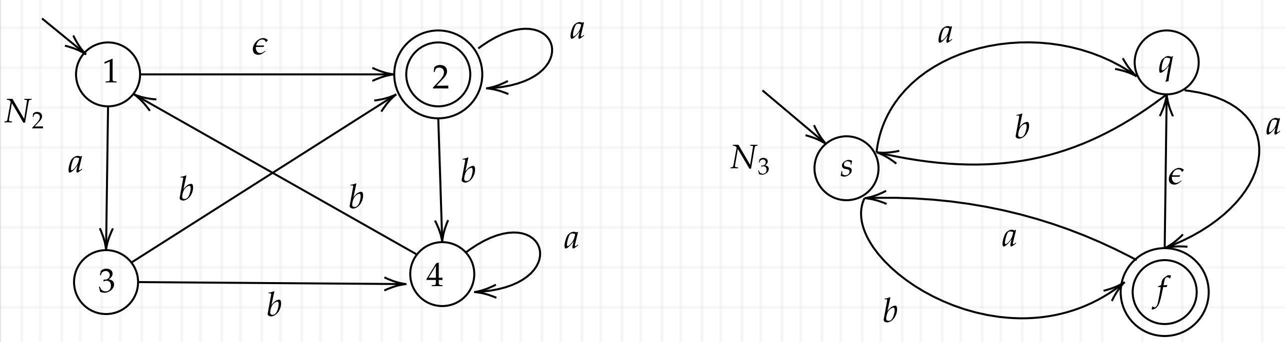
CSE491/596 Lecture 21 Sept. 2020: Applications of the Myhill-Nerode Theorem

[first take any Qs on HW. Instead of a morning office hour, I will do the same thing I usually do for programming projects, which is to be online 10pm--11:30pm for last-minute questions and help with any logistical glitches. This is also 7:30am--9:30am in India. I have also made the deadline the same "overnight stretchy" as for programming projects. Since I had a question about the NFA in problem 2, I will go over it from the picture at left, with Problem 3's NFA at right:



δ = { (1, ϵ, 2),(1, a, 3),(2, a, 2),(2, b, 4),
(3, b, 2),(3, b, 4),(4, a, 4),(4, b, 1) }

δ = {(s, a, q),(s, b, f),(q, b, s),
(q, a, f),(f, a, s),(f, ϵ, q)}.

One question was whether it is OK to make the arc from state 2 to state 4 in N_2 bi-directional. The general answer is *no*, but in this case...]

We have proved only one direction of the Myhill-Nerode Theorem: L has an infinite PD set $\implies$ L is nonregular, but this is the direction to apply for nonregularity proofs. Those proofs can all be made to follow a "script":

Take $S =$ ___________. [Observe S is infinite---this is usually immediately clear.]
Let any $x, y \in S$ $(x \neq y)$ **be given**. Then we can write $x =$ ________ and $y =$ ________ where _________ [and **without loss of generality**, __________].
Take $z =$ ________.
Then $L(xz) \neq L(yz)$ because _____________________________________
____.
Because x, y are an arbitrary pair of strings in S, this shows that S is PD for L, and since S is infinite, it follows that L is nonregular by the Myhill-Nerode Theorem.

I have colored the words **take** and **let...be given** separately to show how they express the logical quantifiers in the formal statement of this direction of MNT:

If there exists an infinite set S such that **for all** distinct $x, y \in S$ **there exists** $z \in \Sigma^*$ such that $L(xz) \neq L(yz)$, **then** L is nonregular.

The difference is that *you* have control of choices in the existential parts, but in the "for-all" parts you have to be prepared for all possibilities. There is a habit to use "let" in both situations, but this can be confusing. [Give humorous story about how both "let" and "any" are self-contradictory words in English, but they are OK together with "...be given."]

Thus to prove a given L nonregular we have to "act out" the proof---and the above is our script. The first example also illustrates the optional "**w.l.o.g.**" clause.

I. $L = \left\{x \in \{s, d\}^* : \#s(x) \geq \#d(x)\right\}$.

Take $S =$ __s^*__. Clearly S is infinite.

Let any $x, y \in S$ $(x \neq y)$ **be given**. Then we can write $x =$ __s^m__ and $y =$ __s^n__ where ____$m, n \geq 0$____ and **wlog.**, ____$m < n$____.

Take $z =$ __d^n__.

Then $L(xz) \neq L(yz)$ because ___$xz = s^m d^n \notin L$ since m is less than n by the "wlog." provision, whereas $yz = s^n d^n \in L$___.

Because x, y are an arbitrary pair of strings in S, this shows that S is PD for L, and since S is infinite, it follows that L is nonregular by the Myhill-Nerode Theorem.

Note that this L is not the same as the language of "spears-and-dragons with unlimited saving of spears" because e.g. the string "ds" belongs to this L despite the spear coming too late in the other. But the *proof* is exactly the same. The fun is that not only do these proofs become fairly automatic once you get comfortable with the script, they are often like re-usable code.

[Here and/or with *reductions*, I used to say for fun that this can be an exception to the university rule against recycling an old answer for a new assignment, even when it was your answer. I even used to sing a relevant section of the Tom Lehrer song "Lobachevsky" which you can find recently linked at https://gilkalai.wordpress.com/2020/08/29/to-cheer-you-up-in-difficult-times-11-immortal-songs-by-sabine-hossenfelder-and-by-tom-lehrer/. But an upsurge in academic integrity violations made this all stop being funny about 15 years ago... Go over the "0,1/3,2/3,1" philosophy...]

II. $L = \{x \in \{a,b\}^* : x^R = x\}$, where x^R means x reversed, e.g., $abbab^R = babba$. [What is ϵ^R?] That is, L is the language of strings that are **palindromes** and has the standard name PAL.

Take S = __a^*b__. Clearly S is infinite.

Let any $x, y \in S$ $(x \neq y)$ **be given**. Then we can write x = __$a^m b$__ and y = __$a^n b$__ where ____$m, n \geq 0$____ and $m \neq n$.

Take z = __a^m__.

Then $L(xz) \neq L(yz)$ because ___$xz = a^m b a^m \in$ PAL but $yz = a^n\, b\, a^m$ which is not in PAL because $m \neq n$ and the single b prevents any other possible way of "parsing" yz as a palindrome___.

Because x, y are an arbitrary pair of strings in S, this shows that S is PD for L, and since S is infinite, it follows that L is nonregular by the Myhill-Nerode Theorem.

We did not need the "wlog." provision this time---but you can always take it even if you don't need it. We also could have started with $S = a^*$ and made the b the first char in z. Why did I put the b "up front" in S? It is to emphasize its importance and help avoid a common mistake of forgetting it altogether. The mistake (in this case---it pops up in others too) is to think that $a^m \cdot a^n$ is not a palindrome whenever $m \neq n$. That may be true with your breakdown but there could be others. E.g. $a^3 a^5 = a^4 a^4$ which is now clearly a palindrome. Indeed, a^{m+n} is *always* in PAL.

III. $L = \{x \in \{(,)\}^* : x \text{ is balanced}\}$. What does "balanced" mean? [Discuss if time allows, but this will be more important a week from now.] This language is often called BAL. It is in fact "isomorphic to" the language of the "unlimited spears" and dragons game *when you win only if you leave the dungeon with zero spears*. E.g., if you are holding 5 spears, you need 5 'closing dragons" to balance out. [Re-use one of the scripts above interactively to finish up.]