Implementing Argument Alternations
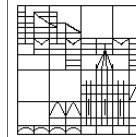
Miriam Butt (Konstanz)
and
Tracy Holloway King (PARC)

ESSLLI 2006, August, Malaga
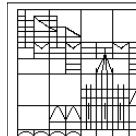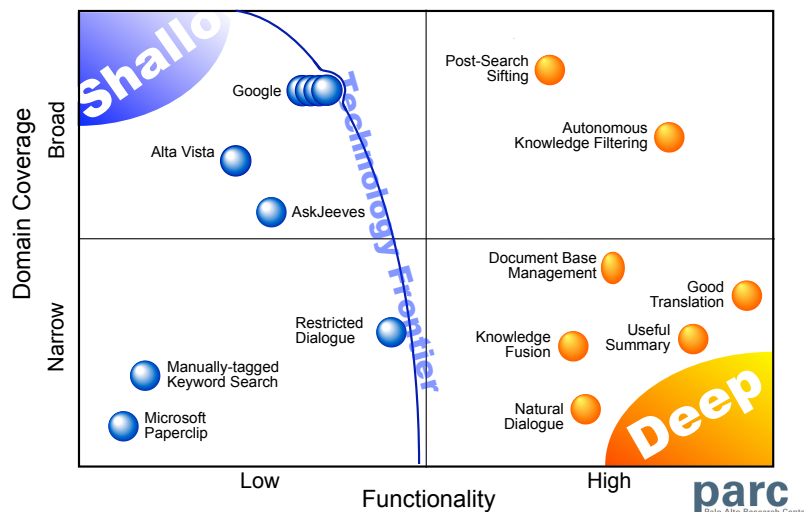
parc
Palo Alto Research Center

Overview

- Day 1:
  - Lexical Rules
    (Passive, Dative Shift, Locative Alternation)
  - Intro to LFG/XLE
- Day 2: Argument Addition (Benefactives, Applicatives)
- Day 3: V-V Complex Predicates (Urdu)
- Day 4: Causatives (Urdu, Romance)
- Day 5: Serial Verbs

parc
Palo Alto Research Center

Context: Language Engineering

parc
Palo Alto Research Center

Deep Grammar Engineering

- Draws on theoretical linguistics, software engineering
  Theoretical linguistics => papers
      Generalizations, universality, idealization (competence)
  Software engineering => programs
      Coverage, interface, QA, maintainability, efficiency, practicality
- Grammar engineering
      Grammar::Theory = Program::Programming language
      Reflect linguistic generalizations
      Respect special cases of ordinary language
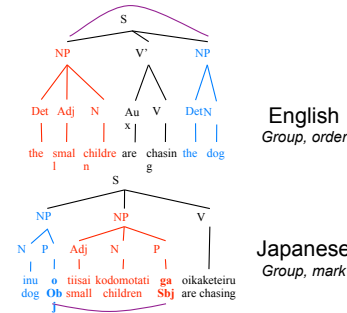   Deal with large-scale interactions
      Theory/practice trade-offs

parc
Palo Alto Research Center

Slide 1
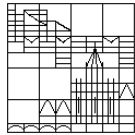
LFG/XLE

- LFG: a theory of grammar
- XLE: a parsing/generation engine for LFG grammars

Slide 2

Different patterns code same meaning

The small children are chasing the dog.

English
Group, order

Japanese
Group, mark

Slide 3

Different patterns code same meaning

The small children are chasing the dog.

LFG theory: minor adjustments on universal theme

English
Group, order

Japanese
Group, mark

Warlpiri
Mark only

Pred	'chase<Subj, Obj>'
Tense	Present
Subj	Pred children / Mod small
Obj	Pred dog

Chase(small(children), dog)

Slide 4
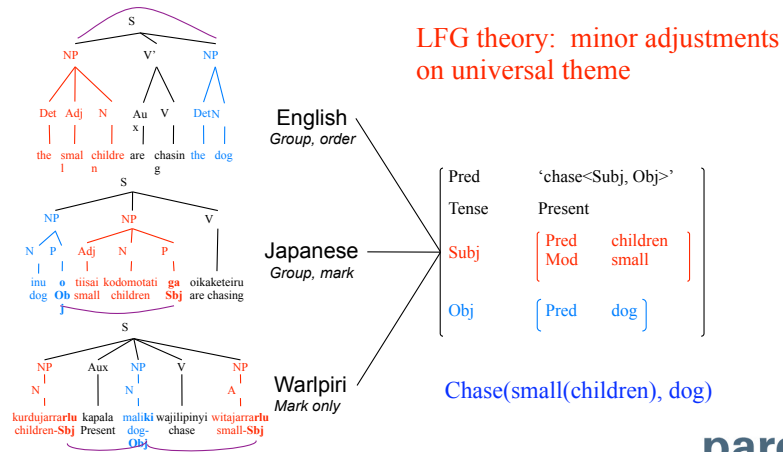
LFG architecture   *Modularity*

C-structures and f-structures in piecewise correspondence.

SUBJ	PRED 'John' / NUM SG
PRED	'like<SUBJ,OBJ>'
TENSE	PRESENT
OBJ	PRED 'Mary' / NUM SG

Formal encoding of order and grouping

Formal encoding of grammatical relations
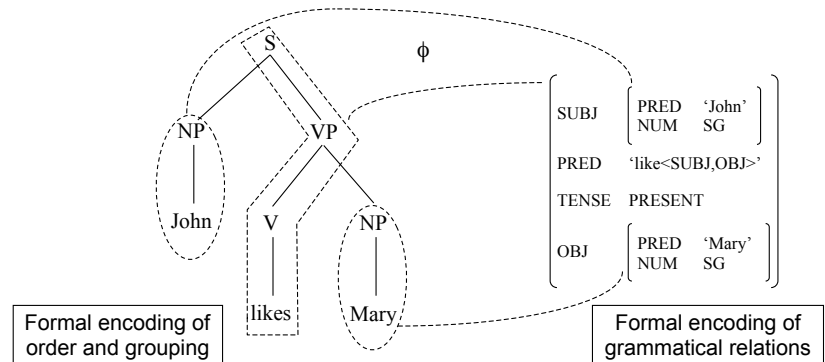
LFG grammar

Rules

S → NP VP
(↑ SUBJ)=↓   ↑=↓

VP → V NP
↑=↓   (↑ OBJ)=↓

Lexical entries

John: NP (↑ PRED)='John'
(↑ NUM)=SG

likes: V (↑ PRED)='like<SUBJ, OBJ>'
(↑ SUBJ NUM)=SG
(↑ SUBJ PERS)=3

- Context-free rules define valid c-structures (trees).
- Annotations are instantiated at tree nodes to give equational constraints that corresponding f-structures must satisfy.
- Satisfiability of constraints determines grammaticality.
- F-structure is solution for equations (if satisfied).

parc Palo Alto Research Center

Rules as well-formedness conditions

S →	NP	VP
	(↑ SUBJ) =↓	↑=↓
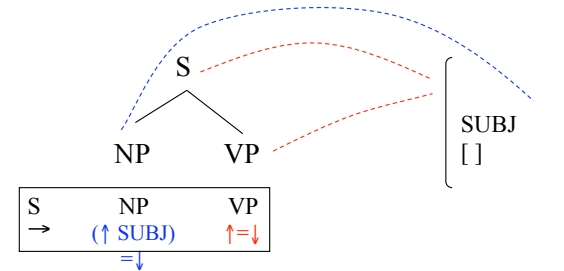
A tree containing S over NP - VP is OK if
F-unit corresponding to NP node is SUBJ of f-unit corresponding to S node
The same f-unit corresponds to both S and VP nodes.

parc Palo Alto Research Center

Inconsistent equations = Ungrammatical

What's wrong with They walks ?

S →   **f**(NP   VP
(↑ SUBJ) **s**   **f**
=↓   ↑=↓ **v**

they   walks
s (↑ NUM)   **v** (↑ SUBJ NUM)
=PL   =SG
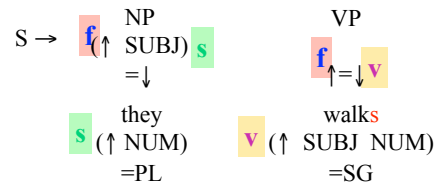
Let **f** be the f-structure of the sentence
s be the f-structure of the Subject
v be the f-structure of the verb

Then (substituting equals for equals):
(f SUBJ) = s and (s NUM)=PL
=> (f SUBJ NUM)=PL

f = v and (v SUBJ NUM)=SG
=> (f SUBJ NUM)=SG

(f SUBJ NUM)=PL and (f SUBJ NUM)=SG
=> SG=PL => FALSE

If a valid inference chain yields FALSE, the premises are unsatisfiable.

parc Palo Alto Research Center

Pargram project

- Large-scale LFG grammars for several languages
  - English, German, Japanese (Korean), French, Norwegian, Chinese, Turkish, Arabic, Hungarian
  - Cover real uses of language--newspapers, documents, etc.
- Parallelism: test LFG universality claims
  - Common c- to f-structure mapping conventions
    - (unless typologically motivated variation)
  - Invariant underlying f-structures
    - Permits shared disambiguation properties, Glue interpretation premises
  - All grammars run on PARC software (XLE)
- International consortium of linguists
  - PARC, Stuttgart, Fuji Xerox, Konstanz, Bergen, Sabanci, Oxford, Oman
  - Sustained effort--full-week meetings twice a year…~10 years!
  - Contributions to linguistics and computational linguistics: books and papers
  - Each group is self-funded, self-managed

parc Palo Alto Research Center

History

- Started in 1994
  - English (PARC)
  - French (XRCE, now PARC)
  - German (IMS-Stuttgart)
- Biannual meetings
  - Alternating between Palo Alto and Europe/Japan
- 1998: Japanese started (Fuji Xerox)
- 1999: Norwegian started (University of Bergen)
- 2000: Urdu (Konstanz)
- 2002: Danish started (Copenhagen)
- 2003: Korean (PARC) porting experiment
- 2004: Welsh, Malagasy (Essex, Oxford)
          Chinese (PARC)
- 2005: Arabic (Oman), Turkish (Sabanci), Hungarian

parc
Palo Alto Research Center

Goals

- Practical
  - Create a capability/platform for NL applications
    » translation, information retrieval, ...
  - Develop discipline of grammar engineering
    » what tools, techniques, conventions make it easy to develop and maintain broad-coverage grammars?
    » how long does it take?
    » how much does it cost?
- Theoretical
  - Refine and guide LFG theory through broad coverage of multiple languages
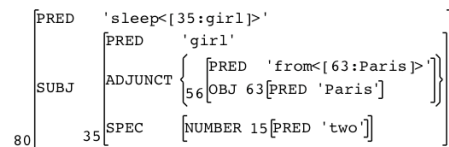 - Refine and guide the algorithms and implementation (XLE)

parc
Palo Alto Research Center

Parallel f-structures (where possible)

"Es schlafen zwei Mädchen aus Paris."

"Deux filles de Paris dorment."
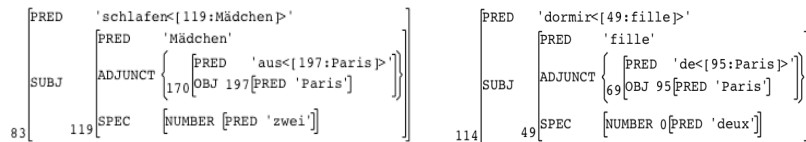
"Two girls from Paris are sleeping."

parc
Palo Alto Research Center


…but different c-structures

parc
Palo Alto Research Center
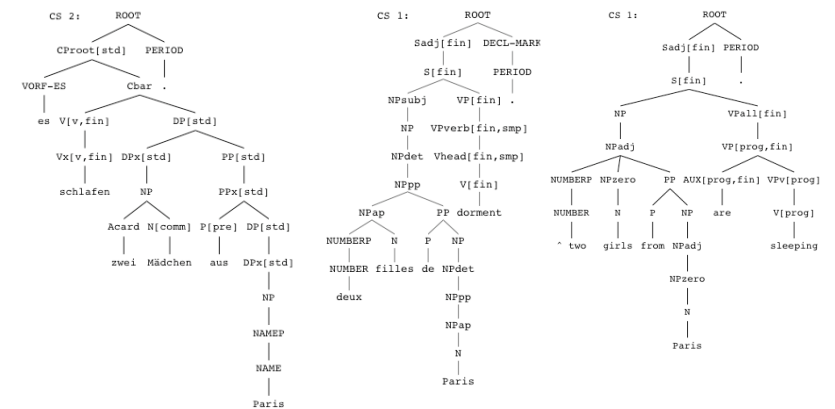
Pargram grammars

	#Rules	#States	#Disjuncts
German	251	3,239	13,294
English*	388	13,655	55,725
French	180	1,747	12,188
Japanese (Korean)	56	368	2,012

* English allows for shallow markup: labeled bracketing, named-entities

parc Palo Alto Research Center


Engineering results

- Grammars and Lexicons
- Parallel f-structures across languages
- Grammar writer's cookbook
- New practical formal devices
  - Complex categories for efficiency   NP[nom] vs. NP: (↑ CASE) = NOM
  - Optimality marks for robustness
    - enlarge grammar without being overrun by peculiar analyses
  - Lexical priority: merging different lexicons

parc Palo Alto Research Center
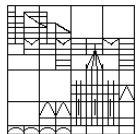
Argument Alternations

- Some Argument Alternations are quite well understood (e.g., passive)
- Others, like complex predicates, causatives or serial verbs are still the subject of intensive linguistic investigation.
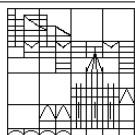- This means that grammar engineering also faces quite a few challenges.

parc Palo Alto Research Center
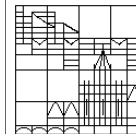
Passive and the Lexicon

Wide Spread View:

A dumping ground for exceptions:

"A kind of appendix to the grammar, whose function is to list what is unpredictable and irregular about the words of a language" (Kiparsky)

parc Palo Alto Research Center

Passive and the Lexicon

Bresnan's (LFG's) view:

A repository of linguistic generalizations —

Rules relating lexical items are a prime locus of syntactic generalizations
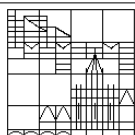

Passive and the Lexicon

Transformational View of Passives:

Active is basic: *The farmer has killed the duckling.*

Passive is derived:
   The duckling has been killed by the farmer.

If *S*1 is a grammatical sentence of the form
   *NP*1 - *Aux* - *V* - *NP*2,
then the corresponding string of the form
   *NP*2 - *Aux* + *be* + *en* - *V* - *by* + *NP*1
is also a grammatical sentence.

Passive and the Lexicon

Problems:

<u>Very</u> language specific

magister   pueros   laudat
teacher.Nom boy.Acc.Pl praise.3Sg
'The teacher praises the boys.'

Not all verbs passivize:
   The duckling has been fallen by the farmer.

Clearly what verbs passivize (and how) is tied to **lexical** information.

Passive and the Lexicon

Next Try:

The nominal that is in the accusative case in an active clause is in the nominative in the 'corresponding' passive clause. (Perlmutter and Postal 1983:7)

magister   pueros   laudat
teacher.Nom boy.**Acc**.Pl praise.3Sg
'The teacher praises the boys.'

pueri   a magistro   laudantur
boy.**Nom**.Pl by teacher.Abl   praise.Pass.3.Pl
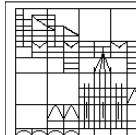'The boys are praised by the teacher.'
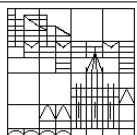
Passive and the Lexicon

Still not good enough crosslinguistically:

West Greenlandic:

gimmi-p miiraq   kii-va-a
dog-Erg   child.**Abs**   bite-3Sg-3Sg
'The dog bit the child.'

miiraq   gimmi-mik   kii-tsip-puq
child.**Abs**   dog-Instr   bite-Pass-3Sg
'A child has been bitten by the/a dog.'

parc

Passive and the Lexicon

Consensus in Relational Grammar, Role and Reference Grammar and LFG:
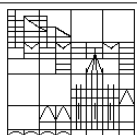
Passive Rule must
  a) be stated in terms of Grammatical Functions (these look different in different theories)

  b) associated with individual lexical items.

Sample LFG Version:
(^ SUBJ) $\Rightarrow$ (^ OBL-AG)
(^ OBJ) $\Rightarrow$ (^ SUBJ)

parc

Passive and the Lexicon

Grammatical Functions in LFG:

SUBJ, OBJ, OBJ-?, OBL-?, ADJUNCT, COMP, XCOMP

Practical Implementation

parc