Common Sense.
For Computers.

THINGS
THINKING

Be Human

Artificial Intelligence in Requirements Engineering

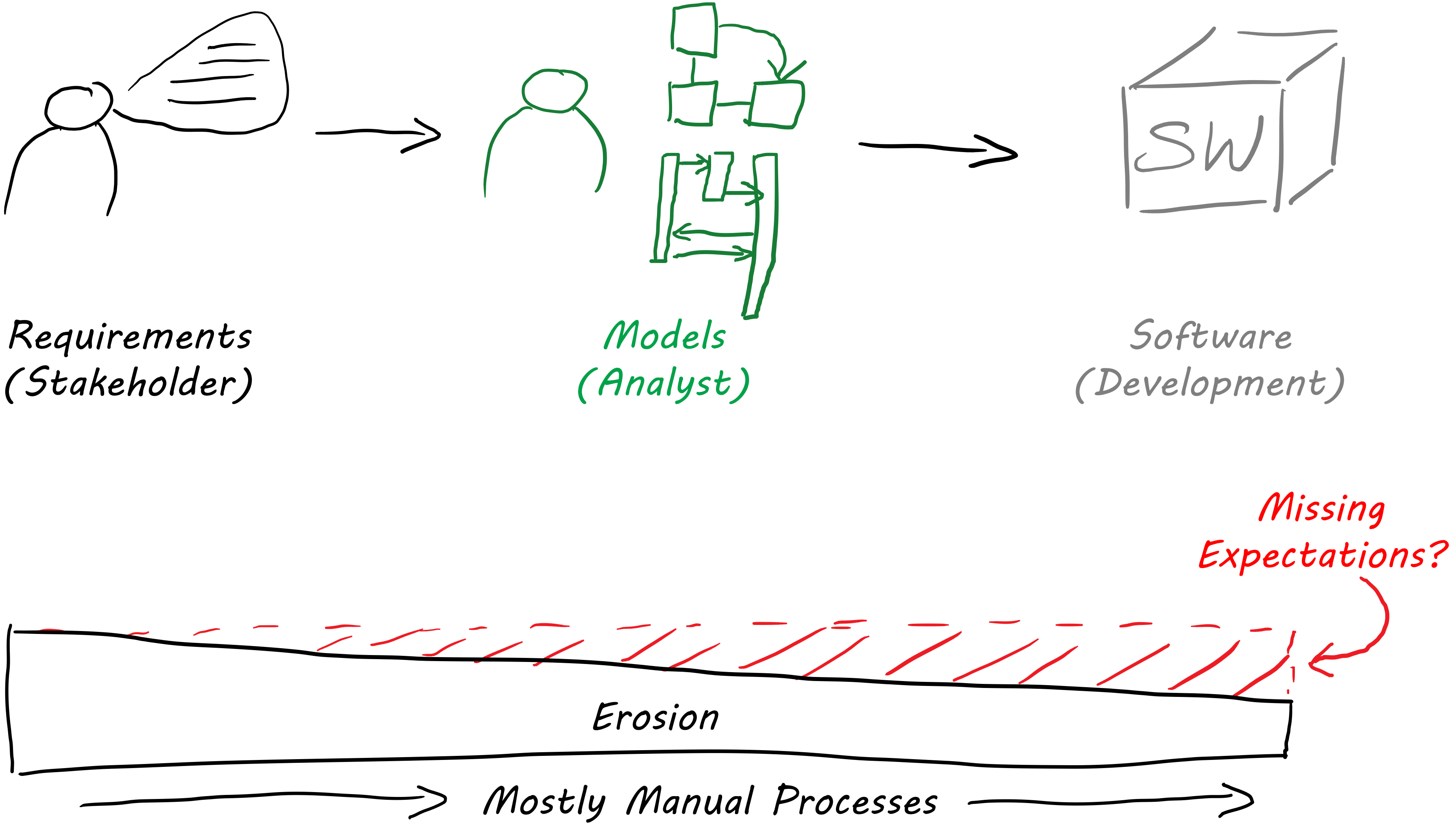
Dr. Mathias Landhäußer
Dr. Sven J. Körner

95% of Requirements are Recorded in Natural Language

Getting the facts right is not enough!

[MFI04]

Challenge: from Requirements to Software



AI domain of expertise is very limited to whatever universe we train them on.

Most of the systems, you show them [..] unusual situations [..] and they will say complete garbage about it.

They don't have common sense.

Yann LeCun, Facebook AI

What Is Common Sense?

- The trophy does not fit into the suitcase, because it is too big.

- The trophy does not fit into the suitcase, because it is too small.

Why do *you* know, what "it" refers to?

“People remember errors committed by AI, but forget human errors”

Linguistic Flaws in Requirements

- Linguistic Flaws
  - Generalization
    - Presuppositions
    - Incomplete comparatives and superlatives
    - Modal words (possibilities)
    - Modal words (necessities)
    - Incompletely specified process words
  - Distortion
    - Quantifiers
    - Incompletely specified conditions
    - Nouns without reference index
  - Deletion
    - Nominalizations

[RUPP]

No Tools, Just Rules?

http://www.sxc.hu/photo/1269809

Problem

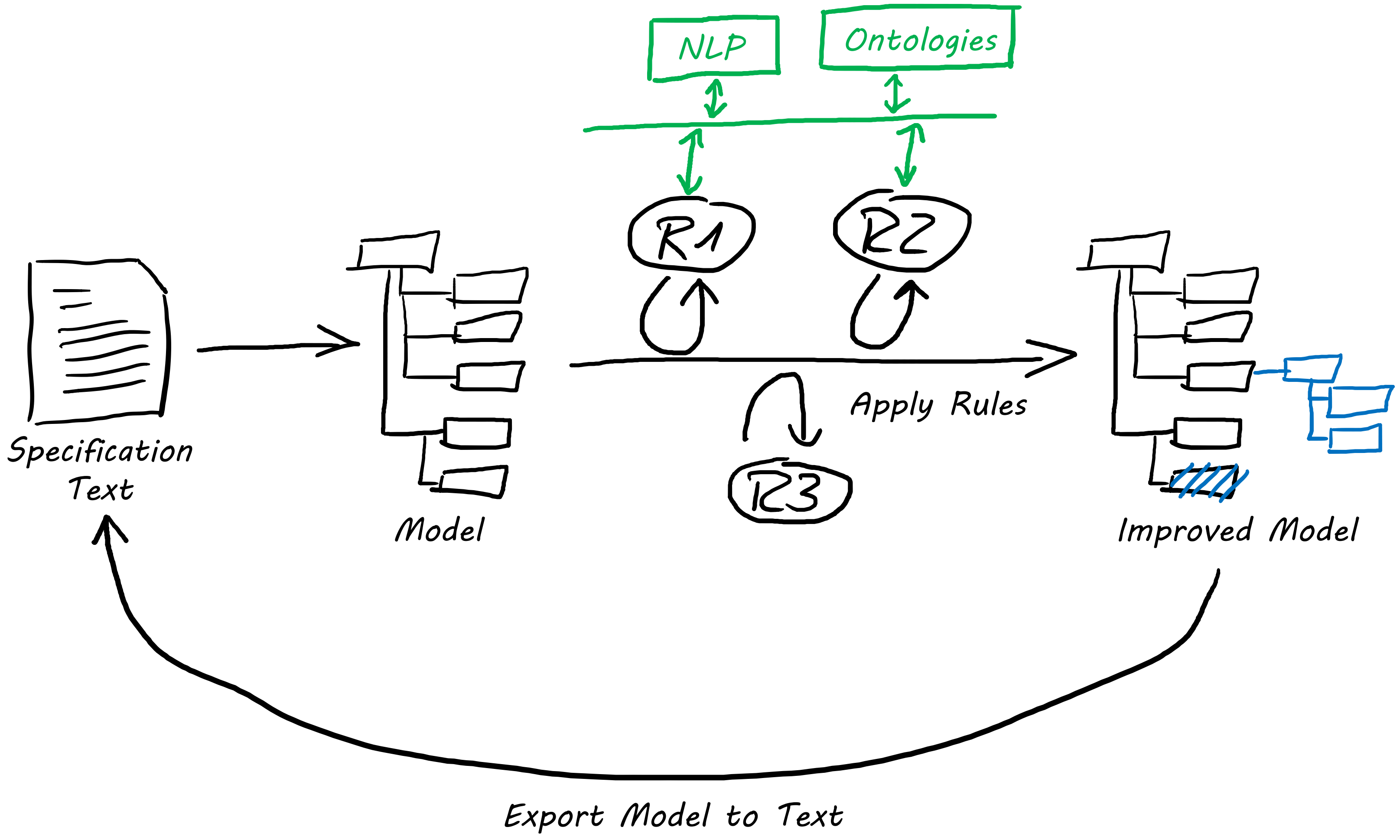
Cost	Fail	43% of all errors in IT and engineering projects lead back to wrong specifications.
Resources	Human	Today, errors based on meaning and understanding must be solved by humans.
Complex	Natural Language	Comprises 95% of all specifications. Also, natural language is the means of choice for anybody to communicate with computer systems.
Understand	Semantics	Are a key aspect to cognitive computing challenges which **cannot** be solved with machine learning (neural networks) and statistical methods. 80% of data today is „dark“. By 2020, 93% of data will be „dark“.

REPLACING

DRUDGE

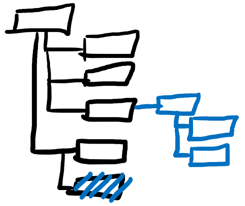
WORK

RESI: The Technical Approach



Technology Details

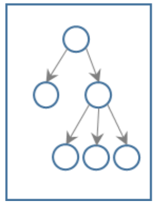
Semantic Processing



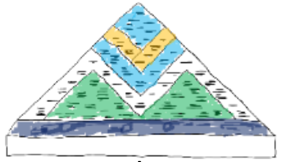
Semantic Model

The Semantic Model is an annotated parse tree, enriched with thematic/semantic role labeling and further semantic information to semantic concepts

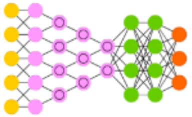
Annotated parse tree



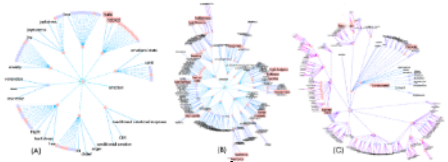
The information layer comprises ontologies, knowledge graphs, lexicons, statistics, and NN to challenge semantics from above deducted model



Ontologies

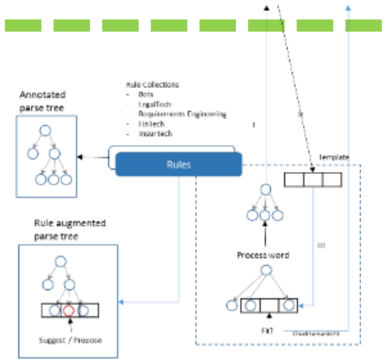


Neural Nets

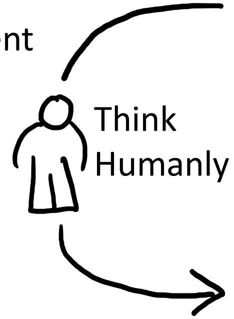


Lexicon/Statistics

The decision layer uses “common sense” to make meaning of the semantic model and augments it. This rule processing is an n-tier approach to solving semantic queries.



Semantic Rule collections to augment
- Bots (Virtual Assistants)
- LegalTech
- Requirements Engineering
- FinTech + Tax + Auditing
- InsurTech
- RetailTech



Why Ontologies?

Ontologies offer world knowledge to a computer system.
They provide semantics and therefore the meaning of a sentence.

RESI Integrated into ProContext's ProcuctManager

Edit requirement "UR-35"

* Type: Requirement for use

Example: " The user must be able to recognise by the system at first sight which order must be worked on as the next. "

* Requirement:

If an appointment is cancelled, the system must display a message.

body

Custom ID / source:

Belongs to keyword: Not selected

If an[1] appointment is cancelled[2], the[3] system[4] must display[5] a[6]message[7].[8]

Unclear determiners

Please specify a determiner for:

1: "an" is an unclear determiner.

3: "the" is an unclear determiner.

6: "a" is an unclear determiner.

Unspecified process words

There seem to be details missing for the following process words:

2: "cancelled" seems to be incompletely specified.

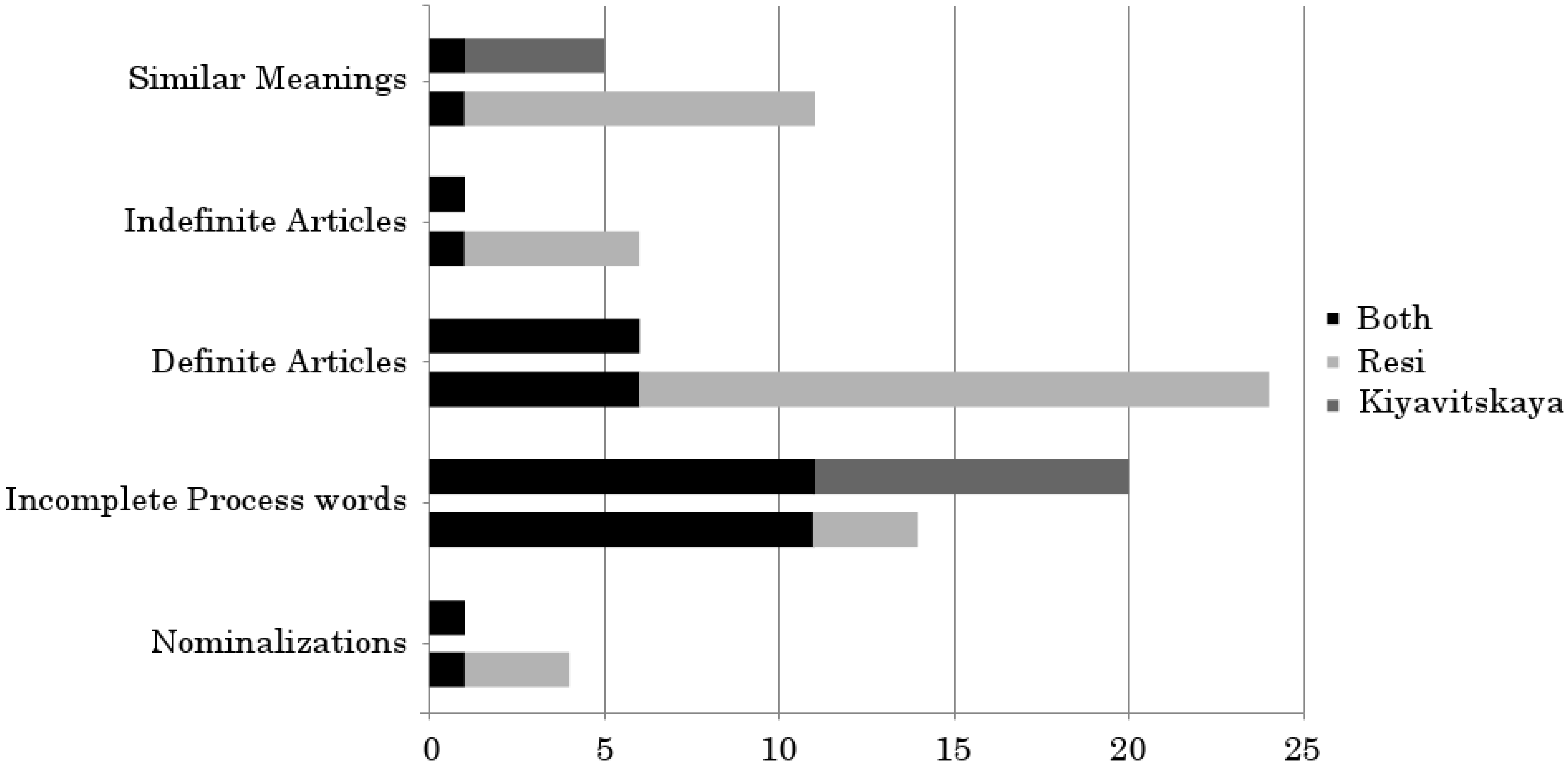
Missing arguments: Performed by, Evaluee-Direct, Purpose in event

5: "display" seems to be incompletely specified.

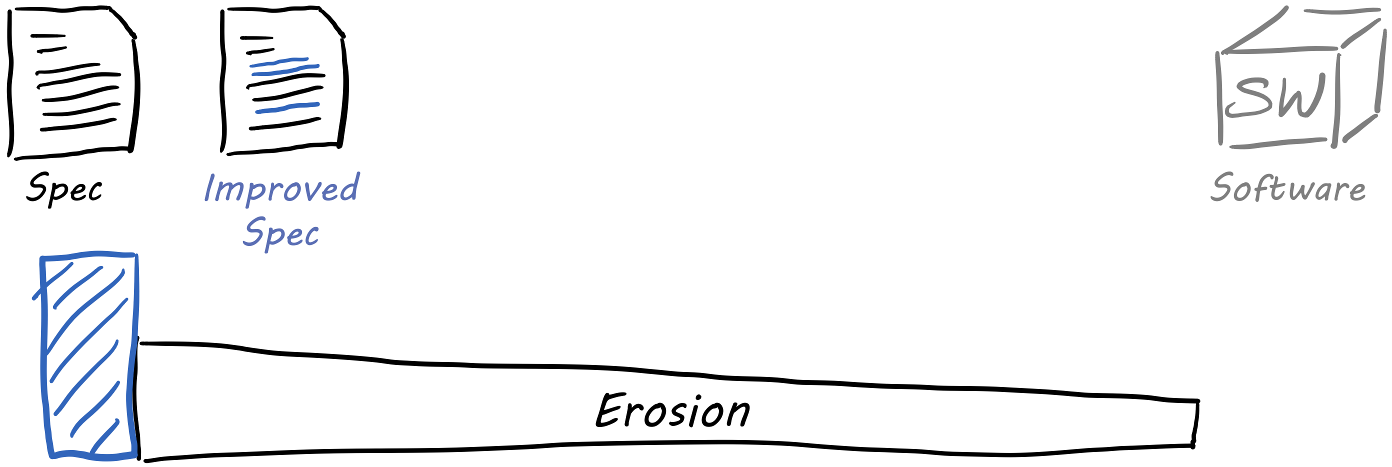
Missing arguments: Sender of info

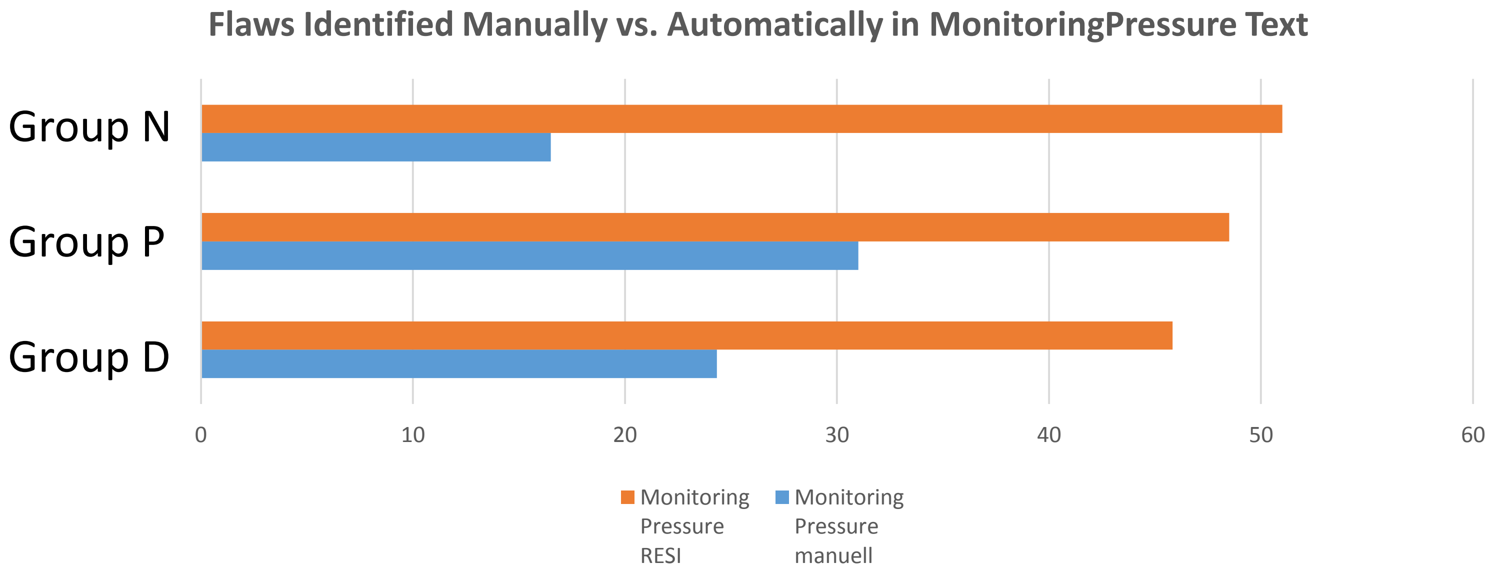
Detected arguments: Instrument-Generic (system)

Evaluation – Results I



Even Non-Professionals Can Improve Specs!





if it’s not working

it better be the customer’s fault

You can observe a lot by watching.

Yogi Berra

Threats to Validity / Issues / Problems

- Internal Validity: case studies in research show the validity of the approach in known use-case scenarios and specifications
- External validity: first results come from demonstrators, but we need to gather more data to being able to make a real statement
- No answer to the question: When can we ignore flaws, when are they important?
  Integrating into everyday workflows (IBM Doors, Jira, PTC, Polarion)
- Biggest problem:
  - finding real-life requirements
  - finding companies that are willing to share their experience in RE openly

, theory

“

IN THEORY,
THEORY AND PRACTICE
ARE THE SAME.
IN PRACTICE,
THEY ARE NOT

“

– Albert Einstein –

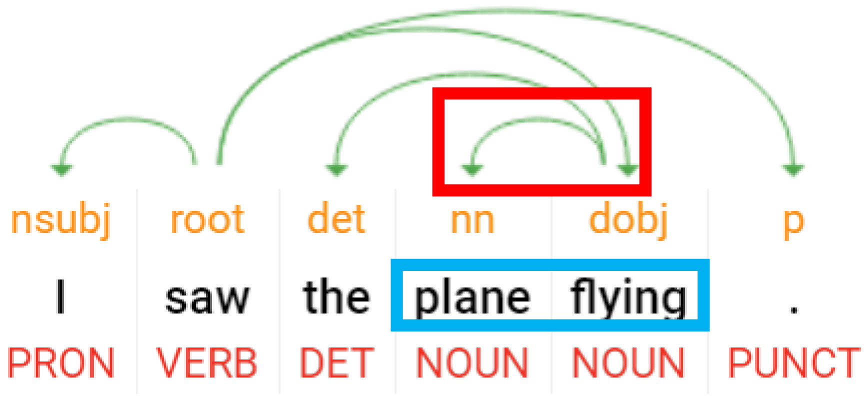
我看到飞机飞行。

Wǒ kàn dào fēijī fēixíng.

我看到飞机飞行。

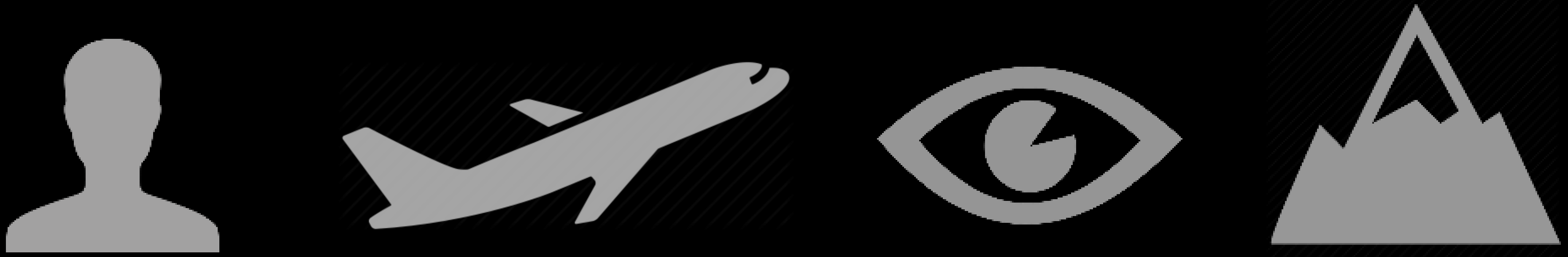
I saw the plane flying.

How Google et al. Work

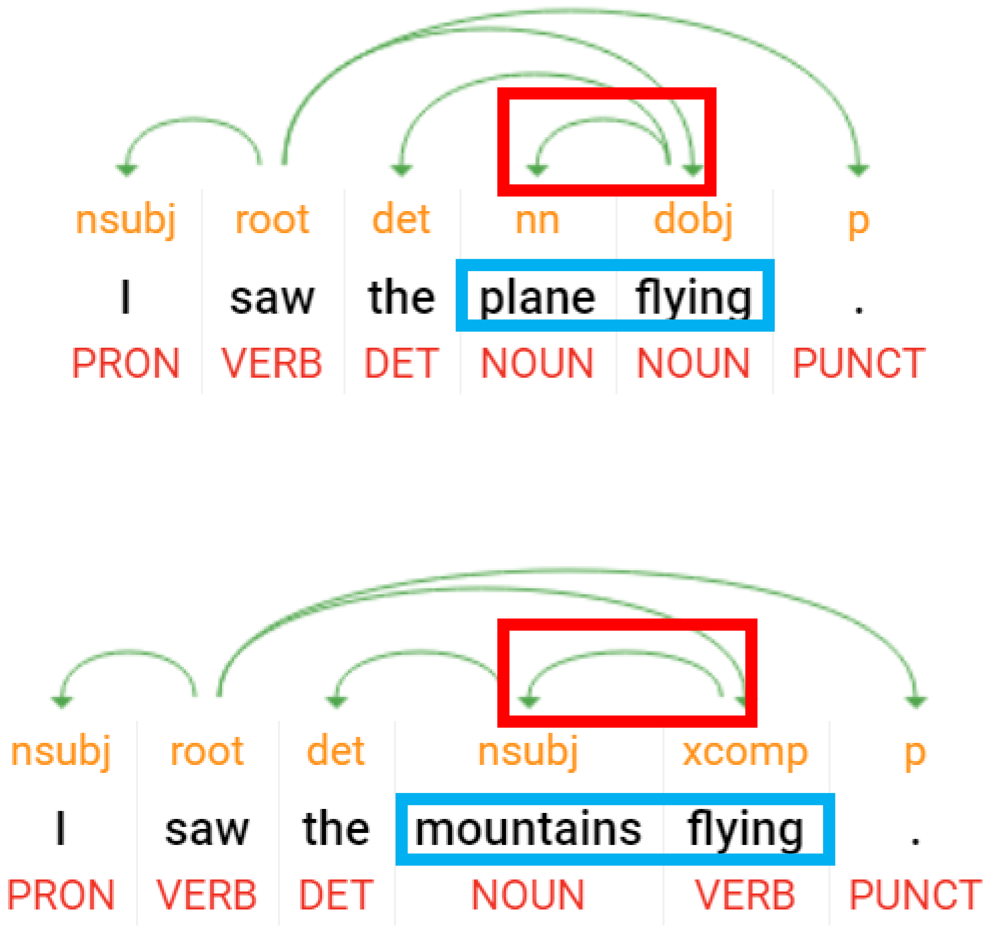


mountains

I saw the ~~plane~~ flying.



How Google et al. Work



Three Main Approaches to AI

Statistics

- Better for non-complex relationships in data
- Can rate results with confidence
- Deals with uncertainties
- Fast for not-so-complicated systems
- Expensive training
- Parametric model requires statistical knowledge
- Error prone in parameter estimation

Machine Learning / Deep Learning

- Ability to detect complex nonlinear relationships between dependent and independent variables
- Works great for perception already today
- Easily implemented (i.e. in multicore processors or systems with GPUs)
- Needs Supervised Learning (which limits the machine power through mankind)
- Does not work with low sample size
- Black box (rather difficult to interpret and to explain/to rebuild)
- Retraining is hard (retraining for backpropagation is problematic)
- Can‘t do a priori

Semantics

- Understands the meaning of natural language
- Complements statistical and ML approaches
- Can justify
- Works a priori
- Needs (linguistic) experience
- Computing power
- Quality depends on ontology (semantic knowledge database)
- Not a one-stop shop (complements other approaches)

A Little Brain Teaser

Killing

BAD

Killing Bacteria

GOOD

Failing to Kill Bacteria

BAD

Never Failing to Kill Bacteria

GOOD

Understanding the meaning of text continues to require knowledge of who produced it and who it is aimed at.

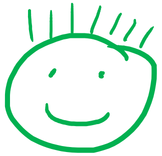
DeNom

Special Treatment for Nominalizations

Nominalizations: Problematic yet often overlooked

- Nominalizations can lead to serious problems during development

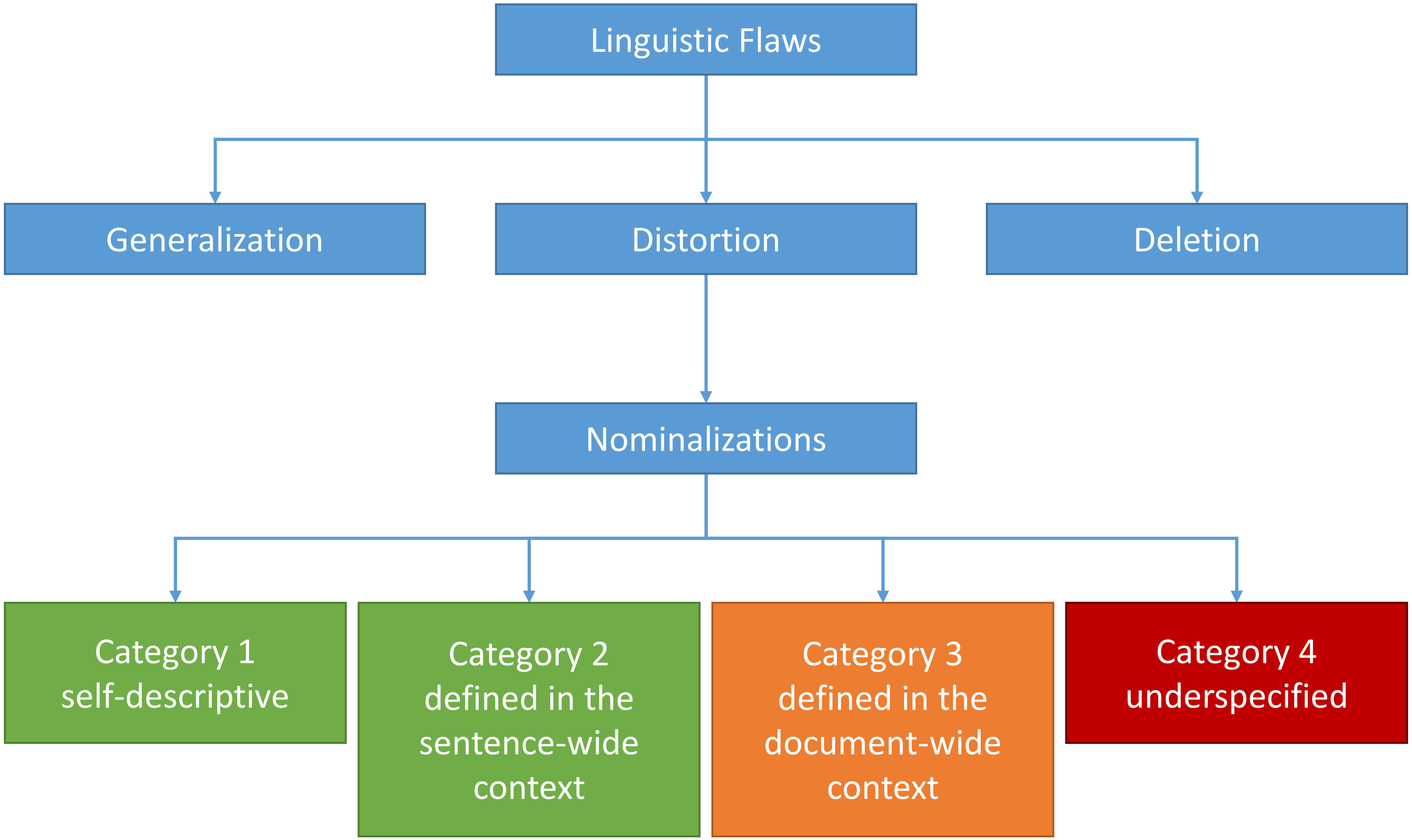
- A requirements engineer’s writing rule:
  Though shall not use nominalizations!



- Inspection rule: Find and eliminate all nominalizations!
  - Can be identified automatically using RESI [RESI]
  - RESI is picky and produces many warnings
  - Effort to high for real-world scenarios [RESI@Automotive]

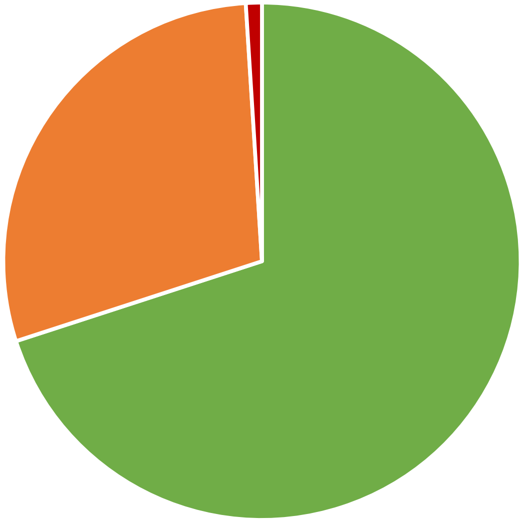
[DeNom]

Not All Nominalizations are Problematic



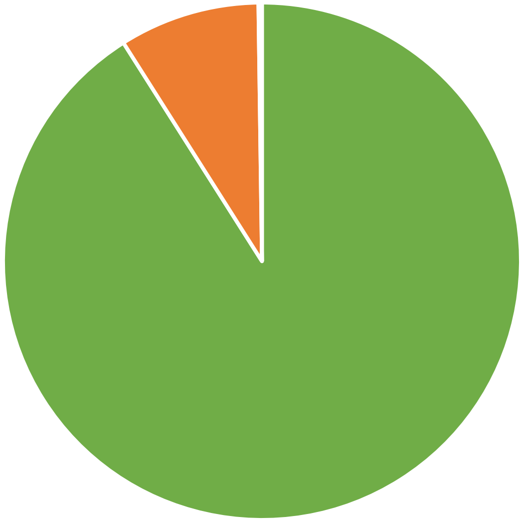
[DeNom]

Fun Fact: Most Nominalizations are OK!



Fully Manual Study:
5 specifications
>40,000 words
356 nominalizations in total

0 % Category 1 (!)
70 % Category 2
29 % Category 3
1 % Category 4

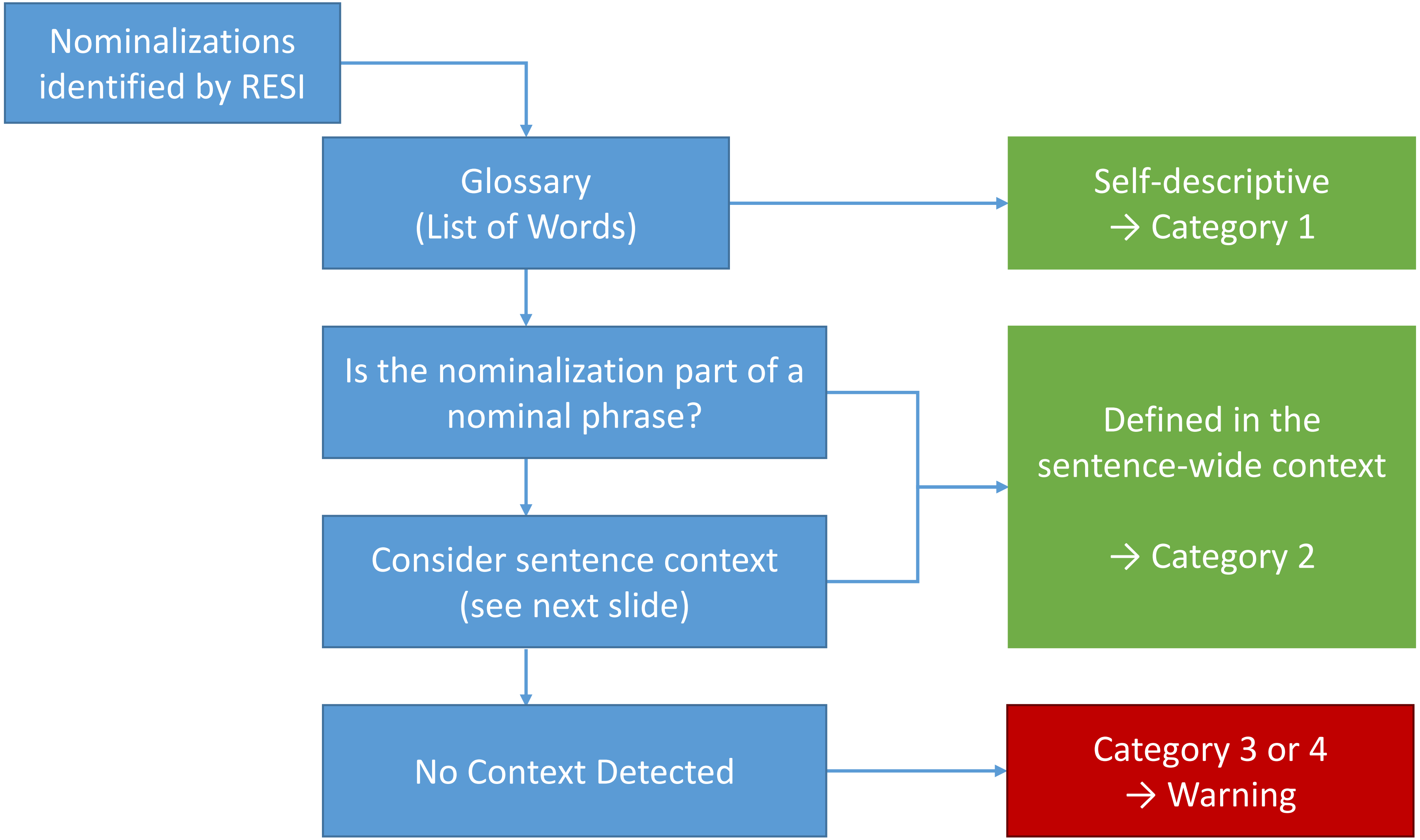


Half-automated Study:
6 specifications
>33,000 words
499 nominalizations detected

0 % Category 1 (!)
83 % Category 2
8 % Category 3
0.2 % Category 4
+ some false positives

[DeNom]

Automatic Categorization



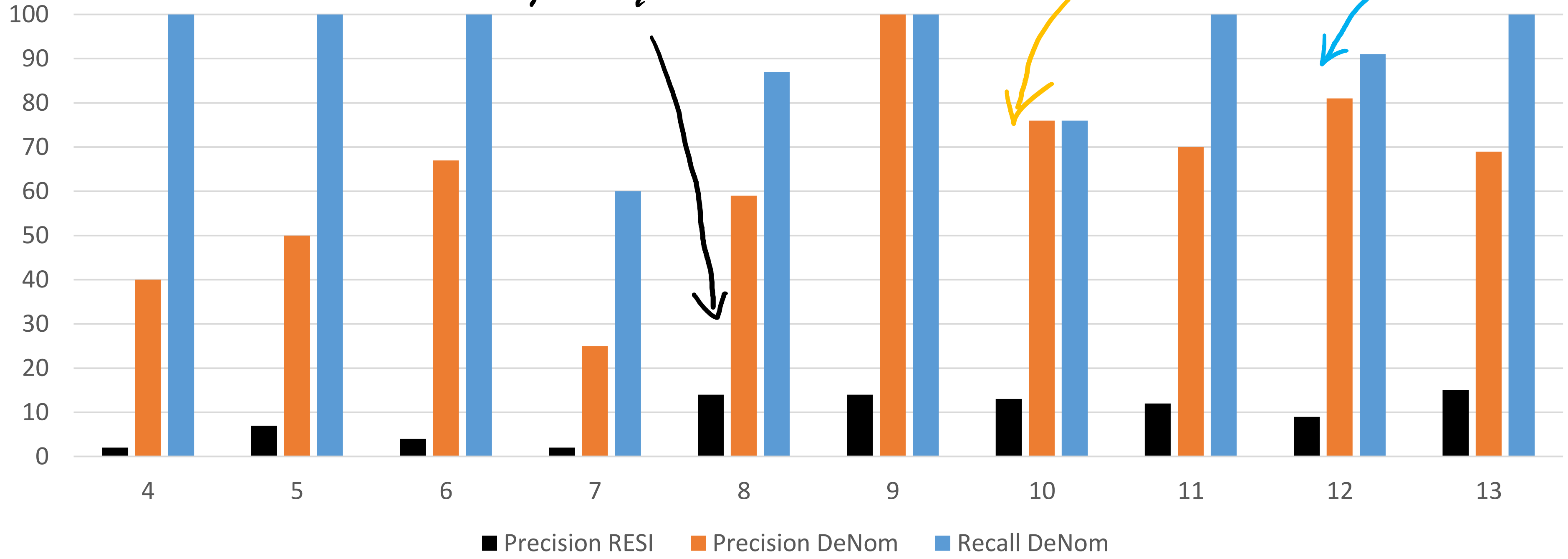
[DeNom]

Evaluation

RESi
ø 8%

DeNom
ø 65%

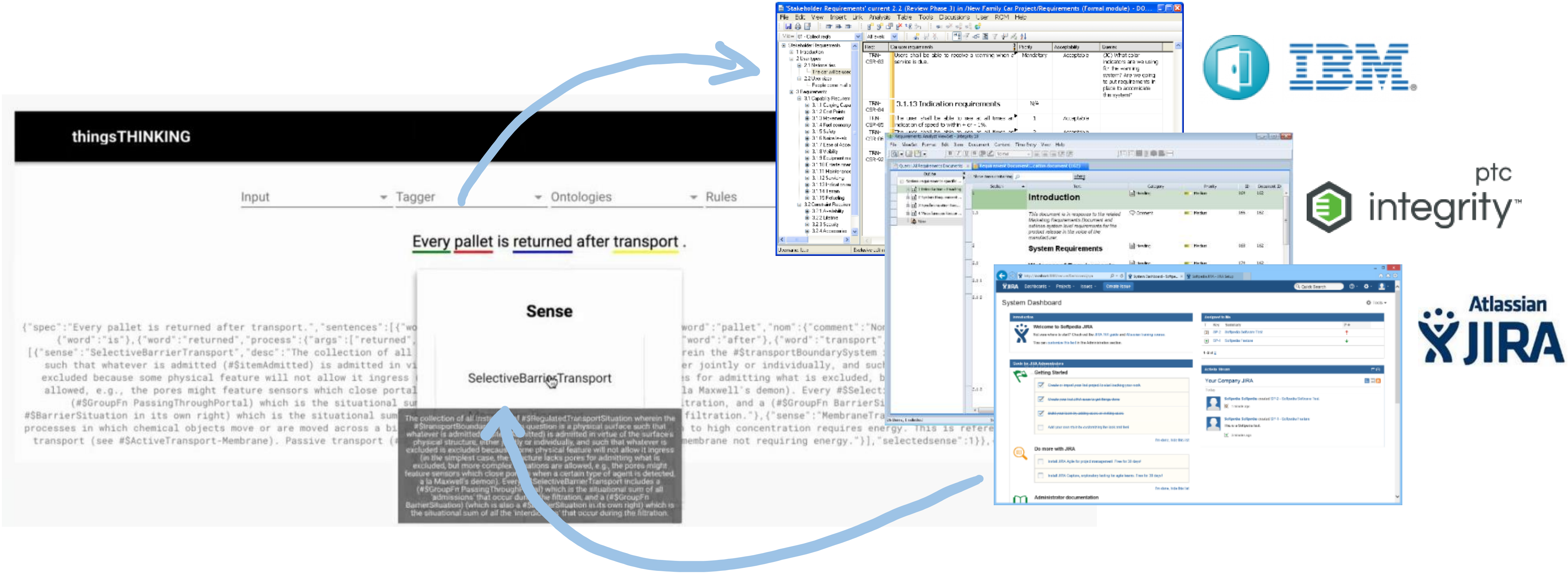
DeNom Recall
ø 88%



- 10 specifications, >59,000 words
- 1,136 nominalizations
  - only 84 of them are problematic
  - DeNom shows 129 warnings
- Precision of RESI on average: 8% (F_1=15%)
- Precision of DeNom on average: 65% (with a recall of 88%, F_1=75%)

[DeNom]

Product: Interactively Disambiguate Requirements Specifications



References

- [Brooks 1987] Frederick P. Brooks, Jr. No Silver Bullet: Essence and Accidents of Software Engineering, Computer, Vol. 20, No. 4 (April 1987) pp. 10-19.
- [MFI04] Mich, L.; Franch, M. & Inverardi, P. N.: Market research for requirements analysis using linguistic tools. Requirements Engineering, Springer, 2004, 9, 40-56
- [Rupp] C. Rupp, Requirements-Engineering und -Management. Professionelle, iterative Anforderungsanalyse für die Praxis, 5., aktualisierte und erw. Aufl. Hanser Fachbuchverlag, 2009.
- [RESI] S. J. Körner and T. Brumm, “Natural Language Specification Improvement with Ontologies,” International Journal of Semantic Computing (IJSC), vol. 03, no. 04, pp. 445–470, 2010.
- [PassiveVoiceMyth] J. Krisch and F. Houdek, “The Myth of Bad Passive Voice and Weak Words -- An Empirical Investigation in the Automotive Industry,” in RE’15:23rd IEEE International Requirements Engineering Conference, Ottawa, Ontario, Canada, 2015.
- [RESI@Automotive] S. J. Körner, M. Landhäußer, and W. F. Tichy, “Transferring Research Into the Real World: How to Improve RE with AI in the Automotive Industry,” in 1st International Workshop on Artificial Intelligence for Requirements Engineering, 2014.
- [DeNom] Landhäußer, Mathias, Sven J. Körner, Jan Keim, Walter F. Tichy, and Jennifer Krisch. “DeNom: A Tool to Find Problematic Nominalizations Using NLP.” In 2nd International Workshop on Artificial Intelligence for Requirements Engineering. Ottawa, Canada, 2015.

Contact Us

thingsTHINKING GmbH

Eisenhutstraße 14
76703 Kraichtal, Germany

Dr. Mathias Landhäußer
+49 175 1970 705
mathias@thingsTHINKING.net

www.thingsTHINKING.net