(R) Disambiguate This! (1/3) [5 points]

One important (and often tricky) task in machine translation is *disambiguation*: identifying which sense of a word is being used in a sentence. Consider the following sentences:

1) The old sing.
2) The singers are old.

In 1), "old" is used as a noun, while in 2), "old" is used as an adjective. Computers must be able to deduce which meaning of the word is intended in order to properly label these sentences for translation. Here's an example:

1)	the	old	are	singing
	`the.DET`	**`old.N`**	`be.VRB`	`sing.VRB`

2)	the	singers	are	old
	`the.DET`	`singer.N`	`be.VRB`	**`old.ADJ`**

Here's a brief explanation of the above syntax:

- The lowercase word before the first period is the *lemma*—the base form of the word.
- The uppercase word after the period is a tag which marks the part of speech.
- The following tags are available: `DET` for determiners (broadly, words that come before a noun, like "this," "your," or "the"); `N` for nouns; `PRN` for pronouns (e.g. "I," "me," or "you"); `VRB` for verbs; `ADJ` for adjectives; `ADV` for adverbs; `PREP` for prepositions (words like "about").

Unlike us, computers are not automatically able to tell that the word *old* is a noun in the first sentence but an adjective in the second. Therefore, we must write rules to determine the correct tag for such words.
Here's an example rule, which is written in a syntax known as *constraint grammar:*

```
      old: SELECT N if (+1 VRB)
```

This selects the noun (`N`) form of the word `old` if the next (`+1`) word is tagged as a verb (`VRB`), and does nothing otherwise. Note that negative numbers may be used to select previous words, as in this rule:

```
      old: SELECT ADJ if (-1 DET)
```

Here's one more rule, which selects the verb form of the word "desert" in every case.

```
      desert: SELECT VRB
```

Before using the rules, our computer system first tags all words that only have one possible part of speech. It then handles the rules in top-down order, applying each rule in turn to every word (from left to right) in the sentence that still has more than one possible tag. Beware: if no rule makes a decision, the system will crash!

(R) Disambiguate This! (2/3)

Below are some sentences in English containing the ambiguous word “her”, which can either serve as a determiner (‘her dress’, represented as `her.DET`) or as a pronoun (‘I saw her’, represented as `her.PRN`).

1.	I	see	her	now.		
	`I.PRN`	`see.VRB`	**`her.PRN`**	`now.ADV`		
2.	Her	son	is	tall.		
	`her.DET`	`son.N`	`be.VRB`	`tall.ADJ`		
3.	The	girl	hears	her	daughter	today.
	`the.DET`	`girl.N`	`hear.VRB`	**`her.DET`**	`daughter.N`	`today.ADV`

4. The dog looks at her.
5. The girl is her friend.
6. I am her daughter.
7. The cat saw her dog yesterday.
8. You walk with her.
9. The boy likes her.
10. A giraffe sees her now.
11. I give her flowers.
12. Her tall daughter is smart.
13. The cat examines her quizzically.
14. I am her older sister.
15. Her orange cat likes me.

R1. For sentences 4-15, indicate, on your answer sheets, whether “her” is being used as a determiner or as a pronoun.

R2. All of the above sentences—except one—can be disambiguated using three rules. List these three rules in your answer sheets, remembering that rule order matters. Assume that all words other than “her” only have one possible tag.

R3. Which sentence is disambiguated incorrectly? Explain, in your answer sheets, why it would be difficult to create a rule that would successfully tag this sentence.

(R) Disambiguate This! (3/3)

Below are some sentences in Sranan Tongo (an English-based creole language with influences from Dutch, Javanese, Hindustani, and Chinese, which is the national language of Suriname) with their translations in English.

Mi lobi den singi.	"I love the songs."
Den lobi yu singi.	"They love your songs."
Den lobi mi.	"They love me."
Mi singi abra yu lobi.	"I sing about your love."
Den lobi dati mi singi.	"They love that I sing."
Yu lobi mi sisa.	"You love my sister."
Mi lobi yu.	"I love you."

Note that no Sranan Tongo words change from their lemma forms. For example, the lemma form of sisa is `sisa.N`. Most of the words in these examples have parts of speech that were also present in the English examples on the previous page, but there are also two additions: “abra” should be tagged `PREP` (preposition) and “dati” should be tagged `COMP` (complementizer).

As you can see, disambiguation is much harder in Sranan Tongo than in English, as many words have multiple meanings. For example, “lobi” can mean "love" (noun) or "to love" (verb).

To deal with this level of difficulty, we need more powerful rules. Here's a rule that uses some additional syntax available for Sranan Tongo:

```
PRN/DET: SELECT DET if (-1 [VRB]) and (+1 PRN)
```

This selects the determiner (`DET`) form of a word that could be either a pronoun `(PRN)` or a determiner (`DET`) if the previous (`-1`) word could possibly be a verb (`[VRB]`) and the next word has been confirmed to be a pronoun (`PRN`) and does nothing otherwise. Specifically, the notation we are adding is the slash / (which can only be used before the colon, not after it); the brackets [ ] (which can only be used after the colon, not before it); and the word “and” (but not “or”)
Of course, you can still use the syntax given in the previous section.

R4. Write a set of rules that would successfully disambiguate the above sentences in your answer sheets. Hint: you should need no more than 5 rules. Recall that, before the rules are applied, all words with only one possible part of speech are tagged with that part of speech.

(R) Disambiguate This! Answer Sheet

(R) Disambiguate This!

R1	4	DET	PRN	5	DET	PRN	6	DET	PRN
	7	DET	PRN	8	DET	PRN	9	DET	PRN
	10	DET	PRN	11	DET	PRN	12	DET	PRN
	13	DET	PRN	14	DET	PRN	15	DET	PRN

R2

R3

R4