

---

# A Best-Fit Approach to Productive Omission of Arguments

---

Eva Mok & John Bryant

University of California, Berkeley

International Computer Science Institute

---

# Simplify grammar by exploiting the language understanding process

- Omission of arguments in Mandarin Chinese
- Construction grammar framework
- Model of language understanding
- Our best-fit approach

# Productive Argument Omission (in Mandarin)

1 

<i>ma1+ma</i>	<i>gei3</i>	<i>ni3</i>	<i>zhei4+ge</i>
mother	give	2PS	this+CLS

 ■ Mother (I) give you this (a toy).

2 

<i>ni3</i>	<i>gei3</i>	<i>yi2</i>
2PS	give	auntie

 ■ You give auntie [the peach].

3 

<i>ao</i>	<i>ni3</i>	<i>gei3</i>	<i>ya</i>
EMP	2PS	give	EMP

 ■ Oh (go on)! You give [auntie] [that].

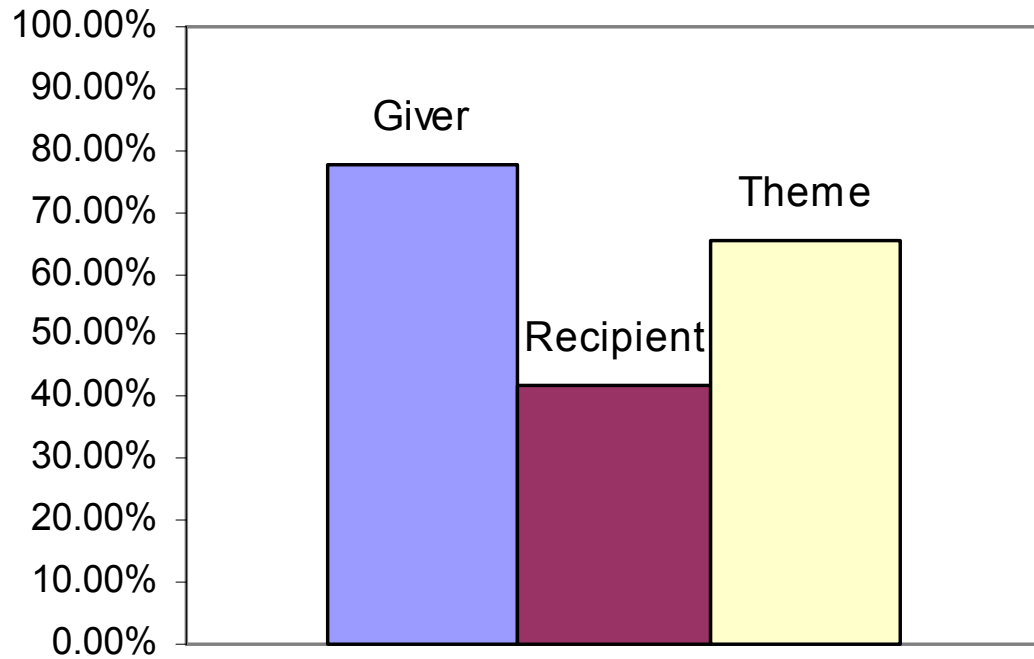
4 

<i>gei3</i>
give

 ■ [I] give [you] [some peach].

# Arguments are omitted with different probabilities

**% omitted (98 total utterances)**



All arguments omitted: 30.6%

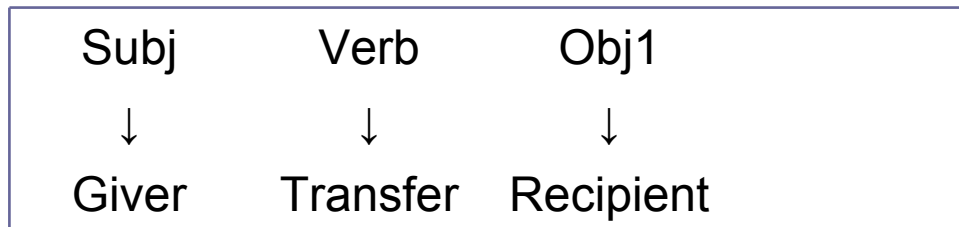
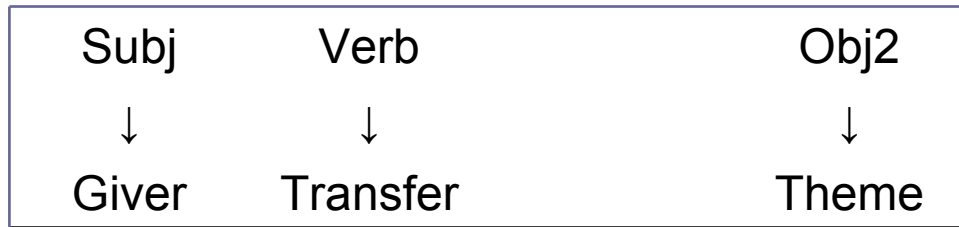
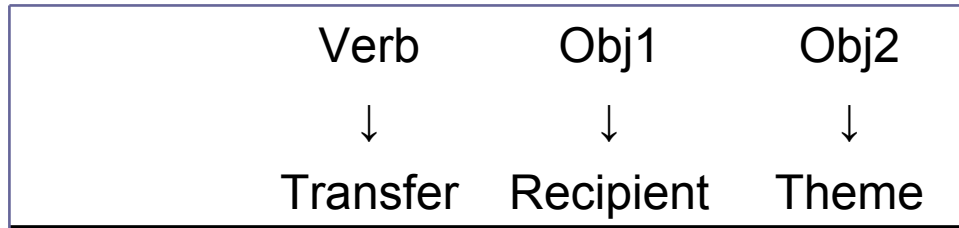
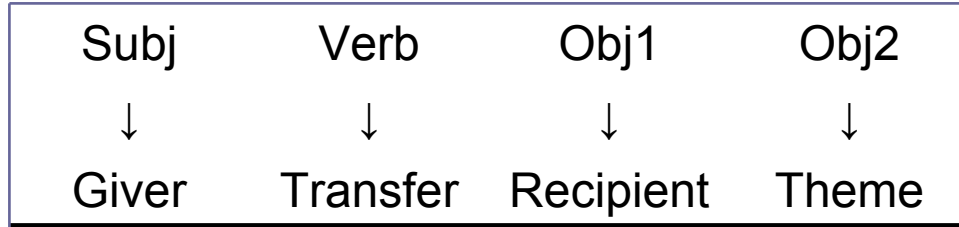
No arguments omitted: 6.1%

---

# Construction grammar approach

- Kay & Fillmore 1999; Goldberg 1995
- Grammaticality: form **and** function
- Basic unit of analysis: construction, i.e. a pairing of form and meaning constraints
- Not purely compositional
- Implies early use of semantics in processing
- Embodied Construction Grammar (ECG) (Bergen & Chang, 2005)

# Problem: Proliferation of constructions



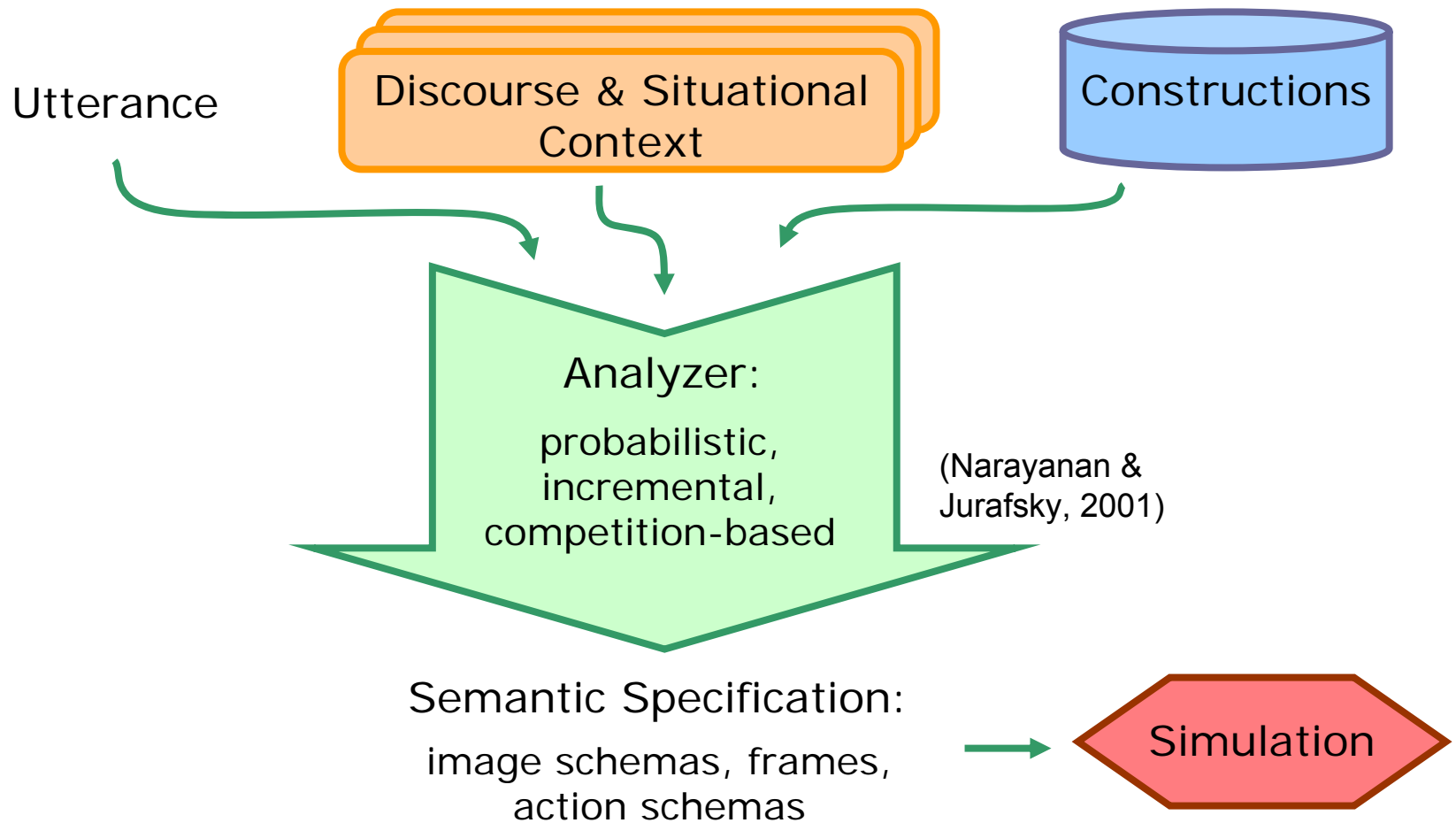
...

# If the analysis process is smart, then...

Subj	Verb	Obj1	Obj2
↓	↓	↓	↓
Giver	Transfer	Recipient	Theme

- The grammar needs only state one construction
- Omission of constituents is flexibly allowed
- The analysis process figures out what was omitted

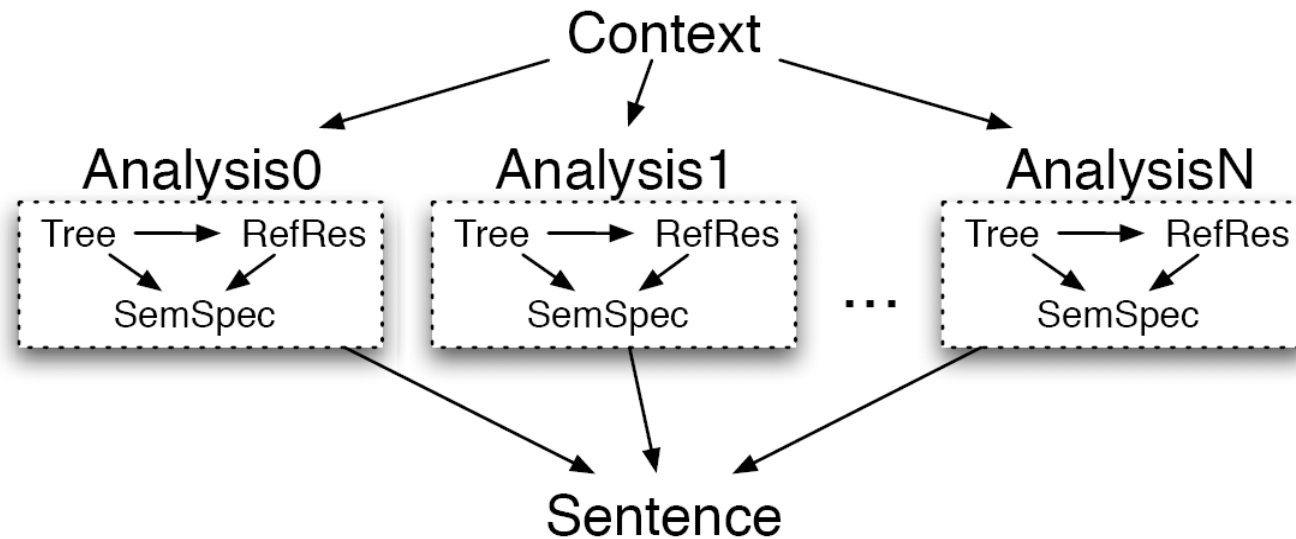
# Best-fit analysis process takes the burden off of the grammar representation



# Competition-based analyzer finds the best analysis

- An analysis is made up of:
  - A constructional tree
  - A set of resolutions
  - A semantic specification

The best fit has the highest combined score



---

# Combined score that determines best-fit

- **Syntactic Fit:**

- Constituency relations
- Combine with preferences on non-local elements
- Conditioned on syntactic context

- **Antecedent Fit:**

- Ability to find referents in the context
- Conditioned on syntactic information, feature agreement

- **Semantic Fit:**

- Semantic bindings for frame roles
- Frame roles' fillers are scored

# Analyzing *ni3 gei3 yi2* (You give auntie)

Two of the competing analyses:

<i>ni3</i>	<i>gei3</i>	<i>yi2</i>	omitted
↓	↓	↓	↓
Giver	Transfer	Recipient	Theme

<i>ni3</i>	<i>gei3</i>	omitted	<i>yi2</i>
↓	↓	↓	↓
Giver	Transfer	Recipient	Theme



## ■ Syntactic Fit:

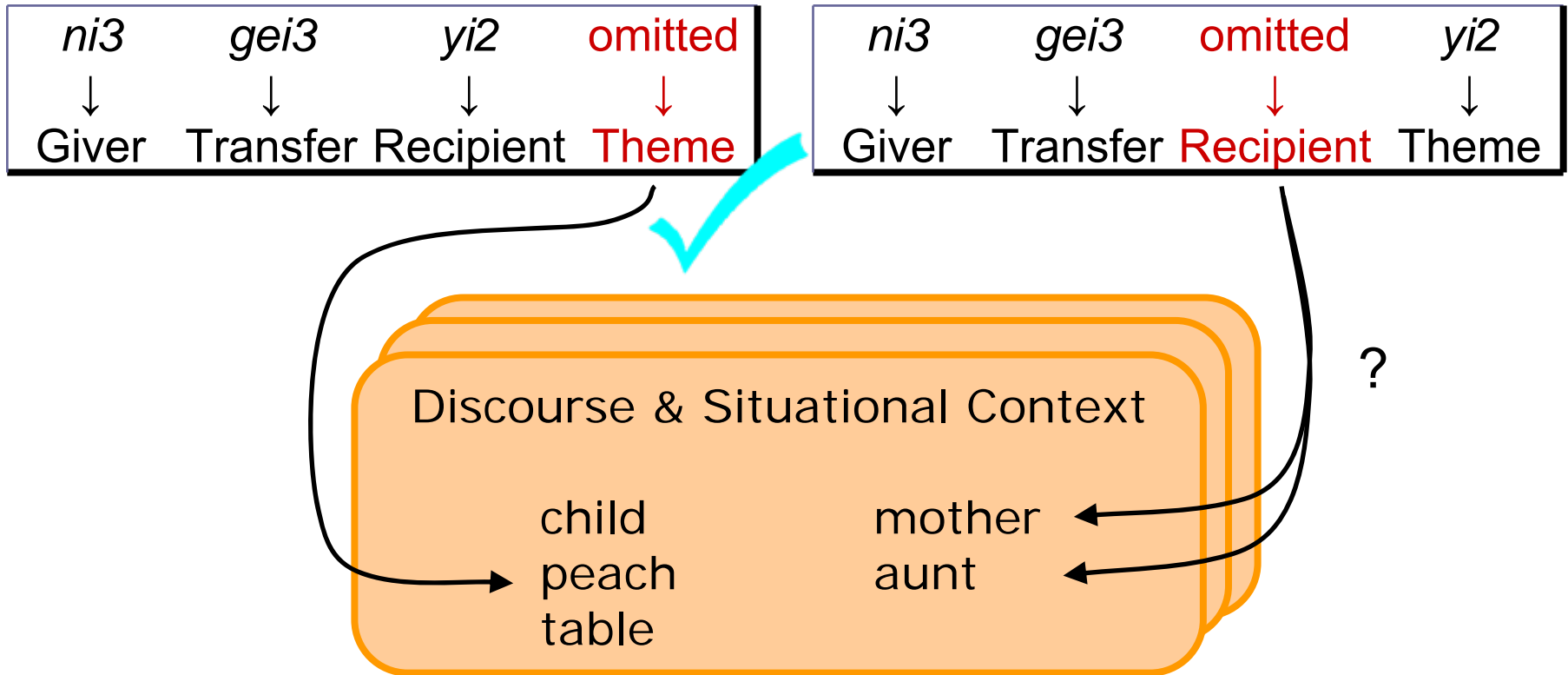
- $P(\text{Theme omitted} \mid \text{ditransitive cxn}) = 0.65$
- $P(\text{Recipient omitted} \mid \text{ditransitive cxn}) = 0.42$

$$(1-0.78)*(1-0.42)*0.65 = 0.08$$

$$(1-0.78)*(1-0.65)*0.42 = 0.03$$

# Can the omitted argument be recovered from context?

## ■ Antecedent Fit:



How good of a theme is a peach?

How about an aunt?

■ Semantic Fit:

<i>ni3</i>	<i>gei3</i>	<i>yi2</i>	omitted
↓	↓	↓	↓
Giver	Transfer	Recipient	Theme
↓	↓	↓	↓
child	give	aunt	peach

<i>ni3</i>	<i>gei3</i>	omitted	<i>yi2</i>
↓	↓	↓	↓
Giver	Transfer	Recipient	Theme
↓	↓	↓	↓
child	give	mother?	aunt



Giver (usually animate)  
Recipient (usually animate)  
Theme (usually inanimate)

The best-fit score combines syntactic, antecedent and semantic fit scores

	omitted Theme	omitted Recipient
syntactic	✓	
antecedent	✓	
semantic	✓	
winner	✓	



---

# Leverage process to simplify representation

- The processing model is complementary to the theory of grammar
- By using a competition-based analysis process, we can:
  - Find the best-fit analysis with respect to constituency structure, context, and semantics
  - Eliminate the need to enumerate allowable patterns of argument omission in grammar
- This is currently being applied in models of language understanding and grammar learning.

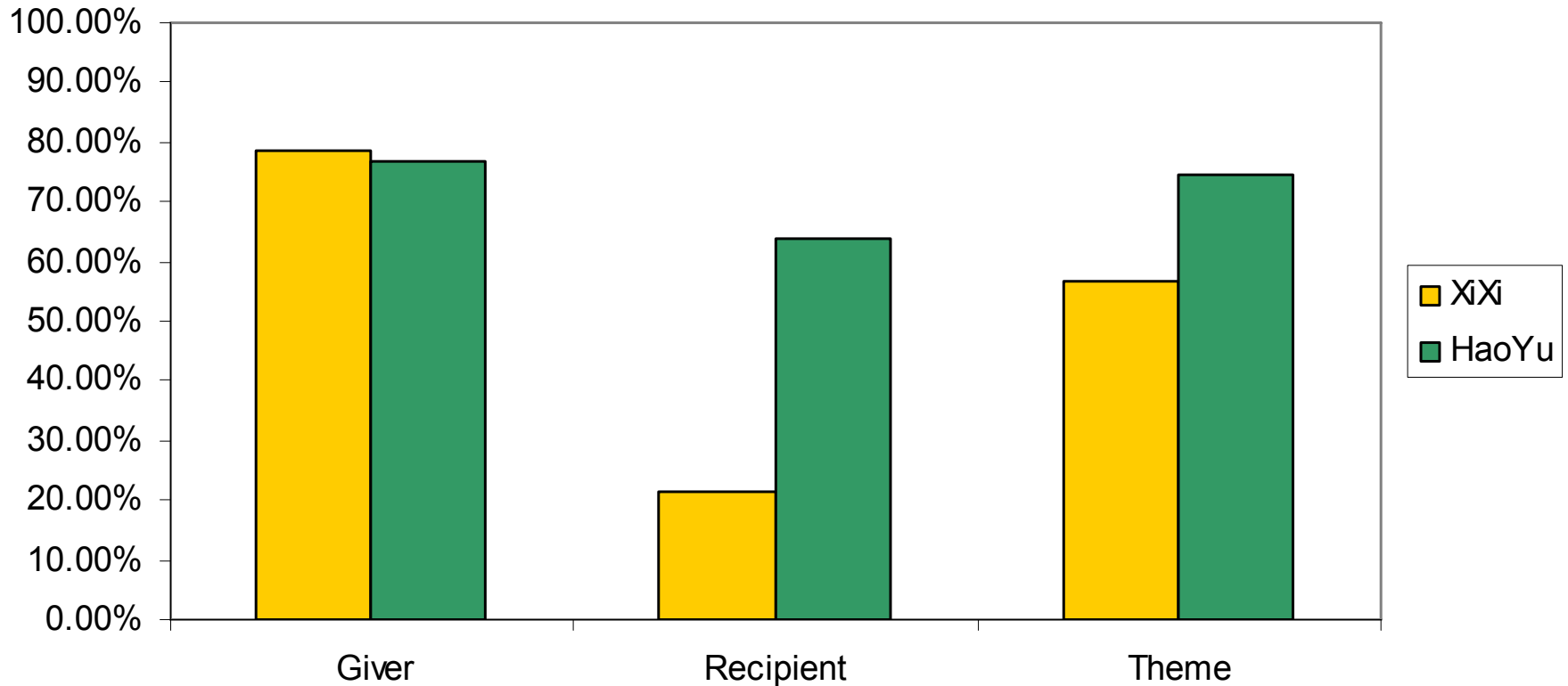
# Using frame and lexical information to restrict type of reference

The Transfer Frame	
Giver	
Recipient	
Theme	
Manner	Purpose
Means	Reason
Place	Time

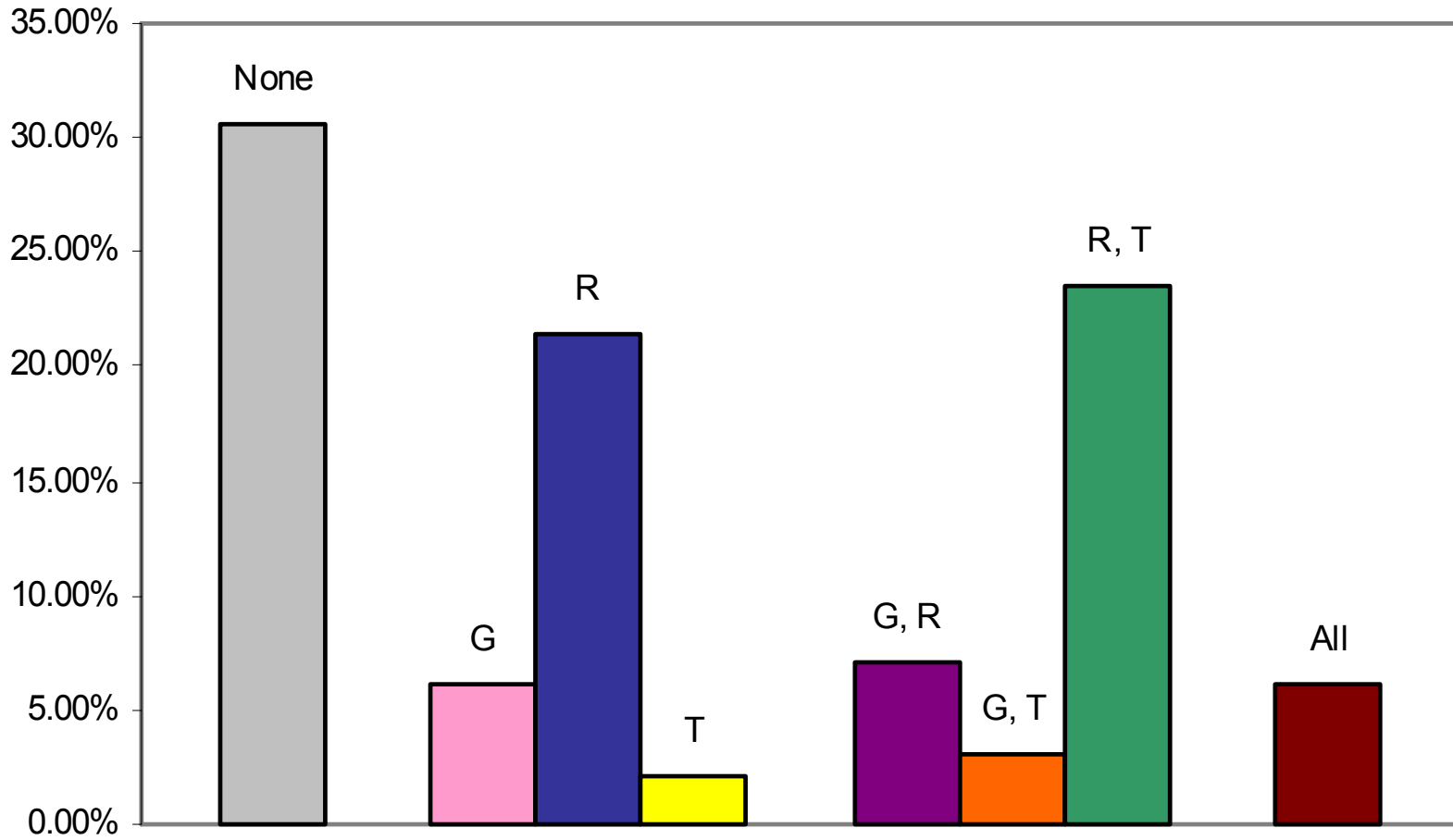
Lexical Unit <i>gei3</i>	
Giver	(DNI)
Recipient	(DNI)
Theme	(DNI)

# Individual Variation in Omission Patterns

**% omitted by each caretaker**



# Patterns of Argument Presence



# Individual Variations in Patterns of Argument Presence

	None	G	R	T	G,R	G,T	R, T	G,R,T
XiXi	14%	2%	31%	2%	10%	4%	31%	6%
HaoYu	49%	11%	11%	2%	4%	2%	15%	6%