LING 227: Homework 1

Paul Fletcher-Hill

February 5, 2015

Part 1

1. Give examples of two different kinds of errors in the output

The first type of error that is seen in the output relate to "S" phonemes. For example, "AE D + M IH + N AH S + T ER + IH NG" was mispared as "AE D + M IH + N AH + S T ER + IH NG".

The second type of error that is seen relates to "ly" endings. For example, "AH + SH UH + R AH D + L IY" was misparsed as "AH + SH UH + R AH + D L IY".

2. Why do these errors happen?

These errors represent anomolies to the syllabification rules outlined in the assignment. The incorrectly parsed "AE D + M IH + N AH + S T ER + IH NG" is a direct result of the "S" rule (Always put "S" with the following syllable). If the "S" defaults to a coda consonent rather than an onset consonent, then it is correctly parsed as "AE D + M IH + N AH S + T ER + IH NG".

The correctly parsed "AH + SH UH + R AH D + L IY" is another exception to the specified rules, because it rejects the Onset Maximization method. Following strict Onset Maximization rules, "D" and "L" would both be onset consonents, because "D" has a sonority of 0 and "L" has a sonority of 2. The sonority difference is at least 2.

3. How could the syllabification program be improved to fix these errors?

Defaulting "S" to be a coda consonent and rejecting the Onset Maximization rule when we find consecutive "D L" increases the syllable level accuracy to 0.947879 and word level accuracy to 0.936937, which is better than the 0.916364 and 0.897898 accomplished with the default rules.

Part 2

1. What are the first syllable types the children attempt (syllable types in the earliest target files)? What are the syllable types children attempt last (syllable types only attempted in later files)?

Here are the syllable pattern frequencies for the two children at approximately one year and four months each:

Charlotte (1 year, 4 months, 11 days)
Frequency of CVC : 47.96
Frequency of V : 47.96
Frequency of CV : 4.08

Georgia (1 year, 4 months, 17 days)
Frequency of CCVC : 30.00
Frequency of V : 26.67
Frequency of CV : 23.33
Frequency of CVC : 20.00

Here are the syllable pattern frequencies for the two children at the latest readings:

Charlotte (2 years, 11 months, 16 days)
Frequency of CV : 44.11
Frequency of CVC : 28.05
Frequency of VC : 9.42
Frequency of CVCC : 7.07
Frequency of V : 5.14
Frequency of CCV : 2.36
Frequency of CCVC : 1.93
Frequency of VCC : 1.07
Frequency of CCCVC : 0.43
Frequency of CVCCC : 0.43

Georgia (2 years, 10 months, 11 days)
Frequency of CV : 41.39
Frequency of CVC : 31.44
Frequency of VC : 7.99
Frequency of V : 6.93
Frequency of CVCC : 5.33
Frequency of VCC : 2.49
Frequency of CCVC : 2.31
Frequency of CCV : 1.07
Frequency of CCVCC : 0.53
Frequency of CVCCC : 0.53

2. What are the first syllable types the children actually say (syllable types in the earliest actual files)? What are the syllable types children start saying last? (syllable types appearing only in later actual files)?

Here are the syllables actually said by the two children in the earliest attempts:

Charlotte (1 year, 4 months, 11 days)
Frequency of CV : 50.52
Frequency of V : 48.45
Frequency of CVC : 1.03

Georgia (1 year, 4 months, 17 days)
Frequency of CV : 71.88
Frequency of V : 25.00
Frequency of CVC : 3.12

And here are the patterns actually said in the latest attempts:

Charlotte (2 years, 11 months, 16 days)
Frequency of CV : 49.23
Frequency of CVC : 26.04
Frequency of V : 7.22
Frequency of VC : 6.56
Frequency of CVCC : 5.25
Frequency of CCV : 2.19
Frequency of CCVC : 1.53
Frequency of VCC : 0.88
Frequency of CVCCC : 0.66
Frequency of CCCVC : 0.44

Georgia (2 years, 10 months, 11 days)
Frequency of CV : 47.44
Frequency of CVC : 28.33
Frequency of VC : 9.73
Frequency of V : 7.00
Frequency of CVCC : 3.41
Frequency of CCVC : 2.39
Frequency of CCV : 1.02
Frequency of CVCCC : 0.34
Frequency of VCC : 0.34

3. Comparing the proportions in pairs of actual vs. target files, which syllable types tend to be over-represented in the childrens pronunciations? Which tend to be under-represented?

Just looking at the earliest sample, the "CV" syllable pattern is significantly over-represented, while the "CVC" pattern is significantly under-represented.

4. What generalizations can you make about acquisition of syllable types? Which syllables do children seem to acquire earlier than others?

"CV" is by far the earliest and most (actually) utilized syllable type. Generally, simpler syllable types are acquired first, such as "CV" and "V".

5. How do your generalizations from question (4) above relate to the Implicational Universals in (a)-(e) above? Are all the implications respected? Explain.

It seems that the Implicational Universals support the way syllables are acquired. Syllables that are more marked are under-represented in the actual files. For example, "CV" is less marked than "CVC", "V", and "CCV". Similarly it was uttered at a higher frequency and over-represented in the actual files.