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 subject roles
 AGENT, COGNIZER, EXPERIENCER, THEME
 - other roles
 - AFFECTED, TO-LOC, CO-THEME, SOURCE, PROPERTY, VALUE, OF, ...
 - plus vague role: ASSOC-WITH
- 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



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

Self-score: 3

Anaphora & Reference

- LF-GRAPH includes different referring expressions
 - explicit: THE, PRO, THIS/THAT,
 - implicit: IMPRO "the larger truck" --> larger than what?
- Antecedents are indicated by COREF links
 - Due to limitations of graphical format, COREF links are hard to interpret

((PARAGRAPH : TERMS

((SENTENCE : UTTNUM 0 : TERMS (SPEECHACT V110743 SA_TELL : CONTENT V105911) (F V105911 (:* MOVE GQ) : AGENT V105814 MODS (V106377) : TENSE W:: PAST) ASSESSMENT INTERSTINATE (STATE OF CONTENT OF V105911 : VAL V106902) (F V106377 (:* TO-LOC INTO) : OF V105911 : VAL V106902) (SENTENCE : UTTNUM 1 : TERMS (SENTENCE : UTTNUM 1 : TERMS (SPEECHACT V118227; SA TELL : CONTENT V114920) (F V111920 (:* STATOLECID) : CONTENT V114920) (F V111920 (:* STATOLECID) : CONTENT V114920) (F V111908 (:* PERSON HE) : CONTEXT-REL HE : COREF V105814) (F V112151 (:* READ READ) : THEME V112755) (A V112755 (:* BOOK BOOK)))))))

Self-score on Performance: 3

Word Sense Disambiguation

- 2000+ Ontology = word senses
 - e.g., "take" has 6 senses
- Sense disambiguation techniques
 - subcategorization



generic selectional restrictions
 e.g., LF::TAKE-TIME requires a LF::DURATION
 among those left, hand-set priors preference

Gold Standard Score: Precision 78% recall 68%

Quantification

LF allows arbitrary generalized quantifiers

- e.g., MOST, EVERY, SOME, ALMOST ALL, A, THE ...
- also includes KIND, BARE
- LF graph produces an underspecified scope
 - scopings allowed are equivalent to "practical" MRS

The seven texts had almost no classic quantifier scoping examples!

"every", "each", "any", "many", "few" didn't occur "some", "somewhat", "several" occurred once "a" - 28 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
- "he must not have eaten" ==>

(LF::F V118630 (:* LF::CONSUME W::EAT) :AGENT V118250 :TMA ((W::TENSE W::PRES) (W::NEGATION +) (W::MODALITY (:* LF::MUST W::MUST) (W::PERF +)))

Negated NPs "no dog"

--> (LF::QUANTIFIER V120670 (:* LF::ANIMAL W::DOG) :QUAN W::NONE)

Modal Verbs: "I believe he lied"

 (LF::F V389942 (:* LF::BELIEVE W::BELIEVE) :THEME V389975 :COGNIZER V389509 :TMA ((W::TENSE W::PRES)))

(LF::F **V389975** (:* LF::STATEMENT W::LIE) :AGENT V389966 :TMA ((W::TENSE W::PAST)))

Negations, modals, conditionals, disjunction

Conditionals

(LF::F V278126 (:* LF::PURCHASE W::BUY) :THEME V278194 :AGENT V277681 :MODS (V277435) :TMA ((W::TENSE W::FUT) (W::MODALITY (:* LF::FUTURE W::WILL)))) (LF::PRO V278194 (:* LF::REFERENTIAL-SEM W::IT) :CONTEXT-REL W::IT) (LF::PRO V277681 (:* LF::PERSON W::I) :CONTEXT-REL W::I) (LF::F V277648 (:* LF::POS-CONDITION W::IF) :OF V278126 :VAL V277648) (LF::F V277648 (:* LF::NONVERBAL-EXPRESSION W::SMILE) :AGENT V277481 :TMA ((W::TENSE W::PRES))) (LF::PRO V277481 (:* LF::PERSON W::YOU) :CONTEXT-REL W::YOU)

"If you smile I'll buy it"

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

"he must not have eaten" ==>

(LF::F V118630 (:* LF::CONSUME W::EAT) :AGENT V118250 :TMA ((W::TENSE W::PRES) (W::NEGATION +) (W::MODALITY (:* LF::MUST W::MUST) (W::PERF +)))

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

(LF::A-H2847615(*1 K:F::LANDAVEHEUE:WRUPAKOW)949605(84790))8)) (LF::FV484398 (** LF::MORE:VAL W::LARGE):FUNCTION V104447 (** LF::SIZE:VAL W::LARGE):OTA (** LF::SIZE:VALFW::LARGE)) (** LF::SIZE:VAL W::LARGE):OTA (** LF::SIZE:VALFW::LARGE)) (** LF::SIZE:VAL W::LARGE):OTA (** LF::REFERENTIAL-SEM) (LF::PRO V282816 (** LF::REFERENTIAL-SEM W::THAT))

"The larger truck" "A truck larger than that"

 Comparatives are between a thing (:FIGURE) and another thing (:GROUND)

Self-score on Performance: no comparatives in texts!

Time Expressions

- Grammar has specific rules for temporal expressions
- Time ranges: "in the 1930s"
 - (LF::THE V287214 LF::TIME-RANGE :DECADE 1930)
- Time durations: "30 years"
 - (LF::A V290689 (:* LF::QUANTITY F::DURATION) :UNIT (:* LF::TIME-UNIT W::YEAR) :AMOUNT 30)
- Dates: "July 20, 2006"
 - (LF::THE V303460 LF::TIME-LOC :YEAR 2006 :MONTH (:* LF::MONTH-NAME W::JULY) :DAY 20)
- Complex phrases "In the mid 80s"
 - (LF::F V331924 (:* LF::TIME-SPAN-REL W::IN) :OF V336557 :VAL V332631)
 - (LF::THE V332631 LF::TIME-RANGE :DECADE 80 :MODS (V332611))
 - (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"
 - LF::F V337676 (:* LF::LINEAR-VAL W::HIGH) :IS V337486 :OF V339652)
 - (LF::A V337486 (:* LF::QUANTITY F::LINEAR-S)
 :UNIT (:* LF::LENGTH-UNIT W::METRE) :AMOUNT 125)
- "10 m/s"
 - LF::F V386151 LF::FREQUENCY :VAL V385606 :OF V388082)
 - (LF::VALUE V385606 LF::FREQUENCY :OVER-PERIOD (:* LF::TIME-INTERVAL W::SECOND) :REPEATS V384377)
 - (LF::A V384377 (:* LF::QUANTITY F::LINEAR-S)
 :UNIT (:* LF::LENGTH-UNIT W::METER) :AMOUNT 10)

Self-score on Performance: 3 (didn't know "m/s" and a bug prevented parsing "\$10,000")

Question Interpretation



Questions parsed into literal speech acts

 Broad coverage of question types, wh movement, except for "how ADJ forms: "how long did it take?" Self-score on Performance: 3 Clarity



This is a religious question!

I think graphs are more accessible than logic

Some Evidence:

 It is impossible to hand-build a gold LF for a complex sentence directing in our linear TRIPS LF form

In contrast, building LF 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"	n/a	n/a	65.8%	76.3%
Average	64.1%	69.7%	67.1%	74.1%