-> Corpus

-> Sentences

-> Tokens.

-> Sentence Segmentation:

   -> ?, !, .   ~ Unambigious.

-> Challenges :

? & ! are hard unambigious.

But " . " is not.

-> Abbreviations : Dr. , Mr. , Mrs. , M.p.h. , U.SA

       Numbers.   3.14 , 6.208 , 99.75%.

-> Approach :   Build a classifier.

    ↳ For each " . "
    ↳ Decide End-Of-Sentence / Not End Of Sentence
    ↳ Handwritten rules, Regular Expressions..

-> Decision Tree : E - O - S

Lots of blank lines

Y → E-o-s

N → **Final Punctuation** ? !

    Y → E-o-s

    N → **Final punctuation is period**

etcetera
etc...

    Y → **Ends with abbr oretc**

        Y → Not E-o-s

        N → E-o-s

    N → Not E-o-s
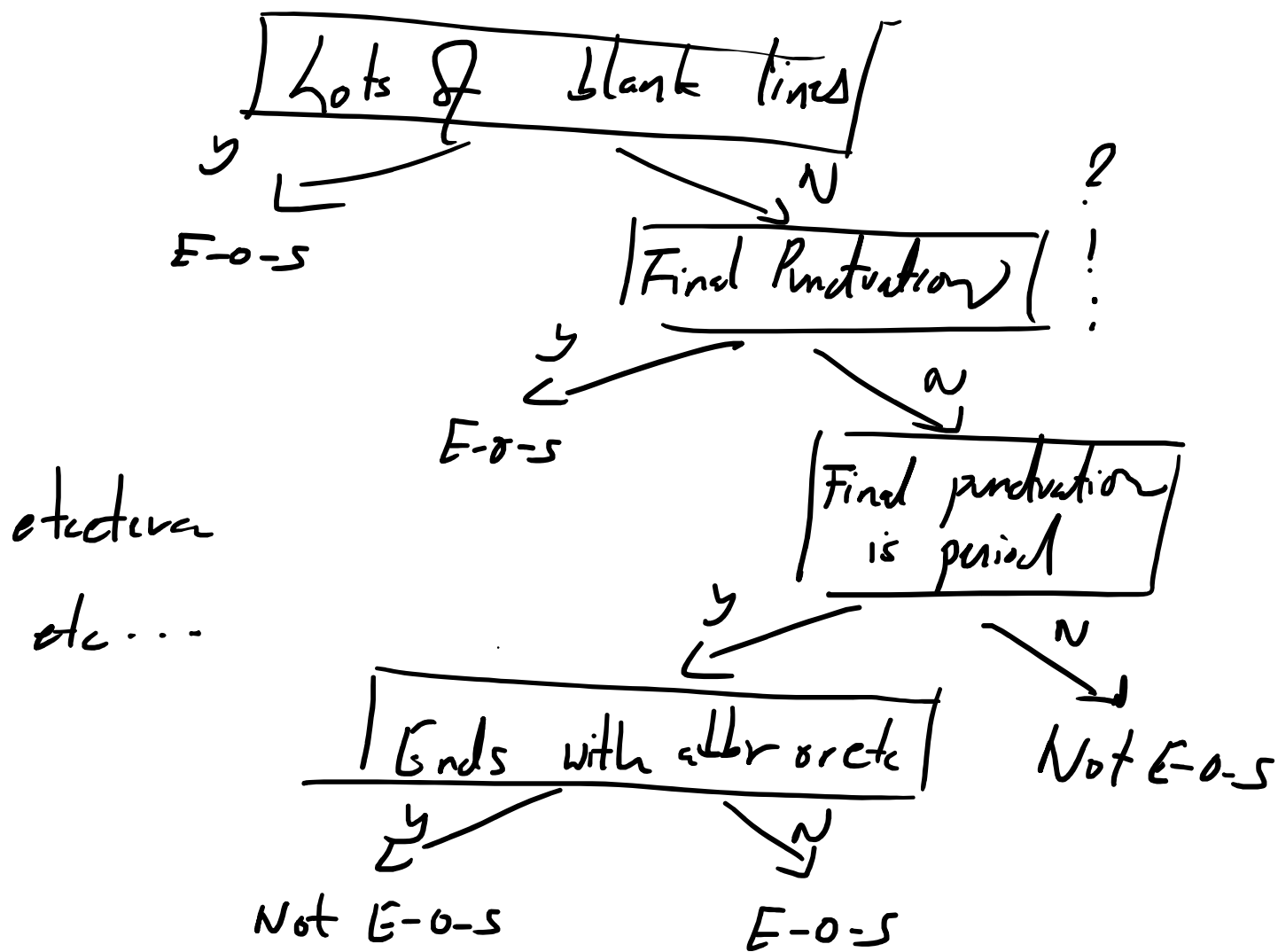
→ Case of word with ".": Upper & Lower, Number

    Dr. Otto Octavius   Dr. norman Osbourn

→ Numeric Features:

Ls length of word

dollars store

dollar store
tienda de opa

→ Does period   lie   at the end of sentence
   or beginning

↳ Dr. Otto Octavius experiments   on himself
   with 4 pair of limbs .

→ Implementing Decision Tree :   Age / DOB

   → If - Else Statement

   → Choosing the features   Dr. Norman Osborn .

   → Numeric Thresholds .

   → Non-Axiomatic Exp

→ Popular Algorithms :   Logic T/F
   ID3, CART, Ch.S