An evaluation of Diagnosis in a Learning Environment for Object-Oriented Modeling

Ludovic AUXEPAULES, Dominique PY

LIUM - Laboratoire d'Informatique de l'Université du Maine (France)

ICALT 2010 – July 7th 2010

Context of this work

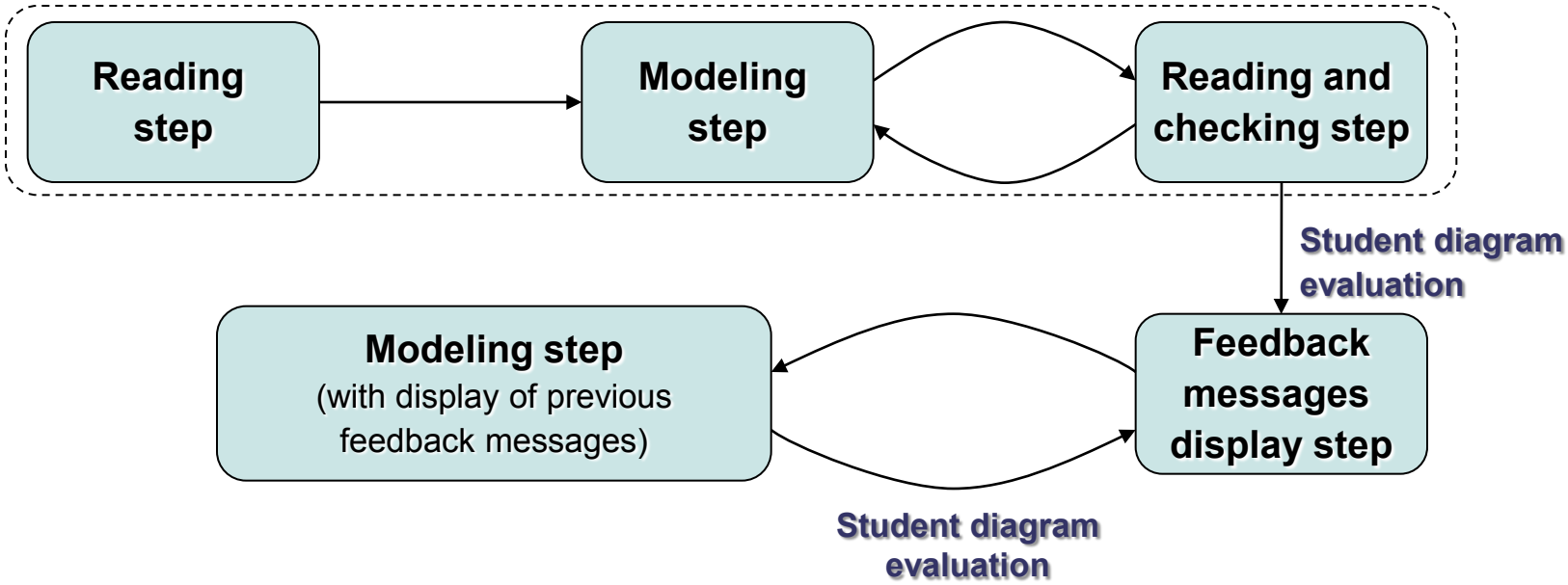
- Project of the LIUM laboratory: « Interaction and knowledge »
- **Participants:** Dominique Py, Mathilde Alonso, Thierry Lemeunier and Ludovic Auxepaules
- **Goal of the project:** designing models, methods and tools for object-oriented modeling learning environments
- **Application:** the *Diagram* environment

Outline

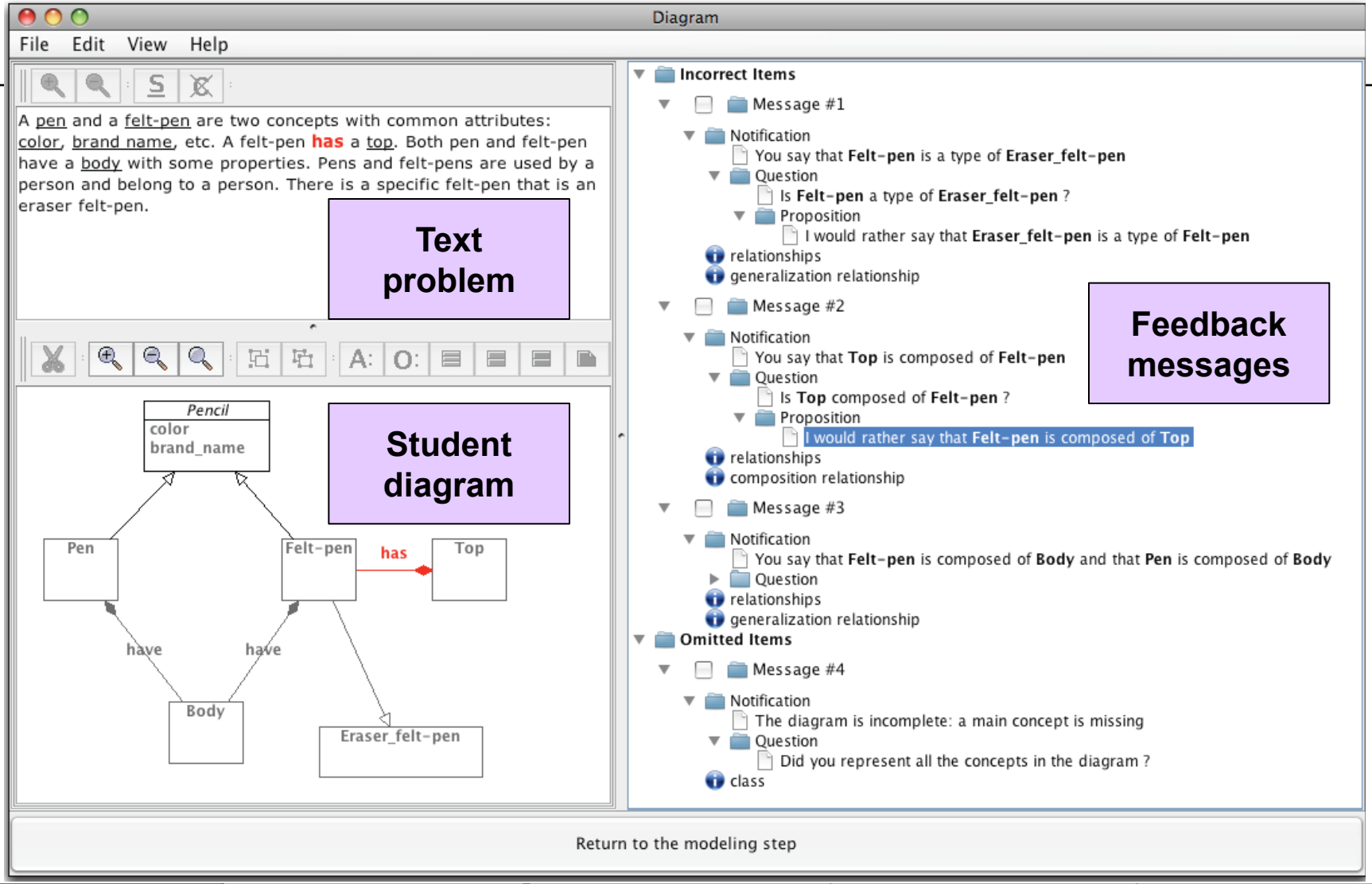
- The *Diagram* environment
- The ACDC matching method
  (*Automatic Class Diagrams Comparator)*
- A diagnosis example produced by ACDC
- An ACDC evaluation with *Diagram* used on ecological context
- Conclusions and perspectives

The *Diagram* environment

- An open UML class diagrams editor
  - Allowing the student to work with the problem text and the diagram together
  - Providing specific interaction modes and help features for novice users
- A three-step method for solving modeling problems



Example of feedback messages in *Diagram*

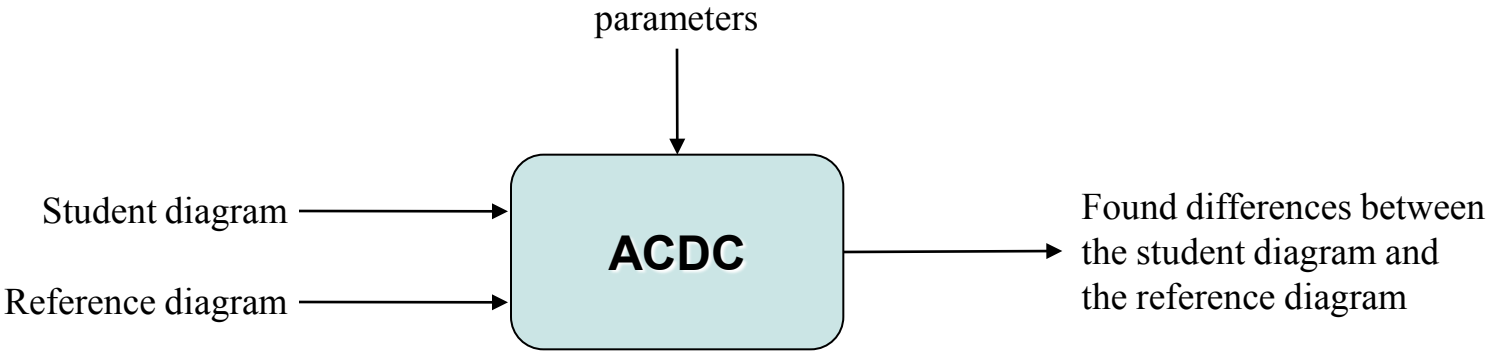


The proposition of diagnosis within *Diagram*

- Modeling and *Diagram* contexts
  - Diagnosis in our context: the system ability to analyse student's answers
  - No pedagogical solver in open-ended domains like modeling
  - Calculus time should **be fast enough** for synchronous pedagogical feedbacks
- A diagnosis based on a models matching method
  - Comparison of the **diagram built by the learner** with **a reference model** supplied by an expert
  - Production of **a differences list** between the models (no errors)

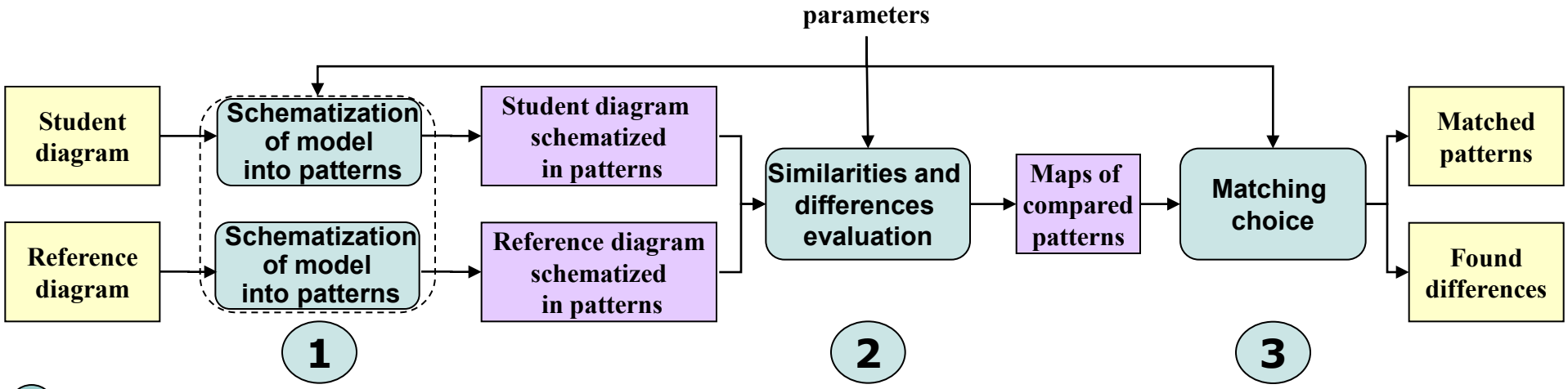
ACDC matching method (*Automatic Class Diagrams Comparator*)

- An **automatic** and **customizable** matching method [Auxepaules 09]



- An hybrid matcher that combines
  - String-based similarity of namespaces and type similarity constraints
  - Element-level and structure-level matching techniques [Shvaiko & Euzenat 05]
- Univalent or multivalent mappings at all model granularity levels
  - It fully or partially matches **one or more** structures of a model to **one or more** structures of another model and mutually

3 sequential steps of ACDC matching method



1. Schematization of inputs models into structural patterns (simple or complex)

   → Trees and graphs algorithms : search, cover and sort of roots, leafs, ways…

2. Evaluation of local similarities and differences of each patterns couple per type

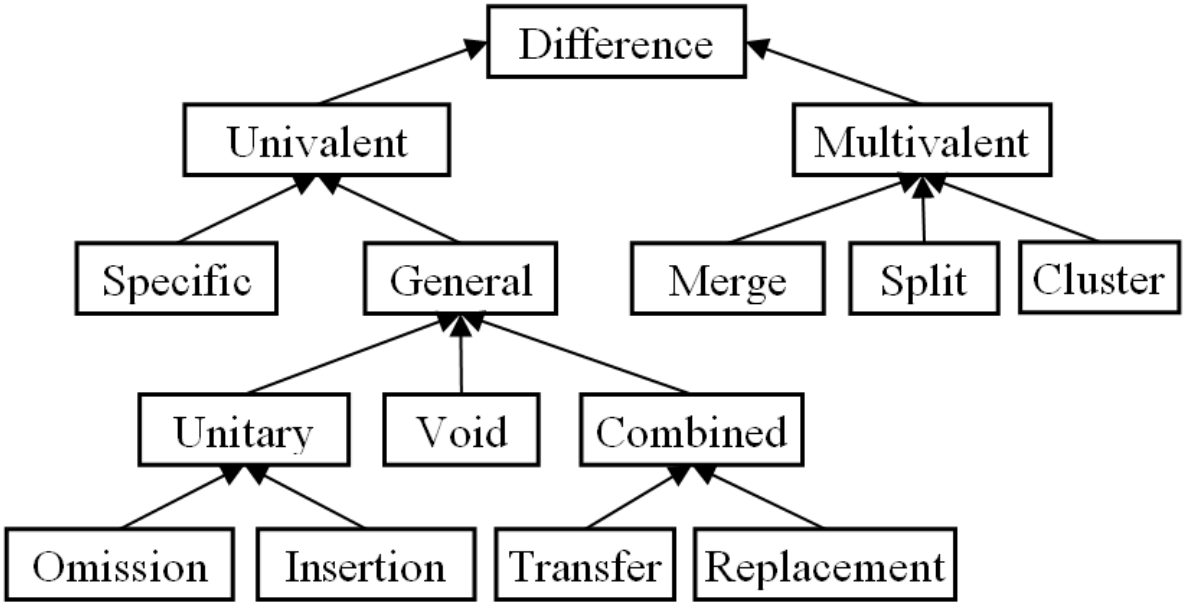
   → Similarity function that combines lots of criteria: names, context, specific properties…

3. Choice of one mapping of models patterns and differences

   → Greedy process without backtracking

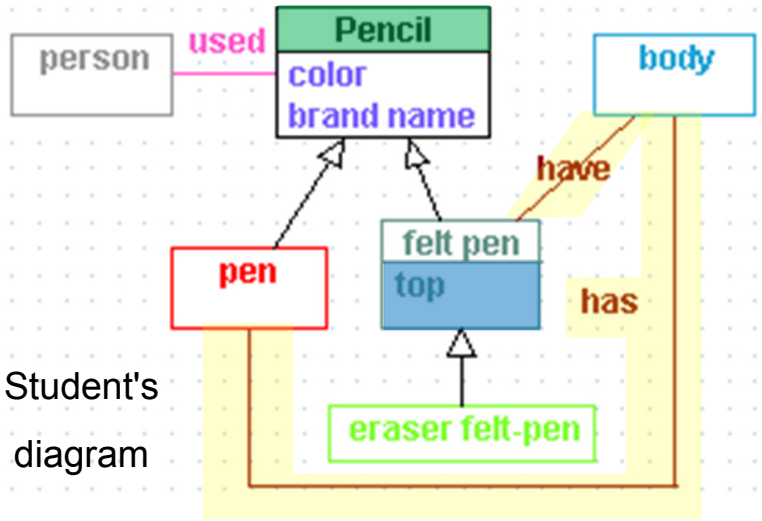
Our differences taxonomy [Auxepaules 09]

- **Univalent** difference: partial match of two single patterns (1:1 matching cardinality)
- **Multivalent** difference: partial match of a patterns group (n:m matching cardinality)

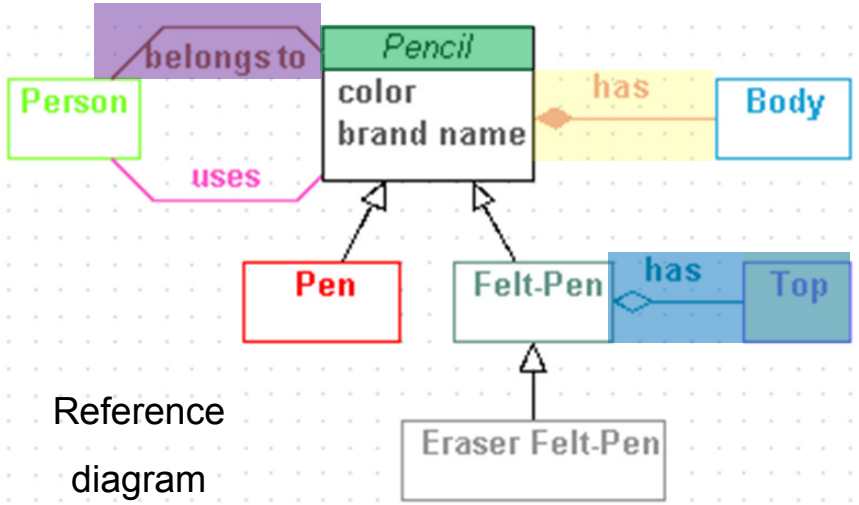


- Those differences are converted into pedagogical differences for elaborating feedback messages [Py *et al.* 08]

A diagnosis example produced by ACDC



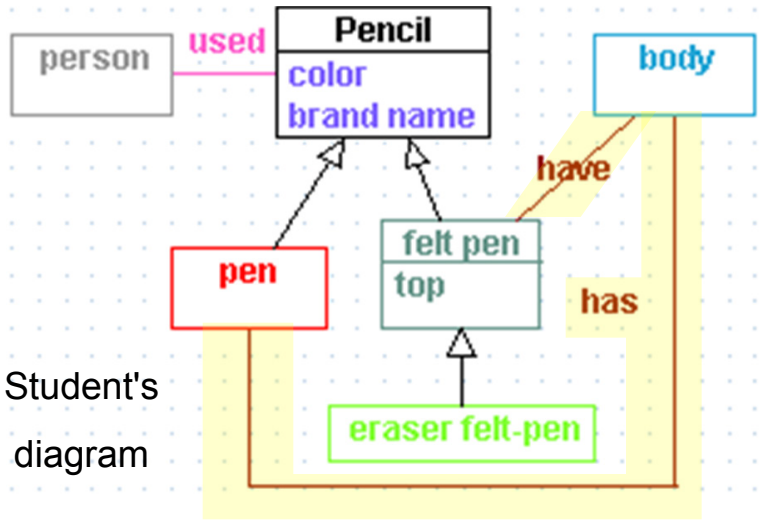
Student's diagram



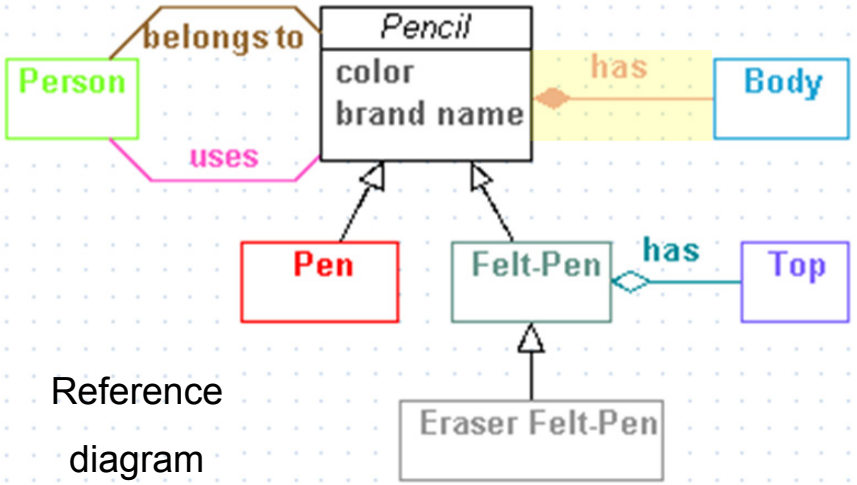
Reference diagram

- 12 full univalent matches (without structural differences)
- 4 main content and structural differences used by *Diagram* to produce pedagogical feedback messages

Example of a compound difference



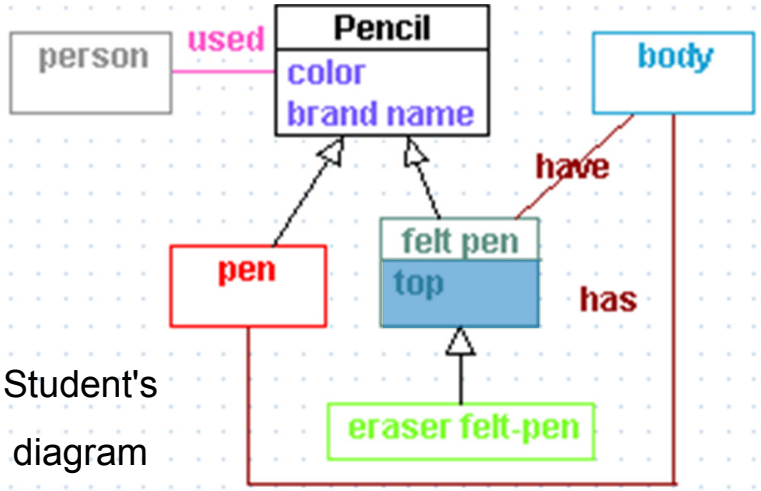
Student's diagram



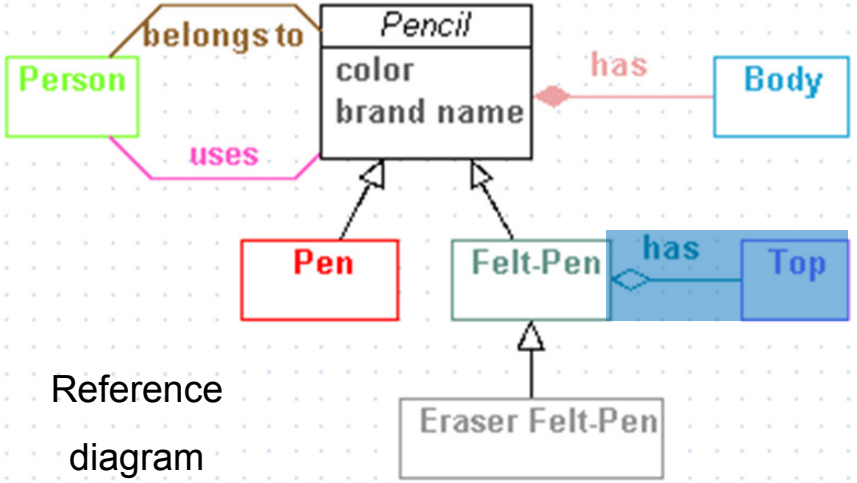
Reference diagram

Differences identified by ACDC	Pedagogical feedbacks
{have (body---felt pen) has (body---pen)} SPLIT {has (Body---*Pencil*)}	Duplication and transfer of a relationship
{have (body---felt pen)} TRANSFER LOWER {has (Body---*Pencil*)}	
{has (body---pen)} TRANSFER LOWER {has (Body---*Pencil*)}	

Example of an other compound difference



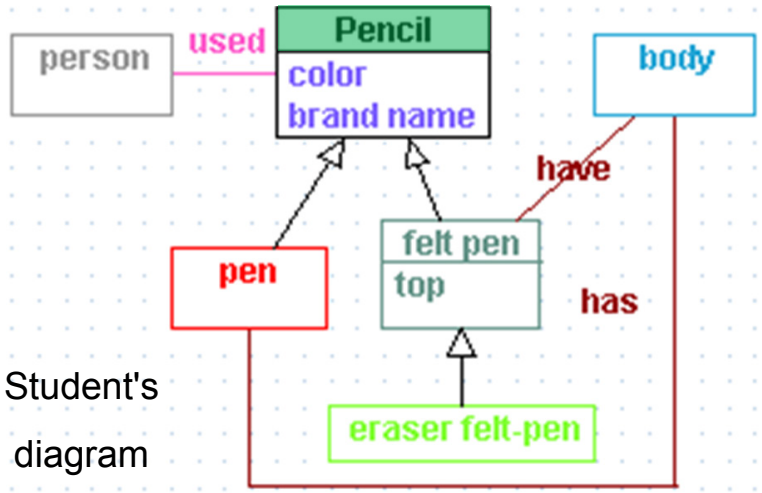
Student's diagram



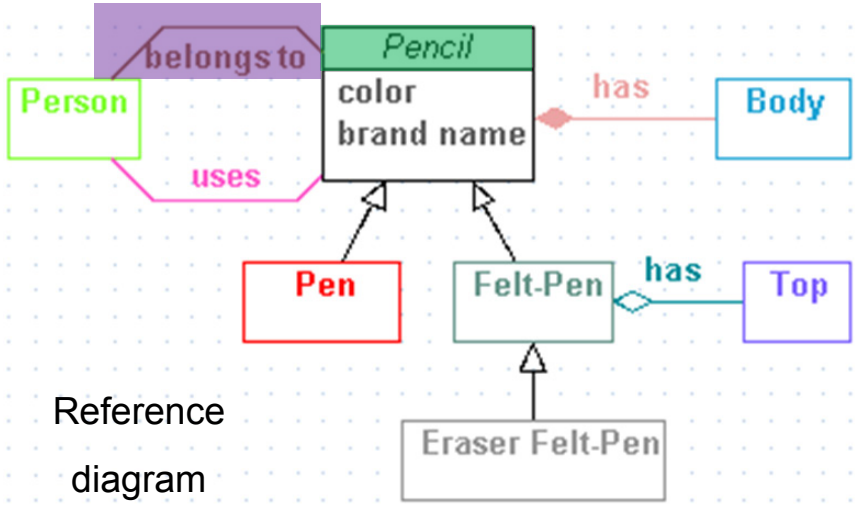
Reference diagram

Differences identified by ACDC	Pedagogical feedbacks
{felt pen::top} REPLACEMENT {Top} | {felt pen::top} INCOMPATIBLE_NATURE {Top}	Misrepresentation of a class and omission of linked elements
OMISSION {belongs to (Person---*Pencil*)}	

Example of two simple differences



Student's diagram



Reference diagram

Differences identified by ACDC	Pedagogical feedbacks
OMISSION {belongs to (Person---*Pencil*)}	Omission of a relationship
{Pencil} VOID {*Pencil*} | {Pencil} NOT_ABSTRACT_TO_ABSTRACT {*Pencil*}	Misrepresentation of a class (a class instead of an abstract class)

1st evaluation of ACDC within *Diagram*

- Previous evaluations
  - Tests of usability, interaction and pedagogical functionalities of *Diagram* (without diagnosis system) [Alonso *et al.* 08]
  - Off-line evaluation of ACDC (not in *Diagram*) [Auxepaules *et al.* 2008]
- Protocol of the evaluation of ACDC within *Diagram*
  - 18 novice students in 2nd year of University (DEUST)
  - 4 practice sessions of 3 hours of modeling in *Diagram*
  - Reference diagrams built by OOM teacher
  - Sessions, built diagrams, diagnosis calls and feedbacks have been recorded
- Evaluation of feedback messages reported in [Alonso & Py 09]

Match quality measures

- Comparison of the automatically identified matches (ACDC outputs) with the manually determined real matches
  - ***False positives A :*** matches needed but not identified by the matcher
  - ***True positives B :*** correct matches automatically predicted by the matcher
  - ***False negatives C :*** matches falsely proposed or mistyped by the matcher

- Measures of quality [Do & Rahm 07] [Giunchiglia *et al.* 07] [Melnik *et al.* 02]
  - ***Precision*** = $|B| / (|B| + |C|)$ → an **accuracy** or **fidelity** measure
  - ***Recall*** = $|B| / (|B| + |A|)$ → a **completeness** measure
  - ***F-Measure*** → an harmonic mean of *Precision* and *Recall*
  - ***Overall*** = $(|B| - |C|) / (|B| + |A|)$ → an evaluation of the post-match effort needed for adding missed matches and removing false ones

Results at the diagram level

Number and % of diagnosis calls where	*Precision*	*Recall*	*F-Measure*	*Overall*
Quality Measure Result = 1 (best)	103 71.6%	60 41.7%	**55** **38.2%**	55 38.2%
0.85 ≤ Quality Measure Result < 1	**30** **20.8%**	**65** **45.1%**	**78** **54.2%**	59 41%
0.7 ≤ Quality Measure Result < 0.85	**11** **7.6%**	**19** **13.2%**	10 6.9%	22 15.3%
Quality Measure Result < 0.7	0 0%	0 0%	1 0.7%	8 5.5%

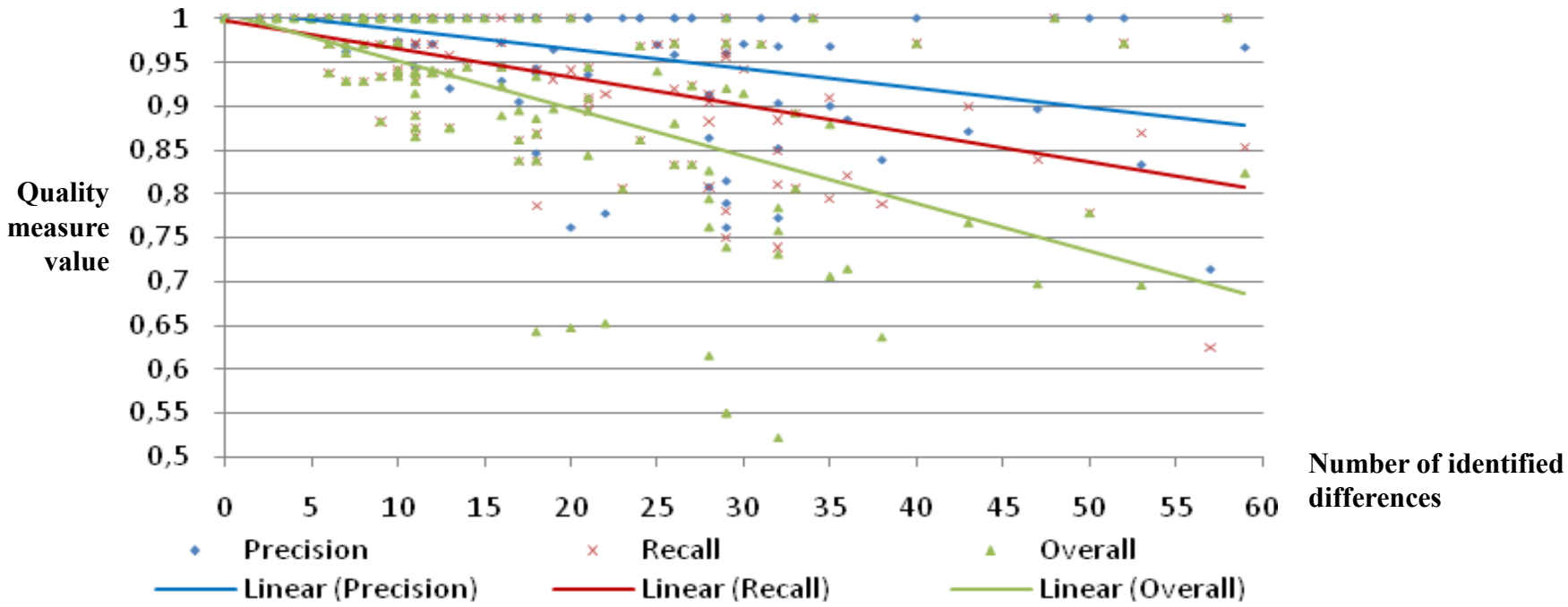
- Good results for the 144 diagnosis calls (144 different student diagrams)
  - **38%** of diagnosis calls outputs are perfect
  - More than **92%** of diagnosis calls outputs are relevant at 85%
  - At least one mismatch on only **28%** of 144 diagnosis calls
  - At least one omited match on **58%** of 144 diagnosis calls
- Diagnosis calculi time from 0.2s to 6s with an average of 2s

Results by type of structural differences

Differences	A: false negatives	B: true positives	C: false positives	Precision	Recall
Split	43	71	23	0.76	0.63
Merge	6	47	22	0.68	0.89
Void	31	3069	52	0.98	0.99
Replacement	62	416	33	0.93	0.87
Transfer	141	133	18	0.88	0.49
Total	**283**	**3736**	**148**	**0.96**	**0.93**

- Most of matches are **strict** (without difference of structure)
- Results related to the biggest structure alterations of models are well
  - **70%** of **multivalent matches** are identified and **72%** are correct

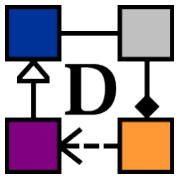
Results with regard to number of differences



- *Precision* and *Recall* trends of ACDC decrease according to the number of differences between compared diagrams
- *Precision* results are better than *Recall* Results
  - A **wrong match can replace one or more correct matches**

Conclusions and perspectives

- An evaluation overview shows that ACDC worked fairly well but requires to be improved in some specific situations
- Use of several reference diagrams to analyze the student's diagram
  - Complete / part of diagrams and correct / erroneous diagrams
  - ACDC can already compare more than two models and choose one reference at the end of the 2nd step of similarities and differences patterns evaluation
- Extend to others kinds of model or domains
  - Direct transposition for Entity-Relationship models of Data-Bases
- Use of ACDC system outputs in other context : teacher needs



Thank you for your attention