# Parallel Computation in a Free Merge World 

Sandiway Fong<br>University of Arizona<br>with Jason Ginsburg<br>Osaka Kyoiku University

Acknowledgement: Dr. Nobuyoshi Asai, U. of Aizu, Japan for the test platform

## Contents

- Linguistic Framework and Combinatorics
- Parallelizing the Framework and Results
- Feedback: Improving the Framework
- Parallelism: Job size

Linguistic Framework and Combinatorics

## Phrase Structure Computation


[Joint work with Jason Ginsburg]
Theoretical basis:

- Chomsky (2007) and Oishi (2015)
- nominal and determiner phrase structure (n*/d*-root) parallels verbal phrase structure ( $v^{*}$-root)
- Pair Merge (PM) analysis:
- <Determiner, Noun>
- forced by non-head determiner
- cf. *\{\{d, root $\},$ nn, root $\}\}$
- (unlabeled Set Merge (SM))
- Relabeling: Cecchetto \& Donati (2015)
- my friend $=\{\mathrm{me},\{$ 's, friend $\}\}$
- friend of mine $=\{$ friend, $\{\mathrm{me},\{' \mathrm{~s}$, friend $\}\}\}$


## Phrase Structure Computation



- Example input: (a list of heads):
[friend, n, [me, n, 's, d*], n*,[the, d]]
- Combinatorial Task: recursively apply operations:

1. External Set Merge (ESM): form $\{H, \alpha\}$; SO: $\alpha$, Input: [ $H, .$.
2. Internal Set Merge (ISM $(\beta))$ : form $\{\beta, \alpha\}$; SO: $\alpha$ and $\beta \subset^{+} \alpha$
3. External Pair Merge (EPM): form $\langle H, \alpha\rangle$; SO: $\alpha$, Input: [H,..]
4. Internal Pair Merge (IPM): form $\langle\alpha, \beta\rangle$; SO: $\alpha$ and $\beta \subset+\alpha$ with constraints such as:
5. $\quad{ }^{*}<\beta[!\mathrm{F}], \alpha>$ where ! $\mathrm{F}=$ unvalued feature F
6. ${ }^{*} \operatorname{ISM}\left(\beta_{i}\right) \operatorname{ISM}\left(\beta_{i}\right)$; i.e. can't ISM same $\beta_{i}$ twice, etc.

- Example output:
<\{the, d\}, \{\{friend, <\{\{\{me, n\}, \{\{me, n\}, 's\}\}, d*\}, \{friend, n\}>\}, n*\}>
- Example of questions answered by computation:
A. is this the shortest derivation? YES
B. are there other possible derivations?

YES, only longer ones...

## Manually Guided Derivation...



## Combinatorics for example ${ }^{\dagger}$

+ naïve version

Log10(\# SOs Generated) vs. \# Operations


- logscale y-axis:
- e.g. $6=10^{6}=$ million
- 15 operations deep:
- 25 million SOs generated
- 1 convergent SO
- (see previous slide)
- 16 operations deep:
- 250 million SOs
- 3 spurious SOs
- (see next slides)


## Parallelizing the Framework and Results

## Two stages of parallel processing

## Example:



- Stage 1: breadth-first derivation tree search (BFS):
- $\mathrm{SO}_{1} . . \mathrm{SO}_{6}$ are incomplete SOs that can be expanded further
- represent dead-ends
- go as deep as necessary to generate the number of starter SOs needed
- example: going 10 deep nets us 1743 SOs



## Step 2: Run threads in parallel


our concern: load balancing

- threads binned by job size
- 1743 jobs (threads)
- produced by initial BFS to 10 operations deep
- each job (go 6 deep)
- 10x range in job size observed:
- 50,000 SOs to 600,000 SOs
- 72\% of jobs small:
- belong to the 3 smallest bins, i.e. 0-150,000 SOs


## Parallel Speedup



16 CPUs: with HTT, 32 logical cores

- Runtime:
- Single thread: 759 (secs)
- 32 threads: 57 (secs)

13x speedup observed

- Speedup: Amdahl's Law
- (theoretical limit)



## Is Hyper Threading (HTT) useful?




- Hyper Threading Technology:
- each core has two sets of registers
- hide memory latency
- Test platform:
- Intel Xeon E5-2687W HTT-capable (2U), 128GB RAM
- total of 16 cores ( 32 logical cores)
- region 8-32 threads:
- shortest overall runtimes are all achieved by with HTT
- region 18-32 threads:
- averages about $11.4 \%$ improvement over no HTT
- region 8-16 threads:
- no HTT is $5.5 \%$ better


## With Workspace (WS) Precomputation

- Results shown earlier, e.g. 57 (secs), were actually computed on a nonnaïve model
- region $10^{7}-10^{8}$ SOs ("wall"):
- too much for the test platform: approx. 4.5 hours CPU time
- Non-naïve model:
- pre-compute sub-Workspace (WS) SOs
- \# operations required reduced
- free Merge then is substantially easier
"walk back from the wall"



## Workspace (WS) Precomputation

- Example:
instead of
[friend, n, [me, n,'s, d*], n*, [the, d]]
actually compute with
[friend, $n,\left\{\{\{m e, n\},\{\{m e, n\}, ~ ' s\}\}, d^{*}\right\}, n^{*}$, \{the, d\}]
- i.e. use pre-computed mappings:

1. $\left.\operatorname{dme}_{\left.d^{*}\right\}}, \mathrm{n}, \mathrm{s}, \mathrm{d}^{*}\right] \mapsto\{\{\{\mathrm{me}, \mathrm{n}\},\{\{\mathrm{me}, \mathrm{n}\}, \mathrm{s}\}\}$,
2. $\quad[$ the,$d] \longmapsto\{$ the, $d\}$

- Results:
- Depth 6: \#SOs: 2,324; 1 solution
- Depth 7: \#SOs: 18,202; 2 solutions, etc..



## Workspace (WS) Precomputation:



## Improve the Framework



- Parallel processing allows us to discover 5 extra analyzes at depth 9 (out of $\approx 10^{6}$ SOs) 10 x quicker...


## Extra Analyses Uncovered: Depth 7 \& 8



Depth = 7
Analysis:
Extraneous ISM of \{friend, n$\}$ to the edge of friend
$<\phi, \phi>$ because
\{friend, n \} and friend have identical $\phi$-features


Depth = 8


|  |  |  |  |
| :---: | :---: | :---: | :---: |
| Step | Branch | Op | SO |
| 1 | - | - | friend |
| 2 | 1 | esm | \{friend,n!case\} |
| 3 | 3 | esm |  |
| 4 | 7 | ism |  |
| 5 | 5 | ism | \{\{friend,n!case\},\{friend,\{\{friend,n!case\},\{\{\{me,n\},\{\{me,n\},16\}\},d*\}\}\}\} |
| 6 | 6 | esm |  |
| 7 | 1 | epm |  |

Spellout heads: [the,friend,of,'s,me]
Final output: [the.friend,of,mine]

| LIs:[friend, n !case, $\{$ \{ $\left.\{\mathrm{me}, \mathrm{n}\},\{\{\mathrm{me}, \mathrm{n}\}, \mathrm{s}\}\}, \mathrm{d}^{*}\right\}, \mathrm{n}^{*}$ !case, , the,d\}] Derivation \#1 |  |  |  |
| :---: | :---: | :---: | :---: |
| Step | Branch | Op | SO |
| 1 | - | - | friend |
| 2 | 1 | esm | \{friend,n!case\} |
| 3 | 3 | m |  |
| 4 | 8 | ism | \{\{friend,n!case\}, ,\{frriend,n!case\},\{\{\{me,n\},\{\{me,n\},s\}\},d*\}\}\} |
| 5 | 7 | ism |  |
| 6 | 6 | ism | \{\{friend,n!case\},\{friend, $\left\{\right.$ friend,n!case\}, $\left\{\left\{\right.\right.$ friend,n!case\}, $\left.\left.\left.\left.\left\{\{\{\mathrm{me}, \mathrm{n}\},\{\{\mathrm{me}, \mathrm{n}\}, \mathrm{s}\}\}, \mathrm{d}^{*}\right\}\right\}\right\}\right\}\right\}$ |
| 7 | 6 | esm | \{\{\{friend, n$\},\left\{\right.$ friend, $\{$ \{friend, , n$\},\left\{\{\right.$ friend, n$\left.\left.\left.\left.\left.\},\left\{\{\{\mathrm{me}, \mathrm{n}\},\{\{\mathrm{me}, \mathrm{n}\}, \mathrm{s}\}\}, \mathrm{d}^{*}\right\}\right\}\right\}\right\}\right\}\right\}, \mathrm{n}^{*}$ !case $\}$ |
| 8 | 1 |  | $<\{$ the, d$\},\left\{\left\{\{\right.\right.$ friend, n$\}$, friend, $\left\{\{\right.$ friend, n$\},\left\{\{\right.$ friend, n$\left.\left.\left.\left.\},\left\{\{\{\mathrm{me}, \mathrm{n}\},\{\{\mathrm{me}, \mathrm{n}\}, \mathrm{s}\}\}, \mathrm{d}^{*}\right\}\right\}\right\}\right\}\right\}, \mathrm{n}^{*}!$ case $\}>$ |
| Spellout heads: [the,friend,of,'s,me] Final output: [the,friend, of, mine] |  |  |  |
|  |  |  |  |

## Extra Analyses Uncovered: Depth 9



## Improve the Framework: Theory Adjustment



## [Joint work with Jason Ginsburg]

- block licensing of extraneous analyses
- Previously:
- all Case valuation done through Agree
- 's analyzed as a pair: root 's + d* (categorizer)
- Now:
- distinguish Inherent from Structural Case
- Inherent Case does not involve $\phi$-features: means $\langle\phi, \phi>$ labeling not available
- Structural Case involves $\phi$-feature valuation, and Nom (or Acc) Case for C/T/ (or $\mathrm{v}^{*} / \mathrm{R}$ )
- 's analyzed as a single re-categorizing head: i.e. $\mathrm{n} \rightarrow \mathrm{d}$


## Improve the Framework: Theory Adjustment



## Improve the Framework: Combinatorics



- Orange line: adjusted theory
- one solution @ 6
- no extraneous solutions @710
- fewer SOs hypothesized
- Blue line: original theory
- one solution @ 6
- one solution @ 7, 8
- five solutions @ 9


## Parallelism: Job size

## Parallel Processing Task Size



- Example:
- say we want to search to depth 11 in parallel
- What is the best way to divvy up the search?
- We can perform the same search by expressing:
- 27 threads, each 7 deep
- 121 threads, each 6 deep
- 610 threads, each 5 deep
- 3750 threads, each 4 deep
- Tradeoff:
- thread overhead vs. load balancing
- (task size not a constant)
- RAM wrt. \# active threads limits task size


## Parallel Processing Task Size: Results



- Conditions:
- blue line: 16 CPUs used (no HTT); 16 active threads
- green line: same 16 CPUs + HTT; 32 active threads
- RAM: 128GB capacity
- Best results:
- HTT on
- 610 threads (from 6 deep initially), each job is 5 deep
- used $\approx 30 G B$ RAM
- cf. 4.7 split used $\approx 88 \mathrm{~GB}$
- cf. 6.5 split used $\approx 15 \mathrm{~GB}$


## Conclusions

- Application is parallel-friendly
- search: multiple possible operations
- speed-up results: 13x on 32 logical cores
- Speed-up allows us to search deeper
- beyond a basic analysis

- Improve the theory
- eliminate extraneous analyzes

Appendix

## Extra Analyses Uncovered



## Extra Analyses Uncovered



## Extra Analyses Uncovered




## Extra Analyses Uncovered



## Extra Analyses Uncovered




Depth $=9$

