## A self-evaluation of LF graphs

## Predicate-argument structure

- uses about 20-30 thematic roles for arguments classified by their semantic relation to the predicate
* e.g., common subjectroles
- AGENT, GOGNZER EXPERIENGER THEME
- other roles
- AFEGTED, 1 O QQQ COTHEME SOURCE PROPERTVVAEUEOF
plus vague role ASSOC-WM
- Note most adjuncts treated as modifiers
- He put it . In the box on the box, beside the box, to the right of the box, five feet away from the box,

Gold Standard Arc Score: Precision 61\% Recall 58\%

## Attachment Correctness

: SPICES-HERBS HERB) (KIND (:* VEGETABLE VEGETABLE)) (IMPRO REFERENTIAL-SEM)

- Attachments indicated by MOD dependencies
-Text 6 was our worst paragraph We only got one of these right (the easy onel)


## Self-score: 3

## Anaphora \& Reference

* LF-GRAPH includes different referring expressions
* explicit: THE * : PRO *, THIS/HAT,
- implicit: IMPRO the arger truck - > larger than what?
- Antecedents are indicated by GOREFlinks
- Due to limitations of graphical format, COREF links are hard to interpret
((PARAGRAPH:TERMS
((SENTENCE:UTTNUM 0:TERMS
((SPEECHACT.V110743 SA:TELL:CONTENT:V105911)
Assessment hies System cid mot do intra-clausal

referemcessinglenter clausal handled retivelyavinn

(PRO V111908 (*.PERSON HE):CONTEXT-RELHEGOBEF V105814)
(F V112151 (: READ READ):THEME V112755)
(A V112755 (: BOOKBOOK) )) )) ))


## Self-score on Performance: 3

## Word Sense Disambiguation



- subcategorization
- generic selectional restrictions
O. g. LFETAKE-TIME Requires a LF:DURATION
among those left hand-set priors preference


## Gold Standard Score: Precision 78\% recall 68\%

## Quantification

- LF allows arbitrary generalized quantifieis
* e.g., MOST, EVERY SOMEA MOST AL A, ATHE
* also includes KIND, BARE
- LF graph produces an unoerspecifiea scope
- scopings allowed are equivalent to practical MRS

The seven texts had almost no classic quantifier scoping examples!
Uevery each any nany few didn't occur some somewhat several occurred once 26 times, the 50 times

Gold score on Specifiers: Precision 76\% Recall 67\%

## Negations, modals, conditionals, disjunction

- Negation and Modal Auxiliaries.
- captured but no scoping attempted
$\because$ he must not have eaten: $=$ ?
(LF:FV118630 (EFOONSUMEW EAT) AGENTV118250
STMA (W\&TENSE W\&PRES) (WMEGATION + )
(WMODALTY (E LPMUST WMMUST) (WPPERF + ))
- Negated NPS ho ©og
V. $\rightarrow$ (LFiQUANTIEERV120670 (ELEANIMA W:DOG) QUAN W*NONE)

Modal Verbs: "l believe he lied":
2 (LF: V389942 (EF:BELEVEW:BELEVE) THEME V389975:COGNIZER V389509:TMA (W:TENSE W:PRES))
(LF:F V/389975 ( $(*$ LF: STATEMENT W: $:$ LIE) :AGENT V389966 :TMA ((W:TENSE W:PAST))

# Negations, modals, conditionals, disjunction 

- Conditionals
(LF:F V278126 (* LFSPURCHASE W/3BUY) THHEMEV278194 AGENTV27.7681

(LF:PRO V278194 (*: LF: REFERENTAE SEM W:T):CONTEXIREEW:IT)


(LF: E V277648 (\% LF NONVERBALEXPRESSION WWMMLE) AGENT V277481:TMA
(W:TENSE W:SRESI))
(LF:PROV277481: (6: FRPERSONWMOU) CONTEXTREWYOU)
If you smile bll buy th
Only one explicit conditional in texts - we missed it due to unknown word "negligible"

Self-score on Performance: 2

## Tense and Aspect

* We extract the information reliably but process no further
$\leadsto$ Che must not have eaten" $->$
(LF:F V118630 (*F CONSUMEWGEAT) AGETV118250
CTMA (WATENSE WAPRES) (WHNEGATION+) (WHODALIY ( (GFRMUST W\&MUST) (WAPERF+))


## Self score on what we do: 4

## Plurals



- Explicit representation of sets introduced by plurals - separation of set modifiers vs element modifiers
- Constructed sets from conjunctions
- Set based concepts. "The gun crew"


## Self-score on Performance: 4

## Comparison Phrases

 (EF: F VR827398 (at LE:MERE VA WWHARE) FUNGTN


"The larqer truck than that

- Comparatives are between a thing ( gum) and another thing (cos)


## Time Expressions

* Grammar has specific rules for temporal expressions
* Time ranges: "in the 1930s"
- (LF:THE V287214 LF:TIMERANGEDECADE 1930)
- Time durations: 30 years"
- (LF:A V290689 (*LF:QUANTITY:FURATION) UNIT (*EF:TIME UNT W:YEAR):AMOUNT:30)
"Dates: July 20 2006"
- (LF:THEV303460 EFTME LOCOYEAR2006:MONTH (UE LFMONTHNAMEW:UULYODAY:20
"Complex phrases In the mid 80s"
W. (EF:V331924 ( (* LF:TIME SPANREL W:IN) :OF V336557 :VAL V332631)
n. (LF:THE V332631 LFTIMERANGEDECADE 80 :MODS (V332611))

4. (LF::F.V332611:(*:LF:STAGE-VAL W:.MID):OF V332631)

Self-SCOre: 4 (missed in the past 30 years")

## Measurement Expressions

* Measurements produce instances of QUANTITY FREQUENCY etc
" "125 m high"

 UNIT:(6:LEMENGTH:UNIT:WMMETBE):AMOUNT:125)
, $10 \mathrm{~m} / \mathrm{s}^{\prime}$
(LF V3861514 YREQUENGY VAEV385606OFVB88082)
- (LFAVALUE V385606 LFMBEQUENCY

OVER:PERIOD:(GESTME NTERVALW:SECOND) :REPEATS V384377.

- (LFA V38437Z: (ELEUANTITYF:LINEAR-S)

UNIT( (6 LEFEENGTH:UNTT W:METER) :AMOUNT 10)
Self-score on Performance: 3 (didn't know " $\mathrm{m} / \mathrm{s}$ " and a bug prevented parsing $\$ 10,000^{\prime \prime}$ )

## Question Interpretation

(SPEECHACT WH-QUESTION)

(THE (:* TIME-INTERVAL DURATION))

WHAT

## "What is the duration of the fall?"

- Questions parsed into literal speech acts
- Broad coverage of question types, wh movement, except for Show ADJ forms. "how long did it take?" Self-score on Performance: 3


## Clarity



This is a religious question

## I think graphs are more accessible thanlogic

Some Evidence:

- It is mpossible to hand build a gold LF for a complex sentence directing in our linear TRIPS LF form
- In contrast, building $E$ graphs is quite intuitive

Self-score: vote your heart!

## Performance Against Gold Standards

| Text | Base System |  | Final System |  |
| :--- | :---: | :---: | :---: | :---: |
|  | Prec | Recall | Prec | Recall |
| 1 "physics" | $70.1 \%$ | $70.1 \%$ | $70.7 \%$ | $76.0 \%$ |
| 2 "cancer" | $62.1 \%$ | $71.9 \%$ | $62.8 \%$ | $72.8 \%$ |
| 3 "dining" | $86.7 \%$ | $90.4 \%$ | $90.8 \%$ | $94.6 \%$ |
| 4 "dogs" | $63.0 \%$ | $68.6 \%$ | $63.8 \%$ | $67.7 \%$ |
| 5 "guns" | $55.0 \%$ | $64.0 \%$ | $60.3 \%$ | $69.5 \%$ |
| 6 "gardens" | $47.4 \%$ | $53.6 \%$ | $56.2 \%$ | $62.1 \%$ |
| 7 "wind" | $\mathrm{n} / \mathrm{a}$ | $\mathrm{n} / \mathrm{a}$ | $65.8 \%$ | $76.3 \%$ |
| Average | $64.1 \%$ | $69.7 \%$ | $67.1 \%$ | $74.1 \%$ |

