Interactive Learning and Cross-Modal Binding – A Combined Approach

Henrik Jacobsson[1], Nick Hawes[2], Danijel Skočaj[3], Geert-Jan Kruijff[1]

[1] Language Technology Lab, DFKI GmbH, Germany
[2] School of Computer Science, University of Birmingham, UK
[3] Faculty of CIS, University of Ljubljana, Slovenia

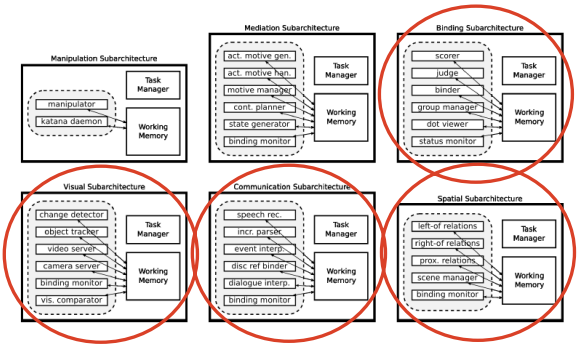
Symposium on Language and Robots 2007
Aveiro, Portugal, 12 December 2007

Overview

- Introduction
  - Binding
  - Learning
- Interplay between binding and learning
  - Explicit and implicit learning
  - Co-learning
  - Negation / Unlearning
- Different modes of learning
- Our learning method
- Experimental results
- Integrated system
  - Demo
- Conclusions and work in progress

CoSy Architecture Schema

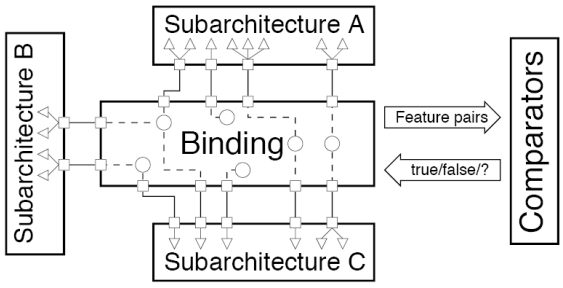
- CoSy Architecture Schema Toolkit (CAST)



- cf. Henrik Jacobsson et. al., Crossmodal Content Binding in Information-Processing Architectures

Binder

- cf. Henrik Jacobsson et. al., Crossmodal Content Binding in Information-Processing Architectures

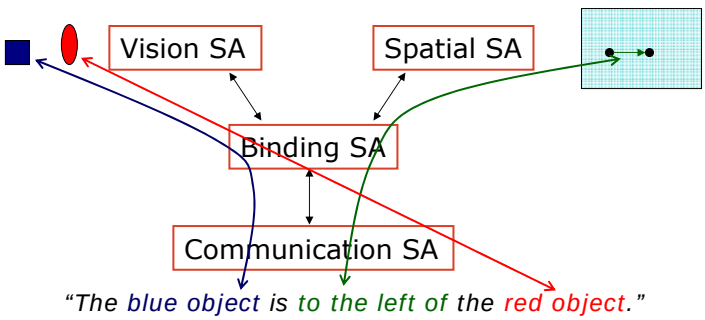


Learning

- All of the knowledge cannot be input to the system by hand => **Learning**
- The information has to be gradually acquired, processed and structured => **Continuous learning**
  - open-ended, life-long, incremental, on-line learning
  - gradually enlarging the ontology
  - No separation on the training and recognition stage
    - the training and recognition are performed in the loop
- The information is provided by a tutor or other modalities => **Interactive cross-modal learning**
  - Communication with the tutor, dialogue
  - Developmental learning, dynamical scaffolding, graded curriculum
  - Co-learning, implicit learning, unlearning
  - Active learning

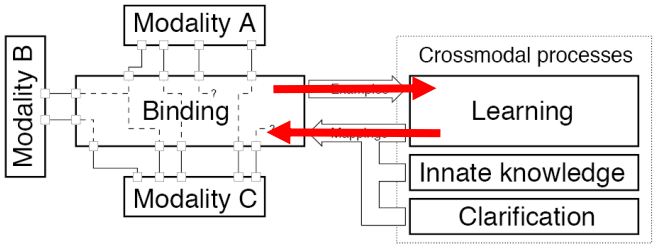
Learning cross-modal associations

- Learning of basic qualitative visual concepts
  - Learning of visual attributes
  - Learning of spatial relationships
- Symbol grounding
- Finding associations between words describing the visual and spatial concepts and automatically extracted visual and spatial features.



Cross-modal binding and learning

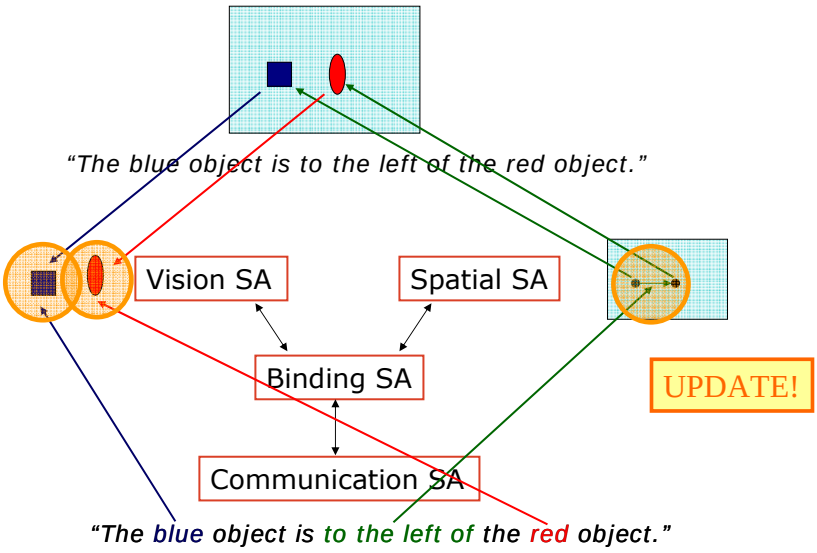
- Interplay between cross-modal binding and interactive learning
  - Learner improves mappings used by binder
  - Binder generates training examples for learner
  - In an incremental and interactive way



Explicit / implicit learning

- Three aspects of intention of a communicative act:
  - Assertions
    - "This is a blue thing."
    - Tutor ascribes new information to the referent.
    - Additional information is a communicative goal (primary purpose) -> **explicit learning**.
    - Use only salience for binding. Update.
  - Command
    - "Put the blue ball to the left of the red cube."
    - Tutor is referring to objects.
    - Additional information is given indirectly as a side effect -> **implicit learning**.
    - Use binder to resolve the referent. Update if reliable.
  - Questions
    - "Is this object blue?"
    - Tutor asks for information.
    - No additional information is given.
    - Use salience for binding. Recognize. Reply. Update if reliable.

Co-learning



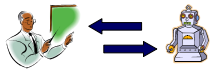
Negation / unlearning

- H: "This is not blue."
- Update the current representations with negative information (negative examples).
- Especially important in an incremental setting and when the system tries to learn autonomously (without tutor supervision)
  - Error propagation
  - Requires error recovery
- Unlearn with false positives

Different modes of learning

Tutor-driven approach

Tutor-supervised approach

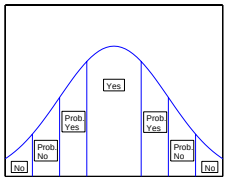


Exploratory approach

- **Tutor-driven** — **TD** – Correct interpretation always given by tutor.
- **Tutor-supervised (conservative)** — **TSc** – Asks tutor for correct interpretation when not completely sure.
- **Tutor-supervised (liberal)** — **TSl** – Asks only when its recognition is very unreliable.
- **Exploratory (conservative)** — **EXc** – Updates when it thinks it's definitely correct.
- **Exploratory (liberal)** — **EXl** – Updates also when it thinks it's probably correct.

Updating rules

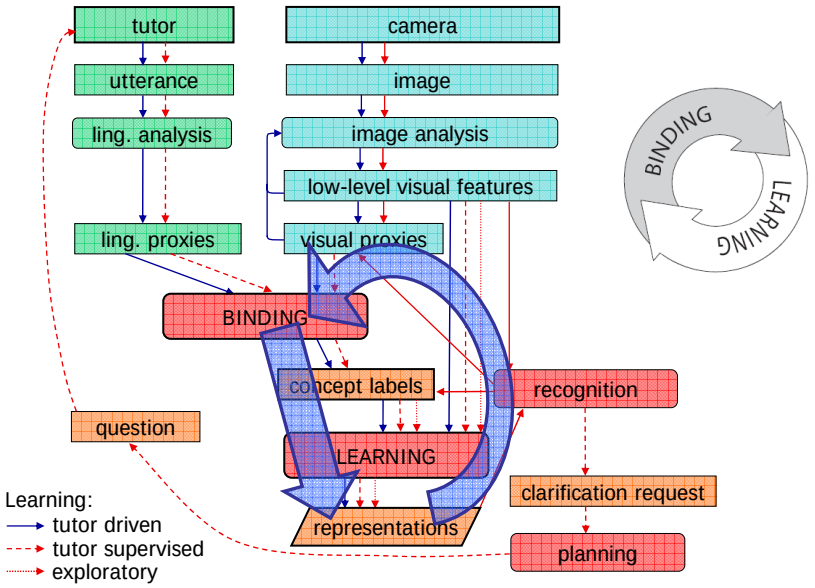
- Recognition
  - reliability of the answer



- Update table:

	YES	PY	PN	NO	DK
TD	ask	ask	ask	ask	ask
TSc	upd	ask	ask	/	ask
TSl	upd	upd	/	/	ask
EXc	upd	/	/	/	/
EXl	upd	upd	/	/	/

Learning process

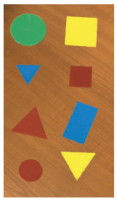


Learning method

- Statistical learning of basic visual concepts.
- Find a subset/combination of extracted features that determines visual attributes/spatial relations
  - **Consistency** of the feature values extracted from the objects labeled with the same concept label.
  - **Specificity** of these values with respect to the rest of the feature values related to other concept labels.
- Each attribute value is associated with one extracted visual feature
- It is modeled with a distribution of the values of the best extracted visual feature (using **KDE distribution**)
- For redundancy, all feature values are modeled with Gaussian distribution to enable switching association assignments
- Selection of the best feature is based on the Hellinger distance between probability distributions
- Incremental updating of representations
- Facilitates unlearning

Experimental setup

- Learning and recognition of visual attributes
  - 6 visual features
    - 3 appearance features (Hu,Sa,In)
    - 3 shape features (Ar,Pr,Cm)
  - 10 visual attributes
    - 4 colors (Rd,Gr,Bl,Yl)
    - 2 sizes (Sm,Lr)
    - 4 shapes (Sq,Cr,Tr,Rc)



  - Interactive system
  - Quantitative evaluation
    - Evolution of the recognition score through time
    - Scoring table:

	YES	PY	PN	NO	DK
YES	1	0.5	-0.5	-1	0
NO	-1	-0.5	0.5	1	0

Dialogue example

'H: This is a red, small, and square object.'
'R: OK.'
'H: What do you see?'
'R: I don't know this object.'
'H: It is green, small, and triangular.'
'R: OK.'
'H: What do you see?'
'R: I see a small and square object.'
'H: Yes, that's correct.'
'R: OK.'
'H: It is also blue.'
'R: OK.'
'H: What do you see?'
'R: I see a green object. I think it is also large.'
'H: Yes, that's correct.'
'R: OK.'
'H: It is also rectangular.'
'R: OK.'

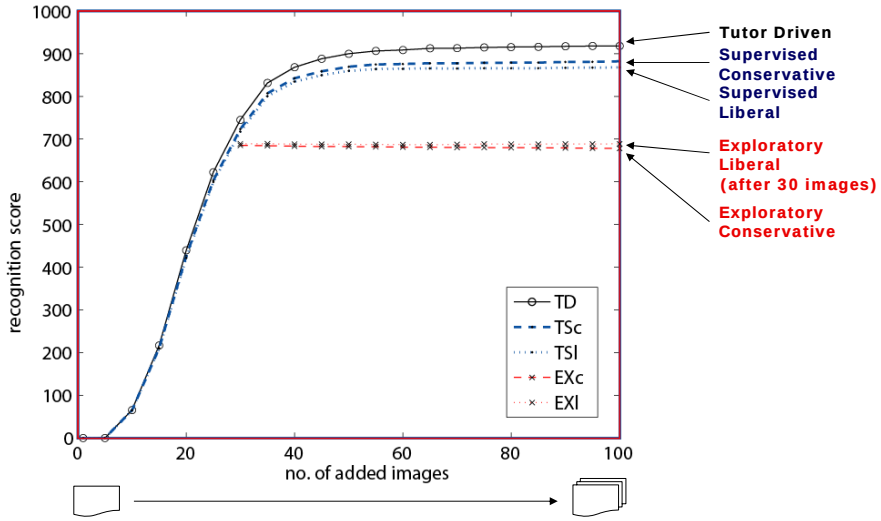
Tutor driven

'R: I don't know this object. What is it like?'
'H: It is blue, small, and rectangular.'
'R: OK.'
'R: Is this object large?'
'H: Yes, that's correct.'
'R: OK.'
'H: It is also yellow and circular.'
'R: OK.'
'R: Is this object red?'
'H: No, it is not red.'
'R: OK.'
'H: It is yellow.'
'R: OK.'
'H: What do you see?'
'R: I see a red, small, and triangular object.'

Tutor supervised

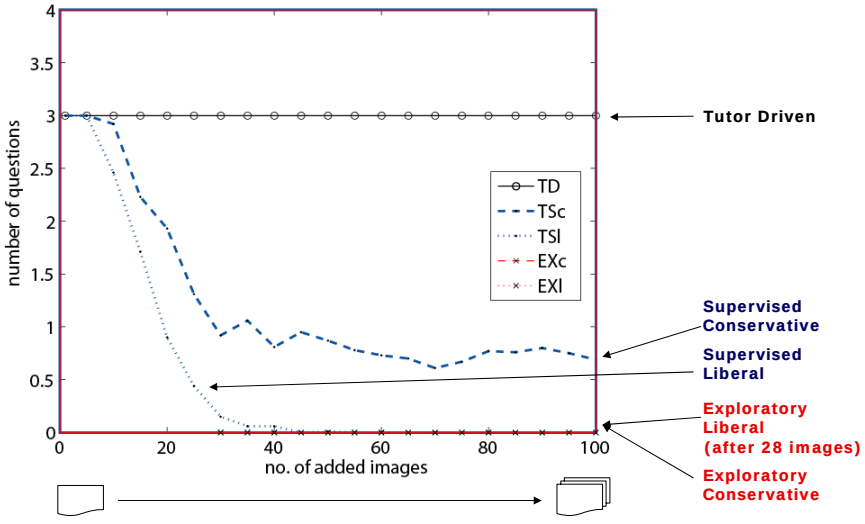
Exp. results – Appearance properties

- Recognition score



Exp. results – Appearance properties

- Number of questions

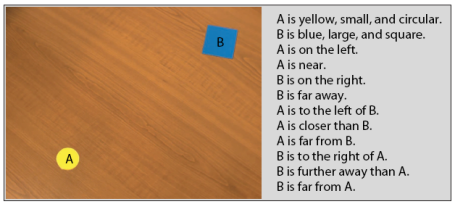


Learning of spatial relations

- 2 objects (A and B)
- 5 features (x,y,dx,dy,d)
- 11 spatial relations:
  - TL: to the left of: "A is to the left of B"
  - TR: to the right of: "A is to the right of B"
  - CT: closer than: "A is closer to me than B"
  - FT: further away than: "A is further away from me than B"
  - NT: near to: "A is near to B"
  - FF: far from: "A is far from B"
  - OL: on the left: "A is on the left"
  - OR: on the right: "A is on the right"
  - IM: in the middle: "A is in the middle"
  - NR: near: "A is near"
  - FA: far away: "A is far away"

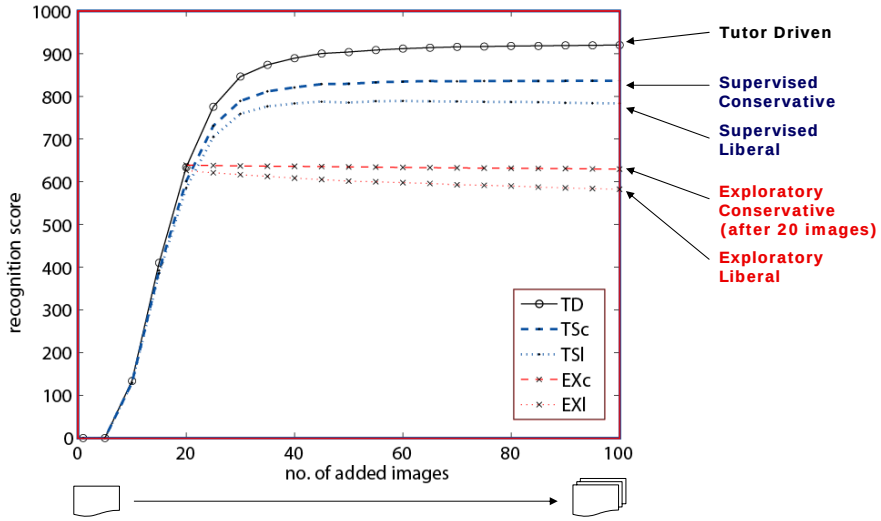
Two examples

- Automatic scene description



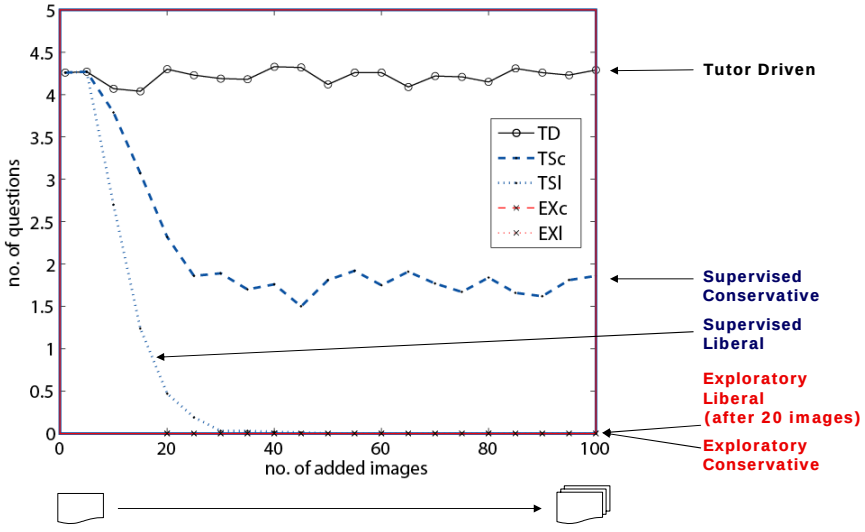
Exp. results – Spatial relations

- Recognition score



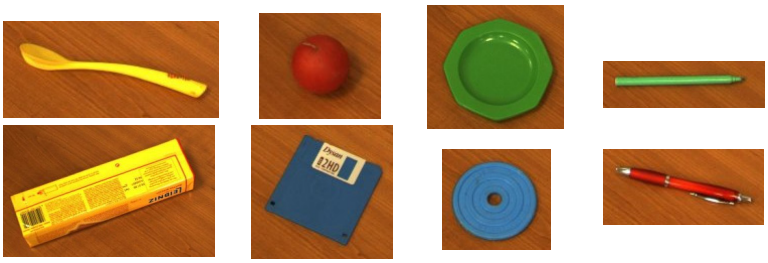
Exp. results – Spatial relations

- Number of questions

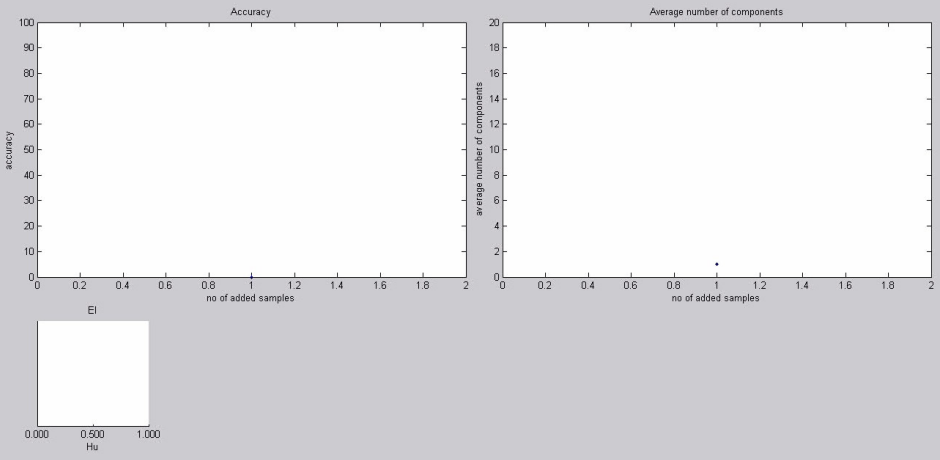


Experimental setup

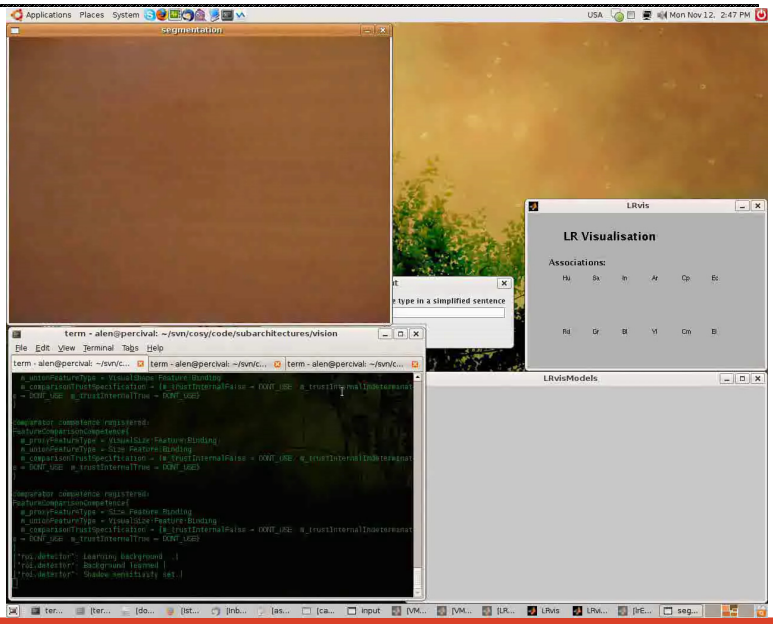
- Learning and recognition of visual properties of everyday objects
- 6 visual features
  - 3 appearance features (Hu,Sa,In)
  - 3 shape features (Ar,Co,Ec)
- 6 visual properties
  - 4 colors (Rd,Gr,Bl,Yl)
  - 2 shapes (Cm,El)



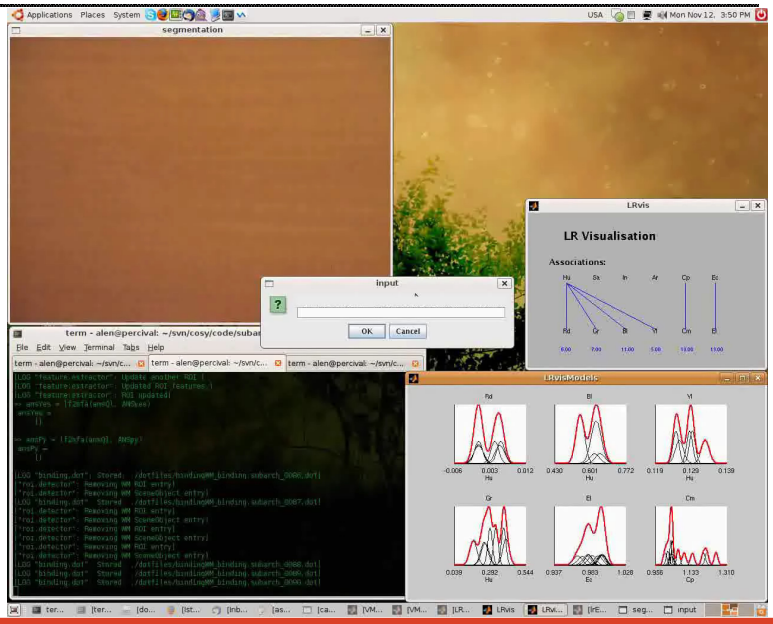
Experimental results



Demo - learning



Demo - unlearning



Conclusions

- A system for **continuous interactive** building of **cross-modal associations** between low level modality-specific features and amodal high level concepts
  - A unified framework for learning object basic properties and spatial relations
  - Facilitates unlearning and co-learning
  - Based on reconstructive representations (KDE)
  - Different modes of learning
  - Mixed initiative learning, implicit learning
- Interplay between cross-modal binding and interactive learning
  - Learner improves mappings used by binder
  - Binder generates training examples for learner
  - In an incremental and interactive way

The end