer points than the person who made the turkey soup.
x  is(Hattie, turkey soup)  (Actual)
   fewer(Hattie, turkey soup, 7)  (Expected)
4. The chef who scored 42 points made the cashew tofu.
5. The person who made the cajun chili scored 7 fewer points than Bertha.

../data/puzzlesEasy/puzzle25
1. Peter Peck was either the person sent out on April 4th or the person who covered the earthquake.
2. Randy Reid left 2 days after the person who covered the flooding.
x  is(Randy Reid, flooding)    (Actual)
   after(Randy Reid, flooding, 2 days)  (Expected)
3. Peter Peck covered the royal wedding.
4. Sid Snyder covered the mass protests.

../data/puzzlesEasy/puzzle26
1. The recipe that cooks at 325 degrees requires 10 minutes less baking time than the meatloaf.
x  is(325 degrees, meatloaf)   (Actual)
   less(325 degrees, meatload, 10 minutes)    (Expected)
2. The chicken cutlet requires 5 minutes less baking time than the candied yams recipe.
x  is(chicken cutlet, candied yams)  (Actual)
   less(chicken cutlet, candied yams, 5 minutes)    (Expected)
3. The mac and cheese requires 5 minutes less baking time than the dish that cooks at 340 degrees.
x  is(mac and cheese, 340 degrees)   (Actual)
   less(mac and cheese, 340 degrees, 5 minutes)  (Expected)
4. The recipe that cooks at 325 degrees bakes somewhat longer than the recipe that cooks at 410 degr
x  is(325 degrees, 410 degrees)   (Actual)
   longer(325 degrees, 410 degrees)  (Expected)

../data/puzzlesEasy/puzzle27
1. Bai Yun was born in January.
2. Chi Chi was Wang Yu's baby.
3. Ling Ling was born sometime after Tian Tian's baby.
4. Po Lang's baby was born sometime after Ling Ling.
x  is(Po Lang, Ling Ling)   (Actual)
   after(Po Lang, Ling Ling)   (Expected)
5. Wang Yu's baby was born 1 month after Yuan Zai's baby.
x  is(Wang Yu, Yuan Zai)    (Actual)
   after(Wang Yu, Yuan Zai, 1 month)    (Expected)

../data/puzzlesEasy/puzzle28
1. "Kip and Ken" was due 2 weeks before Sherrie's book.
2. The book due on September 15 was "Ohio Haunts".
3. "Stars Below" was due 2 weeks after Wayne's book.
4. Latasha's book was due on September 15.


## DECENT
77% success -  96 of 124 total statements
```

9.3.2 Moderate Puzzles

```
../data/puzzlesModerate/puzzle01
1. The Adagio III is 3 feet shorter than the Irish Pride.
2. The 25 ft vessel is owned by Betsy Becker.
3. The White Squid is either the 25 ft boat or the 31 ft boat.
4. Ann Alvarado's sailboat is somewhat shorter than Betsy Becker's vessel.
5. Of the Irish Pride and the 28 ft vessel, one is owned by Ernesto Ellis and the other is owned by

../data/puzzlesModerate/puzzle02
1. William will make 15000 dollars less than the actor filming on September 8th.
2. William is either the man making $115000 or the person filming on September 1st.
3. The actor filming on September 1st will make 15000 dollars more than Jim.
4. The person making $85000 will film on September 15th.
5. Bruce will film on September 29th.

../data/puzzlesModerate/puzzle03
1. North Bridge is the girder bridge.
2. Gorem Bridge is 3000 feet longer than the girder bridge.
3. Of Avery Bridge and North Bridge, one is the suspension bridge and the other is 8500 ft.
4. Gorem Bridge is 1500 feet shorter than Bay Bridge.
5. The 11500 ft bridge is either the truss structure or Bay Bridge.

../data/puzzlesModerate/puzzle04
1. The traveler flying from Qatar is leaving 1 hour later than Terrance.
x  is(Qatar, Terrance)   (Actual)
   later(Qatar, Terrance, 1)   (Expected)
2. Claude is leaving 1 hour later than Jeanne.
x  is(Claude, Jeanne)    (Actual)
   later(Claude, Jeanne, 1)    (Expected)
3. The person leaving at 3:30 pm is either the traveler flying from Jamaica or the person flying fro
x  is(3:30 pm, xor(Jamaica, Peru))   (Actual)
   is(3:30 pm, xor(Jamaice, Peru))   (Expected)
4. Pablo is flying from Peru.

../data/puzzlesModerate/puzzle05
1. The four animals are the 14-year-old tortoise, the pitch belly tortoise, the black neck tortoise
2. Of the 32-year-old one and the two rimmed tortoise, one is named Methuselah and the other is name
3. Toredo is 36 years younger than the pitch belly tortoise.

../data/puzzlesModerate/puzzle06
1. Of the 153 ft. tree and the 85 year-old tree, one is Roger's Oak and the other is Nolan's Pine.
x  not(153 ft, 85 year, Roger's Oak)    (Actual)
   is(xor(153 ft, 85 year), xor(Roger's Oak, Nolan's Pine))  (Expected)
2. Zeke's Spruce is 94 years old.
3. Evan's Fir is 6 feet shorter than Zeke's Spruce.
x  is(Evan's Fir, Zeke's Spruce)  (Actual)
   shorter(Evan's Fir, Zeke's Spruce, 6)   (Expected)
4. Evan's Fir is 3 feet shorter than Roger's Oak.
x  is(Evan's Fir, Roger's Oak)    (Actual)
   shorter(Evan's Fir, Roger's Oak, 3)  (Expected)
5. Evan's Fir is either the 94 year-old tree or the 79 year-old tree.

../data/puzzlesModerate/puzzle07
1. The 25 ft long whale is either the sei whale or Brent.
2. The sei whale is Herbert.
3. The fin whale is 25 ft long.
4. The sei whale is 2 feet longer than the beluga whale.
5. Of the 25 ft long specimen and the right whale, one is Adam and the other is Brent.
```

```
../data/puzzlesModerate/puzzle08
1. The Super Tubers used the catapult.
2. The team that made it to 135 ft was either the contestants that used the vacuum cannon or the tea
3. Of the group that made it to 150 ft and the group that made it to 165 ft, one was the Super Tuber
4. The Russets used the trebuchet.
5. The Super Tubers landed somewhat short of the Russets.
x  is(Super Tubers, Russets)   (Actual)
   short(Super Tubers, Russets)   (Expected)
6. The group that used the vacuum cannon was either the Mad Mashers or the contestants that made it

../data/puzzlesModerate/puzzle09
1. The exoplanet orbiting star BDF 198 is 4 light years closer to us than Xidat.
2. Ereph is 2 light years closer to us than the planet orbiting star HV 491.
x  is(Ereph, HV 491)  (Actual)
   closer(Ereph, HV 491, 2)    (Expected)
3. The planet orbiting star BDF 198 is 2 light years closer to us than Llyrak.
x  closer(BDF 198, Llyrak, 2)  (Actual)
   closer(BDF 168, Llyrak, 2)  (Expected)
4. Of Xidat and Cheelia, one orbits star MKP 427 and the other is 45 light years from earth.
x  is(xor(Xidat, Cheelia), xor(MKP 427, 45 light years))  (Actual)
   not(Xidat, Cheelia, MKP 427, 45 light years)  (Expected)

../data/puzzlesModerate/puzzle10
1. Joann will start sometime before the interpreter who will be working with Patrick.
x  is(Joann, Patrick)    (Actual)
   before(Joann, Patrick)   (Expected)
2. The person who will be working with Nelson will start 2 hours after the interpreter who will be w
x  is(Nelson, Patrick)   (Actual)
   after(Nelson, Patrick, 2)   (Expected)
3. Frederick will start work at 11:00am.
4. Of Ira and the person whose assignment begins at 9:00am, one will be working with Tyler and the o

../data/puzzlesModerate/puzzle11
1. Of Wade's computer and Jack's build, one has a 128 GB hard drive and the other has 6 GB of RAM.
2. Rex's build has a hard drive that is 256 GB larger than the build with 6 GB of RAM.
2 [NER]. Rex's build has a hard drive that is 256 GB larger than the build with 6 GB of RAM.
   larger(Rex, 6 GB, 256 GB)   (Expected)
3. Wade's system has a hard drive that is 128 GB larger than the system with 1 GB of RAM.
x  not(Wade, 128 GB, 1 GB)  (Actual)
   larger(Wade, 1 GB, 128 GB)  (Expected)
4. Rex's system has 3 GB of RAM.

../data/puzzlesModerate/puzzle12
1. The city in Berkshire County has 3000 fewer people than the town in Hampshire County.
x  is(Berkshire, Hampshire)    (Actual)
   fewer(Berkshire, Hampshire, 3000)    (Expected)
2. The city in Essex County has more people than Charles City.
3. Of the town with a population of 28000 and the city with a population of 37000, one is in Essex C
4. Quasqueton has 3000 fewer people than the city in Nantucket County.
x  is(Quasqueton, Nantucket)   (Actual)
   fewer(Quasqueton, Nantucket, 3000)   (Expected)

../data/puzzlesModerate/puzzle13
1. Wang Yu's baby was born 2 months before Den Ping.
x  is(Wang Yu, Den Ping)    (Actual)
   before(Wang Yu, Den Ping, 2 months)  (Expected)
2. Of Wang Yu's baby and Nan Sheng's baby, one was Hua Mei and the other was born in April.
```

```
3. Ling Ling was born sometime before Wang Yu's baby.
4. Tai Shan's baby was born 2 months after Po Lang's baby.

../data/puzzlesModerate/puzzle14
1. The four pandas were Bai Yun, the baby born in February, Tai Shan's baby and the baby born in Jan
2. Nan Sheng's baby was born 2 months before Gao Gao.
3. Of Den Ping and Bai Yun, one was born in February and the other was Tian Tian's baby.
4. Bai Yun was either the offspring born in March or the baby born in January.

../data/puzzlesModerate/puzzle15
1. Of the one discovered in 2008 and the one discovered in 2010, one is Sporrin and the other was fi
2. Casputi was discovered 1 year before the one Dean Courtis discovered.
x  is(Casputi, Dean Courtis)   (Actual)
   before(Casputi, Dean Courtis, 1 year)   (Expected)
3. The one discovered in 2011 is Gostroma.
4. The one Klein discovered was found in 2010.

../data/puzzlesModerate/puzzle16
1. Gail Garrett's project begins sometime after Al Anderson's project.
2. The job starting in May will be headed by Irma Ingram.
3. The York Court project is either the job starting in July or Al Anderson's project.
4. The Clara Street project starts in April.
5. Of the job starting in June and the job starting in July, one will focus on Norway Court and the

../data/puzzlesModerate/puzzle17
1. Mae started with the Reti Opening.
2. The player who started with the Giuoco Start played 1 game after Luis.
x  is(Giuoco Start, Luis)   (Actual)
   after(Giuoco Start, Luis, 1 game)    (Expected)
3. The person who played second was either Tara or the person who started with the Evans Gambit.
4. Rose played 1 game before the person who started with the Evans Gambit.
x  is(Rose, Evans Gambit)   (Actual)
   before(Rose, Evans Gambit, 1 game)   (Expected)

../data/puzzlesModerate/puzzle18
1. Charlie will start sometime after the person who will be working with Nelson.
x  is(Charlie, Nelson)   (Actual)
   after(Charlie, Nelson)   (Expected)
2. Of the person whose assignment begins at 10:00am and Frederick, one will be working with Nelson a
3. The interpreter whose assignment begins at 11:00am is either Ira or the interpreter who will be w
4. The person who will be working with Nelson will start 1 hour after Ira.
x  is(Nelson, Ira)    (Actual)
   after(Nelson, Ira)    (Expected)

../data/puzzlesModerate/puzzle19
1. "Disco Mania" came out sometime before the machine made by Pinnefore.
x  is(Disco Mania, Pinnefore)  (Actual)
   before(Disco Mania, Pinnefore)    (Expected)
2. Of the machine released in 1977 and the machine made by Tarco Inc., one is "Bowling Alley" and th
3. The machine made by Dow Games came out 1 year after "Bowling Alley".
x  is(Dow Games, Bowling Alley)   (Actual)
   after(Dow Games, Bowling Alley, 1 year)    (Expected)
4. The system made by Tarco Inc. came out sometime after the machine made by Pinnefore.
x  is(Tarco Inc, Pinnefore)    (Actual)
   after(Tarco Inc, Pinnefore)    (Expected)

../data/puzzlesModerate/puzzle20
1. Of Cate Carlson and Bev Baird, one ended up with 9000 votes and the other is the teacher.
```

2. The academic finished 1000 votes behind Gilda Gray.
3. The four candidates were Bev Baird, the person who received 9000 votes, the architect and the can
4. Bev Baird was either the person who received 9500 votes or the politician who received 8500 votes

DECENT
71% success - 62 of 87 total statements

9.3.3 Challenging Puzzles

```
../data/puzzlesChallenging/puzzle01
1. Oliver paid 1 dollar more than Irene.
2. The one who got the root beer paid 1 dollar more than Irene.
3. Salvador paid 1 dollar more than the one who got the iced tea.
x  is(Salvador, iced tea)   (Actual)
   more(Salvador, iced tea, 1)    (Expected)
4. Oliver paid less than the one who got the orange soda.
x  is(Oliver, orange soda)  (Actual)
   less(Oliver, orange soda)   (Expected)

../data/puzzlesChallenging/puzzle02
1. The journalist who covered the earthquake left 1 day after the journalist who covered the royal w
x  is(earthquake, royal wedding)  (Actual)
   after(earthquake, royal wedding, 1 day)    (Expected)
2. The journalist who covered the solar eclipse was either the journalist sent out on April 7th or N
3. Of the reporter sent out on April 7th and the reporter sent out on April 5th, one was Orin Olsen
4. Linda Lott left 2 days after Maddy Moore.

../data/puzzlesChallenging/puzzle03
1. Mt. Calhoun is 260 feet shorter than the peak first summited by Art Aguilar.
x  is(Mt. Calhoun, Art Aguilar)   (Actual)
   shorter(Mt. Calhoun, Art Aguilar, 260)  (Expected)
2. The mountain first summited by Chip Carson is taller than Mt. Sloan.
x  is(Chip Carson, Mt. Sloan)  (Actual)
   taller(Chip Carson, Mt. Sloan)    (Expected)
3. Mt. York is 260 feet taller than Mt. Calhoun.
4. Of the peak first summited by Eddie English and Mt. Larsen, one is 14600 feet tall and the other

../data/puzzlesChallenging/puzzle04
1. Of "Bowling Alley" and the game made by Waverly Toys, one was released in 1974 and the other was
2. "Archer Quest" came out 1 year after "Aladeen".
x  is(Archer Quest, Aladeen)   (Actual)
   after(Archer Quest, Aladeen, 1)   (Expected)
3. "Meteor Rise" is either the machine released in 1974 or the game made by Hayco Inc..
4. The system made by Pinnefore came out 1 year before "Aladeen".
x  is(Pinnefore, Aladeen)   (Actual)
   before(Pinnefore, Aladeen, 1)  (Expected)

../data/puzzlesChallenging/puzzle05
1. Of the campaign produced by Eddie Evans and the ad produced by Gilda Gamble, one is the direct ma
x  not(Eddie Evans, Gilda Gamble, 1075)    (Actual)
   is(xor(Eddie Evans, Gilda Gamble), xor(direct mailer, 1075))    (Expected)
2. The billboard, the ad with 775 responses and the ad produced by Gilda Gamble are all different ca
3. The radio spot has brought 150 fewer responses than the ad produced by Julie Jordan.
x  fewer(radio spot, Julie Jordan)   (Actual)
   fewer(radio spot, Julie Jordan, 150)    (Expected)
4. The tv commercial is either the campaign produced by Hal Hopkins or the campaign with 775 respons

../data/puzzlesChallenging/puzzle06
```

```
1. The Utah native is older than the Illinois native.
2. Shari is either the Illinois native or the South Dakota native.
3. The Georgia native is 1 year younger than the Illinois native.
4. Tim is younger than Noel.
5. Of Jason and Tim, one is 110 years old and the other is a native of Utah.

../data/puzzlesChallenging/puzzle07
1. Ollie has 6 fewer wins than the woman with 8 losses.
x  is(Ollie, 8 losses)   (Actual)
   fewer(Ollie, 8 losses, 6)   (Expected)
2. The four people are Diane DeNiro, the fighter with 33 wins, the woman with 2 losses and the boxer
3. Of the fighter with 7 losses and the boxer with 33 wins, one is Annie Axel and the other is Vicki
4. The boxer with 7 losses is either Vicki or the woman with 27 wins.

../data/puzzlesChallenging/puzzle08
1. Of the person who received the $35000 scholarship and the student who received the $40000 scholar
2. The student who will major in Theology was awarded $5000 less than the person who will major in P
x  is(Theology, Philosophy)    (Actual)
   less(Theology, Philosophy, $5000)    (Expected)
3. Martha was awarded $10000 more than the student who will major in Finance.
x  is(Martha, Finance)   (Actual)
   more(Martha, Finance, $10000)  (Expected)
4. The student who will major in Theology was awarded some amount more than Sarah.
x  is(Theology, Sarah)   (Actual)
   more(Theology, Sarah)    (Expected)
5. Delores will major in Finance.

../data/puzzlesChallenging/puzzle09
1. The person who works in the blue section served 25 fewer riders than Guy.
2. Brett is either the employee who served 100 riders or the person who works in the purple section.
3. The worker who served 125 riders works in the pink section.
4. Isaac served 50 more riders than the person who works in the purple section.
x  is(Isaac, purple)  (Actual)
   more(Isaac, purple, 50)  (Expected)
5. Guy is either the worker who served 100 riders or the worker who works in the purple section.

../data/puzzlesChallenging/puzzle10
1. Of the 75 million year old fossil and the 78 million year old fossil, one is the rotosaurus and t
2. Of the pilodontus and the $1250 fossil, one is 78 million years old and the other is 69 million y
3. Of the 75 million year old fossil and the $750 fossil, one is the rotosaurus and the other is the

../data/puzzlesChallenging/puzzle11
1. Meadowgrove will be worked on sometime before Tall Pines.
x  is(Meadowgrove, Tall Pines)    (Actual)
   before(Meadowgrove, Tall Pines)   (Expected)
2. The one Michael will dust on June 4th is either Hazelwood or the farm in Milbridge.
3. Of the one in Milbridge and Iron Hill, one will be dusted on June 7th and the other will be duste
4. The farm in Upper Lake is either the business Michael will dust on June 5th or Hazelwood.
5. The one in Defiance will be worked on 2 days before the business in Paradise.
x  is(Defiance, Paradise)   (Actual)
   before(Defiance, Paradise, 2 days)   (Expected)

../data/puzzlesChallenging/puzzle12
1. The bottle from Le Havre was sent 7 years after Amelia's letter.
2. Olive's message was sent 7 years after the missive from Bournemouth.
3. Carmen's letter was sent 7 years after Amelia's missive.
4. Carmen's missive was sent sometime before the missive from Brighton.
```

../data/puzzlesChallenging/puzzle13
1. Of the student who gave the presentation on President Van Buren and the student who gave the pres
2. The presenter who spoke for 8 minutes was either Loretta or the student who gave the presentation
3. Andy spoke for 14 minutes.
4. The presenter who gave the presentation on President Monroe was either Georgia or the presenter w
5. Ed spoke for a somewhat longer time than the presenter who gave the presentation on President Van
x is(Ed, Van Buren) (Actual)
   longer(Ed, Van Buren)   (Expected)

../data/puzzlesChallenging/puzzle14
1. Of the system with 2 GB of RAM and the computer with 512 GB of hard drive space, one is Greg's an
x not(512 GB, Greg, Philip)   (Actual)
   is(xor(2 GB, 512 GB), xor(Greg, Philip))   (Expected)
2. Nathaniel's computer has a hard drive that is 256 GB smaller than the computer with 3 GB of RAM.
x not(Nathaniel, 256 GB, 3 GB)   (Actual)
   smaller(Nathaniel, 3 GB, 256 GB) (Expected)
3. The four computers are Kerry's computer, the computer with 512 GB of hard drive space, the build
4. The system with 2 GB of RAM is either Philip's build or the system with 256 GB of hard drive spac
4 [NER]. The system with 2 GB of RAM is either Philip's build or the system with 256 GB of hard driv
   is(2 GB, xor(Philip, 256 GB)) (Expected)

../data/puzzlesChallenging/puzzle15
1. The one Cal Barron discovered was discovered 1 year before the one Nolan discovered.
x is(Cal Barron, Nolan)   (Actual)
   before(Cal Barron, Nolan, 1)   (Expected)
2. The comet Cal Barron discovered is either the one discovered in 2008 or Sporrin.
3. Of Gostroma and the comet discovered in 2010, one was first seen by Irv Horton and the other was
4. Of the one discovered in 2009 and the comet discovered in 2010, one was first seen by Nolan and t

../data/puzzlesChallenging/puzzle16
1. Of Victor's letter and the letter that was sent in 1962, one was from Torquay and the other was f
2. The missive from Plymouth was sent 14 years after the bottle from Torquay.
3. Robert's letter was sent 7 years after Warren's letter.
4. The message that was sent in 1969 is either Warren's missive or the bottle from Bournemouth.

../data/puzzlesChallenging/puzzle17
1. Of the winner who won the prize in economics and Betty Brewer, one won in 1976 and the other won
2. The winner who won in 1964 is either Ada Alvarez or the winner who won the prize in economics.
3. The person who won the prize in chemistry won her prize 8 years before the winner who won the pri
x is(chemistry, literature)   (Actual)
   before(chemistry, literature, 8 years) (Expected)
4. Fay Ferguson won her prize before Mandy Marsh.
5. The winner who won the prize in medicine is either the winner who won in 1968 or Ada Alvarez.

../data/puzzlesChallenging/puzzle18
1. Of the patient with the 10:00am appointment and the patient suffering from hip pain, one is Ronni
2. The person suffering from migraines has an appointment 1 hour before the patient suffering from f
x is(migraines, foot pain)   (Actual)
   before(migraines, foot pain, 1 hour)   (Expected)
3. The person suffering from migraines has an appointment sometime after Victor.
4. Of Dave and the person with the 12 noon appointment, one is complaining about foot pain and the o

../data/puzzlesChallenging/puzzle19
1. The 1913 masterpiece is "Girl at Sea".
2. "Lost in Time" was painted sometime after "Willow Bend".
3. Jenadije's masterpiece is either "Girl at Sea" or the 1897 piece.
4. Quekal's masterpiece was painted 8 years after Jenadije's painting.
5. Of Xesobe's piece and Jenadije's piece, one was painted in 1897 and the other is "Hoxley Hills".

```
../data/puzzlesChallenging/puzzle20
1. Of the business owned by Mr. Ortiz and the business owned by Mr. Kirk, one will open on May 5th a
2. Books and More will open 3 days after the business owned by Mr. Kirk.
3. The business opening on May 11th is either Frank's Figs or the business owned by Mr. Ortiz.
4. Of the business opening on May 8th and Frank's Figs, one is owned by Mr. Austin and the other is


## DECENT
73% success -  63 of 86 total statements
```

9.4 Solver: Successful Example

Here, the files and command-line output are shown for successfully solving an example logic grid puzzle.

```
$ logicSolver -r false -i puzzles/game1
[ 'xor(Dustin, hang gliding, 2004)',
  'after(kayaking, camping, 2 years)',
  'after(Zachary, kayaking)',
  'before(camping, James, 2 years)' ]
+--------------+--------+-------+--------+---------+
|              | Dustin | James | Yvonne | Zachary |
|--------------+--------+-------+--------+---------|
| camping      | x      | x     | O      | x       |
|--------------+--------+-------+--------+---------|
| cycling      | x      | x     | x      | O       |
|--------------+--------+-------+--------+---------|
| hang gliding | O      | x     | x      | x       |
|--------------+--------+-------+--------+---------|
| kayaking     | x      | O     | x      | x       |
+--------------+--------+-------+--------+---------+
+------+--------+-------+--------+---------+
|      | Dustin | James | Yvonne | Zachary |
|------+--------+-------+--------+---------|
| 2001 | x      | x     | O      | x       |
|------+--------+-------+--------+---------|
| 2002 | O      | x     | x      | x       |
|------+--------+-------+--------+---------|
| 2003 | x      | O     | x      | x       |
|------+--------+-------+--------+---------|
| 2004 | x      | x     | x      | O       |
+------+--------+-------+--------+---------+
+------+---------+---------+--------------+----------+
|      | camping | cycling | hang gliding | kayaking |
|------+---------+---------+--------------+----------|
| 2001 | O       | x       | x            | x        |
|------+---------+---------+--------------+----------|
| 2002 | x       | x       | O            | x        |
|------+---------+---------+--------------+----------|
| 2003 | x       | x       | x            | O        |
|------+---------+---------+--------------+----------|
| 2004 | x       | O       | x            | x        |
+------+---------+---------+--------------+----------+
Dustin, hang gliding, 2002
James, kayaking, 2003
Yvonne, camping, 2001
Zachary, cycling, 2004

 O puzzles/game1/.
```

And the results of running a number of puzzles.

```
$ logicSolver -q -i ./puzzles
 + ./puzzles/game1
 + ./puzzles/game2
 + ./puzzles/game3
 + ./puzzles/game4
 + ./puzzles/game5
 + ./puzzles/game6
```

```
 + ./puzzles/game7
 + ./puzzles/game8
 + ./puzzles/game9

 + Success. x Failure.

$ logicSolver -q -i ./puzzlesAdvanced
 + ./puzzlesAdvanced/game1
 + ./puzzlesAdvanced/game2
 + ./puzzlesAdvanced/game3
 x ./puzzlesAdvanced/game4

 + Success. x Failure.
```

9.5 Solver: Failure Example

Here, the files and command-line output are shown for unsuccessfully solving an example logic grid puzzle.

```
$ logicSolver -i ./puzzlesAdvanced/game4 -r false
[ 'not(18 minutes, North, A-)',
  'is(D, North)',
  'shorter(Perry, D)',
  'not(14 minutes, D)',
  'more(Blair, Churchill, 2 minutes)',
  'xor(8 minutes, Perry, Eden)',
  'xor(B+, Perry, Eden)',
  'xor(Perry, 8 minutes, B+)',
  'xor(Eden, 8 minutes, B+)',
  'less(Anita, Zachary, 8 minutes)',
  'more(Perry, Hazel, 2 minutes)',
  'not(14 minutes, A-, 6 minutes, C+)',
  'not(A-, Blair)',
  'xor(C-, Balfour, Milton)',
  'xor(14 minutes, Balfour, Milton)',
  'xor(Balfour, C-, 14 minutes)',
  'xor(Milton, C-, 14 minutes)',
  'longer(A, C+)',
  'not(Colleen, 12 minutes)',
  'is(10 minutes, B+)',
  'xor(Theodore, 18 minutes, Disraeli)' ]
+---------------------------------------------------------------------------+
|           | Anita | Colleen | Hazel | Milton | Perry | Theodore | Zachary |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| Balfour   |       |         |       | x      |       |          |         |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| Blair     |       |         |       |        |       | x        |         |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| Churchill |       |         |       |        |       | x        | x       |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| Disraeli  |       |         |       |        |       |          |         |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| Eden      |       |         |       |        | x     | x        | x       |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| North     | x     |         | x     | x      | x     | x        |         |
|-----------+-------+---------+-------+--------+-------+----------+---------|
| Pitt      |       |         |       |        |       |          |         |
+---------------------------------------------------------------------------+
+---------------------------------------------------------------------+
|    | Anita  | Colleen | Hazel | Milton | Perry | Theodore | Zachary |
|----+--------+---------+-------+--------+-------+----------+---------|
| A  | x (+)  |         | x     |        | x     |          |         |
|----+--------+---------+-------+--------+-------+----------+---------|
| A- |        |         |       | x      |       |          | x       |
|----+--------+---------+-------+--------+-------+----------+---------|
| B+ |        |         | x     | x      |       | x        | x       |
|----+--------+---------+-------+--------+-------+----------+---------|
| B- |        |         |       |        |       |          |         |
|----+--------+---------+-------+--------+-------+----------+---------|
| C+ |        |         |       | x      |       |          | x       |
|----+--------+---------+-------+--------+-------+----------+---------|
```

C-								
D	x		x	x	x	x (+)		

	Anita	Colleen	Hazel	Milton	Perry	Theodore	Zachary
6 minutes					x		x
8 minutes						x	x
10 minutes			x	x		x	x
12 minutes	x	x	x		x		x
14 minutes	x		x		x		
16 minutes	x		x	x	x	x	
18 minutes	x		x		x		

	Balfour	Blair	Churchill	Disraeli	Eden	North	Pitt
A					x (+)	x	
A-	x	x				x	
B+	x					x	
B-						x	
C+	x					x	
C-						x	
D	x	x	x	x (+)	x	+	x

	Balfour	Blair	Churchill	Disraeli	Eden	North	Pitt
6 minutes		x			x	x	
8 minutes						x	
10 minutes	x					x	
12 minutes					x	x	
14 minutes			x		x	x	
16 minutes	x	x	x	x	x	+	x
18 minutes		x	x	x	x	x	

	A	A-	B+	B-	C+	C-	D
6 minutes	x	x	x		x		x
8 minutes	x		x				x
10 minutes	x (+)	x	+	x	x	x	x
12 minutes			x				x
14 minutes		x	x		x	x	x
16 minutes	x	x	x	x	x	x	+

```
| 18 minutes |        |  x  |  x  |     |  x  |     |  x  |
+------------+--------+-----+-----+-----+-----+-----+-----+
Anita, Balfour, Blair, Churchill, Disraeli, Eden, Pitt, A-, B+, B-,10 C+, C-, 6 minutes,
8 minutes, 10 minutes
Colleen, Balfour, Blair, Churchill, Disraeli, Eden, North, Pitt, A, A-, B+, B-, C+, C-, D, 6
    minutes, 8 minutes, 10 minutes, 14 minutes, 16 minutes,
18 minutes
Hazel, Balfour, Blair, Churchill, Disraeli, Eden, Pitt, A-, B-, C+, C-, 6 minutes,
8 minutes
Milton, Blair, Churchill, Disraeli, Eden, Pitt, A, B-, C-, 6 minutes, 8 minutes, 12 minutes, 14
    minutes,
18 minutes
Perry, Balfour, Blair, Churchill, Disraeli, Pitt, A-, B+, B-, C+, C-, 8 minutes,
10 minutes
Theodore, Balfour, Disraeli, Pitt, A, A-, B-, C+, C-, 6 minutes, 12 minutes,
14 minutes, 18 minutes
Zachary, Balfour, Blair, Disraeli, North, Pitt, A, B-, C-, D,18 14 minutes, 16 minutes, 18
    minutes
```

x ./puzzlesAdvanced/game4/.

9.6 Solver: All Results

Here we display the results of solving 28 Easy, 20 Moderate, and 20 Challenging puzzles using the expected FOPL parses as inputs. Unsuccessfully solved puzzles are marked with an "x" and show the solutions, comparing actual (incorrect) solutions – with incorrect entities noted with "[entityName]" – to the expected (correct) solutions. If one of the solutions was correct, it is shown as is, without an "(Actual)" or "(Expect)" prefix. **NOTE:** all puzzles have 4 solutions relating 3 entities each, since our dataset consists of logic grid puzzles with a 3x4 size.

9.6.1 Easy Puzzles

```
   ../data/puzzlesEasy/puzzle01
   ../data/puzzlesEasy/puzzle02
   ../data/puzzlesEasy/puzzle03
   ../data/puzzlesEasy/puzzle04
   ../data/puzzlesEasy/puzzle05
   ../data/puzzlesEasy/puzzle06
   ../data/puzzlesEasy/puzzle07
   ../data/puzzlesEasy/puzzle08
   ../data/puzzlesEasy/puzzle09
   ../data/puzzlesEasy/puzzle10
 x ../data/puzzlesEasy/puzzle11
   Error: entityA.popRelationships is not a function

   ../data/puzzlesEasy/puzzle12
   ../data/puzzlesEasy/puzzle13
   ../data/puzzlesEasy/puzzle14
   ../data/puzzlesEasy/puzzle15
   ../data/puzzlesEasy/puzzle16
   ../data/puzzlesEasy/puzzle17
 x ../data/puzzlesEasy/puzzle18
   (Actual) first, [], []
   (Expect) first, Kermit, #118

   (Actual) fourth, [], []
   (Expect) fourth, Bridget, #105

   second, Ophelia, #128

   (Actual) third, [Bridget], [#105]
   (Expect) third, Stretch, #126
```

```
 x ../data/puzzlesEasy/puzzle19
   Error: Cannot read property 'prices (dollars)' of undefined

   ../data/puzzlesEasy/puzzle20
   ../data/puzzlesEasy/puzzle21
   ../data/puzzlesEasy/puzzle22
   ../data/puzzlesEasy/puzzle23
   ../data/puzzlesEasy/puzzle24
   ../data/puzzlesEasy/puzzle25
 x ../data/puzzlesEasy/puzzle26
   Error: Cannot read property 'baking times (minutes)' of undefined

   ../data/puzzlesEasy/puzzle27
 x ../data/puzzlesEasy/puzzle28
   (Actual) September 1, [], []
   (Expect) September 1, Muriel, Heaven's Seal

   (Actual) September 15, [], []
   (Expect) September 15, Latasha, Ohio Haunts

   (Actual) September 22, [], []
   (Expect) September 22, Sherrie, Stars Below

   (Actual) September 8, [], []
   (Expect) September 8, Wayne, Kip and Ken


83.0% accuracy (93 of 112)
```

9.6.2 Moderate Puzzles

```
   ../data/puzzlesModerate/puzzle01
   ../data/puzzlesModerate/puzzle02
   ../data/puzzlesModerate/puzzle03
 x ../data/puzzlesModerate/puzzle04
   Error: xorA.popRelationships is not a function

   ../data/puzzlesModerate/puzzle05
   ../data/puzzlesModerate/puzzle06
   ../data/puzzlesModerate/puzzle07
   ../data/puzzlesModerate/puzzle08
 x ../data/puzzlesModerate/puzzle09
   Error: entityA.popRelationships is not a function

   ../data/puzzlesModerate/puzzle10
   ../data/puzzlesModerate/puzzle11
   ../data/puzzlesModerate/puzzle12
   ../data/puzzlesModerate/puzzle13
   ../data/puzzlesModerate/puzzle14
   ../data/puzzlesModerate/puzzle15
   ../data/puzzlesModerate/puzzle16
   ../data/puzzlesModerate/puzzle17
   ../data/puzzlesModerate/puzzle18
   ../data/puzzlesModerate/puzzle19
   ../data/puzzlesModerate/puzzle20
90.0% accuracy (72 of 80)
```

9.6.3 Challenging Puzzles

```
x ../data/puzzlesChallenging/puzzle01
  (Actual) Irene, [iced tea], water, $4.99, [$5.99]
  (Expect) Irene, water, $4.99

  (Actual) Oliver, [iced tea], root beer, [water], $5.99, [$6.99]
  (Expect) Oliver, root beer, $5.99

  (Actual) Perry, iced tea, [orange soda], [root beer], [water], [$4.99], [$5.99], $6.99, [$7.99]
  (Expect) Perry, iced tea, $6.99

  (Actual) Salvador, orange soda, [root beer], [water], [$5.99], [$6.99], $7.99
  (Expect) Salvador, orange soda, $7.99


  ../data/puzzlesChallenging/puzzle02
x ../data/puzzlesChallenging/puzzle03
  (Actual) Mt. Calhoun, [Chip Carson], Eddie English, [Leonard Lara], 14210, [14340]
  (Expect) Mt. Calhoun, Eddie English, 14210

  (Actual) Mt. Larsen, [Art Aguilar], Chip Carson, [Leonard Lara], [14210], 14600
  (Expect) Mt. Larsen, Chip Carson, 14600

  (Actual) Mt. Sloan, [Art Aguilar], [Eddie English], Leonard Lara, [14210], 14340, [14470]
  (Expect) Mt. Sloan, Leonard Lara, 14340

  (Actual) Mt. York, Art Aguilar, [Chip Carson], [Eddie English], [Leonard Lara], 14470, [14600]
  (Expect) Mt. York, Art Aguilar, 14470


x ../data/puzzlesChallenging/puzzle04
  (Actual) Aladeen, [Hayco Inc], [Tarco Inc], [Waverly Toys], 1975, [1976]
  (Expect) Aladeen, Tarco Inc., 1975

  (Actual) Archer Quest, [Hayco Inc], [Tarco Inc], Waverly Toys, 1976, [1977]
  (Expect) Archer Quest, Waverly Toys, 1976

  (Actual) Bowling Alley, [Hayco Inc], Pinnefore, [Tarco Inc], 1974, [1976]
  (Expect) Bowling Alley, Pinnefore, 1974

  (Actual) Meteor Rise, [Hayco Inc], [Pinnefore], [Tarco Inc], [Waverly Toys], [1974], [1975], [197
  (Expect) Meteor Rise, Hayco Inc., 1977


  ../data/puzzlesChallenging/puzzle05
  ../data/puzzlesChallenging/puzzle06
x ../data/puzzlesChallenging/puzzle07
  (Actual) Annie Axel, [8 losses], [33 wins]
  (Expect) Annie Axel, 8, 33

  (Actual) Diane DeNiro, [15 losses], [24 wins]
  (Expect) Diane DeNiro, 15, 24

  (Actual) Ollie, [2 losses], [27 wins]
  (Expect) Ollie, 2, 27

  (Actual) Vicki, [7 losses], [30 wins]
  (Expect) Vicki, 7, 30
```

```
   ../data/puzzlesChallenging/puzzle08
   ../data/puzzlesChallenging/puzzle09
   ../data/puzzlesChallenging/puzzle10
   ../data/puzzlesChallenging/puzzle11
 x ../data/puzzlesChallenging/puzzle12
   (Actual) Amelia, [Bournemouth], Plymouth, 1962, [1969]
   (Expect) Amelia, Plymouth, 1962

   (Actual) Carmen, [Bournemouth], Le Havre, [Plymouth], 1969, [1976]
   (Expect) Carmen, Le Havre, 1969

   (Actual) George, Bournemouth, [Brighton], [Le Havre], [Plymouth], [1962], [1969], 1976, [1983]
   (Expect) George, Bournemouth, 1976

   (Actual) Olive, Brighton, [Le Havre], [Plymouth], [1969], [1976], 1983
   (Expect) Olive, Brighton, 1983


   ../data/puzzlesChallenging/puzzle13
   ../data/puzzlesChallenging/puzzle14
   ../data/puzzlesChallenging/puzzle15
   ../data/puzzlesChallenging/puzzle16
   ../data/puzzlesChallenging/puzzle17
   ../data/puzzlesChallenging/puzzle18
   ../data/puzzlesChallenging/puzzle19
   ../data/puzzlesChallenging/puzzle20
75.0% accuracy (60 of 80)
```