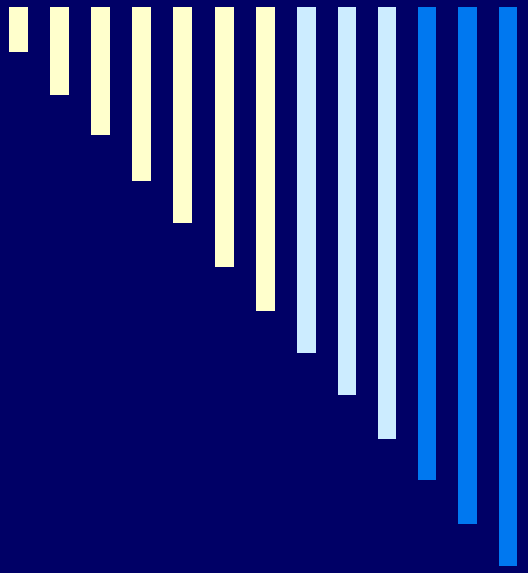
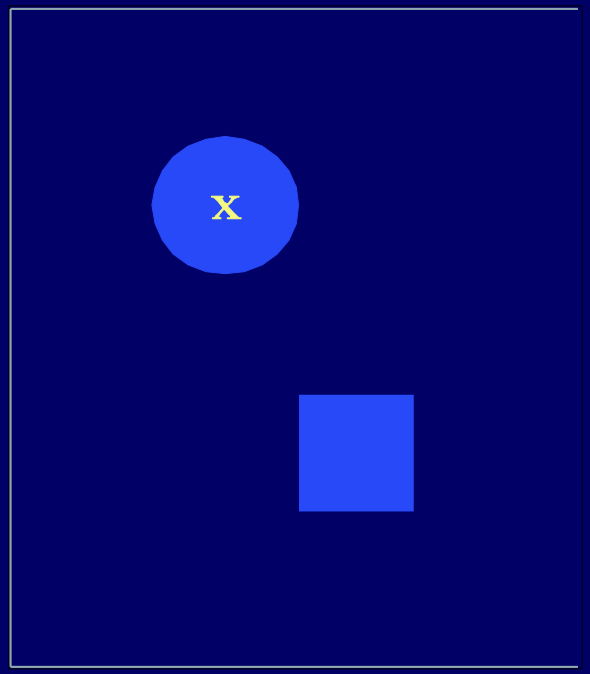


Cognitive Discourse Analysis:
Analysing language data to learn about thoughts and concepts

Thora Tenbrink
Reader in Cognitive Linguistics
Bangor University (Wales, UK)

April 10, 2019

CODA – Analysis of visual scenes

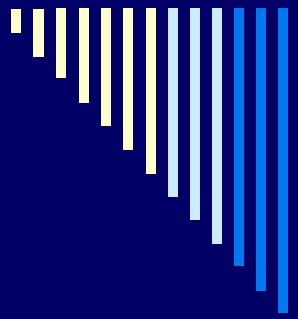


Where is the marked object?

Which one is the marked object?

- "somewhat to the left diagonally above the square"
- "the circle"

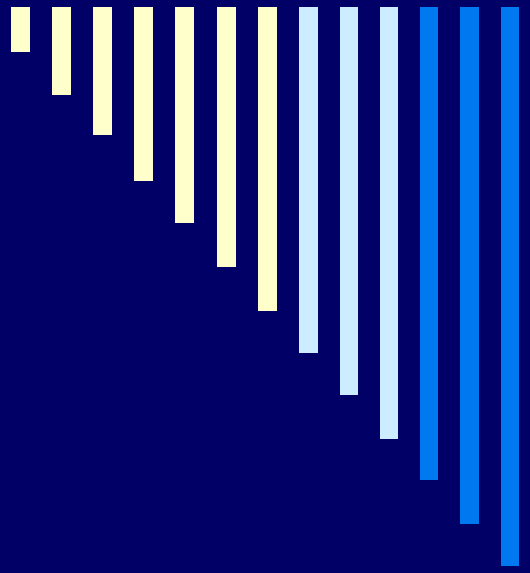
What do people say with respect to the marked object?



Overview

- CODA: Motivation and scope
- Data collection and content analysis
- What to analyse?
- How to analyse: Systematic annotation
- Triangulation and extensions

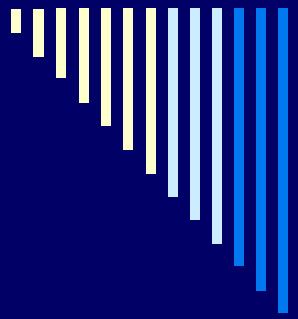
Tenbrink, Thora (in press). Cognitive Discourse Analysis: An introduction. Cambridge University Press.



CODA:
Motivation and scope

Scope of verbal data analysis

- Language as a representation of concepts
  - verbalisation of perceived or remembered scenes or events
- Language as a representation of thought processes
  - verbalisation of complex tasks such as problem solving or decision making



Your projects / interests?

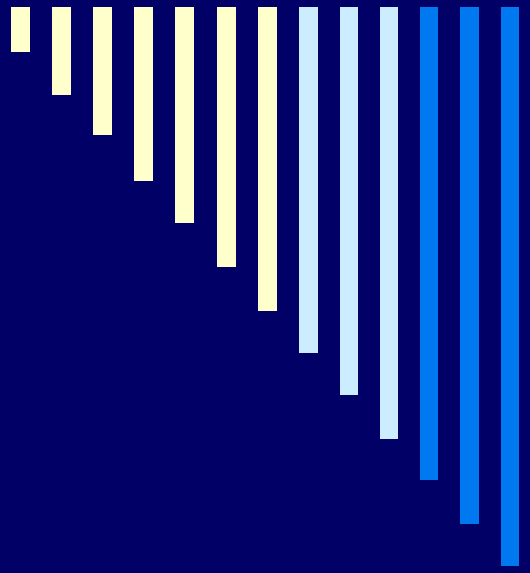
- What kinds of language data have you been dealing with, or could imagine collecting?
- What are possible goals of such data collection and analysis?

Verbal protocol analysis (e.g., for problem solving)

- Content level of analysis: **What** is said?
  - Which cognitive processes and strategies are mentioned explicitly?
  - Are there explanations or reasons for decisions?
  - Do people differ in how they describe their problem solving?

Verbal data analysis (CODA)

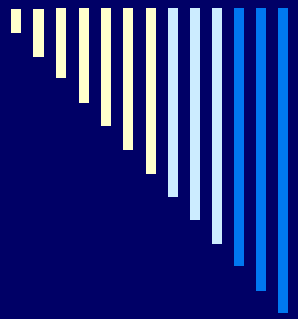
- Enhancing the validity of verbal reports as data
- Beyond content – Linguistic level of analysis:
  How are thoughts expressed?
  - What kinds of linguistic features can be detected in the data?
  - How do they relate to particular strategies and cognitive processes?
- Validation of content-based intuitions
  - Identification of linguistic cues that signal relevant distinctions
  - Operationalisation and quantification across participants



Data collection and content analysis

Some general issues

- You can analyse any kind of text in relation to cognitive aspects
- But not every text will serve your purpose equally well
- If you look at just one text (or different texts produced in different situations), it will be very difficult to draw systematic conclusions
- Empirical studies typically involve at least 15 participants (in each condition) confronted with a carefully designed task



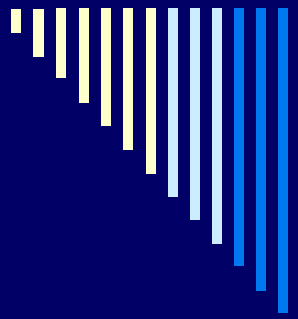
Types of verbal data

- Description
- Interview
- Retrospection
- Introspection
- Direct questions
- Dialogues
- Think Aloud

Description

- Ask participants to *describe…*
  - … what they see with respect to…
  - … how they remember …
  - … how they would deal with…

  - a picture
  - an event
  - a scene
  - a path to a destination …



Description

- Goal: To access verbalisations of internal representations
  - Relative to a currently observed situation
  - Relative to memory
  - Relative to a given context / task
  - …

- Description works well for mental representations
- But what if the thought processes are more complex?

Types of verbal data

- Description
- Interview
- Retrospection
- Introspection
- Direct questions
- Dialogues
- Think Aloud

- What are the main features/ advantages / disadvantages in each of these?

- You can analyse any kind of text in relation to cognitive aspects
- But not every text will serve your purpose equally well
- More ‘natural’ talking often leads to less control & a more challenging analysis process

Summary: Data elicitation methods

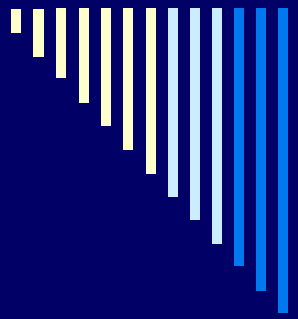
- Distinct communicative purposes …
  - think aloud protocols, written or spoken retrospective reports, instruction, dialogue, …
- … with distinct linguistic structures …
  - more or less explicit, complete, coherent
  - accommodation to addressee / interaction partner
- ... highlight cognitive aspects from different angles

Preparing the data for analysis

- Transcription of spoken language: keep your purpose in mind!
  - no perfectionism – transcribing all details takes an infinite amount of time!
  - handling transcription tools can be equally time-consuming
  - recommended: "f4" simple tool for little money
    - http://www.audiotranscription.de/
- Excel is often sufficient for annotation

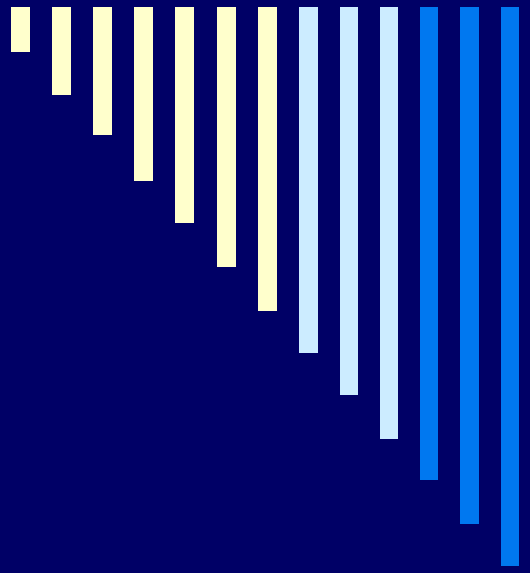
Making sense of content

- Understanding the content is essential
- First step towards a deep & systematic analysis: Develop good intuitions
- Identify useful verbalised content – sometimes people actually answer your research question ‘directly’: extract quotes that can be cited in your report



Segmentation: Possible units

- syntactically complete sentences
  - or parts of sentences?
- semantic completeness – "possible sentence"
- intonation?
- pragmatic unit
  - some subtask is described
- discourse unit
  - all verbalizations pertaining to one task



What to analyse?

Areas where *how things are said* reflects important cognitive aspects

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- What people say reflects their conscious attention
  - Compare:
    Three big black scary cats chase a mouse

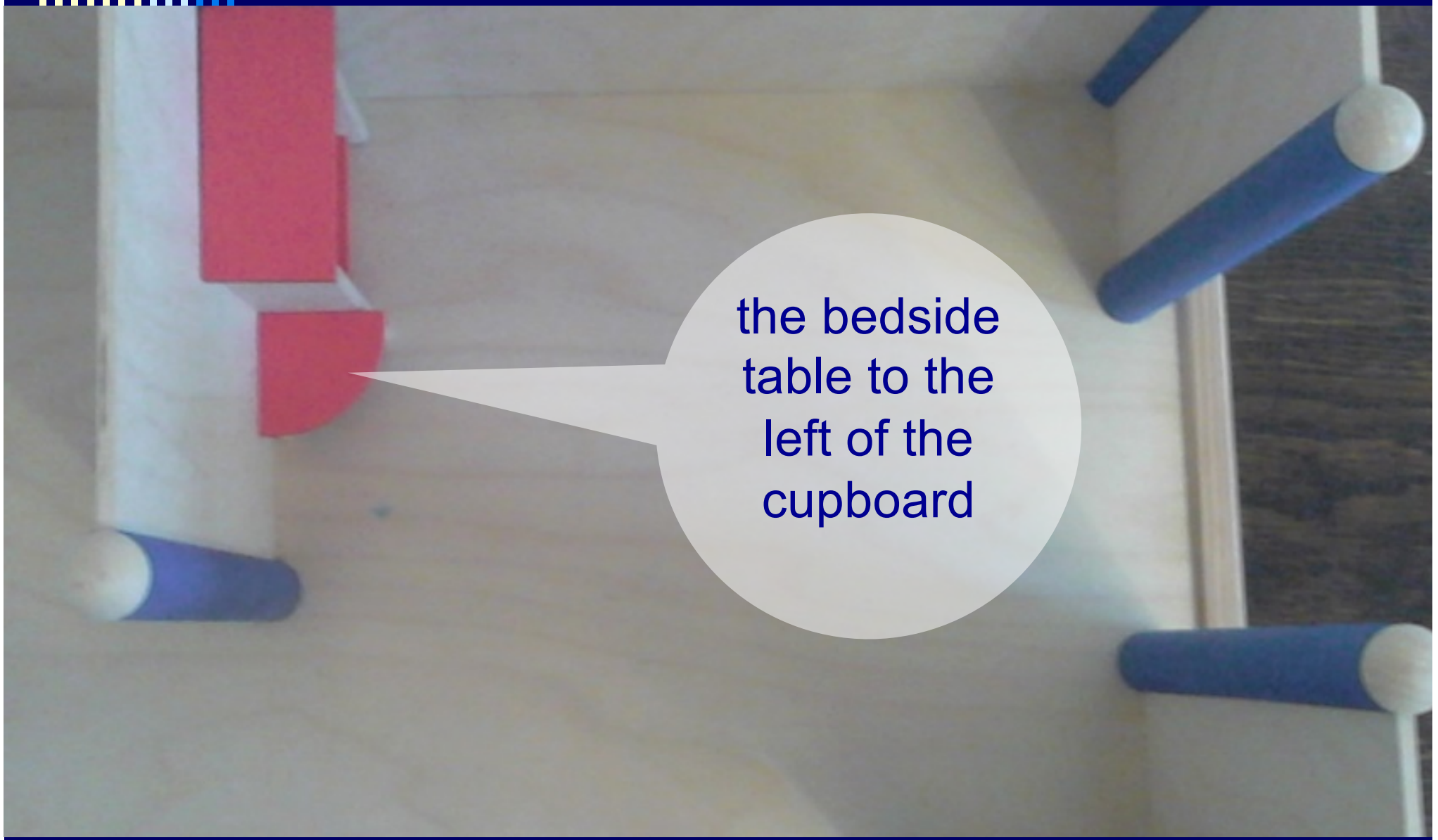
    vs. *There are some cats*
- Quantitative patterns of mention across the linguistic data can lead to insights about the kinds of things attended to at different stages

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- Speaker's perspective can often be derived
  - Compare:
    Three hungry cats chase a mouse

    vs. *A mouse is frightened by three cats*
- Spatial perspectives
- Quantitative patterns of underlying perspectives highlight changes according to situation


the bedside table to the left of the cupboard

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- More or less detail in verbal representations
  - Compare:
    Three cats chase a mouse for 3.5 minutes

  vs. *Some animals at play*
- The number of words used to describe a situation is relevant - but wordiness isn't all
- Granularity can often be related to situational features

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- Speaker's level of certainty is often signaled linguistically
  - Compare: *I think there may have been one or two cats, or a dog*

  vs. *There were three cats and a mouse*
- Certainty can be related to expertise, or to features of the situation where things may be genuinely unclear even for experts

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- Analysis can show what speakers take for granted, or infer from the context, or leave for a partner to infer
  - *This time the mouse doesn't try to escape*
- Presuppositions
- This analysis requires 'reading between the lines' – what's not there, what's the underlying assumption?

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- Verbal protocols reflect transformation processes, e.g. while assembling something
- Objects are assigned functions – a mental transformation
  - *This piece of wood could be a guidepost.*
- Transformational creativity is often verbalized alongside (or before) action

Some key areas

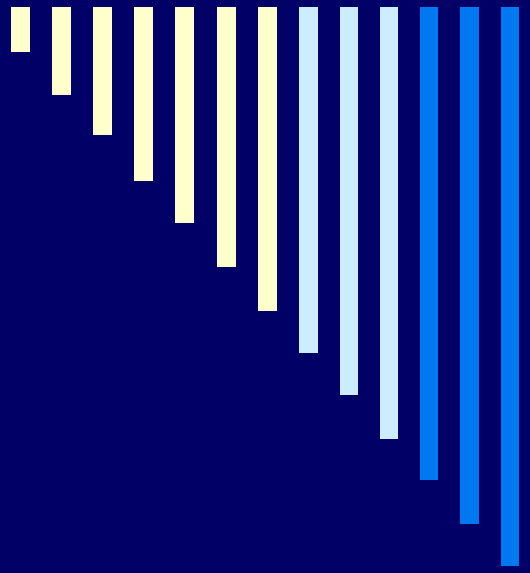
- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

- Problem solvers verbalise their strategies over time
- Think-aloud protocols & retrospective reports
- Different strategies, or stages within strategies, come with different linguistic patterns
- Some strategies are implicit and only recognizable by specific associated linguistic patterns

Some key areas

- Attention
- Perspective
- Granularity
- Certainty
- Inference
- Transformation
- Cognitive Strategies
- Communication

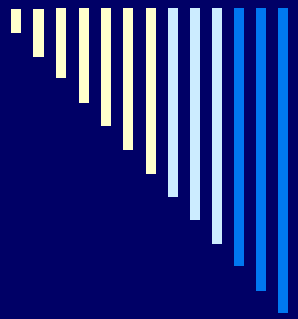
- In dialogue, speakers adjust their verbalisations for the interaction partner
- Underlying assumptions about their knowledge & abilities
- Adaptation, alignment - on the linguistic surface as well as underlying concepts



Step-by-step systematic annotation

Think before annotating

- Keep your research question in mind!
- Develop informed intuitions about your language data
  - What do I think happens in the data?
  - What makes me think this?
- How can I capture this more precisely, prove linguistically? Linguistic analysis:
  - What exactly is it in the data that leads to my intuitions?
    - e.g., "I feel that this person likes Spaghetti, based on how she talks about it!" – what does she say exactly?
  - (How) Is this operationalizable?



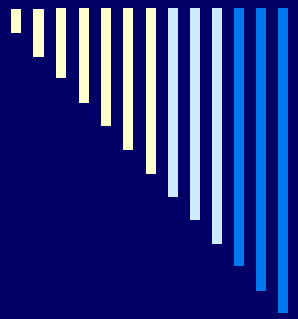
Type-token analysis

- Type-token analysis:
  - Identify *types* of features
  - What do these features express / signify?
  - How frequently does this happen? (*tokens*)
- How does this develop throughout the text?
- What general picture emerges from this analysis?
- If parts of this analysis can be done automatically, the analysis can be pursued over many data sets.

… How to get there?

Data annotation

- Do the annotation in digital format (or transfer systematically if you prefer to start with paper)
- Analysis: an iterative procedure
  - as your analysis procedures become clearer and more refined, you might also reconsider your units
- Consider your analysis categories carefully in relation to your research question
  - you can't analyze all features of your data (too much, often not informative)
- Define your analysis criteria extremely carefully
  - look for a kind of definition that is as unambiguous and general as possible
  - choose a useful level of detail and make this clear in your definition



Annotation definitions

- A definition should contain:
  - a clear description of the feature to be analyzed
  - if possible, a list of options; if not, a list of typical examples (that are not misleading by being exceptional or special in any irrelevant way)
  - an indication whether the options are mutually exclusive
  - typically you need the possibility of choosing "other" (or you'll end up with unique annotations that are not interesting)
  - it may be useful to extract subparts of the original unit for some annotation

Example (step one)

Description unit	Reference to direction
Follow Eastern	none
Right on Grandview	right
Fairly quick left o	
Right on Robbins	
Right on MassAv	
Right on Highlan	
Right on Gray	
Left on Mt Vernor	
Left on Longfello	

References to direction at the start decision point within the description unit: First direction description used, (typically left / right but possibly also West / back / straight / up / down and the like) together with possible modifications such as "veer / sharp / half / a bit". Copy this here (first mention basis – any further descriptions that are syntactically separate get listed in column S). If none is found, code "none".

Example (step two)

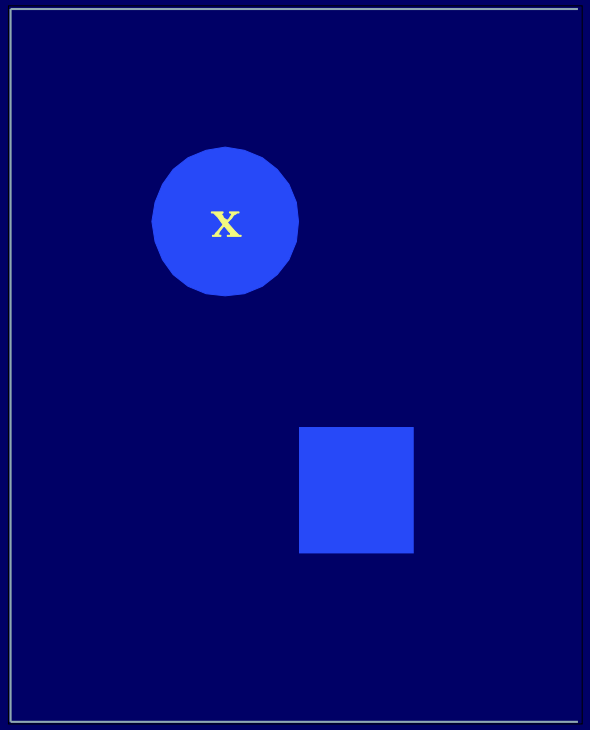
Extract of parts of the original unit

Annotation of these parts of the original unit

- *Coll* on point with scription use *ck / straight* and ch as "veer / sharp / half / a bit". Copy this here (first mention basis – any further descriptions that are syntactically separate get listed in column S). If none is found, code "none".
- s the
- (deviating just a bit from the straight on, as in "veer somewhat to the left")? Code "weak"
- Does it emphasize its standardness (as in "exactly left")? Code "exact"
- Does it make the direction sharper ("a sharp left", which in effect takes the runner back)? Code "sharp"
- Does it combine several different expressions syntactically, as in *northwest*? Code "combined"

Language segment Description of segment (starting from decision point)	DP: characterization of intersection	DP: no. of branches	DP: Reference to direction	DP: Perspective underlying direction term	DP: Specification of direction
Follow Eastern	none	n.a.	none	n.a.	none
Right on Grandview	none	n.a.	right	route	none
Fairly quick left on Hawthorn	none	n.a.	left	route	none
Right on Robbins	none	n.a.	right	route	none
Right on MassAve- stay for a while	none	n.a.	right	route	none
Right on Highland	none	n.a.	right	route	none
Right on Gray	none	n.a.	right	route	none
Left on Mt Vernon	none	n.a.	left	route	none
Left on Longfellow	none	n.a.	left	route	none
Follow the road's sharp curve around back to Eastern	none	n.a.	sharp curve around	other	sharp

What does it mean to be ‘more complex’? How to operationalise this?



Where is the marked object?

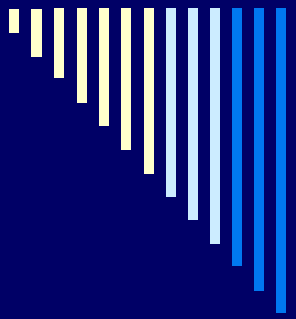
- "somewhat to the left diagonally above the square“

Which one is the marked object?

- "the circle"

Excel annotation of linguistic features

ID	condition	picture nam	Utterance	no. direction	precisifiers	relatum	
240	1	2vreloben.j	oben	1	1	0	
265	1	4reihe.jpg	on the far left	1	2	0	
266	1	2vrelschrae	behind the box to the left	2	1	1	
267	1	3gschraeg.j	at the bottom	1	1	0	
268	1	2ghalbschra	top third of the outlined box	1	1	3	
269	1	3vreloben.j	above the blue square	1	1	1	
270	1	2vschraeg.j	to the left of the square	1	1	1	



Consistency and reliability

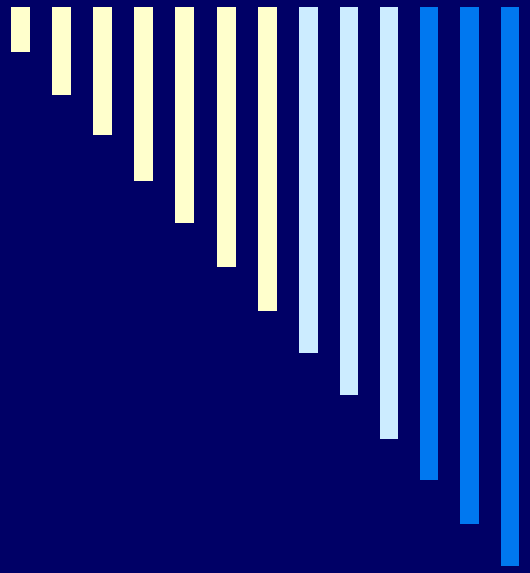
You have a set of unitized and annotated natural language data, for example in an Excel file: One unit per row; one annotation per column. What now?

Before you actually do anything with your data:

- Test your annotations for consistency
  - the same annotations should be found for the same (or equivalent) units:
  - so look at all units that are annotated in the same way
  - in Excel you can use the filter for this purpose
- Check for inter-coder reliability
  - different people should come to the same conclusions (annotations)
  - to start with you can simply check the percentage overlap
  - for publication you need more sophisticated calculations (Krippendorff's Alpha)

Relative frequencies

- Count the number of occurrences of each annotated feature
  - in Excel, this can be done easily by using the filter
- Assess the relative frequency – depending on some relevant measure
  - e.g., no. of words, or no. of units
  - if you want to compare diverse (parts of) texts, and have counted the features in each of them, look for % of a feature relative to
    - no. of words in each (part of the) text
    - no. of units in each (part of the) text
- It is useful to do this in a separate file (or tab)
  - one line for each (part of a) text
  - columns: no. of units / no. of words / no. of features…
  - then: percentages relative to units, or words
  - Excel can do computations automatically if you enter a formula



Triangulation and extensions

Theory based

- The interpretation of language data relates to previous findings / theory motivating the study
- The significance of certain linguistic features is well documented from theory
  - Insights from Cognitive Linguistics show what a particular expression ‘means’ in a cognitive sense

Limits

- Language data can be very complex
  - Analysis may take (too) long
  - The researcher may miss something important
- People may not want to say everything
- People may make things up (‘confabulate’)
- People may misremember
- Not all thought processes get to be reflected in language

Some types of data for triangulation with language

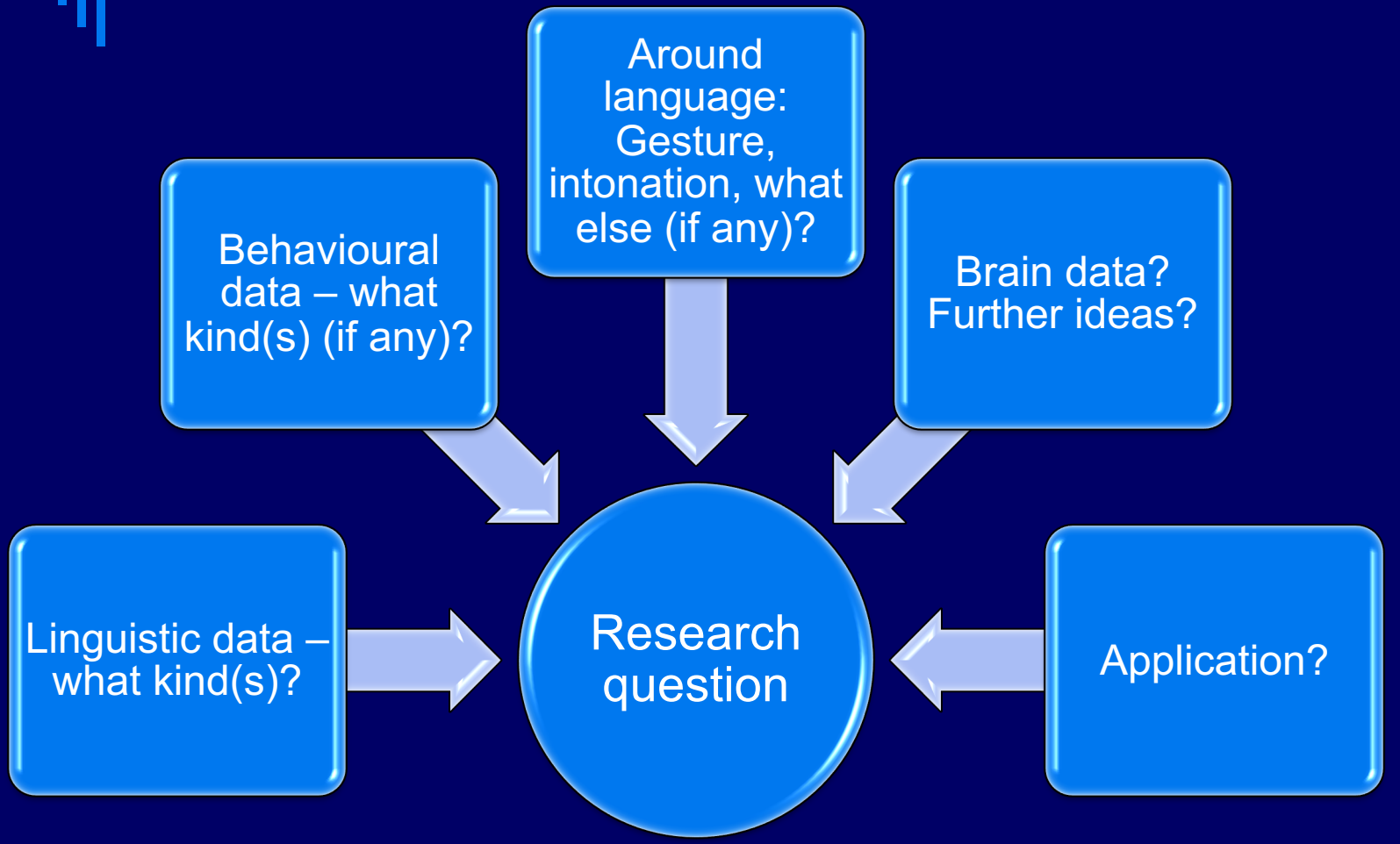
- Around language
  - Gesture, mimics, intonation, pauses, accent, errors, …
- Behaviour
  - Choices (according to task)
  - Action steps (e.g., in a problem solving process)
- Performance
  - Success in a task (according to a criterion)
  - Reaction times
- Memory
- Brain data
  - fMRI, EEG, ERP, …


If I'd known they wanted me to use all this info– I would never have asked for it!
Julie

Extensions

- Language analysis data can be published ‘as such’
- With triangulation a broader audience will be interested
- Findings can feed into theory
  - Cognitive modelling (Gugerty & Rodes, 2007)
  - More general theories about the human mind
- Findings can inform practical applications
  - Human-centred assistance systems
  - Dialogue systems / human-computer (or human-robot) interaction

Data collection



CODA: Analysis of verbal data in Cognitive Science

- Uses methods of DA to learn about
  - how we think
  - how our thoughts are reflected in language
- Linguistic point of view:
  - Gaining insights about the relation between language use and cognition
- Cognitive science point of view:
  - Gaining insights about human cognition by (supporting) linguistic evidence

More here…

http://knirb.net

- Tenbrink, Thora. 2015. Cognitive Discourse Analysis: Accessing cognitive representations and processes through language data. *Language and Cognition 7*:1, 98 – 137.

- Tenbrink, Thora (in press). *Cognitive Discourse Analysis: An introduction.* Cambridge University Press.