

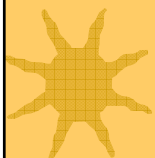
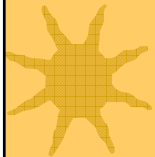
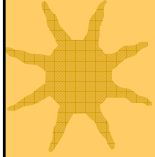


CMSC 723 / LING 645: Intro to Computational Linguistics

November 14, 2007

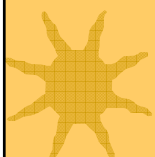
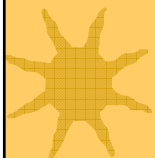
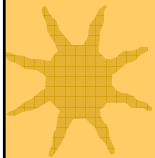
Lecture 11 Additional Notes
Earley Revisited
Prof. Bonnie J. Dorr

Co-instructor: Nitin Madnani
TAs : Hamid Shahri, Alexandros Tzannes



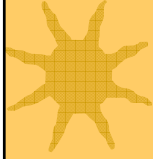
Recall that there are Three Main Sub-Routines of Earley Algorithm

- ★ Predictor: Adds predictions into the chart.
- ★ Completer: Moves the dot to the right when new constituents are found.
- ★ Scanner: Reads the input words and enters states representing those words into the chart.

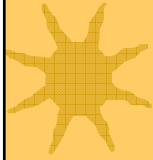




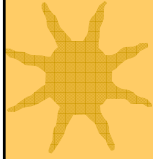
Predictor the way I presented it



★ Intuition: create new state for top-down prediction of new phrase.



★ Applied when non part-of-speech non-terminals are to the right of a dot: $S \rightarrow \bullet VP [0,0]$



★ Adds new states to *current* chart

- One new state for each expansion of the non-terminal in the grammar

$VP \rightarrow \bullet V [0,0]$

$VP \rightarrow \bullet V NP [0,0]$

This is why re-grades were required! By this definition we can never predict Verb \rightarrow * book

★ Formally, let B be a non-POS category:

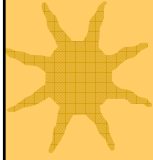
$S_j: A \rightarrow \alpha \bullet B \beta, [i,j]$

$S_j: B \rightarrow \bullet \gamma, [j,j]$

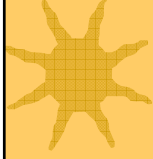
HOWEVER ...



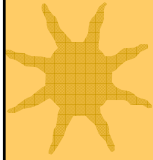
Scanner the way I presented it



★ Intuition: Create new states for rules matching part of speech of next word.



★ Applicable when a word is to the right of a dot: $V \rightarrow \bullet book [0,0]$ 'Book...'



★ Looks at current word in input

★ If match, adds state(s) to *next* chart

$V \rightarrow book \bullet [0,1]$

★ Formally, let B be a word:

$S_j: A \rightarrow \alpha \bullet B \beta, [i,j]$

$S_{j+1}: A \rightarrow \alpha B \bullet \beta, [i,j+1]$

Note that this is why I expected to see predicted lexical rules in the previous state set! But this conflicts with the definition of Predict because we never get a rule like this in the first place!



Example: State Set S_0 for Parsing “Book that flight” using Grammar for L_1

Chart[0]	State	Item Set	Label	Category
S0	$\gamma \rightarrow \bullet S$	[0,0]	Dummy start state	
S1	$S \rightarrow \bullet NP VP$	[0,0]	Predictor	
S2	$S \rightarrow \bullet Aux NP VP$	[0,0]	Predictor	
S3	$S \rightarrow \bullet VP$	[0,0]	Predictor	
S4	$NP \rightarrow \bullet Pronoun$	[0,0]	Predictor	
S5	$NP \rightarrow \bullet Proper-Noun$	[0,0]	Predictor	
S6	$NP \rightarrow \bullet Det Nominal$	[0,0]	Predictor	
S7	$VP \rightarrow \bullet Verb$	[0,0]	Predictor	
S8	$VP \rightarrow \bullet Verb NP$	[0,0]	Predictor	
S9	$VP \rightarrow \bullet Verb NP PP$	[0,0]	Predictor	
S10	$VP \rightarrow \bullet Verb PP$	[0,0]	Predictor	
S11	$VP \rightarrow \bullet VP PP$	[0,0]	Predictor	

J&M do not add the following state to the chart: $Verb \rightarrow * book [0,0]$ which means



Example: State Set S_1 for Parsing “Book that flight”

Chart[1]	State	Item Set	Label	Category
S12	$Verb \rightarrow book \bullet$	[0,1]	Scanner	
S13	$VP \rightarrow Verb \bullet$	[0,1]	Completer	
S14	$VP \rightarrow Verb \bullet NP$	[0,1]	Completer	
S15	$VP \rightarrow Verb \bullet NP PP$	[0,1]	Completer	
S16	$VP \rightarrow Verb \bullet PP$	[0,1]	Completer	
S17	$S \rightarrow VP \bullet$	[0,1]	Completer	
S18	$VP \rightarrow VP \bullet PP$	[0,1]	Completer	
S19	$NP \rightarrow \bullet Pronoun$	[1,1]	Predictor	
S20	$NP \rightarrow \bullet Proper-Noun$	[1,1]	Predictor	
S21	$NP \rightarrow \bullet Det Nominal$	[1,1]	Predictor	
S22	$PP \rightarrow \bullet Prep NP$	[1,1]	Predictor	

This rule appears out of the blue!!! This is WRONG. Does not match Earley (1970). I wrote to J&M and they are changing the book right now.



There are THREE alternatives ... only ONE of which matches the original Earley Description.

* **Approach 1.** J&M's current approach:

- Predict: Lexical Rules are not predicted. Just move to the next state set when the dot is to the left of a POS category.
- Scan: Lexical Rules are inserted in the next state set (out of the blue) with the dot at the right edge.
- Issue: Lexical Rules come out of nowhere, which is not the Earley 1970 specification.

* **Approach 2.** Notes presented by Dorr THIS YEAR:

- Predict: Lexical Rules are predicted with dot at left edge. (POS categories are considered non-terminals.) [Note: This was omitted in this year's slides, but was the intent.]
- Scan: Lexical Rules are carried down into the next state set, with the dot moved over to the right edge.
- Issue: We would have to add a lexical rule for every dictionary entry whose POS matches the input word. Messy! Doesn't match Earley 1970.

* **Approach 3.** Notes presented by Dorr LAST YEAR:

- Predict: Lexical Rules are never added to the chart! Only predict a rule when the dot is to the left-hand side of a non-terminal. (POS categories are considered terminals.)
- Scan: Carry down all rules to the next state set where the POS category on the right of the dot matches the POS of the current input. Move dot over the POS category.
- MATCHES Earley 1970.

To be adopted by J&M 2008



How does that change the chart?

Chart[1]	S12	Verb → book •	[0,1]	Scanner	Completer
	S13	VP → Verb •	[0,1]	Scanner	Completer
	S14	VP → Verb • NP	[0,1]	Scanner	Completer
	S15	VP → Verb • NP PP	[0,1]	Scanner	Completer
	S16	VP → Verb • PP	[0,1]	Scanner	Completer
	S17	S → VP •	[0,1]		Completer
	S18	VP → VP • PP	[0,1]		Completer
	S19	NP → • Pronoun	[1,1]		Predictor
	S20	NP → • Proper-Noun	[1,1]		Predictor
	S21	NP → • Det Nominal	[1,1]		Predictor
	S22	PP → • Prep NP	[1,1]		Predictor