LFG 98
University of Queensland
Brisbane

Optimality Theory Style Constraint Ranking in Large-scale LFG Grammars

Anette Frank, Tracy Holloway King, Jonas Kuhn, John Maxwell

XRCE, Xerox PARC, IMS Stuttgart, Xerox PARC

Overview

– Facing the ambiguity problem
in large-scale grammar development

– Extending LFG's projection architecture to include an Optimality Theory-style preference mechanism

  * Filtering syntactic and lexical ambiguity

  * Reconciling robustness and accuracy

  * Filters for generation

Facing the Ambiguity Problem: Optimality Projection for Preference Ranking

The idea:

– associate f–structures with a set of ‘optimality marks’

– optimality projection:
  an extension of LFG’s projection architecture

$$\begin{tikzcd} & \text{f–structure} \\ \text{c–structure} \arrow[ur, "\phi"] \arrow[dr, "o"'] & \\ & \text{o–structure (set of OT-marks)} \end{tikzcd}$$

OT-marks are defined by o–descriptions off c–structure

– ranking f–structures
  according to a preference ranking of OT-marks

    optimalityranking:
    mark1 mark2 NEUTRAL mark3 mark4.

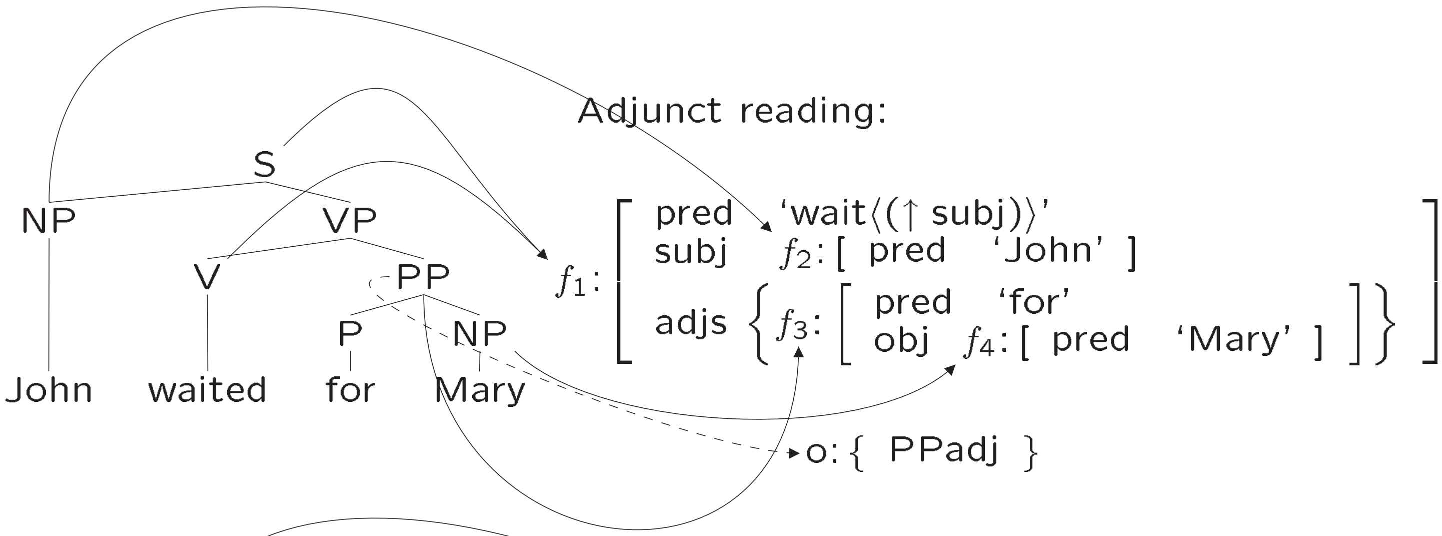
A simple example: PP ambiguities

(1) John waited for Mary.
(2) John waited for three hours.

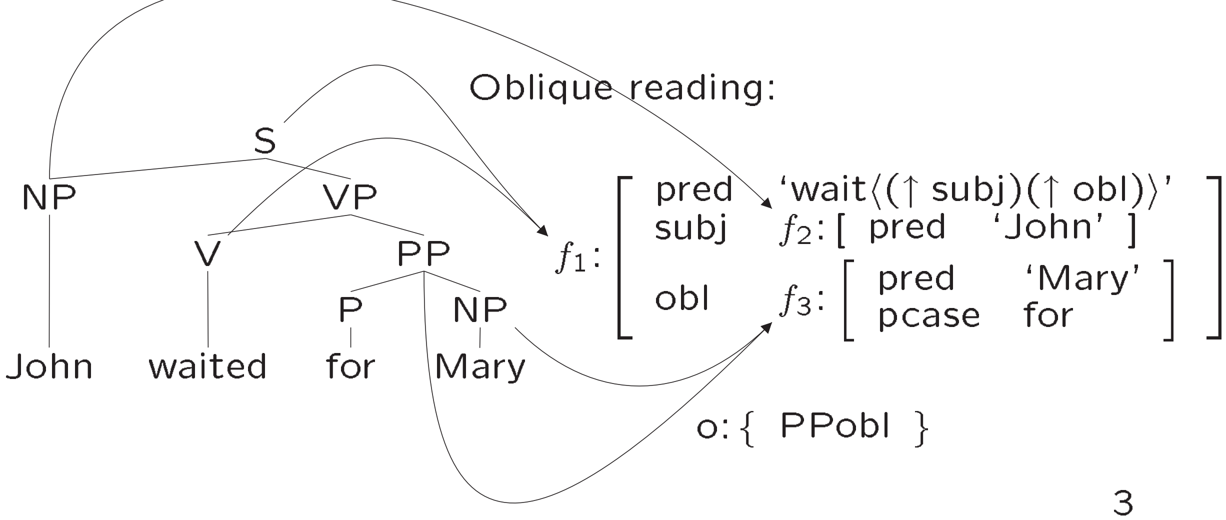
$$\text{VP} \rightarrow \text{V} \left(\begin{array}{c} \text{NP} \\ (\uparrow \text{obj}) = \downarrow \end{array} \right) \left\{ \begin{array}{c} \text{PP}* \\ (\uparrow \text{obl}\ ) = \downarrow \\ \text{PPobl} \in o* \\ \\ \downarrow \in (\uparrow \text{adjuncts}) \\ \text{PPadj} \in o* \end{array} \right\}$$

optimalityranking:
PPobl PPadj NEUTRAL mark3 mark4.

Adjunct reading:



Oblique reading:



Interpretation of ranked constraints

– OT constraints are ranked from highest to lowest

mark1 NEUTRAL mark2 UNGRAMMATICAL mark3 NOGOOD mark4

mark_1 ...	preference marks
NEUTRAL	
mark_i ..	dispreference marks
UNGRAMMATICAL	
mark_j ...	ungrammatical marks
NOGOOD	
mark_z ...	excluding ('nogood') marks

– Partial order: preference marks can be of equal strength

(mark1 mark1a) mark2 NEUTRAL (mark3 mark4) mark5

– Dispreference vs. preference marks:

Optimality Theory:

∗ compare marks of (two or more) analyses
from most dispreferred to least dispreferred;

∗ if analyses differ, choose the one with fewest marks

Here in addition:

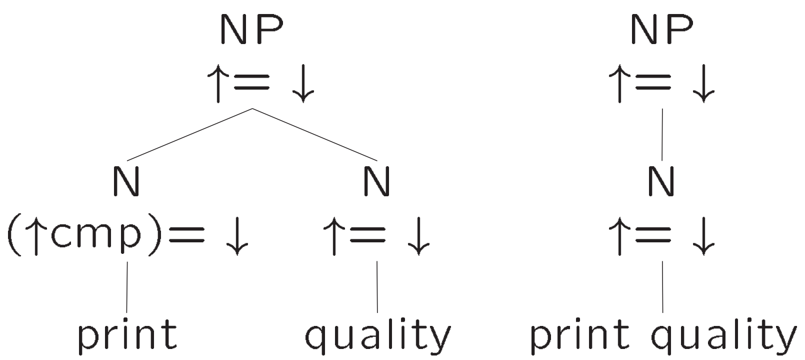
∗ compare marks of remaining analyses
from most preferred to least preferred;

∗ if analyses differ, choose
the ones with greater number of preference marks

Interpretation of Ranked Constraints
Preference Marks

Example – Multiword expressions

(3) a. print quality
  b. Druckqualität
    print-quality
  c. qualité d'impression
    quality of printing



Preferring multiword nouns:

optimalityranking:
mark1 Mwe NEUTRAL mark3 UNGRAMMATICAL ... NOGOOD ...

print quality: N (↑pred)= 'print quality'
          Mwe $\in o*$

(4) The [$_{\mathrm{N}}$ print quality] of this printer is excellent.

(5) I want to [print] [$_{\mathrm{NN}}$ quality documents].

Interpretation of Ranked Constraints
Dispreference Marks

Example – Headless NPs

(6) Meistens kauft die größere Firma [$_{NP}$ die kleinere].
Mostly buys the larger company the smaller
'Usually, the larger company buys the smaller one.'

(7) Nachts fallen helle Farben auf.
at night stand bright colors out

a. Nachts fallen [$_{NP(nom)}$ helle Farben] auf.
'At night, [bright colors] stand out.'

b. Nachts fallen [$_{NP(nom)}$ helle] [$_{NP(dat)}$ Farben] auf.
'At night, [bright ones] strike [colors].'
'At night, [colors] notice [bright ones].'

Dispreferring headless NPs:

optimalityranking:
mark1 NEUTRAL AheadNp UNGRAMMATICAL ... NOGOOD ...

$$
\text{NP} \longrightarrow \left\{ \begin{array}{ccc}
\text{(DET)} & \text{A} & \\
(\uparrow\text{spec}){=}\downarrow & \downarrow\in(\uparrow\text{mod}) & \\
 & (\uparrow\text{pred}){=}\text{'pro'} & \\
 & \text{AheadNp} \in o* & \\
 & & \\
| \quad \text{(DET)} & \text{(A)} & \text{N} \\
(\uparrow\text{spec}){=}\downarrow & \downarrow\in(\uparrow\text{mod}) & \uparrow{=}\downarrow
\end{array} \right\}
$$

Interpretation of Ranked Constraints Ungrammatical Marks

– only succeed if no grammatical analysis –

Example – Agreement Mistakes

(8)	Les	moteurs	sont	grand.
	the	motors.masc-pl	are	big.masc-sg
	'The motors are big.'			

$$
\begin{array}{ll}
\text{+Sg ANUM} & \{\ (\uparrow\text{subj num})=\ \text{sg} \\
& \ \mid (\uparrow\text{subj num})\neq\ \text{sg} \\
& \quad \text{NoAdjAgr} \in o*\ \}.
\end{array}
$$

optimalityranking:
mark1 NEUTRAL ... UNGRAMMATICAL NoAdjAgr NOGOOD ...

Ungrammatical marks only succeed in absence of grammatical analysis

(9)	Les	enfants	sont	heureux.
	the	children.masc-pl	are	happy.masc-sg/pl
	'The children are happy.'			

Interpretation of Ranked Constraints
Nogood Marks

– never succeed –

Example – Corpus-specific constructions

(10) 67. I would like to meet on Wednesday.

$$
\begin{array}{llll}
S \longrightarrow & \text{(NUMBERP)} & \text{NP} & \text{VP} \\
 & (\uparrow \text{tag}){=}\downarrow & (\uparrow \text{subj}){=}\downarrow & \uparrow{=}\downarrow \\
 & \text{VM-Corpus} \in o{*} & &
\end{array}
$$

optimalityranking:
(VM-Corpus NEUTRAL) ... UNGRAMMATICAL ... NOGOOD .

Going back to other corpora:

optimalityranking:
NEUTRAL ... UNGRAMMATICAL ... NOGOOD VM-Corpus.

Applications in large–scale Grammars
Spurious Ambiguity

– Coordination –

(11) a. [$_{NP}$ the cousins] and [$_{NP}$ a sister]
b. a [$_{N'}$ brilliant singer] but [$_{N'}$ mediocre guitarist]
c. the many [$_{N}$ friends] and [$_{N}$ relatives] of Peter's

(12) a. [$_{NP}$ [$_{NP}$ [[cats]]] and [$_{NP}$ [[dogs]]]]
b. [[$_{N'}$ [$_{N'}$ [cats]] and [$_{N'}$ [dogs]]]]
c. [[[$_{N}$ [$_{N}$ cats] and [$_{N}$ dogs]]]]

Assigning ranked preferences to attachment levels:

$$\mathrm{NP} \rightarrow \mathrm{NP} \quad \underset{\mathrm{Level1} \in o*}{\mathrm{CONJ}} \quad \mathrm{NP}$$

$$\mathrm{N'} \rightarrow \mathrm{N'} \quad \underset{\mathrm{Level2} \in o*}{\mathrm{CONJ}} \quad \mathrm{N'}$$

$$\mathrm{N} \rightarrow \mathrm{N} \quad \underset{\mathrm{Level3} \in o*}{\mathrm{CONJ}} \quad \mathrm{N}$$

optimalityranking:
Level3 Level2 Level1 NEUTRAL ... UNGRAMMATICAL ... NOGOOD

Counterexamples:

(13) old men and women
a. [old [$_{N}$ [$_{N}$ men] and [$_{N}$ women]]]
b. [$_{NP}$ [$_{NP}$ old men] and [$_{NP}$ women]]

Applications in large–scale Grammars
Filtering of Syntactic Ambiguity

– PP ambiguities –

(14) John waited for Mary.

$$\mathrm{VP} \rightarrow \mathrm{V} \left(\begin{array}{c} \mathrm{NP} \\ (\uparrow \mathrm{obj}) = \downarrow \end{array} \right) \left\{ \begin{array}{c} \mathrm{PP}* \\ (\uparrow \mathrm{obl}\ ) = \downarrow \\ \mathrm{PPobl} \in o* \\ \\ \downarrow \in (\uparrow \mathrm{adjuncts}) \\ \mathrm{PPadj} \in o* \end{array} \right\}$$

Preference or dispreference?

– Preference for oblique reading:

PPobl NEUTRAL ... UNGRAMMATICAL ... NOGOOD.

(15) John waited {for | four} {hours | ours }.

(16)a. John waited [$_{\text{PPobl}}$ for ours].
b. John waited [$_{\text{PPadj}}$ for ours].
c. John waited [$_{\text{PPobl}}$ for hours].
d. John waited [$_{\text{PPadj}}$ for hours].
e. John waited [$_{\text{NPadj}}$ four hours]. (dispreferred)

– Dispreference for adjunct reading:

mark1 NEUTRAL PPadj UNGRAMMATICAL ... NOGOOD.

(17)a. [$_{\text{NP}}$ Fruit flies] [$_{\text{V}}$ like] [$_{\text{NP}}$ a banana]
b. [$_{\text{NP}}$ Fruit] [$_{\text{V}}$ flies] [$_{\text{PP}}$ like a banana] (dispreferred)

– Intuitively, oblique readings seem to be preferred:

(18) He painted the wall with cracks.

Applications in large–scale Grammars
Filtering of Syntactic Ambiguity

– PP ambiguities –

Use selectional restrictions as an additional filter?

(19) John waited for Mary.

(20) John waited for three hours.

Disprefer OBLique analysis if PP's concept class does not match the OBLique's selectional restrictions

$$
\begin{array}{ll}
\text{wait: V} & (\uparrow\text{pred})=\text{ `wait}\langle(\uparrow\text{subj})(\uparrow\text{obl})\rangle\text{'} \\
 & (\uparrow_r \text{ arg1-cl})= \{ \text{ cl1 cl2}\} \\
 & (\uparrow_r \text{ arg2-cl})= \{ \text{ cl1 cl5}\} \\
 & \{ \ ((\uparrow\text{obl})_r \text{ cl}) \ll (\uparrow_r \text{ arg2-cl}) \\
 & \ \mid ((\uparrow\text{obl})_r \text{ cl}) \not\ll (\uparrow_r \text{ arg2-cl}) \\
 & \quad \text{NoSelres} \in o* \ \}.
\end{array}
$$

optimalityranking:
mark1 PPobl NEUTRAL NoSelres UNGRAMMATICAL ... NOGOOD.

Still counterexamples ...

(21) Il renonce au voyage.
He abandons the journey

(22) Il renonce au premier essai.
He abandons the first attempt
He abandons on the first attempt

optimalityranking:
mark1 NEUTRAL NoSelres UNGRAMMATICAL ... NOGOOD.

Applications in large-scale Grammars

– Preference ranking on a flexible basis –

Preference ranking can be designed in a flexible way:

– Interactive disambiguation:
accessing lower-ranked analyses, stepwise

– Specifying strength of reliance on preference ranking:
admit specific number/grades of low-ranked analyses

– Interfacing preference constraints with
(external) knowledge sources for disambiguation

Preference ranking is an invaluable tool

in large scale grammar development

Applications in large–scale Grammars
Filtering of Lexical Ambiguity

– using rich lexica –

(23) a.

N

n_stem	gend_tag	nbr_tag	n_tag
utilisation	+Fem	+Sg	+Noun

b.

+Fem	gend_tag	(↑gend)= fem.
+Sg	nbr_tag	(↑num)= sg.

c. +munknown n_stem (↑pred)= '%stem'.

d. $\left[\begin{array}{l} \text{pred 'utilisation'} \\ \text{gend fem} \\ \text{num sg} \end{array}\right]$

– High potential of lexical ambiguity

est	être +Verb (be)	*si*	si +Conj (if)
	est +Noun (east)		si +Noun (B)
été	être +Verb (be, part.)	*la*	le +Det (the, fem))
	été +Noun (summer)		la +Noun (A)

(24) a.

[$_{S}$ [$_{NP}$	Le	défaut]	[$_{VP}$ [$_{AUX}$	est]	[$_{V}$	corrigé	]]]
	the	error		is		corrected	

'The error is corrected.'

b.

[$_{NP}$	Le	[$_{NMOD}$	défaut	est]	[$_{VPAP}$	corrigé]]
	the		error	east		corrected

'the corrected error east'

– Disprefer domain specific "rare" readings

optimalityranking:
mark1 NEUTRAL RareNoun UNGRAMMATICAL ... NOGOOD.

est:	N	(↑pred)= 'est'	RareNoun $\in o*$.
si:	N	(↑pred)= 'si'	RareNoun $\in o*$.

Applications in large–scale Grammars
Grammar Engineering

– flexible handling of grammar versions –

Creating grammars with switched–off parts of rules

- diagnosing sources of efficiency problems (Kuhn and Rohrer 1997)
- special grammar versions for grammar–based semi-automatic lexicon acquisition (Kuhn, Eckle-Kohler and Rohrer (1998))

Example:
Parsing corpora with hypothetical verb subcategorization frames to determine correct subcategorization

Using UNGRAMMATICAL marks to increase precision:
transitive–reflexive ambiguity marked as ‘ungrammatical’

(25)	Ich	habe	mich	korrigiert.
	I	have	myself	corrected

Applications in large–scale Grammars
Generation

– reconciling robustness and accuracy for generation –

Parsing, but not generating ungrammatical input

(26) a. Les moteurs sont grand.
the motors.masc-pl are big.masc-sg
'The motors are big.'

b. +Sg ANUM $\{$ ($\uparrow$subj num)= sg
$|$ ($\uparrow$subj num)$\neq$ sg
NoAdjAgr $\in o*$ $\}$.

c. optimalityranking:
... NEUTRAL ... UNGRAMMATICAL NoAdjAgr NOGOOD.

Parsing a broad range of constructions – but generating "restricted language"

(27) a. The rear burner, when left on for too long, will tend to overheat.

b.

S $\longrightarrow$	NP	CONJPsub	VP
	($\uparrow$subj)=$\downarrow$	($\uparrow$adjunct)=$\downarrow$	$\uparrow$=$\downarrow$
		ParseOnly $\in o*$	

c. optimalityranking:
... NEUTRAL ... UNGRAMMATICAL ParseOnly NOGOOD.

Conclusions

Facing ambiguity in large-scale grammars:

Integration of OT-style preference mechanism into LFG projection architecture

– Distinct types of ranking constraints
  * preference marks
  * dispreference marks
  * ungrammaticality marks
  * nogood marks

– Various types of application:
  * spurious ambiguities
  * filtering syntactic and lexical ambiguity
  * robustness
  * generation
  * grammar engineering

– Global preference constraints may interact with unrelated analyses

– Preference ranking can be designed on a flexible basis

* interactively access lower–ranked analyses, stepwise
* specify strength of reliance on preference ranking:
  admit specific number/grades of lower-ranked analyses

An invaluable tool in large-scale grammar development!