

# Top-down Chart Parsing: the Earley algorithm

Parsing  
ISCL-BA-06

Çağrı Çöltekin  
ccolt@infsa.uni-tuebingen.de

University of Tübingen  
Seminar für Sprachwissenschaft

Winter Semester 2020/21

version: 27-Nov-2020 12:58

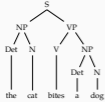
# Parsing so far

- We can formulate parsing as
  - Top-down: begin with the start symbol, try to produce the input string to be parsed
  - Bottom-up: begin with the input, and try to reduce it to the start symbol
- Another aspect of a parser is its directionality. Two choices are:
  - Directional: parses processes the input left to right (right to left is also possible, but rarely used)
  - Non-directional: order is not important, typically require all input to be in memory before processing

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 1 / 14

# Top-down parsing as search



- S → NP VP
- NP → Det N
- VP → V NP
- VP → V
- Det → a
- Det → the
- N → cat
- N → dog
- V → bites

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 2 / 14

# Earley algorithm

- Earley algorithm is a top down (and left-to-right) parsing algorithm (Earley 1970)
- It allows arbitrary CFGs
- Keeps record of constituents that are predicted using the grammar (top-down) in-progress with partial evidence completed based on input seen so far at every position in the input string
- Time complexity is  $O(n^3)$

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 3 / 14

# Earley chart entries (states or items)

Earley chart entries are CF rules with a 'dot' on the RHS representing the state of the rule

- $A \rightarrow \alpha \beta \mid, \mid$  predicted without any evidence (yet)
- $A \rightarrow \alpha \bullet \beta \mid, \mid$  partially matched
- $A \rightarrow \alpha \beta \bullet \mid, \mid$  completed, the non-terminal A is found in the given span

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 4 / 14

# Earley algorithm: an informal sketch

- Start at position 0, predict S
- Predict all possible states (rules that apply)
- Read a word
- Update the table, advance the dot if possible
- Go to step 2
- If we have a completed S production at the end of the input, the input is recognized

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 5 / 14

# Earley algorithm: three operations

- Predictor** adds all rules that are possible at the given state
- Completer** adds states from the earlier chart entries that match the completed state to the chart entry being processed, and advances their dot
- Scanner** adds a completed state to the next chart entry if the current category is a POS tag, and the word matches

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 6 / 14

# Earley parsing example (chart[0])

state	rule	position	operation
0	$\gamma \rightarrow \bullet S$	[0,0]	initialization
1	$S \rightarrow \bullet NP VP$	[0,0]	predictor
2	$S \rightarrow \bullet Aux NP VP$	[0,0]	predictor
3	$NP \rightarrow \bullet Det N$	[0,0]	predictor
4	$NP \rightarrow \bullet NP PP$	[0,0]	predictor
5	$NP \rightarrow \bullet Prn$	[0,0]	predictor

Note: the chart[0] is independent of the input.

- S → NP VP
- S → Aux NP VP
- NP → Det N
- NP → Prn
- NP → NP VP
- VP → V NP
- VP → V PP
- VP → VP PP
- PP → Prp NP
- N → duck
- N → park
- V → duck
- V → ducks
- V → saw
- Prn → she | her
- Prp → in | with
- Det → a | the
- Aux → does | has

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 7 / 14

# Earley parsing example (chart[1])

state	rule	position	operation
6	$Prn \rightarrow she \bullet$	[0,1]	scanner
7	$NP \rightarrow Prn \bullet$	[0,1]	completer
8	$S \rightarrow NP \bullet VP$	[0,1]	completer
9	$NP \rightarrow NP \bullet PP$	[0,1]	completer
10	$VP \rightarrow \bullet V NP$	[1,1]	predictor
11	$VP \rightarrow \bullet V$	[1,1]	predictor
12	$VP \rightarrow she \bullet   her$	[1,1]	predictor
13	$PP \rightarrow \bullet Prp NP$	[1,1]	predictor

- S → NP VP
- S → Aux NP VP
- NP → Det N
- NP → Prn
- NP → NP PP
- VP → V NP
- VP → V
- VP → VP PP
- PP → Prp NP
- N → duck
- N → park
- V → duck
- V → ducks
- V → saw
- Prn → she | her
- Prp → in | with
- Det → a | the
- Aux → does | has

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 8 / 14

# Earley parsing example (chart[2])

state	rule	position	operation
14	$V \rightarrow saw \bullet$	[1,2]	scanner
15	$VP \rightarrow V \bullet NP$	[1,2]	completer
16	$VP \rightarrow V \bullet$	[1,2]	completer
17	$S \rightarrow NP VP \bullet$	[0,2]	completer
18	$NP \rightarrow \bullet Det N$	[2,2]	predictor
19	$NP \rightarrow \bullet NP PP$	[2,2]	predictor
20	$NP \rightarrow \bullet Prn$	[2,2]	predictor

- S → NP VP
- S → Aux NP VP
- NP → Det N
- NP → Prn
- NP → NP PP
- VP → V NP
- VP → V
- VP → VP PP
- PP → Prp NP
- N → duck
- N → park
- V → duck
- V → ducks
- V → saw
- Prn → she | her
- Prp → in | with
- Det → a | the
- Aux → does | has

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 9 / 14

# Earley parsing example (chart[3])

state	rule	position	operation
21	$Det \rightarrow a \bullet$	[2,3]	scanner
22	$NP \rightarrow Det \bullet N$	[2,3]	completer

- S → NP VP
- S → Aux NP VP
- NP → Det N
- NP → Prn
- NP → NP PP
- VP → V NP
- VP → V
- VP → VP PP
- PP → Prp NP
- N → duck
- N → park
- V → duck
- V → ducks
- V → saw
- Prn → she | her
- Prp → in | with
- Det → a | the
- Aux → does | has

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 10 / 14

# Earley parsing example (chart[4])

state	rule	position	operation
23	$N \rightarrow duck \bullet$	[3,4]	scanner
24	$V \rightarrow duck \bullet$	[3,4]	scanner
25	$NP \rightarrow Det N \bullet$	[2,4]	completer
26	$VP \rightarrow V NP \bullet$	[1,4]	completer
27	$S \rightarrow NP VP \bullet$	[0,4]	completer

- S → NP VP
- S → Aux NP VP
- NP → Det N
- NP → Prn
- NP → NP PP
- VP → V NP
- VP → V
- VP → VP PP
- PP → Prp NP
- N → duck
- N → park
- V → duck
- V → ducks
- V → saw
- Prn → she | her
- Prp → in | with
- Det → a | the
- Aux → does | has

C. Çöltekin, MPI / University of Tübingen

Winter Semester 2020/21 11 / 14

## Earley parsing: summary

- Complexity (asymptotic) is the same as CKY
  - time complexity:  $O(n^3)$
  - space complexity:  $O(n^2)$
- Our example shows recognition, we need to maintain back links for parsing
- Again, Earley chart stores a parse forest compactly, but extracting all trees may require exponential time

## Summary

- The early parser is a top-down parser with bottom-up filtering (or, you can also view it the other way around)
- The parser improves over a backtracking parser by
  - dynamic programming: not re-computing the subtrees
  - filtering: not generating hypotheses (predictor) that does not lead to useful
- It can process any CFG (no need for CNF)
- There is a nice relation between CKY and Earley (next week)

Next:

Mon Exercises / assignment / discussion

Thu Deterministic parsing

## An exercise

Construct the CKY and Earley charts for the sentence below

The duck saw her in the park

## Recommended grammar:

S	→ NP VP	PP	→ Prp NP
NP	→ Det N	N	→ park
NP	→ Prn	N	→ duck
NP	→ NP PP	V	→ is
NP	→ NP S	V	→ saw
VP	→ V NP	Prn	→ she
VP	→ V	Prp	→ in
VP	→ VP PP	Det	→ the

## Acknowledgments, references, additional reading material



Earley, Jay (Feb. 1970). "The Efficient Context-Free Parsing Algorithm". In *Communications of the ACM*. 13(2): 121, pp. 94-102. <https://doi.org/10.1145/321077.321078>