Assignment 6: Statistical Parsing and Lexical Semantics

My posse consists of: `null` *Salman Ahmad (saahmad@mit.edu)*

Part II: Semantic Analysis of "`Invest`"

Data Collection

To collect all of the occurrences of "invest" along with their subcategorization frame I used the follow `tregex` query on the supplied subset of the Penn Tree Bank:

```
@VP < ( __< /^[Ii]nvest[esi]$|^[Ii]nvest$/ )
```

The query returned 52 matches in 52 trees.

Analysis

A summary of the data is shown in the next section.

According to Levin the word "invest" is mainly used as an "fulfilling" verb that is used to describe something that will be provided or utilized in the future. "Invest" can be used in two subtle different contexts: it can be used to specify the **act** that is being provided and it can also be used to specify the **thing or item** that is being provided. The verb typically appears in a "NP V NP to / with NP" frame. For example:

- "John invested the land to Sally"

Looking at the subcategories in the Data Tabulation section you can see that not all of the frames follow the "NP V NP to / with NP" subcategory that Levin describes. In fact, many of them do not even include a prepositional phrase at all, for example, subcategories (5, 7, 18) and many others. These usages are not described in Levin at all. Examples of uses of "invest" that are not accompanied by a preposition include:

- "invest internationally"
- "invest elsewhere"

Furthermore, Levin talks at length about the use of the "to" and "with" prepositions that are used with verbs like "Invest". However, this almost never appears in the Penn Tree Bank. Rather, the Penn Tree Bank is dominated by the use of the "in" preposition with invest. There are plenty of examples:

- "invest money in stocks that they 'll need in the near future – for example , for college tuition payments or retirement expenses"
- "invest in common stocks"
- "invest in business"

In summary, Levin primarily describes "invest" being used as an "equip" and "fulfilling" verb while in practice it seems rarely used in the context.

Data Tabulation

The table below displays the number of times a unique subcategorization frame exists in the Penn Tree Bank.

Table 1: Subcategory counts

Subcategory	**Count**
1) [VP (VB invest) PP-CLR]	12
2) [VP (VB invest) NP PP-CLR]	8
3) [VP (VB invest) ADVP-LOC]	2
4) [VP (VB invest) NP PP-LOC-CLR]	2
5) [VP (VB invest)]	2
6) [VP (VBP invest) PP-CLR]	2
7) [VP (VB invest) , SBAR-ADV]	1
8) [VP (VB invest) ADVP-MNR PP-CLR PP-TMP]	1
9) [VP (VB invest) ADVP-MNR]	1
10) [VP (VB invest) NP ADVP-LOC PP-MNR]	1
11) [VP (VB invest) NP ADVP-LOC-CLR]	1
12) [VP (VB invest) NP PP-CLR , NP-2]	1
13) [VP (VB invest) NP PP-CLR , PP-2]	1
14) [VP (VB invest) NP PP-CLR S-PRP]	1
15) [VP (VB invest) NP PP-CLR SBAR-1]	1
16) [VP (VB invest) NP PP-TMP PP-LOC-CLR S-CLR]	1
17) [VP (VB invest) NP S-PRP]	1
18) [VP (VB invest) NP]	1
19) [VP (VB invest) PP-CLR PP-PRP]	1
20) [VP (VB invest) PP-LOC-CLR NP]	1
21) [VP (VB invest) PP-LOC-CLR]	1
22) [VP (VB invest) PP-TMP]	1
23) [VP (VBP invest) ADVP PP-LOC-CLR ADVP-LOC]	1
24) [VP (VBP invest) ADVP-LOC]	1
25) [VP (VBP invest) PP-LOC-CLR]	1
26) [VP (VBZ invests) ADVP PP-CLR]	1
27) [VP (VBZ invests) ADVP PP-LOC-CLR]	1
28) [VP (VBZ invests) ADVP PP-LOC]	1
29) [VP (VBZ invests) NP NP-TMP PP-CLR SBAR-ADV]	1
30) [VP (VBZ invests) NP PP-CLR]	1