

Copies and Language: Insights from Computational Linguistics

Greg Kobele

StuTS+TaCoS

Universität Leipzig

Introduction

provides means of understanding copy constructions

Makes particularly salient

External syntax

- distribution in clause

Doesn't have much to say

about **internal syntax** of copies, or about **how** copies are made

The question

What is the internal syntax of copies?

- what do copies look like
- what kinds of things can be copied
- ...and what kind of mechanism makes copies

What copying patterns are
there in NL?

Chinese A-not(A) Questions

1. Bill xihuan wo
“Bill likes me.”

Chinese A-not(A) Questions

1. Bill xihuan wo
“Bill likes me.”
2. Bill bu xihuan wo
“Bill doesn't like me.”

Chinese A-not(A) Questions

1. Bill xihuan wo
“Bill likes me.”
2. Bill bu xihuan wo
“Bill doesn't like me.”
3. Bill xihuan bu xihuan wo
“Does Bill like me?”

Chinese A-not(A) Questions

1. Bill xihuan wo
“Bill likes me.”

2. Bill bu xihuan wo
“Bill doesn’t like me.”

3. Bill xihuan bu xihuan wo
“Does Bill like me?”

1. Stan zai zher
“Stan is here.”

Chinese A-not(A) Questions

- | | |
|---|---|
| 1. Bill xihuan wo
“Bill likes me.” | 1. Stan zai zher
“Stan is here.” |
| 2. Bill bu xihuan wo
“Bill doesn’t like me.” | 2. Stan bu zai zher
“Stan isn’t here.” |
| 3. Bill <u>xihuan</u> <u>bu xihuan</u> wo
“Does Bill like me?” | |

Chinese A-not(A) Questions

1. Bill xihuan wo
“Bill likes me.”
2. Bill bu xihuan wo
“Bill doesn’t like me.”
3. Bill xihuan bu xihuan wo
“Does Bill like me?”

1. Stan zai zher
“Stan is here.”
2. Stan bu zai zher
“Stan isn’t here.”
3. Stan zai zher bu zai zher
“Is Stan here?”

Chinese A-not(A) Questions

1. Bill xihuan wo
“Bill likes me.”
2. Bill bu xihuan wo
“Bill doesn’t like me.”
3. Bill xihuan bu xihuan wo
“Does Bill like me?”
4. *Bill **xihuan** bu **ai** wo
NOT: “Does Bill like me?”

1. Stan zai zher
“Stan is here.”
2. Stan bu zai zher
“Stan isn’t here.”
3. Stan zai zher bu zai zher
“Is Stan here?”
4. *Stan zai **zher** bu zai **nar**
NOT: “Is Stan here?”

Yoruba Verbal Relative Clauses

- Olu ra adie
"Olu bought a chicken"

Yoruba Verbal Relative Clauses

- Olu ra adie
"Olu bought a chicken"
- Adie ti Olu ra (tobi)
"The chicken that Olu bought (is big)"

Yoruba Verbal Relative Clauses

- Olu ra adie
"Olu bought a chicken"
- Adie ti Olu ra (tobi)
"The chicken that Olu bought (is big)"
- Rira adie ti Olu ra adie (ko da)
"The fact that Olu bought a chicken (was bad)"

Yoruba Verbal Relative Clauses

- Olu ra adie
"Olu bought a chicken"
- Adie ti Olu ra (tobi)
"The chicken that Olu bought (is big)"
- Rira adie ti Olu ra adie (ko da)
"The fact that Olu bought a chicken (was bad)"

Yoruba Verbal Relative Clauses

- Olu ra adie
"Olu bought a chicken"
- Adie ti Olu ra (tobi)
"The chicken that Olu bought (is big)"
- Rira adie ti Olu ra adie (ko da)
"The fact that Olu bought a chicken (was bad)"
- *Rira adie ti Olu je nkan
buying chicken that Olu ate something

English X or no X

1. **Cameras** or no **cameras**, the clothes are coming off.

English X or no X

1. Cameras or no cameras, the clothes are coming off.
3. War on terror or no war on terror, I'm not eating these brussel sprouts.

English X or no X

1. Cameras or no cameras, the clothes are coming off.
2. *Cameras or no recording devices, the clothes are coming off.

3. War on terror or no war on terror, I'm not eating these brussel sprouts.
4. *Global struggle against extremism or no war on terror, I'm not eating these brussel sprouts.

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”

1. **Wer** hat dich gestern gekitzelt?
“Who tickled you yesterday?”

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”
2. Wer glaubt Benedict wer dich gestern gekitzelt hat?
“Who does Benedict believe tickled you yesterday?”

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”
2. **Wer** glaubt Benedict **wer** dich gestern gekitzelt hat?
“Who does Benedict believe tickled you yesterday?”

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”
2. Wer glaubt Benedict wer dich gestern gekitzelt hat?
“Who does Benedict believe tickled you yesterday?”
3. Wer glaubt Benedict wer Maria gesagt hat wer dich
gestern gekitzelt hat?
“Who does Benedict believe that Maria said tickled you
yesterday?”

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”
2. Wer glaubt Benedict wer dich gestern gekitzelt hat?
“Who does Benedict believe tickled you yesterday?”
3. **Wer** glaubt Benedict **wer** Maria gesagt hat **wer** dich
gestern gekitzelt hat?
“Who does Benedict believe that Maria said tickled you
yesterday?”

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”
2. Wer glaubt Benedict wer dich gestern gekitzelt hat?
“Who does Benedict believe tickled you yesterday?”
3. Wer glaubt Benedict wer Maria gesagt hat wer dich
gestern gekitzelt hat?
“Who does Benedict believe that Maria said tickled you
yesterday?”
4. *Wer glaubt Benedict **was** Maria gesagt hat wer dich
gestern gekitzelt hat?

1. Wer hat dich gestern gekitzelt?
“Who tickled you yesterday?”
2. Wer glaubt Benedict wer dich gestern gekitzelt hat?
“Who does Benedict believe tickled you yesterday?”
3. Wer glaubt Benedict wer Maria gesagt hat wer dich
gestern gekitzelt hat?
“Who does Benedict believe that Maria said tickled you
yesterday?”
4. *Wer glaubt Benedict was Maria gesagt hat wer dich
gestern gekitzelt hat?

And many more

Vata à li-da zué saká
“We ate rice yesterday”

And many more

Vata li à li-da zué saká
“We ATE rice yesterday”

And many more

Vata li à li-da zué saká

“We ATE rice yesterday”

SLQZ Bgwi'ih Mike lohoh Mike

“Mike looked at himself”

And many more

Vata li à li-da zué saká

“We ATE rice yesterday”

SLQZ Bgwi'ih Mike lohoh Mike ze'cy cahgza' Felipe.

“Mike looked at himself and Felipe did too”

And many more

Vata li à li-da zué saká
“We ATE rice yesterday”

SLQZ Bgwi'ih Mike lohoh Mike ze'cy cahgza' Felipe.
“Mike looked at himself and Felipe did too”

Malagasy Tsy nahavita na inona na inona aho
“I didn't accomplish anything.”

And many more

Vata li à li-da zué saká
“We ATE rice yesterday”

SLQZ Bgwi'ih Mike lohoh Mike ze'cy cahgza' Felipe.
“Mike looked at himself and Felipe did too”

Malagasy Tsy nahavita na inona na inona aho
“I didn't accomplish anything.”

What mechanisms can
generate copies?

Not all formalisms can describe copying

In phonology

- FSOT
- SPE (à la *Kaplan & Kay*)

In syntax

- GPSG
- GB (à la *Rogers*)
- WG

Copying doesn't come *for free*

- What mechanisms **can** generate copies at all?
- separate this question from:
 - hierarchies of functional projections
 - the 'right' structure

How to tell?

WCW

$$\{wcw : w \in \{a, b\}^*\}$$

a sentence is in this language iff

1. it has exactly one c
2. the material to the left of c is identical to the material to its right

if you cannot describe this language
you cannot describe copying

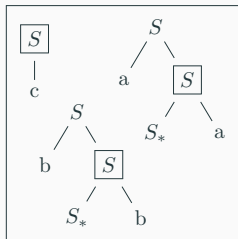
if you can describe this language
I'm interested in what you can do

(Some) mechanisms

- Adjunction
- Indexed rewriting
- Movement
- Copy movement

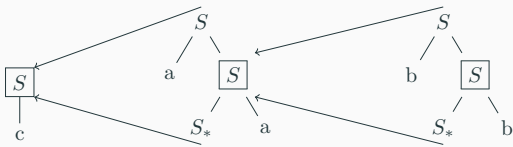
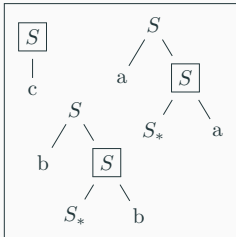
Adjunction

Lexicon



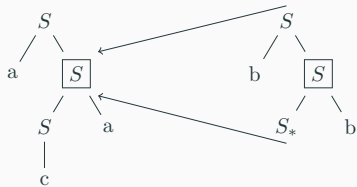
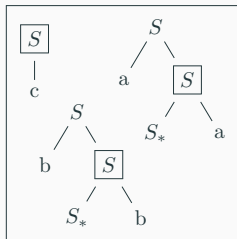
Adjunction

Lexicon



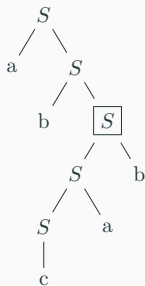
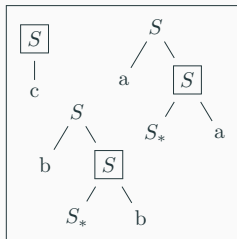
Adjunction

Lexicon



Adjunction

Lexicon



Indexed rewriting

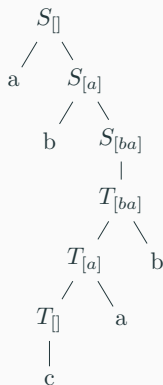
Rules

$$S_{[\alpha]} \rightarrow xS_{[x\alpha]}$$

$$S_{[\alpha]} \rightarrow T_{[\alpha]}$$

$$T_{[x\alpha]} \rightarrow T_{[\alpha]}x$$

$$T_{[]} \rightarrow c$$



Derivation

$$S_{[]} \Rightarrow aS_{[a]} \Rightarrow abS_{[ba]} \Rightarrow abT_{[ba]} \Rightarrow abT_{[a]}b \Rightarrow abT_{[]}ab \Rightarrow abcab$$

Movement

Lexicon

a :: *x *z A z	a :: *A *y x y	ε :: *x *z w
b :: *x *z B z	b :: *B *y x y	c :: *w *y s
		ε :: x z y

Movement

Lexicon

a :: *x *z A z	a :: *A *y x y	ε :: *x *z w
b :: *x *z B z	b :: *B *y x y	c :: *w *y s
		ε :: x z y

a :: *x *z A z

ε :: x z y

Movement

Lexicon

a :: *x *z A z	a :: *A *y x y	ε :: *x *z w
b :: *x *z B z	b :: *B *y x y	c :: *w *y s
		ε :: x z y



Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$

$a :: *A *y x y$



Movement

Lexicon

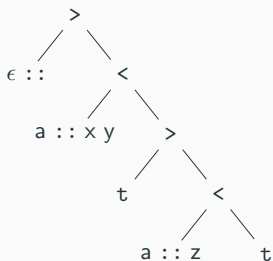
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

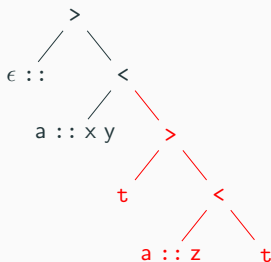
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

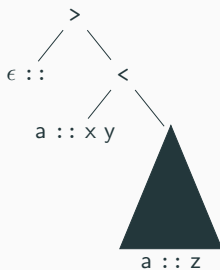
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$

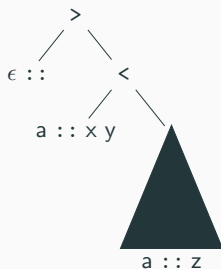


Movement

Lexicon

a :: *x *z A z	a :: *A *y x y	ε :: *x *z w
b :: *x *z B z	b :: *B *y x y	c :: *w *y s
		ε :: x z y

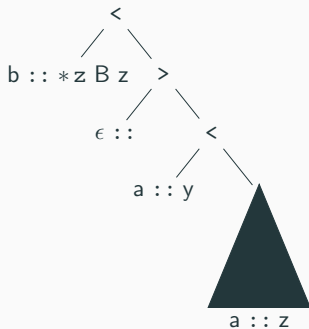
b :: *x *z B z



Movement

Lexicon

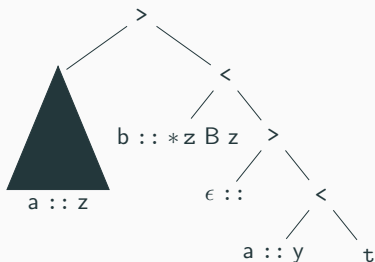
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

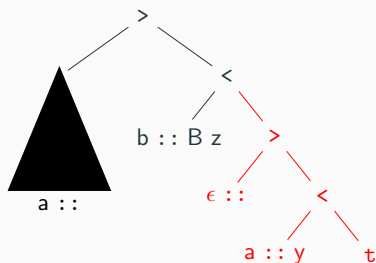
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

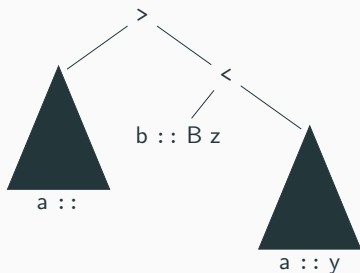
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$

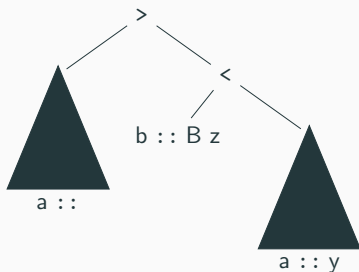


Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$

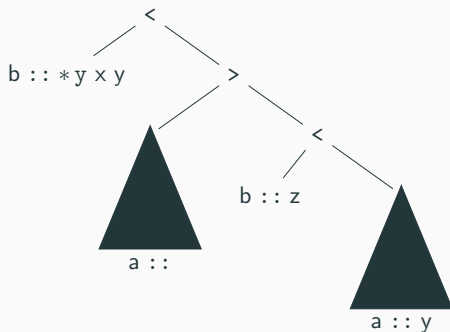
$b :: *B *y x y$



Movement

Lexicon

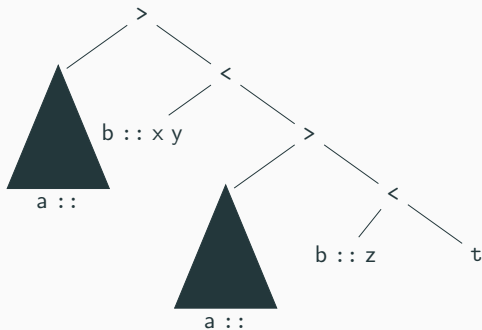
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$

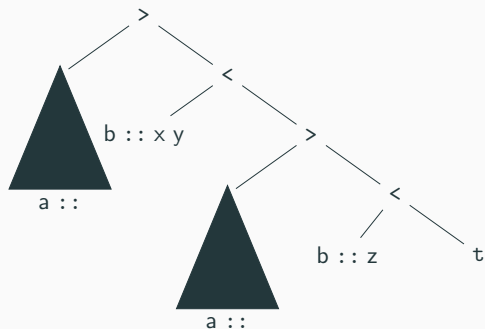


Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$

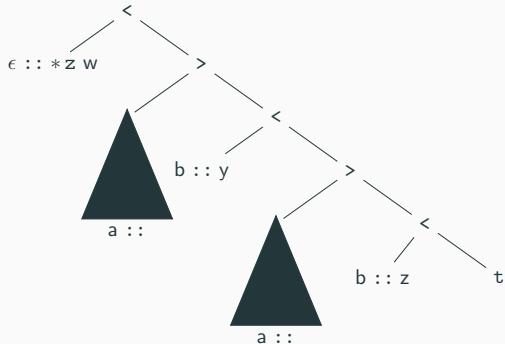
$\epsilon :: *x *z w$



Movement

Lexicon

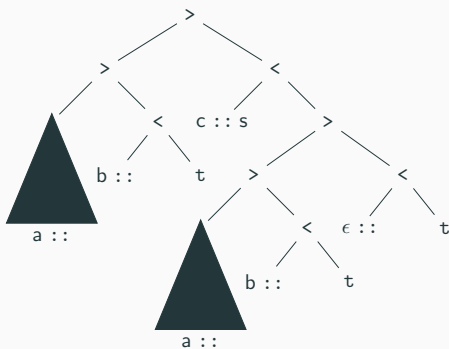
$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Movement

Lexicon

$a :: *x *z A z$	$a :: *A *y x y$	$\epsilon :: *x *z w$
$b :: *x *z B z$	$b :: *B *y x y$	$c :: *w *y s$
		$\epsilon :: x z y$



Copying

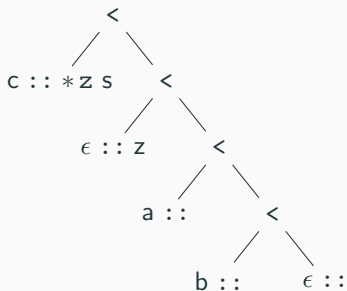
Lexicon

a	::	*x	x	ε	::	x
b	::	*x	x	ε	::	*x s z
c	::	*Ŷ	*z	s		

Copying

Lexicon

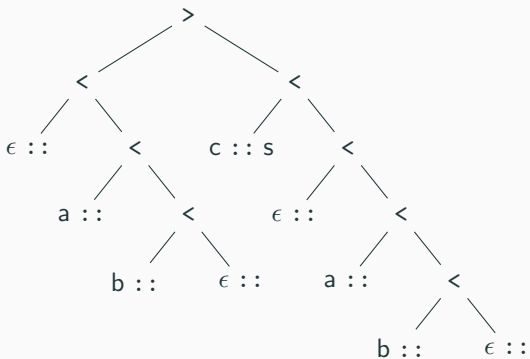
$a :: *x x$	$\epsilon :: x$
$b :: *x x$	$\epsilon :: *x s z$
$c :: *s *z s$	



Copying

Lexicon

$a :: *x x$	$\epsilon :: x$
$b :: *x x$	$\epsilon :: *x s z$
$c :: *s$	$*z s$



Different mechanisms, *same patterns*

- TAG
- CCG
- Tree-local MCTAG
- MCFTG
- LIG

Different mechanisms, *same patterns*

- ACG(2,3)
- CFTG

Different mechanisms, *same patterns*

- MG
- Set-local MCTAG
- ACG(2,>3)
- MCFG
- STR(CFHG)
- MSOT(REG)

Different mechanisms, *same patterns*

- CMG
- almost-linear $ACG(2, >3)$
- PMCFCG

Take home message

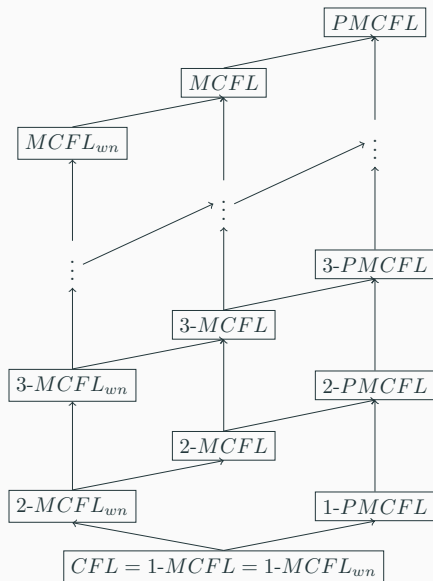
Patterns are real

grammars are a *mode of presentation* of underlying patterns

Meaningful to ask

which patterns are instantiated in NL?

Inclusion diagram



What copying patterns are there?

Separating MCFL_{wn} from MCFL

The copying theorem (Kanazawa & Salvati, 2010)
if $L = \{w_1cw_2 \mid w_1 \in L_0\}$ is a MCFL_{wn} then L_0 is generable by a
non-branching MCFG

Translation: only simple things (L_0) can be copied
 L_0 cannot necessarily involve branching structures

- *give [a good book] [to a nice student]

The first question

Question 1

Are only structurally trivial things copyable?

- gk conjectures: **no!**

A *no* answer means

TAG, CCG, etc are **wrong**

A *yes* answer means

Huge confirmation of a deep prediction of TAG, CCG

Separating MCFL from PMCFL

Semilinearity

intuition: Operations add a fixed, finite amount of structural elements

Translation: copies cannot be copied

$$a \Rightarrow aa \not\Rightarrow aa$$

Exponential growth is just recursive doubling

The second question

Question 2

Can copies be made of copies (of copies ...)?

- gk conjectures: **yes!**

A *no* answer means

Huge confirmation of mild-context sensitivity hypothesis

A *yes* answer means

Hypothesis of mild-context sensitivity (Joshi, 85) is **wrong**

... and with it most restrictive grammar formalisms

The limits of full copying

Non-semilinear patterns
must involve exact copying

Translation: recursive copying must be exact

- only recursive copying is truly **copying**
- non-recursive copying can just as easily generate *opposite* 'copies'

$$wc\bar{w} = \{ac\bar{a}, abc\bar{a}\bar{b}, abbc\bar{a}\bar{b}\bar{b}, \dots\}$$

- let $\bar{a} = b$ and $\bar{b} = a$; can't have

$$a \rightarrow \underline{ba} \rightarrow \underline{abba} \rightarrow \underline{baababba} \rightarrow \underline{abbabaabbaababba}$$

The third question

Question 3

Can we have non-exact (i.e. structure-only) copying in recursive copy constructions?

- gk conjectures: **No!**

What copying patterns are
there in Language?

Chinese A-not(A)

- Ni ting de dong
"You understand"
- Ni ting bu dong
"You don't understand"
- Ni ting de dong ting bu dong
"Do you understand?"

Chinese A-not(A)

- Ni ting de dong
"You understand"
- Ni ting bu dong
"You don't understand"
- Ni ting de dong ting bu dong
"Do you understand?"

Yoruba verbal relatives

- Bode ti Akin ṣubu
Bode push Akin fall
- Titi ti Bode ti Akin ṣubu
pushing rel Bode push Akin fall
- Titi Akin ṣubu ti Bode ti Akin ṣubu
pushing Akin fall rel Bode push Akin fall
- Titi ṣubu ti Bode ti Akin ṣubu
pushing fall rel Bode push Akin fall

Conclusions

- There are deep and fundamental questions that we need to ask about copying
 - is copying always small?
 - is copying only of simplexes?
 - is copying recursive?
- The answers to these will sink various approaches to grammar
- Once we know what patterns there are, we can focus on coming up with an elegant description language

Thank you