From *bedtime* to *spacetime*: The hidden language behind compound words

Chenchen (Julio) Song, cs791@cam.ac.uk

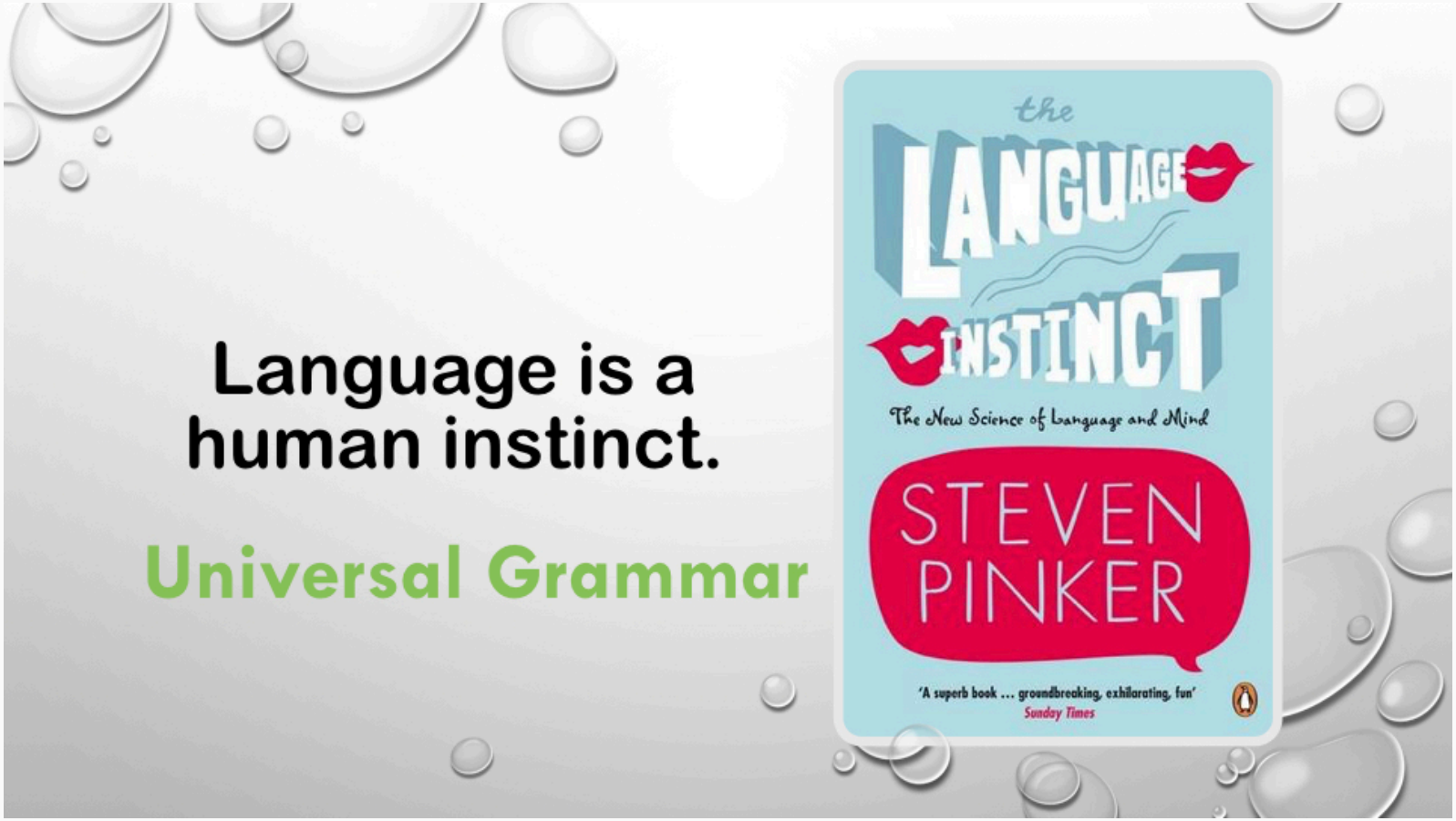
3rd-year PhD in Linguistics

22 Oct 2017, Caius Graduate Research Symposium

Introduction

My PhD project: *Flexibility of Syntactic Categories*

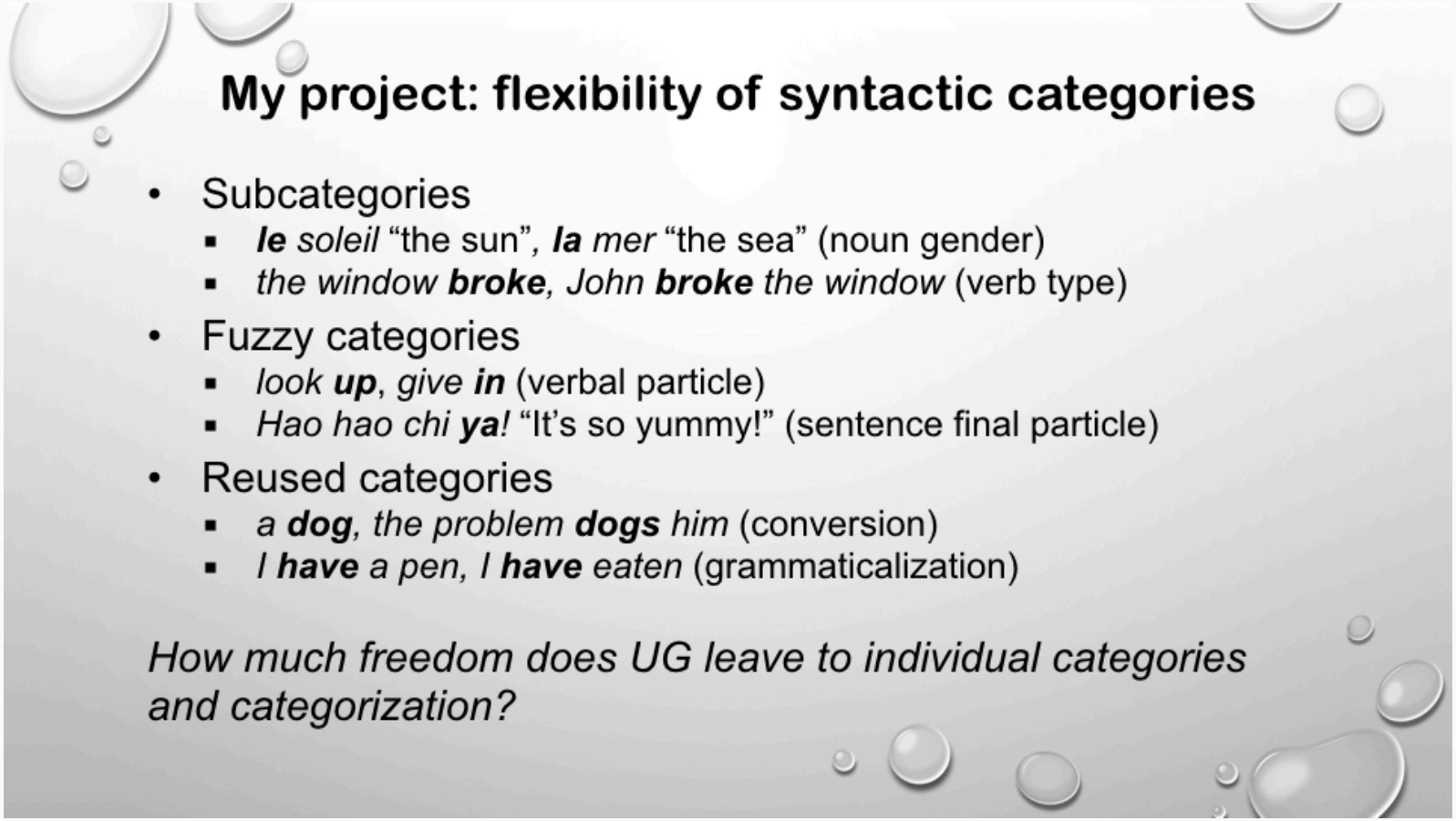
If you came to my MCR talk last term… ☺



Song, C. (2017). Universal Grammar and where to find it. Gonville & Caius MCR talk. May 30.

My PhD project: *Flexibility of Syntactic Categories*

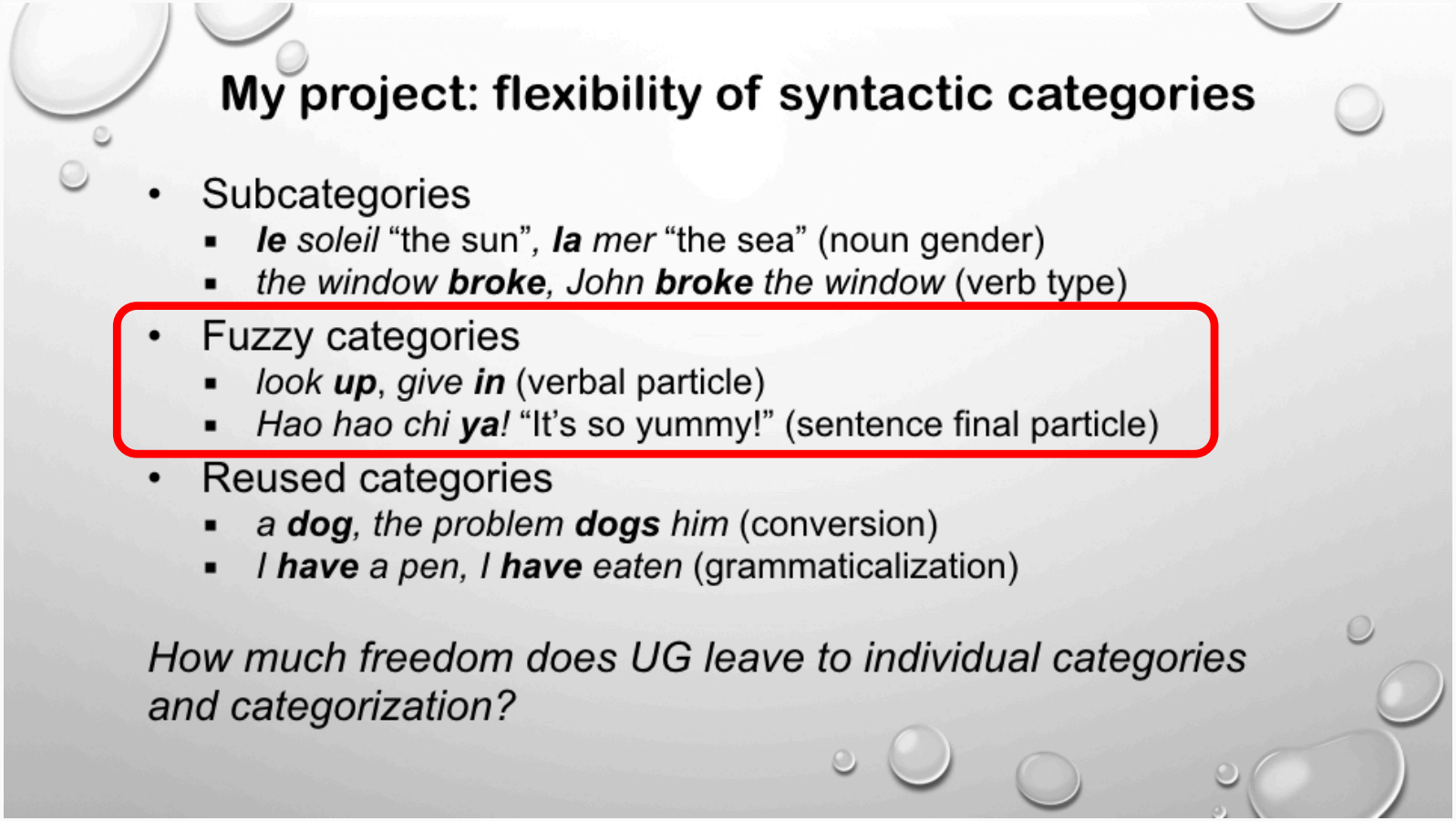
If you came to my MCR talk last term… ☺



Song, C. (2017). Universal Grammar and where to find it. Gonville & Caius MCR talk. May 30.

My PhD project: *Flexibility of Syntactic Categories*

If you came to my MCR talk last term… ☺



Song, C. (2017). Universal Grammar and where to find it. Gonville & Caius MCR talk. May 30.

Compound words

Compound words

bbc.co.uk

Brexit: Has Theresa May promised EU more money?

Laura Kuenssberg
Political editor

2 hours ago | UK Politics

Share

It's all about the money.

The UK and the EU have managed together to make a tiptoe forward in the Brexit talks.

But listening to EU leaders this afternoon it is abundantly clear that unless they change their minds, the UK is going to have to budge on the cash to make enough progress by December to be able to truly get on to the next phase of talks.

Certainly public money is tight in this administration, but frankly a hefty Brexit bill in exchange for a good deal would be the one big payment that the Chancellor Philip

Top Stories

Brexit deadlock talk 'exaggerated' - Tusk
EU leaders say there is not enough progress to start trade talks yet, but they hope to begin in December.
3 hours ago

Anger over Trump's UK crime tweet
2 hours ago

'Loaded' Kodi box trader avoids jail
1 hour ago

Most Read

1. Middlesbrough modified Kodi box trader gets suspended jail term
2. Anger over Donald Trump's UK crime tweet
3. David Blaine under investigation over UK model's rape claim
4. Meteor shower to fill weekend skies
5. MP Clive Lewis sorry for 'unacceptable language' after video
6. Why Tina Turner came out of retirement

Compound words

bbc.co.uk

Brexit: Has Theresa May promised EU more money?

Laura Kuenssberg
Political editor

2 hours ago | UK Politics

Share

It's all about the money.

The UK and the EU have managed together to make a tiptoe forward in the Brexit talks.

But listening to EU leaders this afternoon it is abundantly clear that unless they change their minds, the UK is going to have to budge on the cash to make enough progress by December to be able to truly get on to the next phase of talks.

Certainly public money is tight in this administration, but frankly a hefty Brexit bill in exchange for a good deal would be the one big payment that the Chancellor Philip

Top Stories

Brexit deadlock talk 'exaggerated' - Tusk
EU leaders say there is not enough progress to start trade talks yet, but they hope to begin in December.
3 hours ago

Anger over Trump's UK crime tweet
2 hours ago

'Loaded Kodi box trader avoids jail
1 hour ago

Most Read

1. Middlesbrough modified Kodi box trader gets suspended jail term
2. Anger over Donald Trump's UK crime tweet
3. David Blaine under investigation over UK model's rape claim
4. Meteor shower to fill weekend skies
5. MP Clive Lewis sorry for 'unacceptable language' after video
6. Why Tina Turner came out of retirement

The compounds we've seen

short	longer	even longer
tip-toe	*trade talk*	*week-end sky*
after-noon	*meteor shower*	*Kodi box trader*
dead-lock	*top story*	*European Union (EU) leader*
pro-gress	*crime tweet*	*United Kingdom (UK) Politics*
Middles-brough	*jail term*	*Brexit dead-lock talk*
week-end	*rape claim*	...

The compounds we've seen

short	longer	even longer
tip-toe	*trade talk*	*week-end sky*
after-noon	*meteor shower*	*Kodi box trader*
dead-lock	*top story*	*European Union (EU) leader*
pro-gress	*crime tweet*	*United Kingdom (UK) Politics*
Middles-brough	*jail term*	*Brexit dead-lock talk*
week-end	*rape claim*	...

All made up of a modifier and a head.

The compounds we've seen

short	longer	even longer
tip-toe	*trade talk*	*week-end sky*
after-noon	*meteor shower*	*Kodi box trader*
dead-lock	*top story*	*European Union (EU) leader*
pro-gress	*crime tweet*	*United Kingdom (UK) Politics*
Middles-brough	*jail term*	*Brexit dead-lock talk*
week-end	*rape claim*	...

All made up of a modifier and a head.

The compounds we've seen

short	longer	even longer
tip-toe	*trade talk*	*week-end sky*
after-noon	*meteor shower*	*Kodi box trader*
dead-lock	*top story*	*European Union (EU) leader*
pro-gress	*crime tweet*	*United Kingdom (UK) Politics*
Middles-brough	*jail term*	*Brexit dead-lock talk*
week-end	*rape claim*	...

All made up of a modifier and a head.

The compounds we've seen

short	longer	even longer
tip-toe	*trade talk*	*week-end sky*
after-noon	*meteor shower*	*Kodi box trader*
dead-lock	*top story*	*European Union (EU) leader*
pro-gress	*crime tweet*	*United Kingdom (UK) Politics*
Middles-brough	*jail term*	*Brexit dead-lock talk*
week-end	*rape claim*	...

All made up of a modifier and a head.

Here I only discuss nominal examples.

The modifier can get very long

Germanic languages are very good at this...

German *Donau-dampf-schiff-fahrts-gesellschafts-kapitäns-mütze*

"Danube steamboat shipping company Captain's hat"

Dutch *kinder-carnavals-optocht-voor=bereidings-werkzaamheden-plan*

"children's carnival procession preparation activity plan"

Swedish *Nord-öster-sjö-kust-artilleri-flyg-spanings-simulator-an=läggnings-materiel-under=hålls-upp=följnings-system-diskussions-in=läggs-för=beredelse-arbeten*

"Northern Baltic sea coast artillery reconnaissance flight simulator facility equipment maintenance follow-up system discussion post preparation work"

The modifier can get very long

Uralic languages are no slouches...

Hungarian *vér-nyomás-mérő-készülék-gyártó-üzem*

“blood pressure measuring apparatus producing factory”

Finnish *lento-kone-suihku-turbiini-moottori-apu-mekaanikko-ali-upseeri-oppilas*

“airplane jet turbine engine auxiliary mechanic non-commissioned officer student”

The modifier can get very long

Uralic languages are no slouches…

Hungarian	*vér-nyomás-mérő-készülék-gyártó-üzem* “blood pressure measuring apparatus producing factory”
Finnish	*lento-kone-suihku-turbiini-moottori-apu-mekaanikko-ali-upseeri-oppilas* “airplane jet turbine engine auxiliary mechanic non-commissioned officer student”

And don’t forget their English translations! 😊

Modifier-head compounds seem to be universal

Even in languages very different from English...

Mapuche (language isolate, polysynthetic, central Chile)

ilo-korü	“meat-soup”
wingka-kofke	“whiteman-bread ➙ European bread”
mapu-dungun	“land-words ➙ the Mapuche language”

Mohawk (Iroquoian, polysynthetic, northeastern North America)

kátsche’ káhi	“jug fruit ➙ pear”
akohsá:tens ohsò:kwa’	“horse nut ➙ horse chestnut”
arawé:n’ othè:sera’	“oat flour ➙ oatmeal”

Lieber, R. & P. Štekauer (eds.) (2009). *The Oxford Handbook of Compounding*. OUP.

Modifier-head compounds seem to be universal

...and in languages that have been claimed to be compound-free.

Evenki (Tungusic, agglutinative, China/Russia)

eyji shee	"brick tea"
aaxin jolo	"liver stone ➙ marble"
unaaji ute	"girl son ➙ daughter"

Ainu (language isolate, agglutinative, Japan)

atuy asam	"bottom sea ➙ sea bottom"
kamuy napuri	"mountain god ➙ holy mountain"
supuya kur	"trace smoke ➙ trace of smoke"

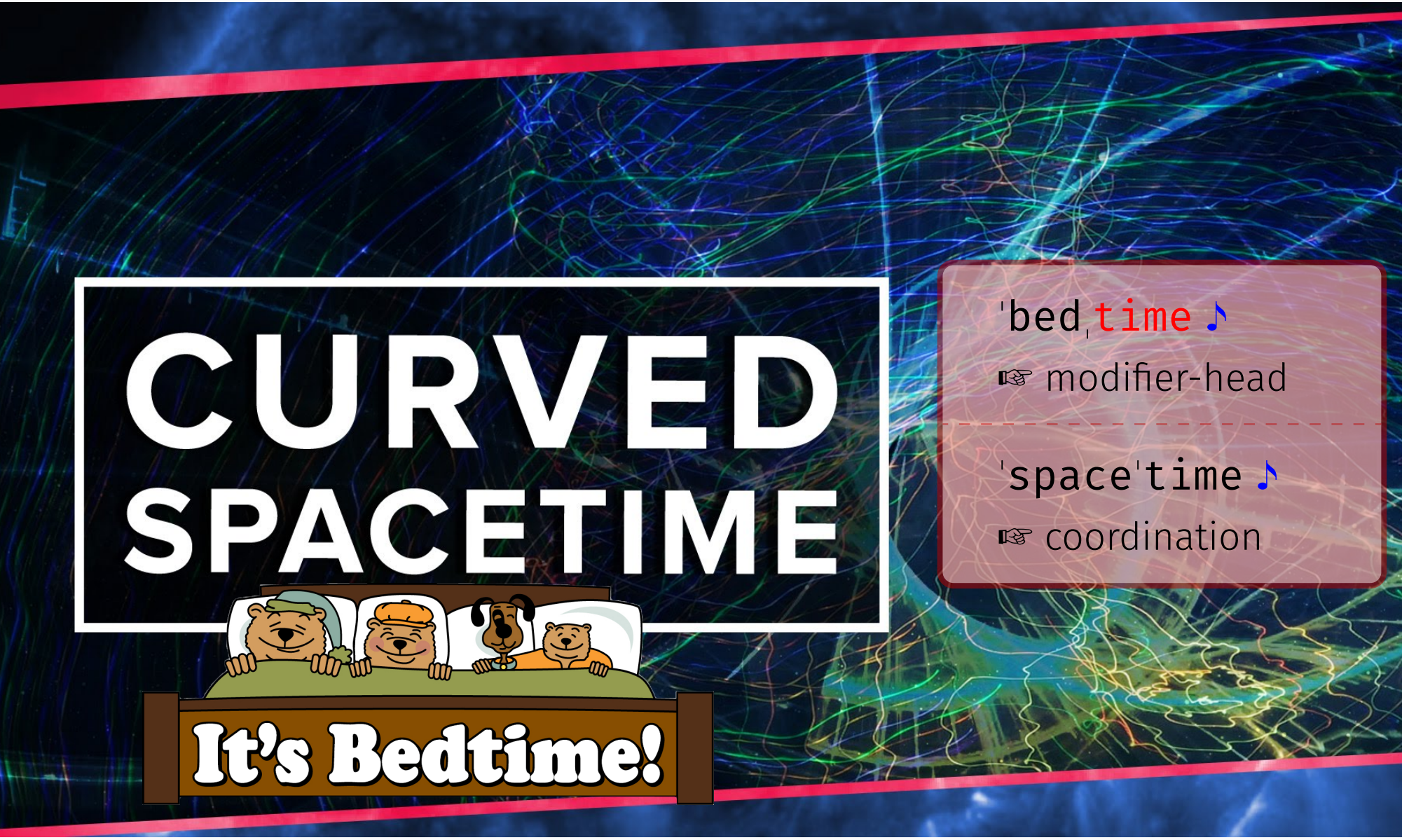
Lieber, R. & P. Štekauer (eds.) (2009). *The Oxford Handbook of Compounding*. OUP.
Hu, Z. & K. Chao (1986). *A Description of the Evenki Language*. Beijing: Publishing House of Minority Nationalities.

So far, so good.

bedtime vs. *spacetime*

Another type of compound

Another type of compound



ˈbedˌtime ♪

☞ modifier-head

ˈspaceˈtime ♪

☞ coordination

Another type of compound

But they are both noun-noun in structure...

'bed,time ♪

☞ modifier-head

'space'time ♪

☞ coordination

How can one structure generate two sound-meaning effects?
A robot wouldn’t know! 😝

Coordination compounds are no less common

English	*poet-doctor, Austria-Hungary*	
Spanish	*casa-quinta*	“house-villa ➙ country house”
	pollera pantalón	“pant-skirt ➙ skirt trousers”
Greek	*andro-jino*	“man-woman ➙ couple”
	alato-pipero	“salt-pepper”
Mandarin	*fù-mǔ*	“father-mother ➙ parents”
	yǔ-zhòu	“space-time ➙ universe”
Erzá Mordvin	*ruća-panar*	“skirt-shirt ➙ clothes”
	čińek-veńek	“day-night ➙ day and night”
Evenki	*ashi edi*	“husband-wife
	axin nexun	“brother brother ➙ friend”

Lieber, R. & P. Štekauer (eds.) (2009). *The Oxford Handbook of Compounding*. OUP.
Wälchli, B. (2005). *Co-compounds and Natural Coordination*. OUP.
Hu, Z. & K. Chao (1986). *A Description of the Evenki Language*. Beijing: Publishing House of Minority Nationalities.

Coordination compounds are no less common

And the numbers!

English	*twenty-one, twenty-two, twenty-three*
French	*vingt-et-un 21, vingt-deux 22, vingt-trois 23*
Japanese	*jū-ichi 11, jū-ni 12, jū-san 13*
Evenki	*jaan emun 11, jaan juur 12, jaan ilan 13*

bedtime vs. *spacetime*

Looking at the structure, spacetime *is the expected compound type.*

'bed,time ♪

☞ asymmetric

'space'time ♪

☞ symmetric

Modification is well-known to be exceptional.

Modification

Modification is exceptional

In syntax, it has a special generative mechanism.

☞ Modification forms ordered pairs:

$$<\alpha, \beta> = \{\{\alpha\}, \{\alpha, \beta\}\} \neq <\beta, \alpha>$$

☞ Everything else forms simple sets:

$$\{\alpha, \beta\} = \{\beta, \alpha\}$$

Examples:

good students	=	<*students*, *good*>
	=	{{*students*}, {*students*, *good*}}
study linguistics	=	{*study*, *linguistics*}

Chomsky, N. (2004). Beyond Explanatory Adequacy. In A. Belletti (ed.) *Structures and Beyond*. OUP, 104–131.

Modification is exceptional

In semantics, it has a special composition rule.

☞ Modification is function conjunction:

$$[\![<\alpha, \beta>]\!] = [\![\alpha]\!] \ \& \ [\![\beta]\!]$$

☞ Everything else is functional application:

$$[\![\{\alpha, \beta\}]\!] = [\![\alpha]\!]([\![\beta]\!]) \text{ or } [\![\beta]\!]([\![\alpha]\!])$$

Examples:

$$\begin{aligned}
[\![\textit{good students}]\!] &= [\![\textit{students}]\!] \ \& \ [\![\textit{good}]\!] \\
&= \lambda x \ . \ x \text{ are students } \& \ x \text{ are good} \\
[\![\textit{study linguistics}]\!] &= [\![\textit{study}]\!]([\![\textit{linguistics}]\!]) \\
&= [\lambda y \lambda z \ . \ z \text{ study } y](\text{linguistics}) \\
&= \lambda z \ . \ z \text{ study linguistics}
\end{aligned}$$

Heim, I. & A. Kratzer (1998). *Semantics in Generative Grammar*. Blackwell.

The exceptionality has been encoded in the mode of combination.

Borer-Chomsky Conjecture

All parameters of variation are attributable to differences in features of particular items in the lexicon.

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (ed.) *The Limits of Syntactic Variation*. Benjamins, 351–374.

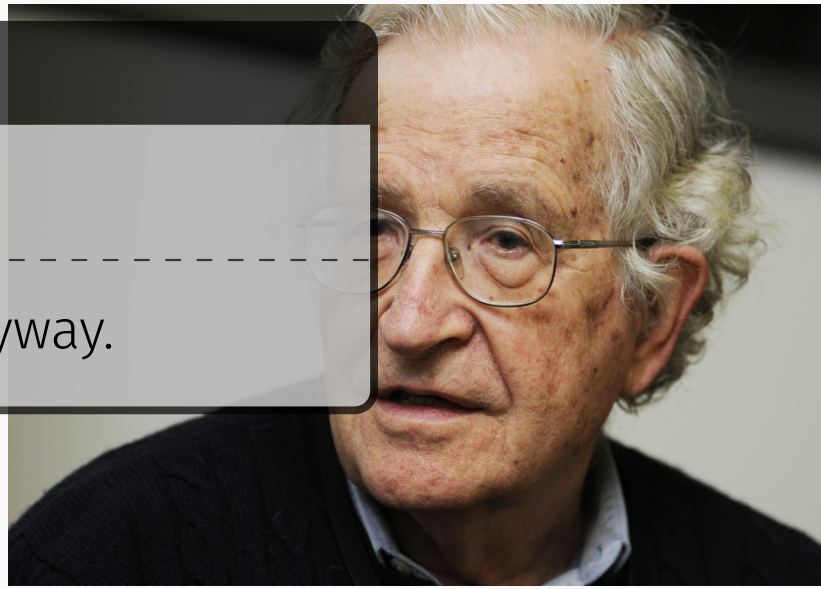
Borer-Chomsky Conjecture

All parameters of variation are attributable to differences in features of particular items in the lexicon.

Why?

Remove burden from UG.

Lexicon has to be learned anyway.



Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (ed.) *The Limits of Syntactic Variation*. Benjamins, 351–374.

Borer-Chomsky Conjecture

All parameters of variation are attributable to differences in features of particular items in the lexicon.

Examples:

- Null subject phenomena $<<$ agreement property on Tense
- Word order variation $<<$ movement property on categories

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (ed.) *The Limits of Syntactic Variation*. Benjamins, 351–374.

Borer-Chomsky Conjecture

All parameters of variation are attributable to differences in features of particular items in the lexicon.

Examples

- Null
- Wor

Parlo/i/a italiano. (Italian, consistent)
**(Je/tu/il) parle(s) français.* (French, no)
Kàn-jiàn le. "Saw." (Mandarin, radical)

Tense

egories

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (ed.) *The Limits of Syntactic Variation*. Benjamins, 351–374.

Can modification be reduced to lexicon too?

A richer metalanguage

The metalanguage behind Humanese

It's a language of categories, e.g. *good students study linguistics.*

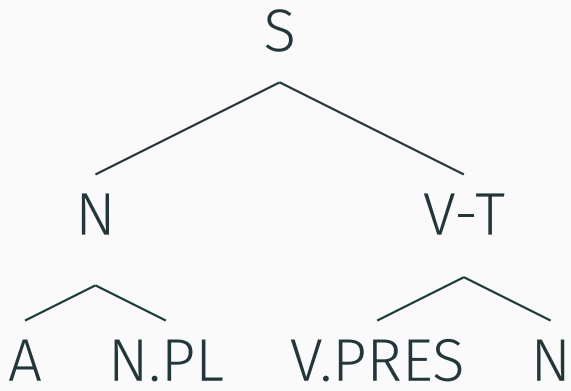
$$[_{S}\ [_{N}\ \text{A N.PL}\ ]\ [_{V\text{-}T}\ \text{V.PRES N}\ ]]$$

The metalanguage behind Humanese

It's a language of categories, e.g. *good students study linguistics.*

$$[_{S}\ [_{N}\ \text{A N.PL}\ ]\ [_{V\text{-}T}\ \text{V.PRES N}\ ]]$$

Or more vividly...



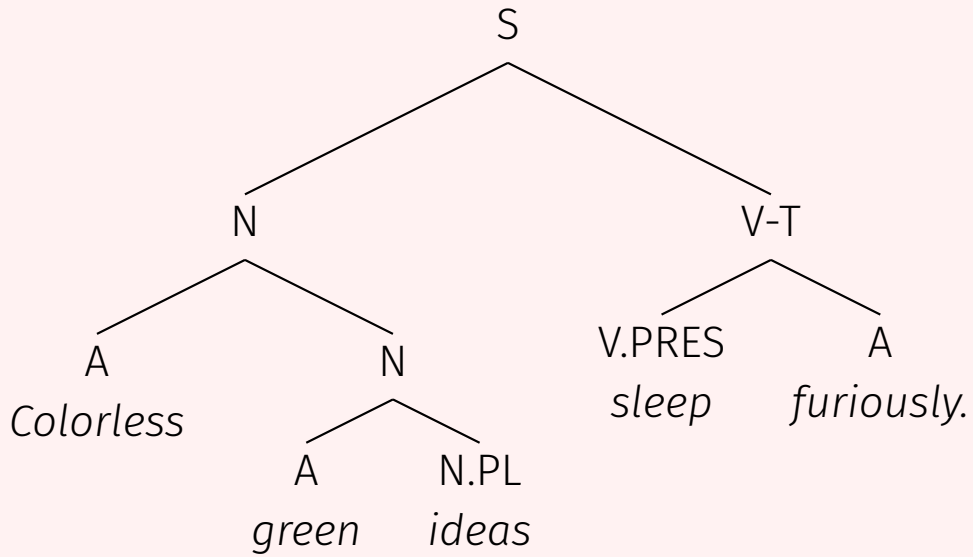
The metalanguage behind Humanese

It's a language of categories, e.g. *good students study linguistics.*

Or more

Grammatical nonsense 😒

Colorless green ideas sleep furiously.



The categories behind compound words

☞ Obviously the concrete categories N(oun), V(erb), A(djective).

e.g. *week-end* N-N, *rape claim* V-N, *dead-lock* A-N.

The categories behind compound words

☞ Obviously the concrete categories N(oun), V(erb), A(djective).

e.g. *week-end* N-N, *rape claim* V-N, *dead-lock* A-N.

☞ Also some abstract categories like &, ∀.

e.g. *Austria-Hungary* N&N, *every-body* ∀N.

The categories behind compound words

☞ Obviously the concrete categories N(oun), V(erb), A(djective).

e.g. *week-end* N-N, *rape claim* V-N, *dead-lock* A-N.

☞ Also some abstract categories like &, ∀.

e.g. *Austria-Hungary* N&N, *every-body* ∀N.

'space'time >> N&N, 'bed'time >> ??

My proposal: a “fuzzy” category

It’s both concrete and abstract.

- Stands for elements otherwise concrete, e.g. *bedtime*, *crime tweet*, *meteor shower*.
- Imposes no category box of its own, e.g. *rape claim*, *dead-lock*, *end-of-term exam*.

My proposal: a “fuzzy” category

It’s both concrete and abstract.

- Stands for elements otherwise concrete, e.g. *bedtime*, *crime tweet*, *meteor shower*.
- Imposes no category box of its own, e.g. *rape claim*, *dead-lock*, *end-of-term exam*.

Its characteristics are clear.

- Turns anything into a modifier.
- Doesn’t change the head’s category.
- Very basic and widespread.

My proposal: a “fuzzy” category

It’s both concrete and abstract.

- S
  c
- I
  dead-lock, *end-of-term exam*.

The idea

[Cat:__]	“I’m looking for a value” 😍
[Cat:__] + N = [Cat: N]	“I become part of a big N” 😇

Its characteristics are clear.

- Turns anything into a modifier.
- Doesn’t change the head’s category.
- Very basic and widespread.

My proposal: a “fuzzy” category

It’s both concrete and abstract

– S

– I

dead-lock, *end-of-term exam*.

Its

The idea

[Cat:__] “I’m looking for a value” 😍

[Cat:__] + N = [Cat: N] “I become part of a big N” 😇

Any visibility?

Linking element:

GER *kapitän-s-mütze*, DUT *carnaval-s-optocht*, SWE *anläggning-s-materiel*

Voicing change:

JAP *te-gami* “hand-paper ➙ letter”, BAS *su-pazter* “fire-corner ➙ hearth”

Predictions

Some textbookish ones...

- Universality: it's the capacity to form a category.
  - children not born with categories (poor UG)
  - but they can create a lot (experience)

Predictions

Some textbookish ones...

- Universality: it's the capacity to form a category.
  - children not born with categories (poor UG)
  - but they can create a lot (experience)
- Flexibility: it allows for a lot of modifier material
  - the same material reused (maximize minimal means)
  - e.g. *baby boom* vs. *babysit*, *shame culture* vs. *shame eat*
  - and complex material is allowed too (recursion)
  - e.g. the Germanic long words...

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry* 36. 1–22.
Biberauer, T. (2017). Factors 2 and 3: A principled approach. Song, C. & J. Baker (eds.) *Cambridge Occasional Papers in Linguistics* 10. 38–65.

A jolly one: we can now distinguish *bedtime* and *spacetime*!
(So can our robot!)

Modification can now be encoded in the lexicon.

'bed,time ♪

☞ Cat-N

'space'time ♪

☞ N-N

Thank you!