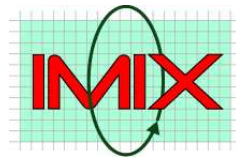


Dialogue management in the IMIX system.

Boris van Schooten

*Human Media Interaction,
IMIX project, VIDIAM subproject
(Interactive Multimodal Information eXtraction)*



Contents

- Introduction

What is IMIX? What is Question Answering? Research Goals.

- Dialogue strategies

Existing strategies. What do users do?

- Handling follow-up questions

Unimodal and multimodal follow-up questions. IR strategies. Resolving referents. Pictures and gestures.

- Work in progress

Multimodal QA Dialogue: IMIX

IMIX: Interactive Multimodal Information eXtraction

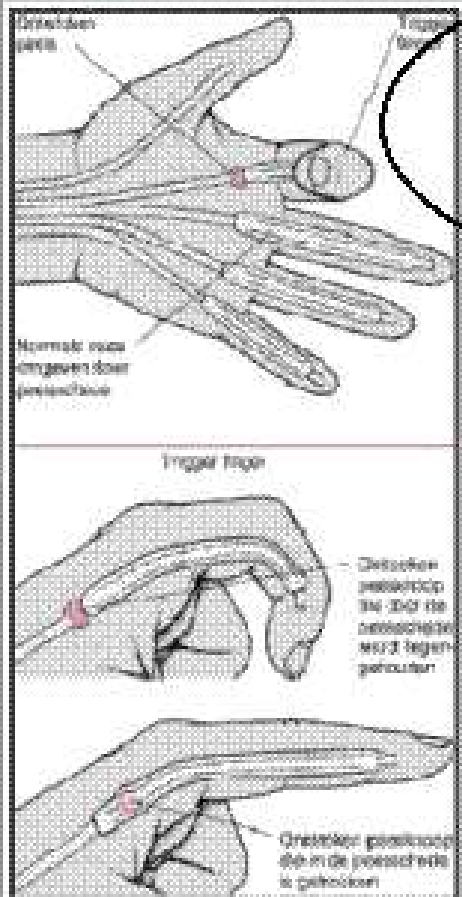
A Question Answering (QA) system, answering natural language questions.

IMIX features:

- QA for casual users in the medical domain
- More effective QA by means of dialogue. The system enables full dialogue tailored to naive/casual users
- Multimodal input with pen, text, and speech; multimodal output with pictures, text, and speech.

IMIX Demonstrator

U tikte in: wat is een trigger finger




Normale pees omgeven door gewrichtsvloeistof

Trigger finger

Dittoeken positie bij DOZ de pees wordt nu tegengehouden

Overtoeken positie bij DOZ in de pees wordt nu gelokt

ook tendovaginitis nodosa, tendovaginitis stenosans, trigger finger of snapping finger, aandoening aan de pees van een vinger.



Spreken en tekenen Tikken en tekenen Nieuwe dialoog Stoppen

Research questions of the IMIX-VIDIAM project

VIDIAM: DIAlogue Management and the VIsual channel

How can dialogue and multimodality give added value to QA?

- What effective dialogue strategies are there?
- How do real users behave?
- What architecture should such a system have?
- Can a generic domain-independent QA Dam module be built?

Potential dialogue strategies

- Handling user follow-up questions

U: And who wrote the scenario ?

U: All types of proteins ?

U: So the answer is "no" ?

U: One minute per what ?

U: What are these blue spots ?

- Handling user evaluation of system answer

U: I didn't mean that

U: What has that got to do with RSI ?

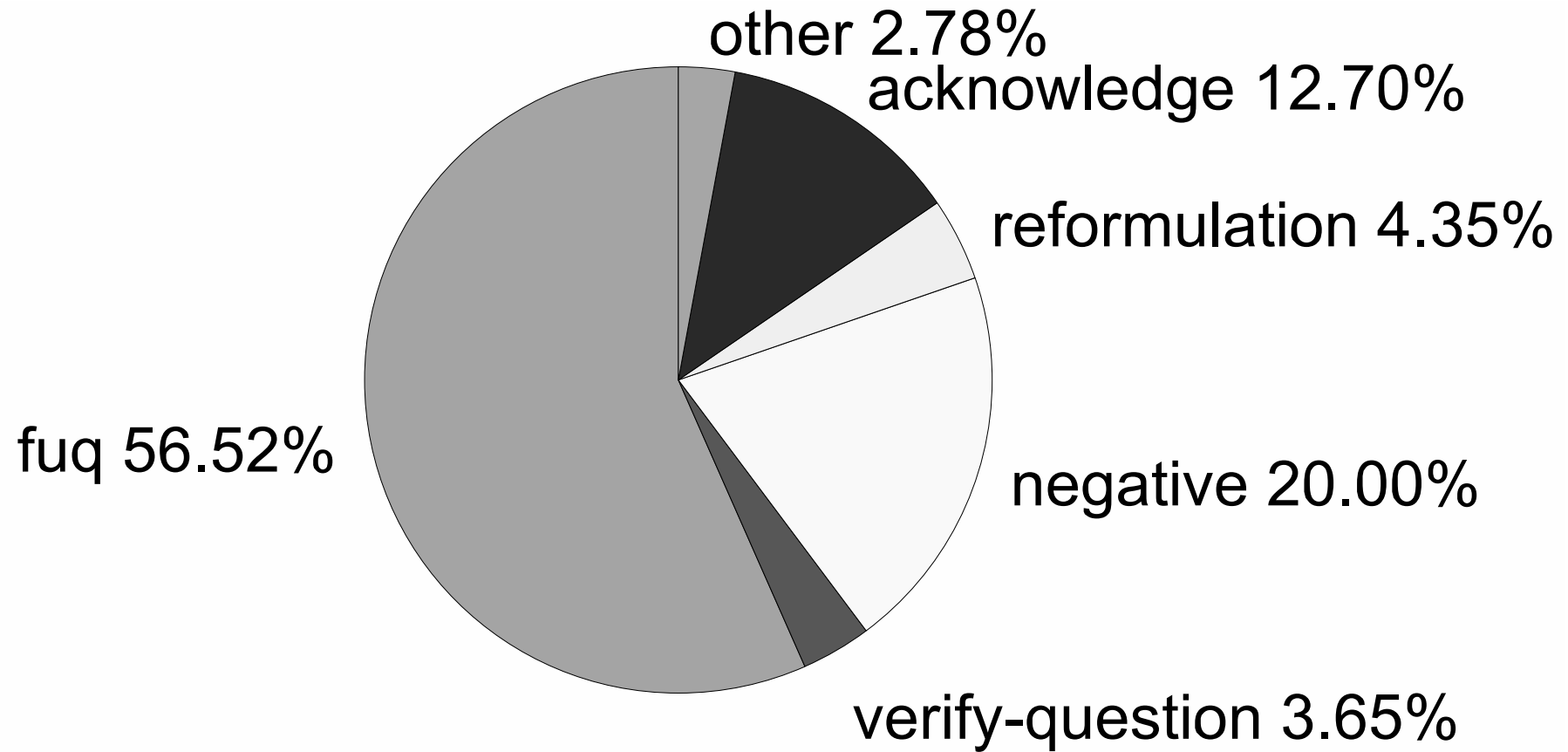
- System clarification to improve questions

S: The question is not specific enough. Perhaps you mean ... ?

S: Do you mean ... or ... ?

S: I can't find an answer. Perhaps I should search for ... or ... ?

User reactions to answers



Handling follow-up questions

Three main strategies:

- Rewriting the follow-up question
 - can be used with any QA system
 - difficult; state of the art is rather limited
- Adding dialogue context to the IR query
 - does not require correct sentence structure
- Searching in previous document set
 - simplest possible implementation, gives good performance
 - for follow-up questions asking for more details on the same topic

Handling follow-up questions: implementation

1 Identify if sentence requires referent

Use cue words, POS tags, absence of domain-specific keywords

2 Collect referents from previous questions and answers

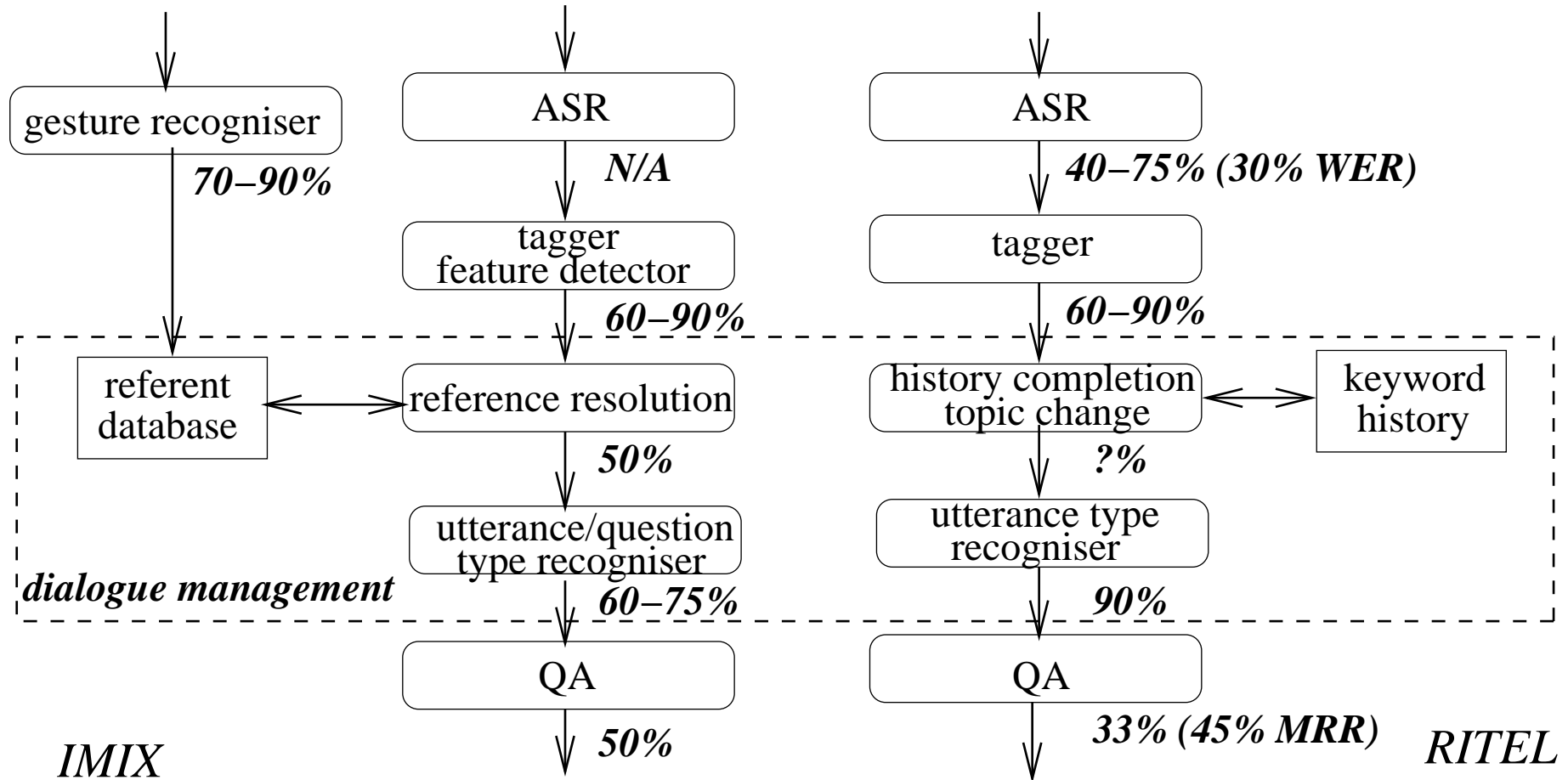
All sufficiently specific phrases are potential referents; use Lappin and Leass salience mechanism

3 Select referent from referent set

Criteria: salience, anaphor-antecedent matching, semantic matching

4 Rewrite sentence to include referent (not required for IR completion method)

Handling follow-up questions: architecture



Handling follow-up questions: Rewriting strategies

- Anaphor

How can I recognise it?

=> *How can I recognise RSI?*

- PP Attachment

What are the symptoms?

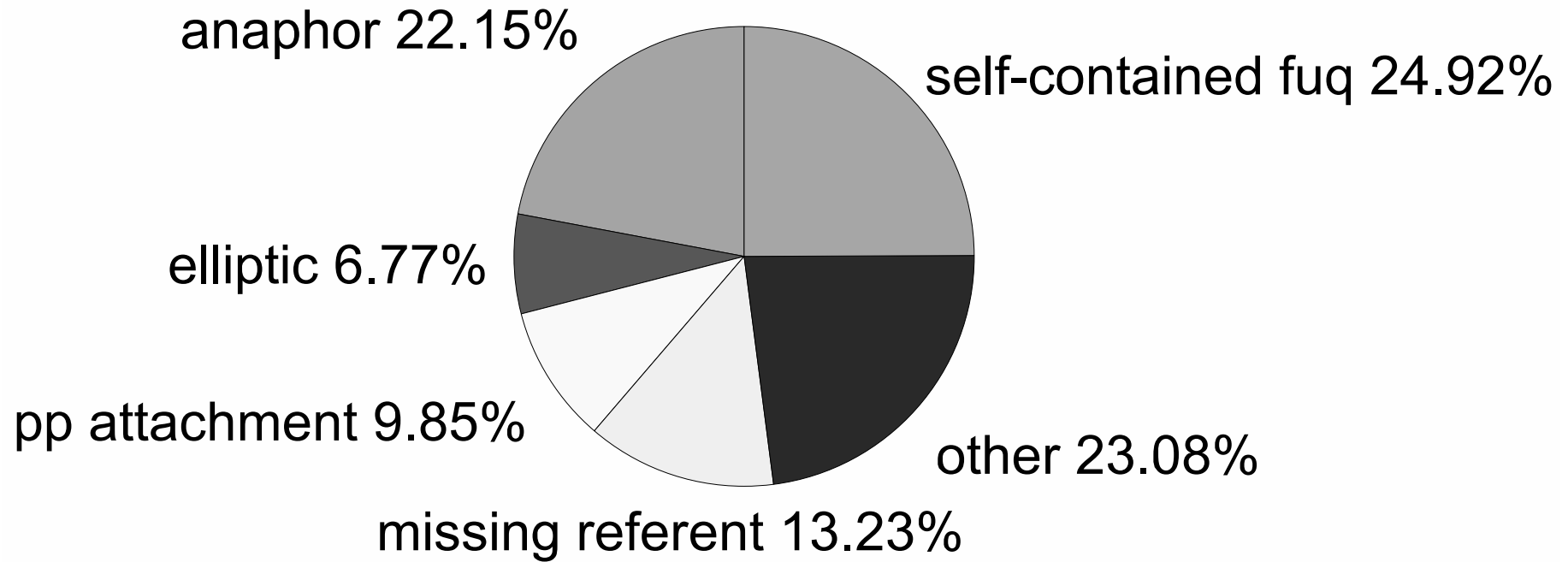
=> *What are the symptoms of RSI?*

- Elliptic

And writing?

=> *And can I get RSI from writing?*

Handling unimodal FQ rewriting potential



Handling unimodal FQ: findings

- Rewriting is only possible for 52% of the non-self-contained questions. The rest could be resolved using IR context completion.
- Rewriting performance still low. Anaphor is the most successful; overall performance = 42%
- Reference resolution: 75% of the referents in the second question corpus refer to a referent in the initial question
- 73% of the question classified as "missing referent" could be answered by the source document of the answer

Handling multimodal FQ

- **Self-contained:** no processing
- **Regular rewritable:** rewrite into a fully syntactic self-contained question using the Anaphor method.

ex: Can you tell me more about this?

(while pointing at a label "salt reabsorption")

rewritten: Can you tell me more about [salt reabsorption in the Bowman capsule]?

- **Regular:** make self-contained by indicating which keywords or utterances should be considered context

ex: What is this shrinkage caused by?

(pointing at a picture of an Alzheimer afflicted brain)

- **Visual element identity:** give the identity of the referred visual element

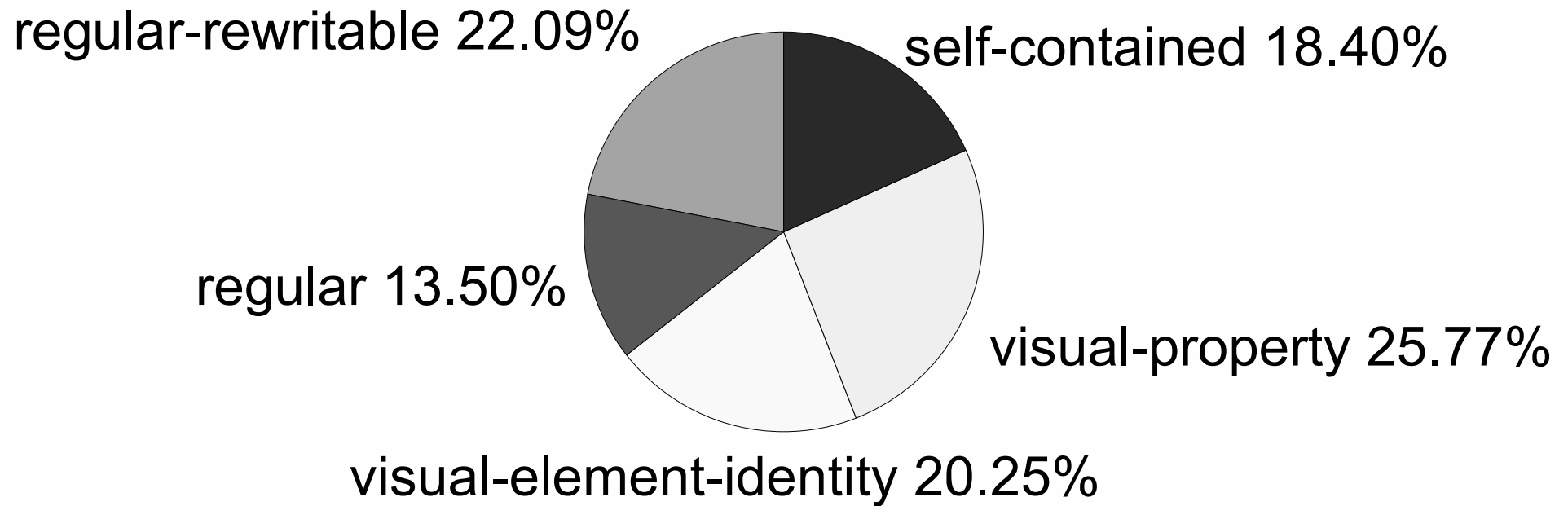
ex: What are these blue circles?

- **Visual property:** give general information about picture

ex: Does the DNA pass through the ribosome as drawn here?

ex: Is it important that the armrests are this high? (pointing at armrests of an office chair)

These are not regular QA questions; they cannot be submitted to a QA engine but have to be addressed in an alternative way.



Occurrence of the different multimodal types

Handling multimodal FQ: resolving visual referents

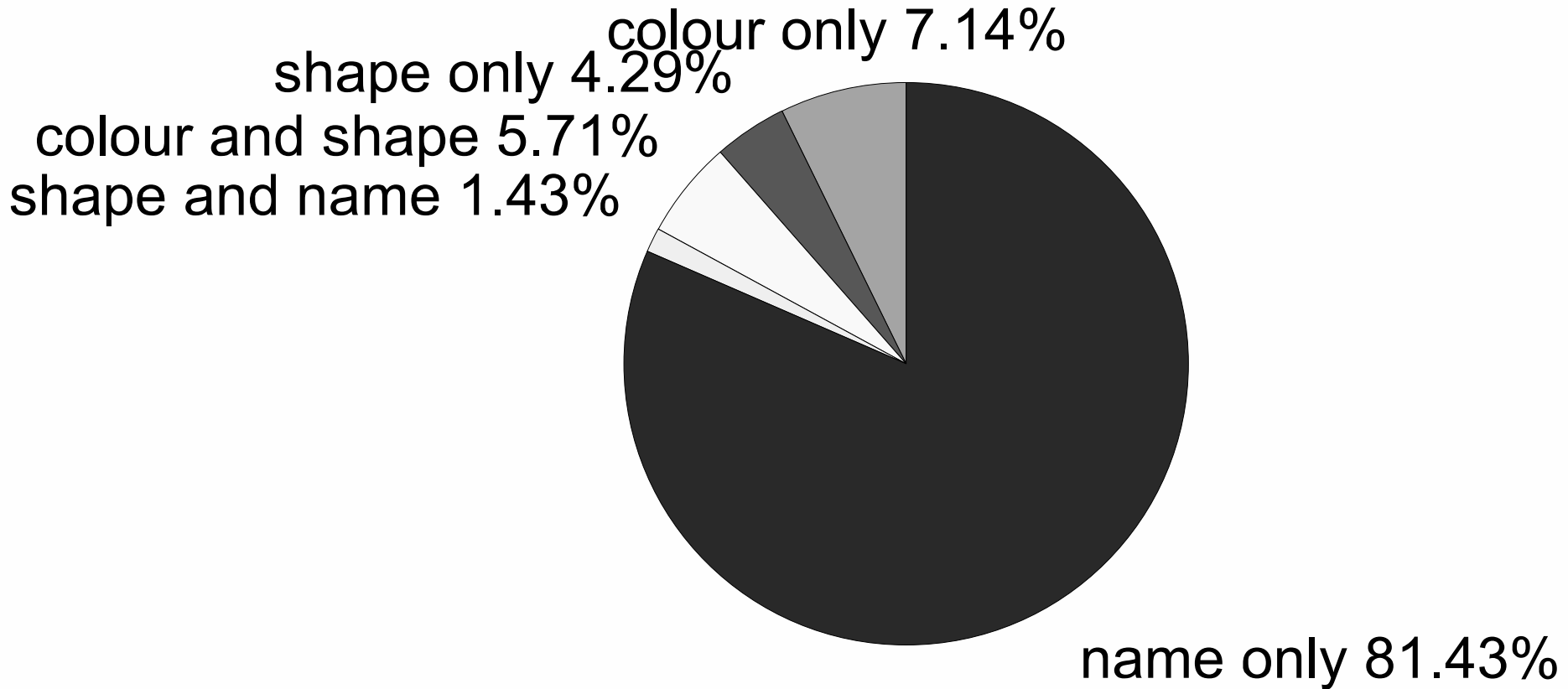
Basic findings:

- all the questions have a linguistic component
- all mouse strokes are part of gestures, each indicating one or more visual elements. Most indicate one visual element.
- most pointing gestures correspond to anaphors (91%)

So: gestures can be viewed as hints to help anaphorically disambiguate the linguistic reference to visual referents.

We distinguish the following hints: **gesture, name, colour, shape**

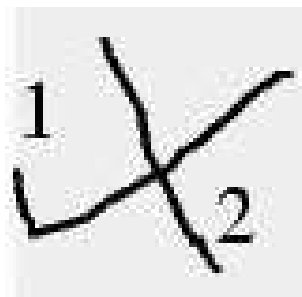
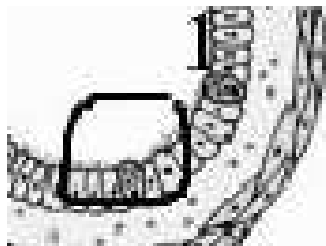
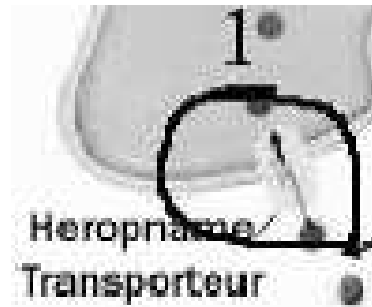
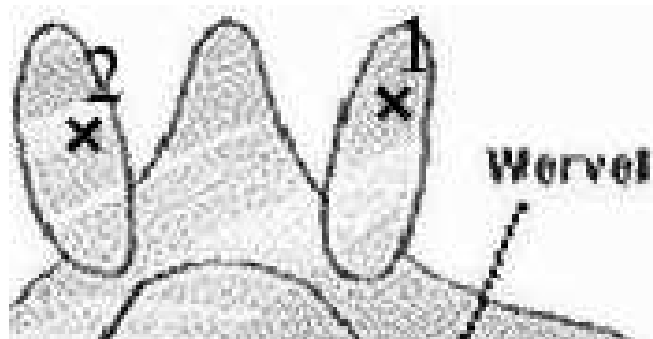
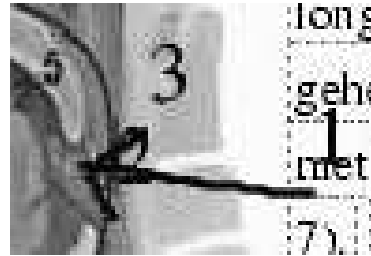
Handling multimodal FQ: Resolving visual referents



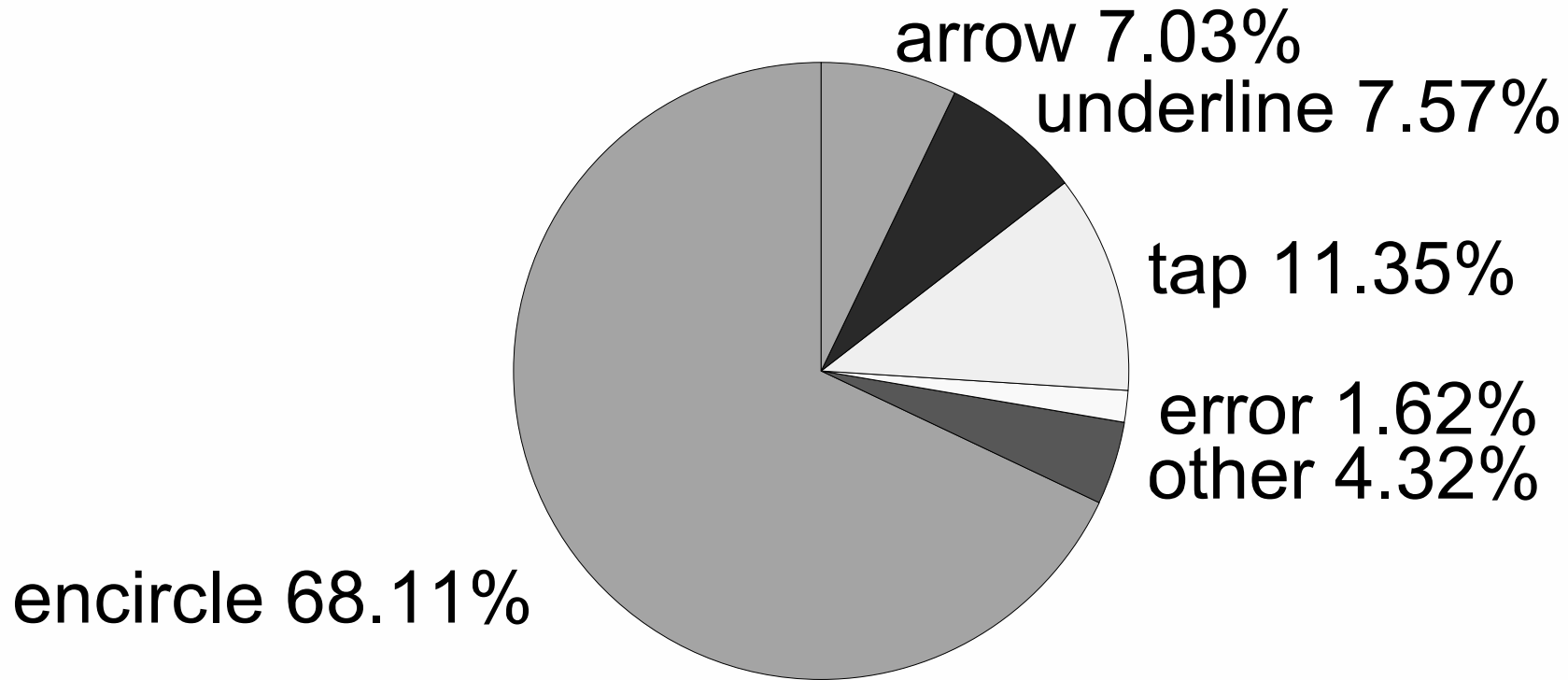
**The hints given by the user to disambiguate
visual anaphors**

Handling multimodal FQ: resolving visual referents

die ziektekiemen kur
de prionziektes. 1

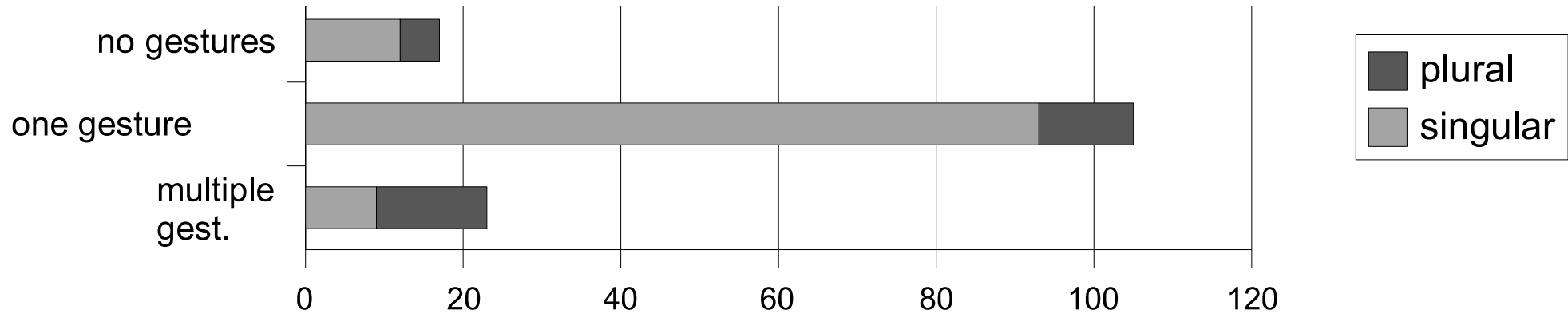


Handling multimodal FQ: resolving visual referents



The types of gestures and their distribution

Handling multimodal FQ: Resolving visual referents



The relation between anaphors, gestures, number of gestures per anaphor, and anaphor multiplicity

Findings:

- Pointing at multiple visual elements is correlated with use of plural anaphors
- Both singular and plural anaphors are regularly used in all cases.

Handling multimodal FQ: Annotating pictures

How should pictures be annotated to be able to handle these follow-up questions?

- annotate all visual elements as fine-grained as possible (which is tractable for schematic pictures)
- enable aggregates of elements to be annotated as separate elements

labels: colour, shape, keywords, function, noun phrase

function label: graphical, annotation-text, callout, flow

Handling multimodal FQ: results

Results of gesture referent identification using this annotation and a very basic gesture matching algorithm:

- 66% correct.
- 16% of the referents were not annotatable
- 6% of the gestures encompass multiple visual elements

Conclusion:

- our annotation scheme proved effective to annotate 84% of the appropriate referents;
- 88% of these could be found by the gesture recogniser

First attempt at follow-up question type classification using machine learning: 40% of the 5 classes plus a 6th non-multimodal class correctly classified.

Most important features:

- number of gestures (0=non-multimodal, 1=regular, 2=visual)
- occurrence of "this" (regular rewritable)
- occurrence of "picture" (visual property)

How to further improve the performance?

Work in progress: confirmation of speech recognition and other uncertain analysis.

Ritel: let system do **implicit confirmation** of relevant information; let user **confirm or disconfirm**.

U: What is the population of the UK ?

*S: You are looking for a number related to population and Uruguay ?
I could not find an answer. Another question ?*

U: No, I meant United Kingdom.

IMIX: let user choose between different options if system is uncertain.

U: What are its symptoms ?

*S: Do you mean 'What are the symptoms of RSI'
or 'What are the symptoms of mouse arm' ?*

U: RSI

=> a more systematic way to cover these and different cases

Some other research items:

- Suggestions. Let the system come up with a list of suggestions (along with the answer). A suggestion list and synonym list module are made available.
- Link between follow-up question and information need. When does the user want more details, when does s/he move to a new topic, when is s/he trying to verify a belief, etc.