CompLing Project Summary

Han Li

May 20, 2022

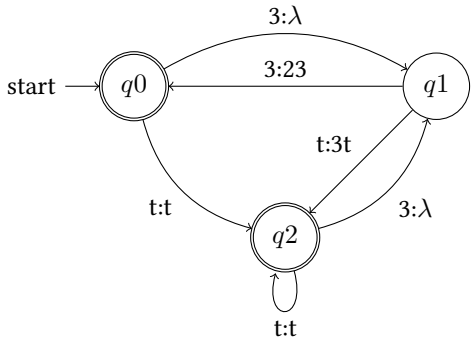
This project aims to construct a finite state transducer of the third tone sandhi process in Mandarin Chinese. This final summary will introduce this phonological process and provide a theoretical description and an implement of a Haskell-based code snippet, and a finite state description.

1 Computational background

The phonological process in question is the third tone assimilation in Mandarin Chinese. Mandarin Chinese has four tones: T1 (high level), T2 (high rising), T3 (circumflex), T4 (high falling). The third tone assimilation is when two T3 are adjacent to each other, the first T3 will change into a T2 [Chen, 2000, Yip, 2002]. This process is described as below in a rule-based grammar:

$$\text{T3} \rightarrow \text{T2} \ / \ \_\text{T3}$$

This process could be modelled by an input strictly local (ISL) function because only T3 and its following T3 will trigger the tone sandhi and there is no more additional information in the string that will contribute to it [Chandlee, 2019]. In this case, a ISL FST can clearly demonstrate this process.



In this FST, there are three states: $q0, q1, q2$. Among three states, $q0$ is the initial state, and $q0, q2$ are the accepting states. The number 3 indicates the third tone T3, and the letter t refers to all other tones T1, T2, T4. When the initial state $q0$ reads a T3, it goes to the $q1$ and will output an empty string λ, otherwise it reads T1,2,4 and goes to the $q2$ and output the same string. In $q1$, the FST will wait for the following string to decide which path to go: if a second T3 occurs, it will go back to $q0$ and output T2T3. The third tone sandhi will be realized in this way. If the second tone is not T3, the state will move from $q1$ to $q2$ and output T3 followed by whatever the input is.

2 Haskell Implement

The second part of this project is to implement this process on Haskell. The code is based on the DFST homework from the class. It consists of three main components: a process function, a transduce function, and an FST of tone sandhi transducer.

The transducer function is not the exactly same as the one from the previous homework. Since the third tone sandhi process is a regressive assimilation, it needs to output an empty string and wait for the next string to decide

what to output. I tried to solve this question but didn't work out in Haskell. To make this work out, I had to reverse the whole string and then process and finally reverse the output again, which has the same result.

```
> transduce :: DFST q b [a] -> [b] -> [a]
> transduce dfst [] = []
> transduce dfst xs = reverse (snd (process dfst (start dfst, prefix dfst) (reverse xs)))
```

Since the reverse function changes the progressive assimilation into a regressive one, the omega function has also changed accordingly:

```
> oSD :: Omega Int Char [Char]
> oSD 0 b = [b]
> oSD 1 b    | b == '3' = "2"
>            | otherwise = [b]
```

For clarification of other parts of code, please refer to the 3tone_sandhi.lhs.

3 Questions

This project is a simple modelling of Mandarin tone sandhi, but there are some questions awaiting answers.

First, how to realize this process without reversing the string is not solved. I tried to make omega function output an empty string λ but this can't process a single tone (input "3" and output an empty string).

Second, tones in many languages including Mandarin, Cantonese, and Thai are "dynamic". Tones with dipping, falling, rising shapes involve a moving from the starting point to the tone target. By simplifying these dynamic tones into R(ising) and F(alling) is actually ruling out many details of pitch trajectory. Another problem brought by simplification is how to address the unspecified tone like the neutral tone T0 in Mandarin. T0 does not have a fixed tone value but will take the pitch value of the tail of the preceding tone (not completely copy the preceding tone). It looks necessary that we need to revisit the tone representations used in the FST.

References

M.Y. Chen. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge Studies in Linguistics. Cambridge University Press, 2000. ISBN 9781139431491. URL https://books.google.com/books?id=D328u70WNgMC.

Moira Yip. *Tone*. Cambridge University Press, 2002.

Jane Chandlee. A computational account of tone sandhi interaction. In *Proceedings of the Annual Meetings on Phonology*, volume 6, 2019.