Linguistic Creativity and Computational Modeling of Language

**Gothenburg, CLASP colloquium,
February 10, 2023**

Sina Zarrieß
Computational Linguistics (CLAUSE)
Bielefeld University
https://clause-bielefeld.github.io/

Joint work with…

Ralf Vogel

Jutta Hartmann

Joana Cholin

(Heads of the planned research initiative)

And more…

Oliver Bott

Petra Wagner

Arndt Riester

Hendrik Buschmeier

Outline

- **Intro: linguistic creativity**

- Modeling linguistic creativity

- Creativity from an NLG perspective

  - Experiments on decoding

- Outlook

  - Further tasks

Emil (9y) and his friend

Ciaosen Digga

Ciao **Brudi** [bruːdi]

- Conventional units:

- ``Ciao''

- ``Moin'' (Northern-German: ``hello")

- Colloquial form: ``Moinsen''

- Conventional units:

- ``Bro'' [broː]

- ``Bruder'' [ˈbʁuːdɐ] („brother")

- -``i" (diminutive for first names)

Originality in every-day language use

- **Newly construed linguistic units**
  - e.g. ``Brudi''
- **New uses of existing units**
  - e.g. ``Querdenker'' (previously: „lateral thinker", now: „anti-vaccination/right-wing person")
- **New linguistic units**
  - e.g. ``Ciaosen''
- **New linguistic solutions to communicative problems**
  - e.g. referring to a tangram shape as an ``ice skater''
- **Changes in a community's linguistic inventory**
  - e.g. [r] instead of [ʁ] in ``Brudi''

Beyond originality: creativity

- Creativity is a far-reaching and interdisciplinary field of research (arts, labor, engineering, …)
- Core criteria for creativity:
  - **Originality:** creative actions result in **new products**
  - **Effectiveness:** creative products serve **some purpose**
  - **Authorship:** crucial in many other domains, but maybe not in language
  - **Intentionality:** maybe not that relevant in language
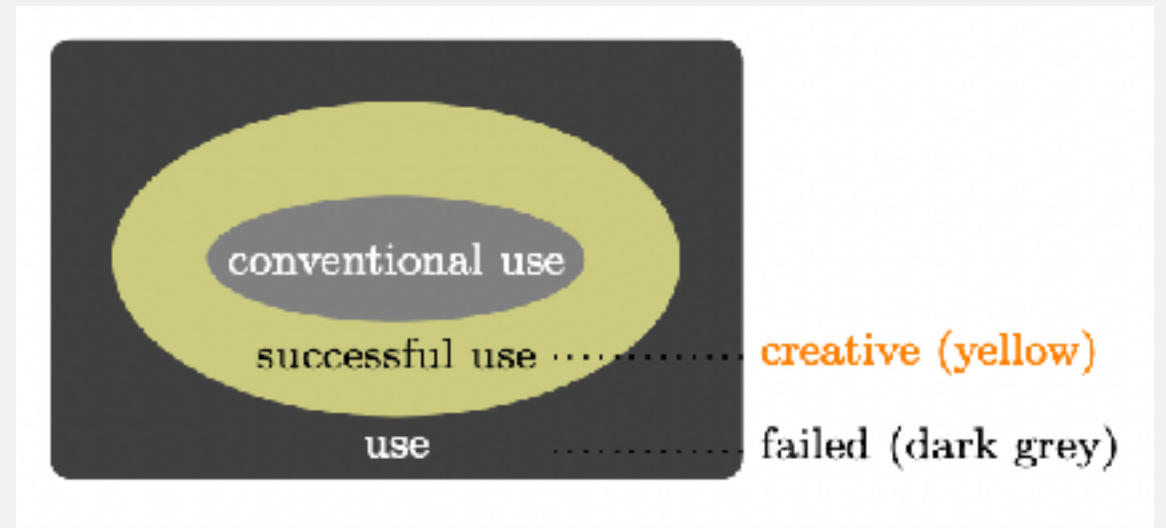
Linguistic creativity

- **Originality**

  - Speakers combine and extend conventionalized units, and they produce and unterstand non-conventionalized units and inventions

- **Effectiveness**

  - Linguistic units are effective when they are communicatively successful, e.g.,

  - … easy to comprehend and process

  - … easy to agree on and reuse in interaction

  - … help to achieve communicative goals

Linguistic creativity *may* result in language change, it is the source of language change, but not every creative unit triggers change

Planned research initiative in BI

- Aim: a theory of language that focusses on the **speaker's expressive freedom**

- Develop empirical and formal approaches to linguistic creativity across domains and levels of language

- Model the linguistic, cognitive, contextual and social **factors that enable and delimit creativity**

Sign A	Interpretation B	Cognition C
creative linguistic units; mechanisms of linguistic creativity; determination of (non-)conventionality; creativity in linguistic form (sound, grammar)	meaning in context; limits of semantic/pragmatic creativity; social significance of creative units; interactive creativity; expressive creativity	processing of creative units and its cognitive conditions; speaker heterogeneity; individual differences; creativity and learning; cognitive modelling

Outline

- Intro: linguistic creativity
- **Modeling linguistic creativity**
- Creativity from an NLG perspective
  - Experiments on decoding
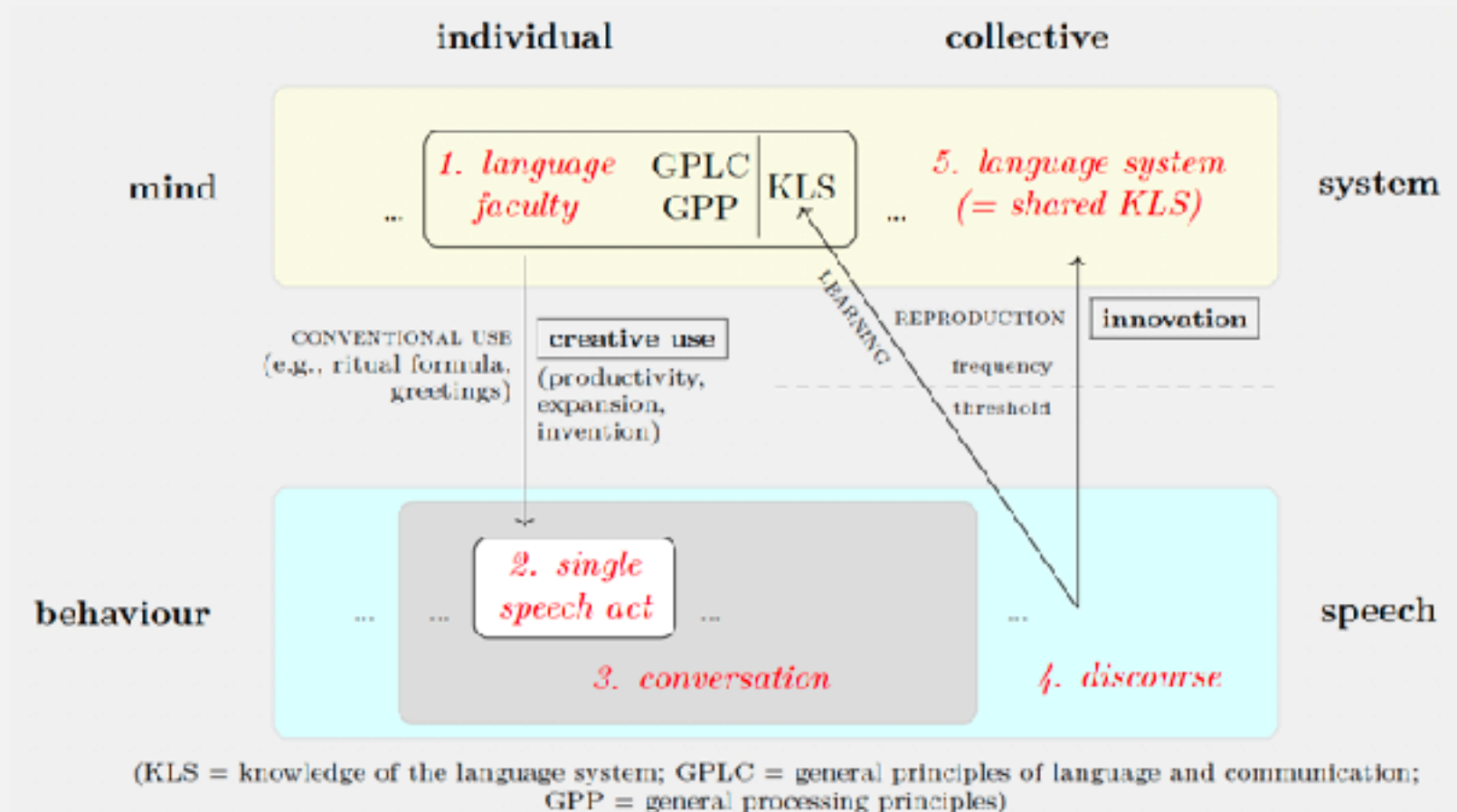- Outlook
  - Further tasks

Linguistic creativity

- Speakers are the „owners" of their language

- … and may use the conventions of their language as „**tools**"

- … but they are not limited to the conventions in their **communicative actions**

- … instead, what they apply in communication, is their linguistic and communicative competence

- … which includes not only the **knowledge of their language system**

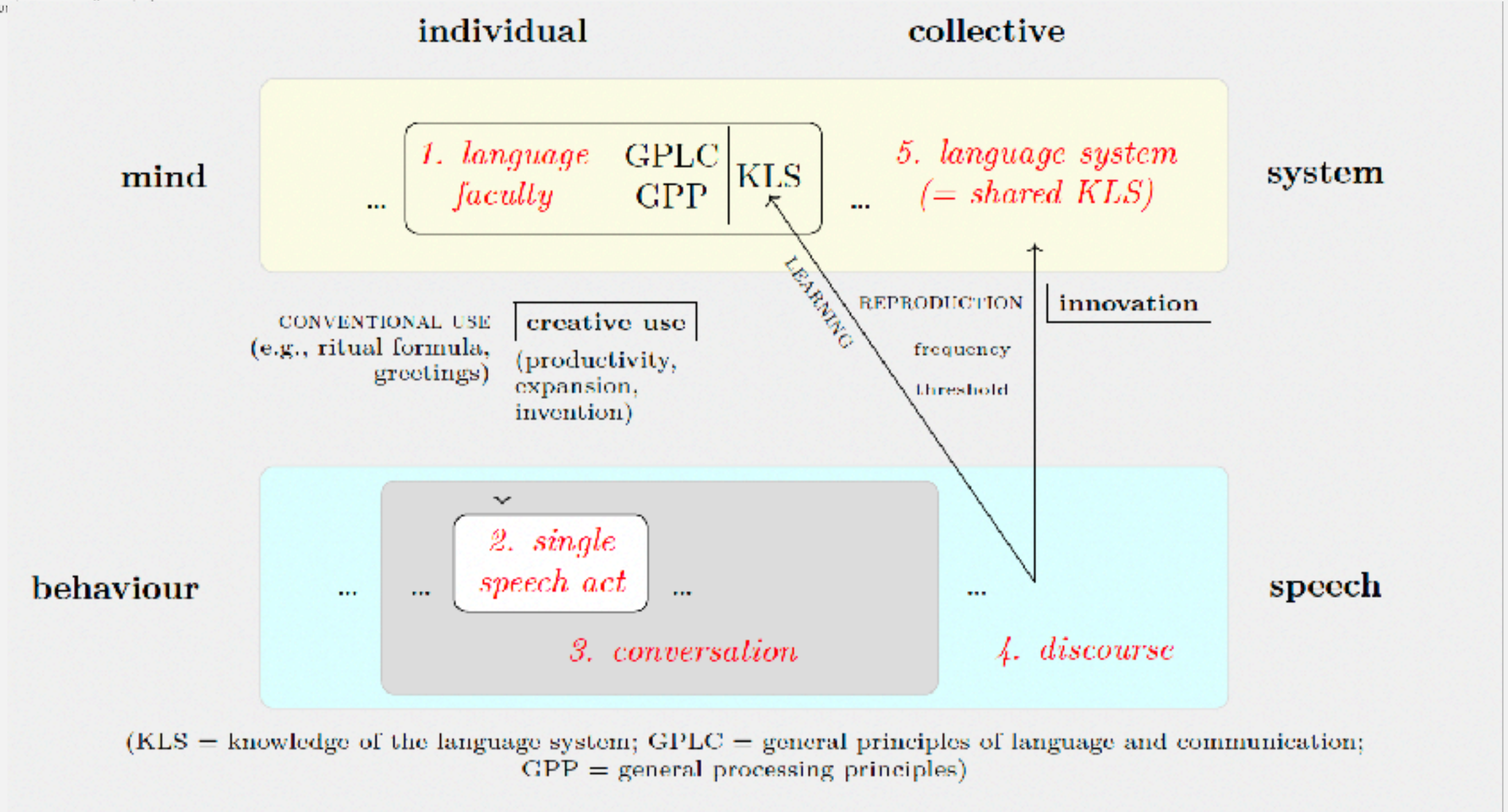- … but also general **cognitive principles**

- Traditional topic in the philosphy of language: Humboldt, de Saussure, Chomsky, a.o.
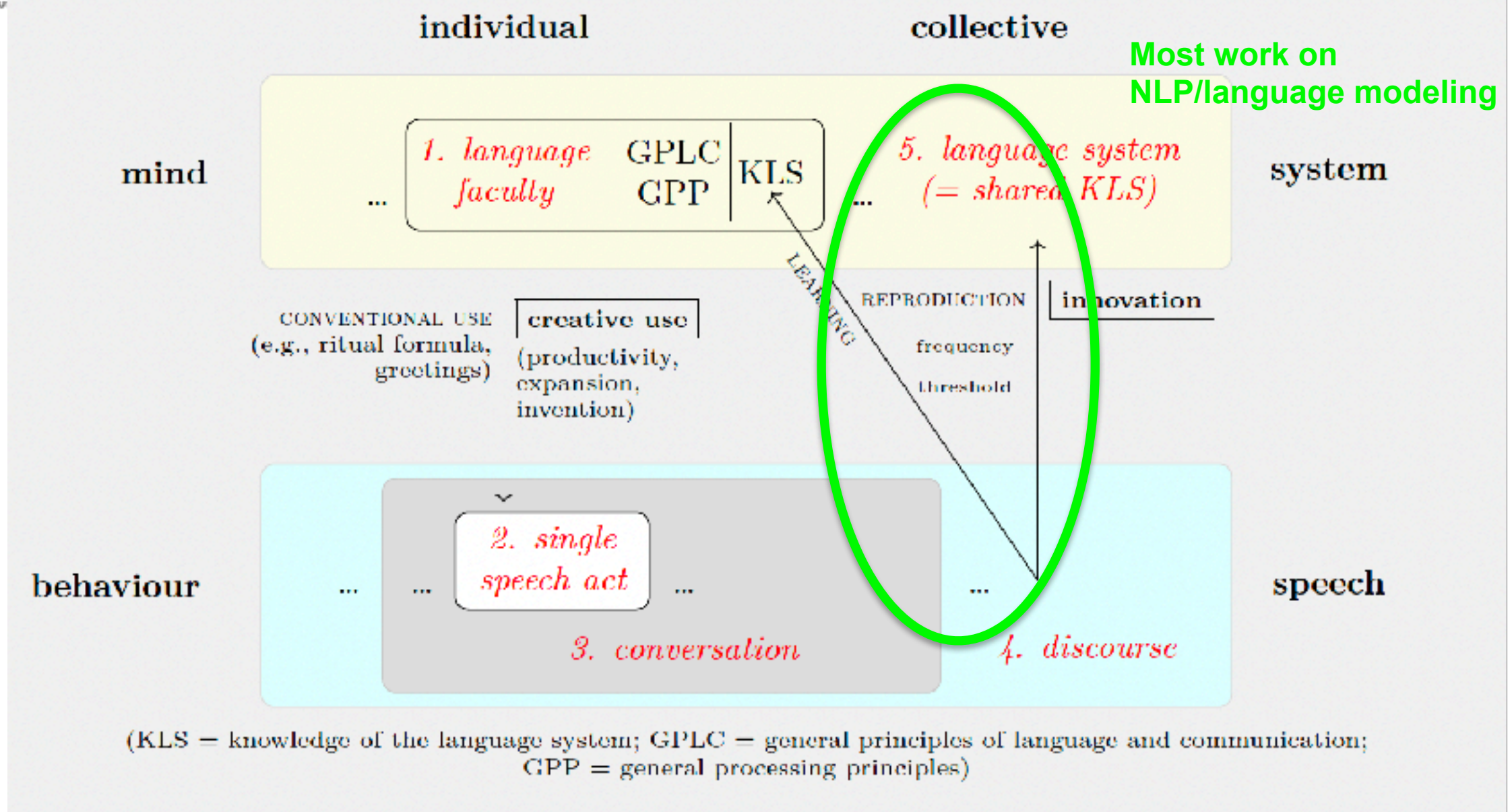
A Neo-Saussurian conception of language (by Ralf Vogel)

- Generative linguistics models the competence of an ideal speaker, as the capacity to **create complex linguistic expressions**
- The Saussurian approach strictly distinguishes the individual and the collective dimension
- This targets the more general capacity to **create the rules** that form the conventions of language

individual   collective

mind

1. language faculty   GPLC GPP KLS

5. language system (= shared KLS)   system

CONVENTIONAL USE (e.g., ritual formula, greetings)   creative use (productivity, expansion, invention)

LEARNING   REPRODUCTION innovation frequency threshold

behaviour   2. single speech act   speech

3. conversation   4. discourse

(KLS = knowledge of the language system; GPLC = general principles of language and communication; GPP = general processing principles)

Where would you put neural language models?

individual   collective

mind   **system**

1. *language faculty*   GPLC   GPP | KLS   ...   5. *language system (= shared KLS)*

CONVENTIONAL USE (e.g., ritual formula, greetings) | creative use (productivity, expansion, invention)

LEARNING   REPRODUCTION | innovation   frequency   threshold

behaviour   **speech**

...   ...   2. *single speech act*   ...

3. *conversation*   4. *discourse*

(KLS = knowledge of the language system; GPLC = general principles of language and communication; GPP = general processing principles)

(KLS = knowledge of the language system; GPLC = general principles of language and communication; GPP = general processing principles)
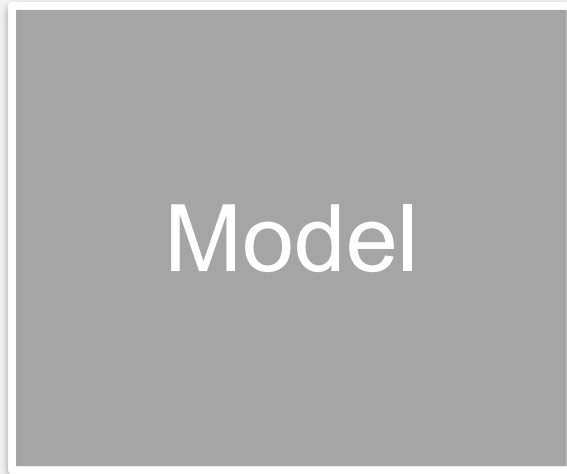
Outline

- Intro: linguistic creativity

- Modeling linguistic creativity

- **Creativity from an NLG perspective**

  - Experiments on decoding neural LMs

- Outlook

  - Further tasks

``Core" Natural Language Generation

Model

Text/ Speech/ Speech Acts

Neural NLG

$$P(y \mid x) = \sum_{j=1}^{J} P(y_j \mid y_1^{j-1}, x, h)$$

Conditional neural language model

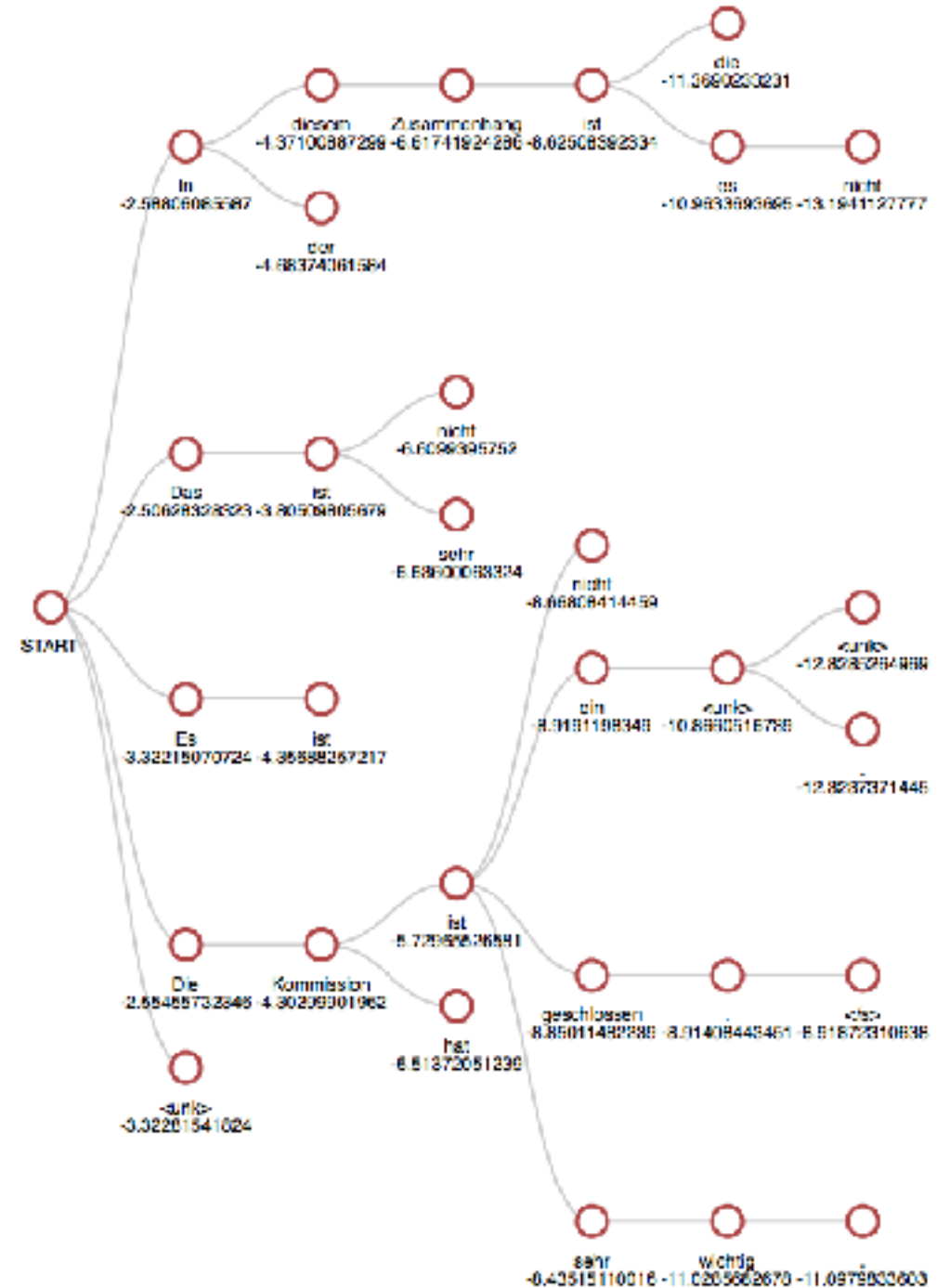*Scores the entire search space, conditioned on input*

$V*$

Default: beam search

Decoding

Infers the output, at test time

Output Sequence

Decoding: Beam search

- Objective: maximize likelihood

- Approximates exhaustive search

- Pruning (partial) sequences with a probability lower than top k sequences

Big challenge for neural NLG: Variability (diversity)

- There are many different ways to talk about the same thing (local diversity).
- Speakers generally use many different words and expressions (global diversity).

Image captions/paragraphs produced by different annotators (Corpus: Visual Genome, Example: Krause et al. 2017)

Sentences
1) A girl is eating donuts with a boy in a restaurant
2) A boy and girl sitting at a table with doughnuts.
3) Two kids sitting a coffee shop eating some frosted donuts
4) Two children sitting at a table eating donuts.
5) Two children eat doughnuts at a restaurant table.

Paragraph
Two children are sitting at a table in a restaurant. The children are one little girl and one little boy. The little girl is eating a pink frosted donut with white icing lines on top of it. The girl has blonde hair and is wearing a green jacket with a black long sleeve shirt underneath. The little boy is wearing a black zip up jacket and is holding his finger to his lip but is not eating. A metal napkin dispenser is in between them at the table. The wall next to them is white brick. Two adults are on the other side of the short white brick wall. The room has

Neural text de-generation

- Holtzmann et al. (2019): beam search generates repetitive, awkward text
- LM decoding should avoid „high probability zones in text"
- Widely discussed in NLG (not so much in MT)

Beam Search Text is Less Surprising
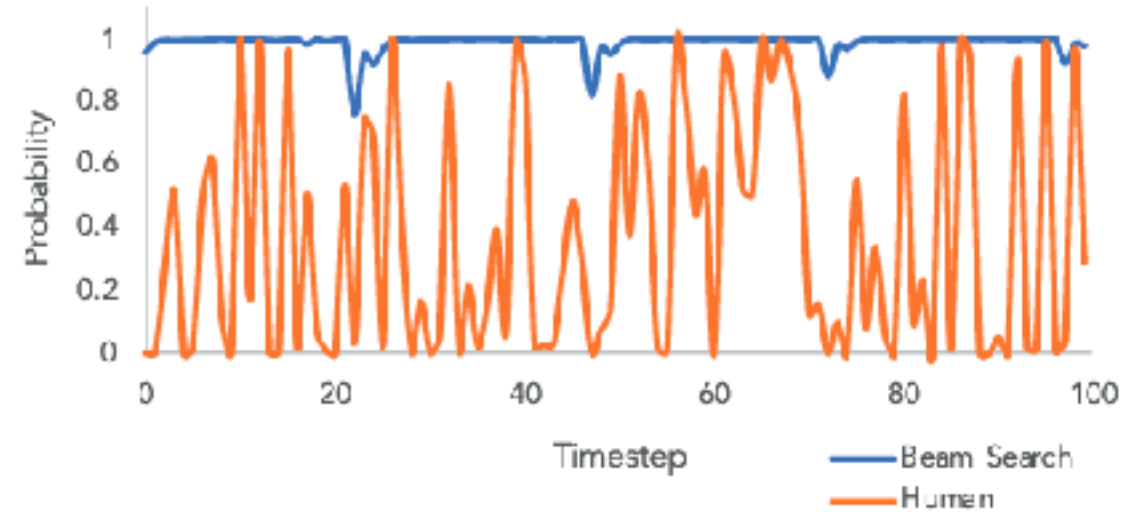
Beam Search

...to provide an overview of the current state-of-the-art in the field of computer vision and machine learning, and to provide an overview of the current state-of-the-art in the field of computer vision and machine learning, and to provide an overview of the current state-of-the-art in the field of computer vision and machine learning, and to provide an overview of the current state-of-the-art in the field of computer vision and machine learning, and...

Human

...which grant increased life span and three years warranty. The Antec HCG series consists of five models with capacities spanning from 400W to 900W. Here we should note that we have already tested the HCG-620 in a previous review and were quite satisfied With its performance. In today's review we will rigorously test the Antec HCG-520, which as its model number implies, has 520W capaci and contrary to Antec's strong beliefs in multi-rail PSUs is equipped...

Decoding for diversity

- Holtzmann (2019): Nucleus sampling

- Increase diversity by increasing randomness

- Pure sampling is risky

- Sample from the top-p portion of the distribution:

- $x \in V^{(p)}$, if $P(x \mid x_{i<t}) \geq p$.
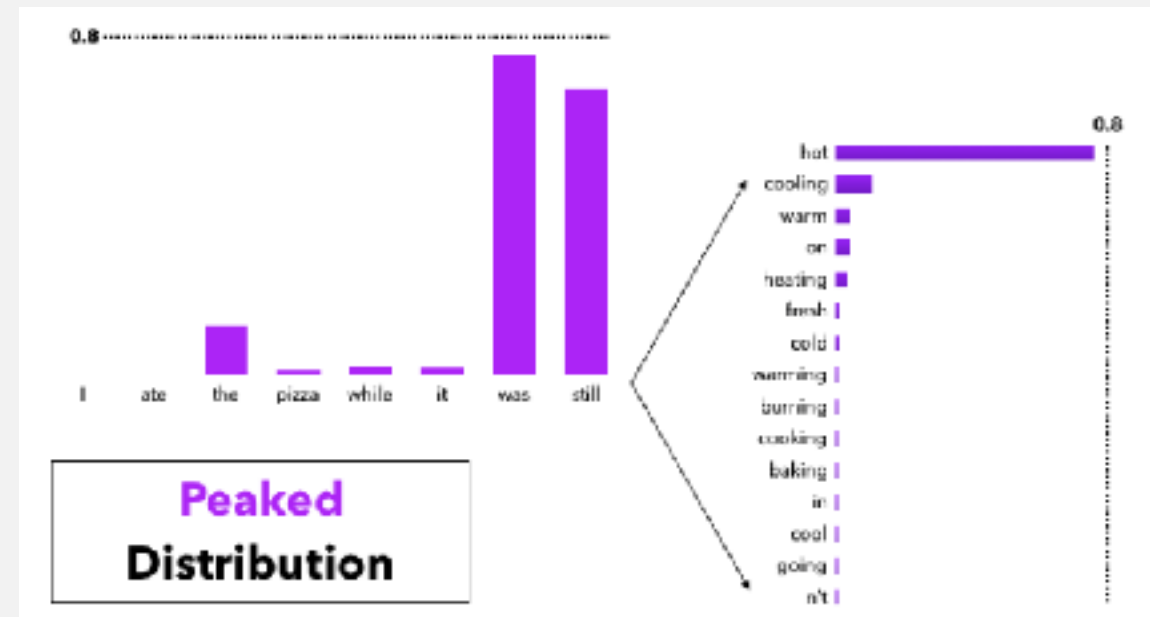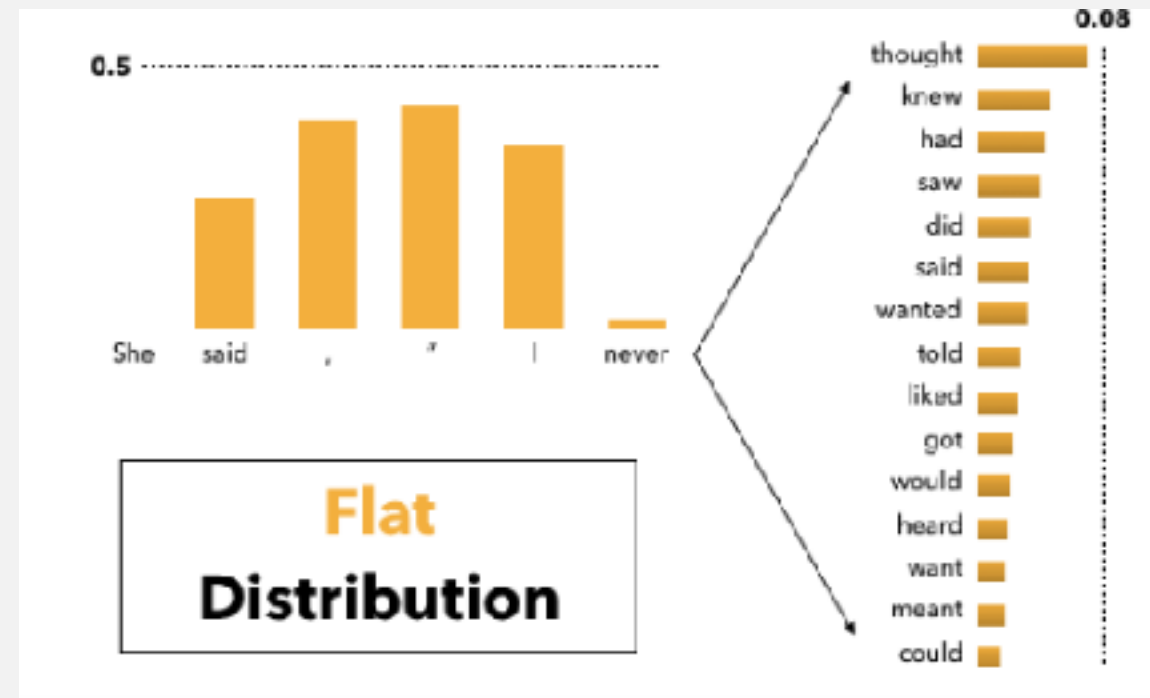
The Quality-Diversity Trade-Off

- Ippolito (2019): Evaluate diverse decoding methods for local diversity

- Diversity is negatively correlated with human-perceived quality

Beam Search
A bus is stopped at a bus stop.
A bus is parked at a bus stop.
A bus stopped at a bus stop in a city.
A bus stopped at a bus stop at a bus stop.
A bus that is parked in front of a building.

Random Sampling
A bus parked at a bus stop at a bus stop.
There is a bus that is at the station.
A man standing by a bus in a city.
A bus pulling away from the train station.
A bus stopped at a stop on the sunny day.

Another challenge for neural NLG: Effectiveness

- In spontaneous communication, speakers collaborate and aim for effectiveness

  (Grice 1975, Clark 1996, Frank & Goodmann 2012,…)

- General goal in NLG: pragmatic appropriateness

„bride"

„woman in light blue background left"

„bride"

Decoding for effectiveness

- Decoding = reasoning in context
- Rational Speech Acts Model for neural NLG, e.g.
  Andreas & Klein, 2016, Cohn-Gordon et al. 2018
- Emitter-surpressor beam search, Vedantam et al. 2017

$$\Delta(I_t, D) = \arg\max_s \sum_{\tau=1}^{T} \sum_{i=1}^{|D|} \log \frac{p(s_\tau \mid s_{1:\tau-1}, I_t)}{p(s_\tau \mid s_{1:\tau-1}, D_i)^{1-\lambda}}$$

„woman in light blue background left"

Rational Speech Acts

- **A formal model** of recursive pragmatic reasoning that can be nicely plugged in at the decoding stage of neural NLG

- Literal listener:

$$L_0(w\,|\,u) \propto \frac{S_0(u\,|\,w) * P(w)}{\sum_{w' \in W} S_0(u\,|\,w') * P(w')}$$

- Pragmatic speaker:

$$S_1(u\,|\,w) \propto \frac{L_0(w\,|\,u)^{\alpha} * P(u)}{\sum_{u' \in U} L_0(w\,|\,u')^{\alpha} * P(u')}$$

- Usually approximated by incremental, word-level or character-level reasoning (Cohn-Gordon et al. 2018)

NLG challenges in a nutshell: Handling the search space

- Variability: use a large search space, consider many potential candidate sequences

- Effectiveness: find a sequence that works well in a particular context
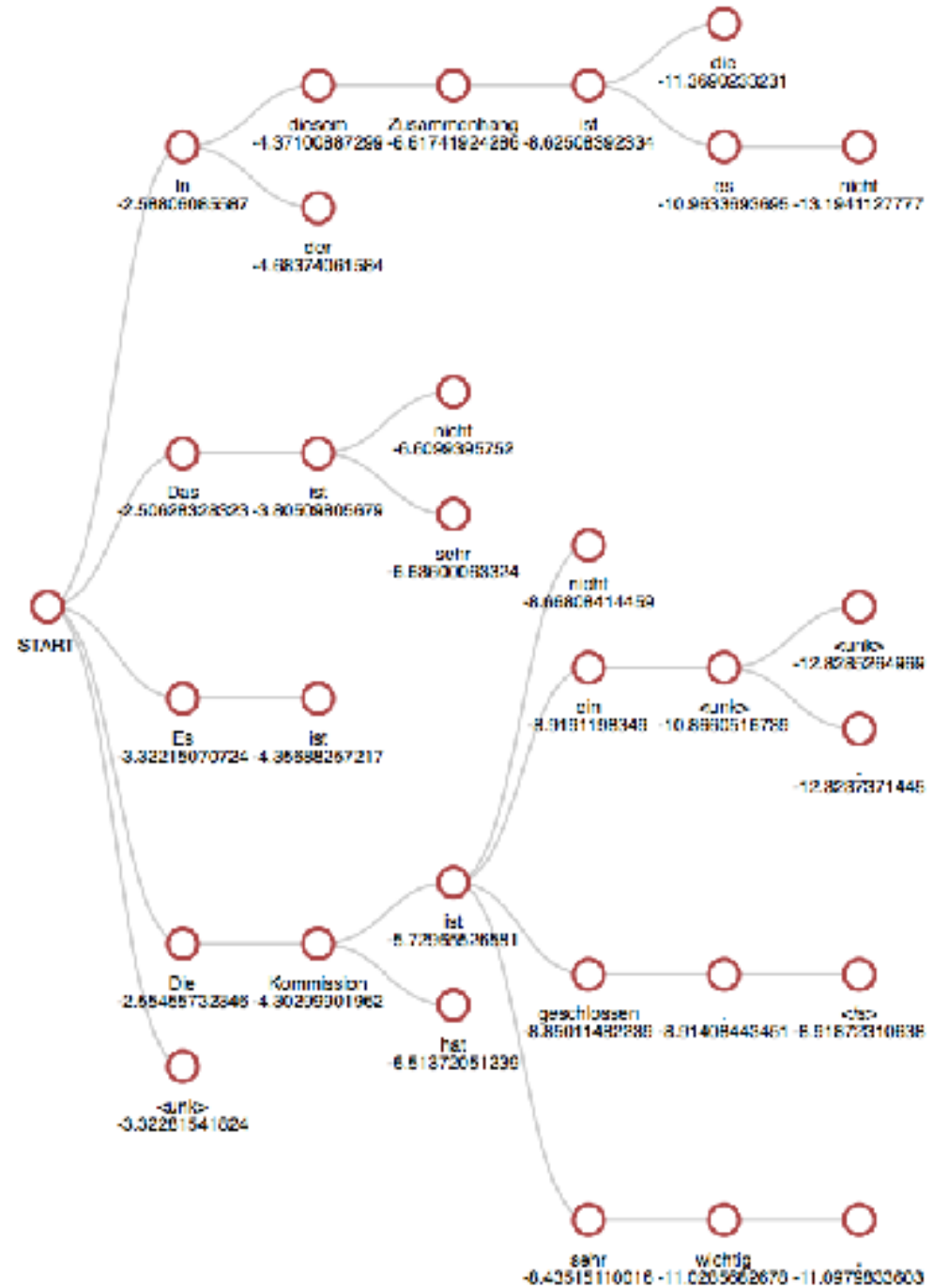
Some papers on decoding

- Attari, Nazia, et al. "Generating Coherent and Informative Descriptions for Groups of Visual Objects and Categories: A Simple Decoding Approach." INLG 2022.

- Zarrieß, Sina, et al. **"Decoding, Fast and Slow: A Case Study on Balancing Trade-Offs in Incremental, Character-level Pragmatic Reasoning**." INLG 2021.

- Schüz, Simeon, et. al. **"Diversity as a by-product: Goal-oriented language generation leads to linguistic variation."** SigDial 2021.

- Zarrieß, Sina, Henrik Voigt, and Simeon Schüz. "Decoding methods in neural language generation: a survey." Information 12.9 (2021): 355.

- …

Diversity as a By-product (SIGDial 2021)

- Compare 3 decoding objectives

  - Likelihood (Beam search)

  - Diversity (Sampling)

  - Effectiveness (RSA)

- Along 3 dimensions

  - Quality/overlap with human captions

  - Global diversity (vocabulary size)

  - Informativeness (acc. image retrieval)

Greedy	a desk with a laptop and a desktop computer
$Nucleus_{p0.7-t1.0}$	a desktop computer sitting on top of a desk
$ES - Beam_{\lambda0.5}$	a cluttered cubicle with multiple computers and monitors
$RSA - Beam_{\alpha1.0}$	an office cubicle with multiple computers and monitors

Diversity as a by-product of effectiveness

- Effectiveness: y- axis (listener identifies the correct target)

- Variability: x-axis (Type-token ratio)

- Effectiveness objective increases global diversity

- Diversity objective does not increase informativeness
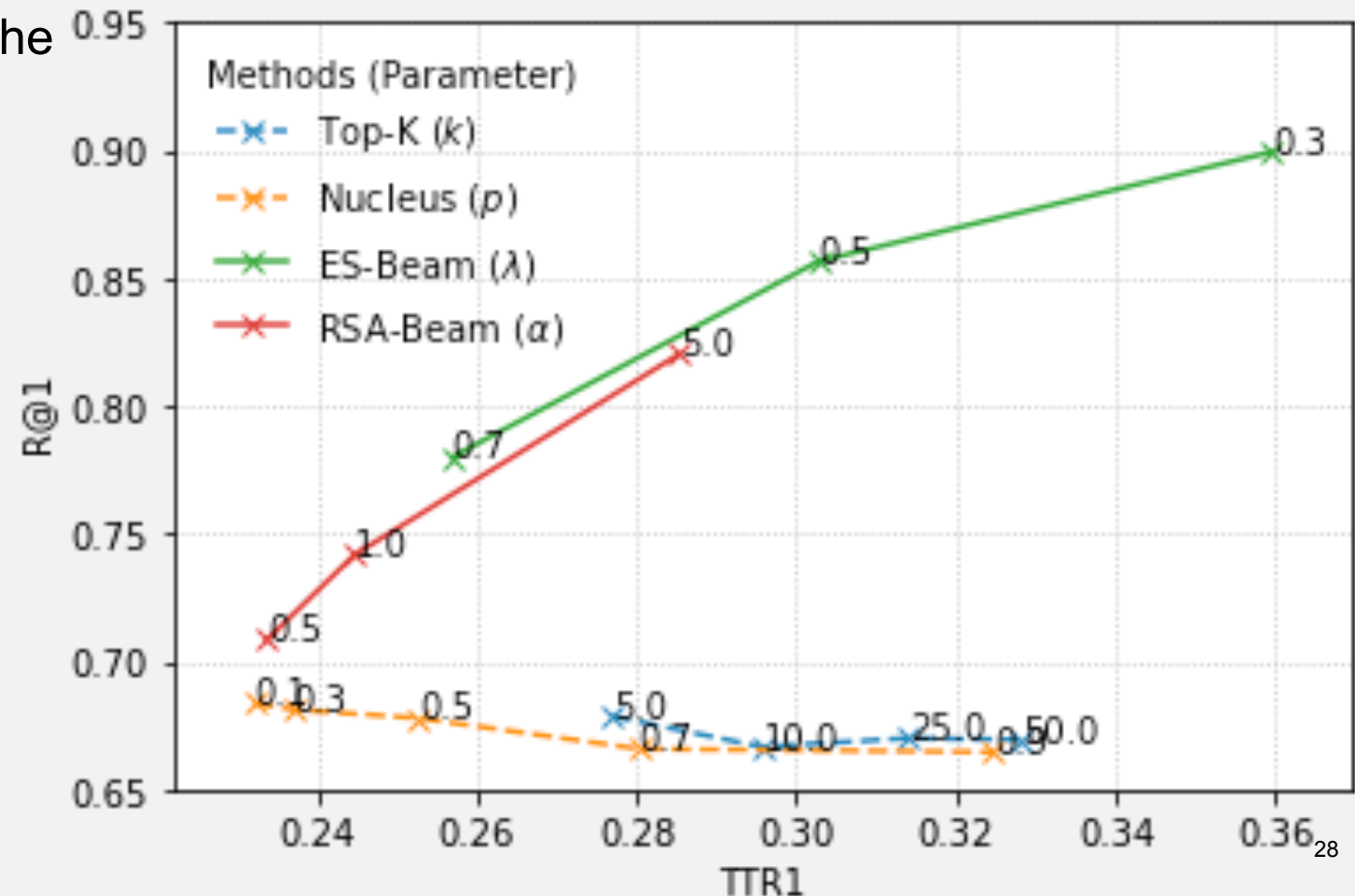
How diverse does it get?

- Frequent words generated too frequently
- Rare words generated too rarely
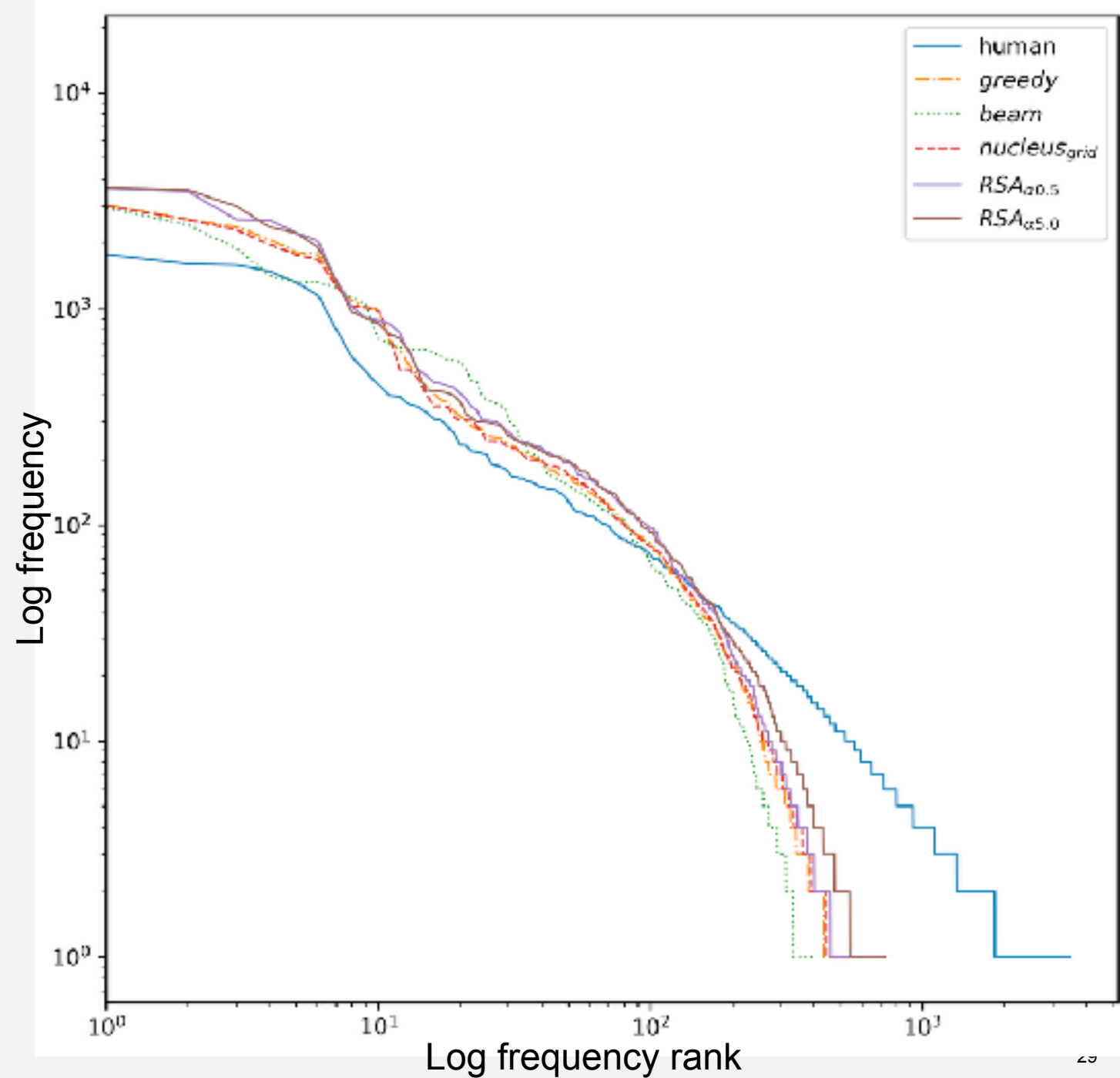
Quality-Effectiveness Trade-Off

- Pragmatic decoding decreases quality

- Degree depends on hyper-parameters

- Comparable to diversity-quality tradeoff

Method	$BLEU_4$	CIDEr	SPICE
Greedy	0.303	0.988	0.188
Beam	0.321	1.020	0.192
Top-K$_{k10-t0.7}$	0.231	0.813	0.168
Top-K$_{k10-t1.0}$	0.173	0.673	0.153
Top-K$_{k25-t0.7}$	0.222	0.785	0.164
Top-K$_{k25-t1.0}$	0.154	0.612	0.144
Nucleus$_{p0.7-t0.7}$	0.276	0.923	0.180
Nucleus$_{p0.7-t1.0}$	0.223	0.779	0.164
Nucleus$_{p0.9-t0.7}$	0.250	0.855	0.174
Nucleus$_{p0.9-t1.0}$	0.165	0.623	0.144
ES-Beam$_{\lambda0.7}$	0.290	0.919	0.179
ES-Beam$_{\lambda0.5}$	0.225	0.727	0.154
ES-Beam$_{\lambda0.3}$	0.088	0.371	0.104
RSA-Beam$_{\alpha0.5}$	0.291	0.951	0.183
RSA-Beam$_{\alpha1.0}$	0.282	0.928	0.180
RSA-Beam$_{\alpha5.0}$	0.235	0.797	0.165

Diversity and/or Effectiveness?

- Diversity may not be needed as an explicit objective (in certain tasks)
- But: there are some issues with pragmatic decoding

|	
Greedy	a desk with a laptop and a desktop computer
$Nucleus_{p0.7-t1.0}$	a desktop computer sitting on top of a desk
$ES-Beam_{\lambda 0.5}$	a cluttered cubicle with multiple computers and monitors
$RSA-Beam_{\alpha 1.0}$	an office cubicle with multiple computers and monitors

Decoding, fast and slow (INLG 2021)

- Idea: Effectiveness should not decrease quality/ fluency of an output

- **Fast** literal speaker

  - Character-level image captioning

- **Slow** pragmatic speaker (Cohn-Gordon et al. 2018)

  - RSA at every character

- **Mixed speaker (S_x):**

  - RSA, but only after a whitespace

S_0 a group of people riding on the backs of horses
S_1 two brown hornes grazing in a fenced grassy field
S_x two horses in a field in front of a field

Results

• The mixed speaker maintains **effectiveness**, according to performance of L_0:

S_0	-	54.1
S_1	1	63.1
S_1	3	68.5
S_1	5	70.4

S_x	1	61.8
S_x	3	64.8
S_x	5	66.9

• The mixed speaker maintains **quality**, according to CIDEr:

S_0	-	54.1	0.778
S_1	1	63.1	0.704
S_1	3	68.5	0.589
S_1	5	70.4	0.481

S_x	1	61.8	0.718
S_x	3	64.8	0.652
S_x	5	66.9	0.606

Another Trade-Off

- S_1 produces more out-of-vocabulary words („hornes")

- S_1 and S_x are locally more repetitive („in a field", „in front of a field")

- Repetitions seem to occur when the language model is „out of meaningful words" for the given input

- … this is when the model should try and be creative!

S_0 a group of people riding on the backs of horses
S_1 two brown hornes grazing in a fenced grassy field
S_x two horses in a field in front of a field

Input

Modeling language use: how to achieve diversity & effectiveness?

Conditional Language Model

Modeling language production: *where/when/how* does pragmatic reasoning happen?

$V*$

Objectives: do we need different, further, multiple, … objectives?

Decoding

Language models & decoding: is this a good architecture?

Output Sequence

Linguistic creativity and computational modeling

(KLS = knowledge of the language system; GPLC = general principles of language and communication; GPP = general processing principles)

- To model creativity, we need to gain a better unterstanding of the ways in which speakers handle and extend their ``linguistic search space", considering **effectiveness**

Outline

- Intro: linguistic creativity

- Modeling linguistic creativity

- Creativity from an NLG perspective

  - Experiments on decoding

- Outlook

  - Further tasks

Linguistic knowledge & individual cognitive diff.

- **Verbal fluency task**: Participants enumerate as many words as possible for a given category in 60sec

- Number/type of responses indicate cognitive performance

- Further metrics: clustering and switching between sub-categories

Animals:

Dog Cat Mouse Horse Donkey Cow Rat Tiger Elephant Rhino Crocodile Marten Giraffe Whale Seal Fish Eel Seal Moray Lion Leopard Cheetah Snake Spider Ant Beetle Bee Bear Anteater Koala

Eel Cat Peacock Horse Tiger Dog Aardvark Hippopotamus

Analyzing Verbal Fluency data with embeddings

- Bottleneck for experimental studies: manual analysis of the sub-categories with predefined sub-category inventories (potentially subjective, time-consuming, etc.)

- Some papers explore word embeddings for scoring responses: distances, path length, etc. in vector space

- Our study (**COLING 2022)** shows that ConceptNet achieves highest correlations with human annotations of clusters and switches in a dataset of German verbal fluency responses for 11 different categories

- We correlate human annotations of **switches between clusters** with **distance in embedding space**
- Correlations are very different across different categories
- BERT does not work at all

Table 2: Pearson Correlation Analysis Results on Total Switch Count and Mean Similarity Scores

Categories	GloVe	fastText	ConceptNet	BERT
animals	−.17, n.s.	**−.25, p.<.05**	**−.24, p.<.05**	−.07, n.s.
body parts	−.19, n.s.	.09, n.s.	**.23, p.<.05**	.30, p.<.01
clothes	.03, n.s.	−.147, n.s.	−.14, n.s.	.04, n.s.
countries	**−.43, p.<.01**	**−.39, p.<.01**	**−.40, p.<.01**	.02, n.s.
fabrics	−.11, n.s.	.01, n.s.	−.19, n.s.	.12, n.s.
groceries	**−.34, p.<.01**	**−.24, p. <.05**	**−.27, p.<.01**	−.08, n.s.
hobbies	−.144, n.s.	.03, n.s.	−.17, n.s.	.058, n.s.
insects	−.19, n.s.	**−.38, p.<.01**	**−.27, p.<.01**	−.13, n.s.
occupations	−.00, n.s.	−.16, n.s.	−.08, n.s.	.058, n.s.
vessels	.17, n.s.	−.01, n.s.	−.03, n.s.	.06, n.s.

https://aclanthology.org/2022.coling-1.16.pdf

Probing language models for verbal fluency

• Controlled sequence production task that tests

• … conceptual-lexical knowledge

• … cognitive processes involved in lexical retrieval („Say as many appropriate words as possible")

• … memory and executive control („Avoid words already produced/not appropriate for the task")

• … see: Shao, Zeshu, et al. "What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults."
Frontiers in psychology 5 (2014): 772.

• This seems related to challenges in neural language generation:

  • … diversity, repetitiveness

  • … hallucination, ``text degeneration''
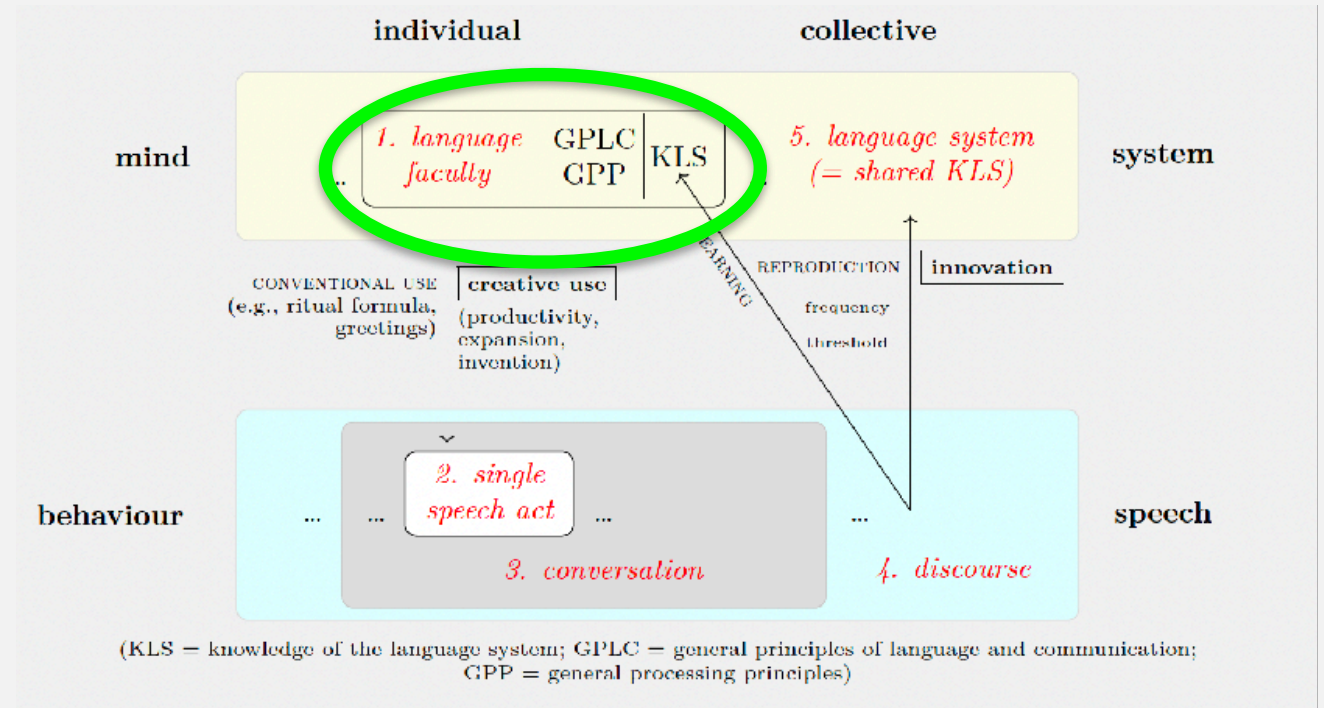
  • … faithfulness to the input, knowledge

Prompting GPT-2 for verbal fluency…

```
PROMPT:
Tiere:
1. Hund
2.
OUTPUT:
Tiere:
1. Hund
2. Katze
3. Waschbär
4. Dromedar
5. Hase
6. Kaninchen
7. Dachs
8. Schneeleopard
9. Hase
11. Fuchs
```

```
PROMPT:
Tiere:
1.
OUTPUT:
Tiere:
1. Ḥurūl
2. Ḥorūd
3. Ḥirumur
4. Ḥeru
5. Ḥayn (Sara,
```

Linguistic creativity and individual differences

- How and where do we account for individual aspects of the language faculty in a computational model of language?

(KLS = knowledge of the language system; GPLC = general principles of language and communication; GPP = general processing principles)

43

Outlook: creativity in dialogue

• Previous studies on interactive reference games focussed on alignment

• Our hypothesis: partners in dialogue can co-create new strategies, rules and conventions on the fly, e.g., when something is difficult to refer to
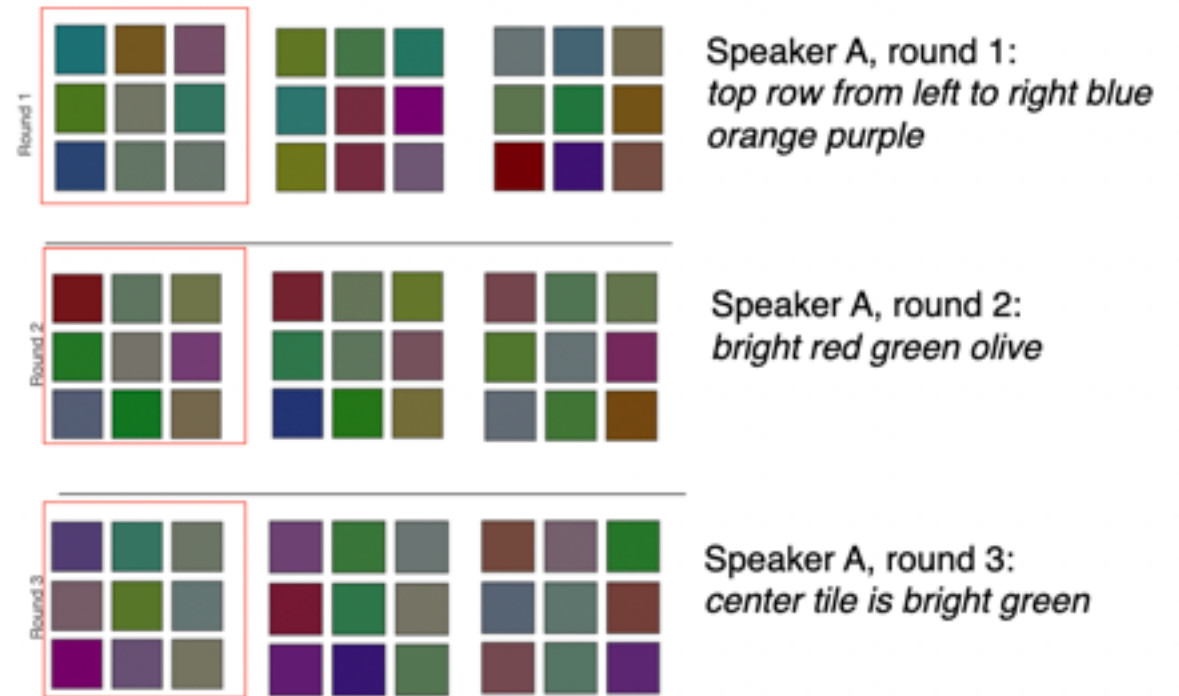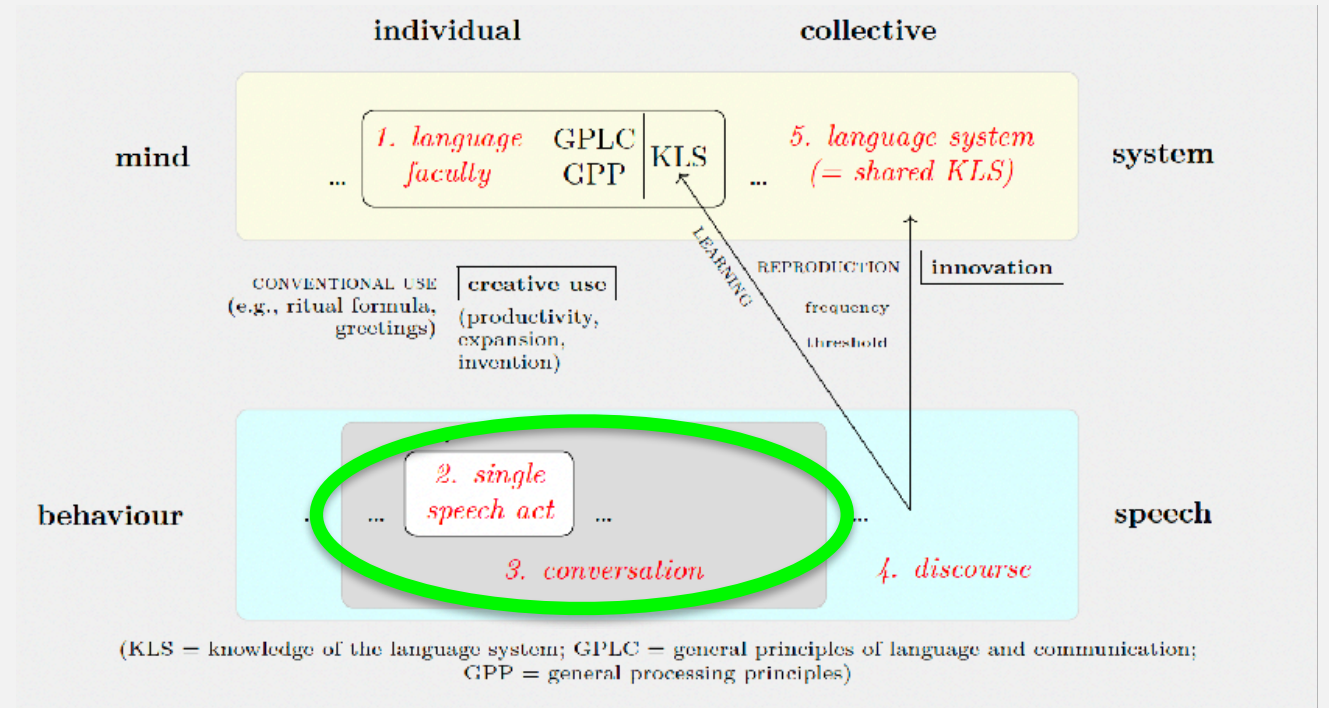
Speaker A, round 1:
top row from left to right blue orange purple

Speaker A, round 2:
bright red green olive

Speaker A, round 3:
center tile is bright green

Figure 1: Examples of referring expressions for color grids (from McDowell and Goodman, 2019)

44

Linguistic co-creativity

- Why and how do speakers get more creative when they work together?

(KLS = knowledge of the language system; GPLC = general principles of language and communication; GPP = general processing principles)

Summary

- Speakers can be **original and effective** in their everyday language use
- We see linguistic creativity rooted in the **individual language faculty** of the speaker, who has knowledge of the language system but can deviate from the conventions of the system
- Current computational models of language generation have a poor account of effectiveness and originality, and general **communicative/cognitive/social principles** are not well understood/ integrated
- Linguistic creativity may be a very fruitful direction for dialogue models, going **beyond alignment and adaptation**