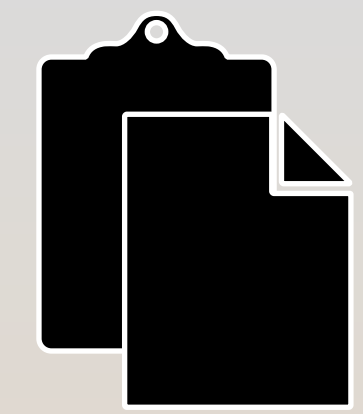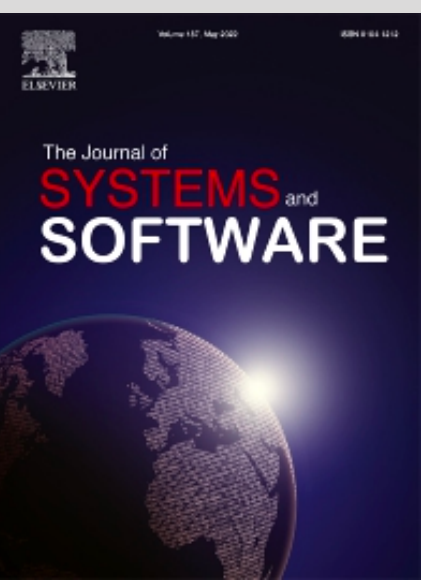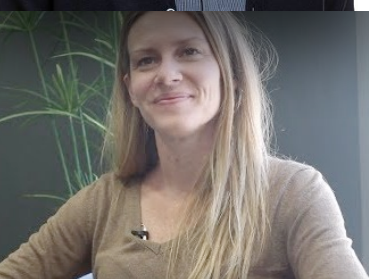ESEM 2023

Journal First

A Decade of Comment Quality Assessment: A Systematic Literature Review

Pooja Rani, Arianna Blasi, Natallia Stulova, Sebastiano Panichella, Alessandra Gorla, Oscar Nierstrasz

University of Zurich, Switzerland
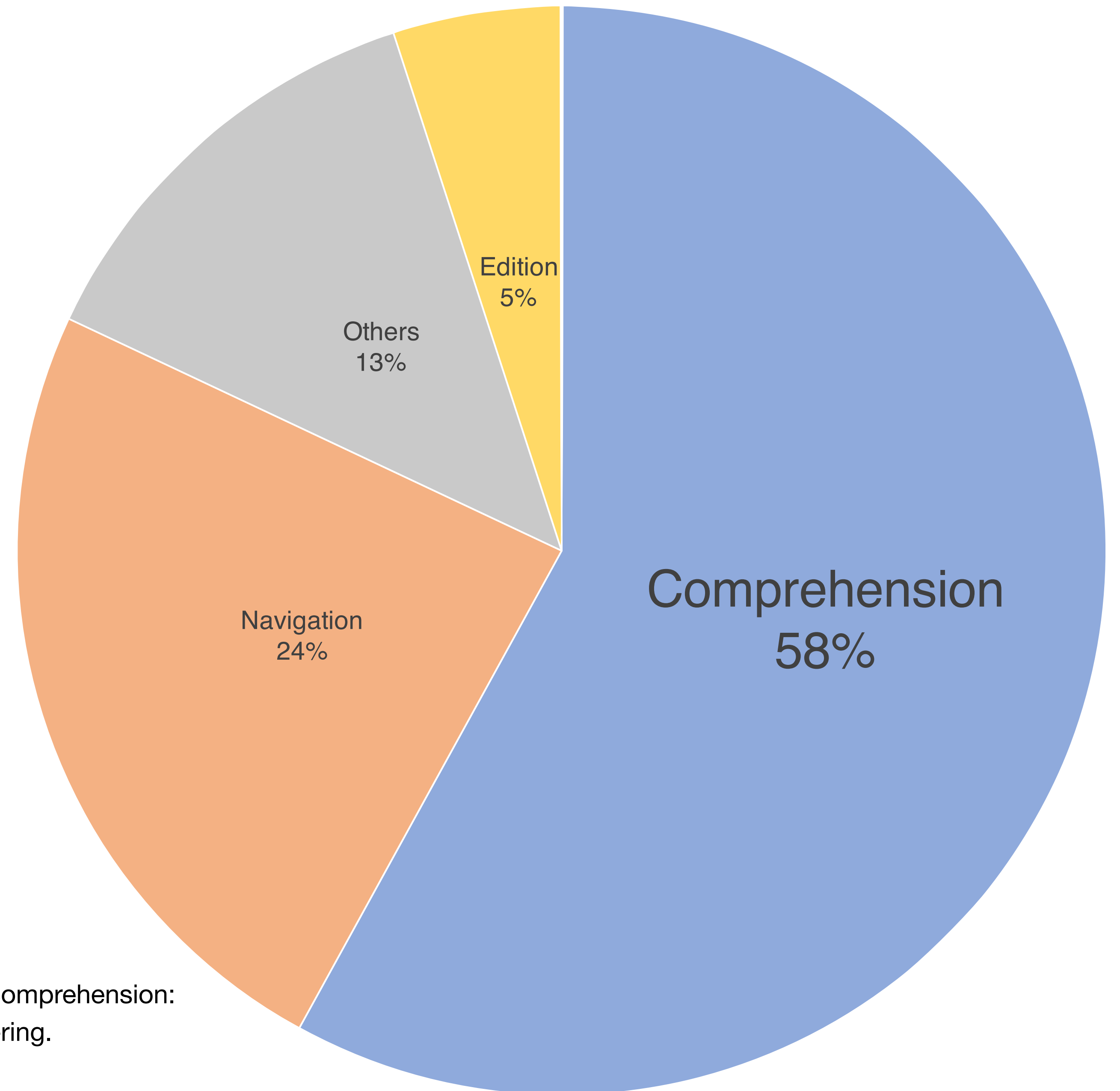
How software developers spend their time?

Xia, X., Bao, L., Lo, D., Xing, Z., Hassan, A. E., & Li, S. (2017). Measuring program comprehension: A large-scale field study with professionals. IEEE Transactions on Software Engineering.

Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {
    …
}
```

Comprehension
58%

Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {
    …
}
```

Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *   Window win = new Window(parent);
 *   win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {
    …
}
```

Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {

    …

}
```

Trustworthy form of documentation

- McMillan et al. 2010

Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {

    …

}
```

Trustworthy form of documentation

- McMillan et al. 2010

High-quality comments support developers in various activities

- Dekel et al. 2009

Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {

    …
}
```

enough?

correct?

consistent?

Other quality attributes?

Is this a high-quality comment?

Quality is a multi-dimensional concept
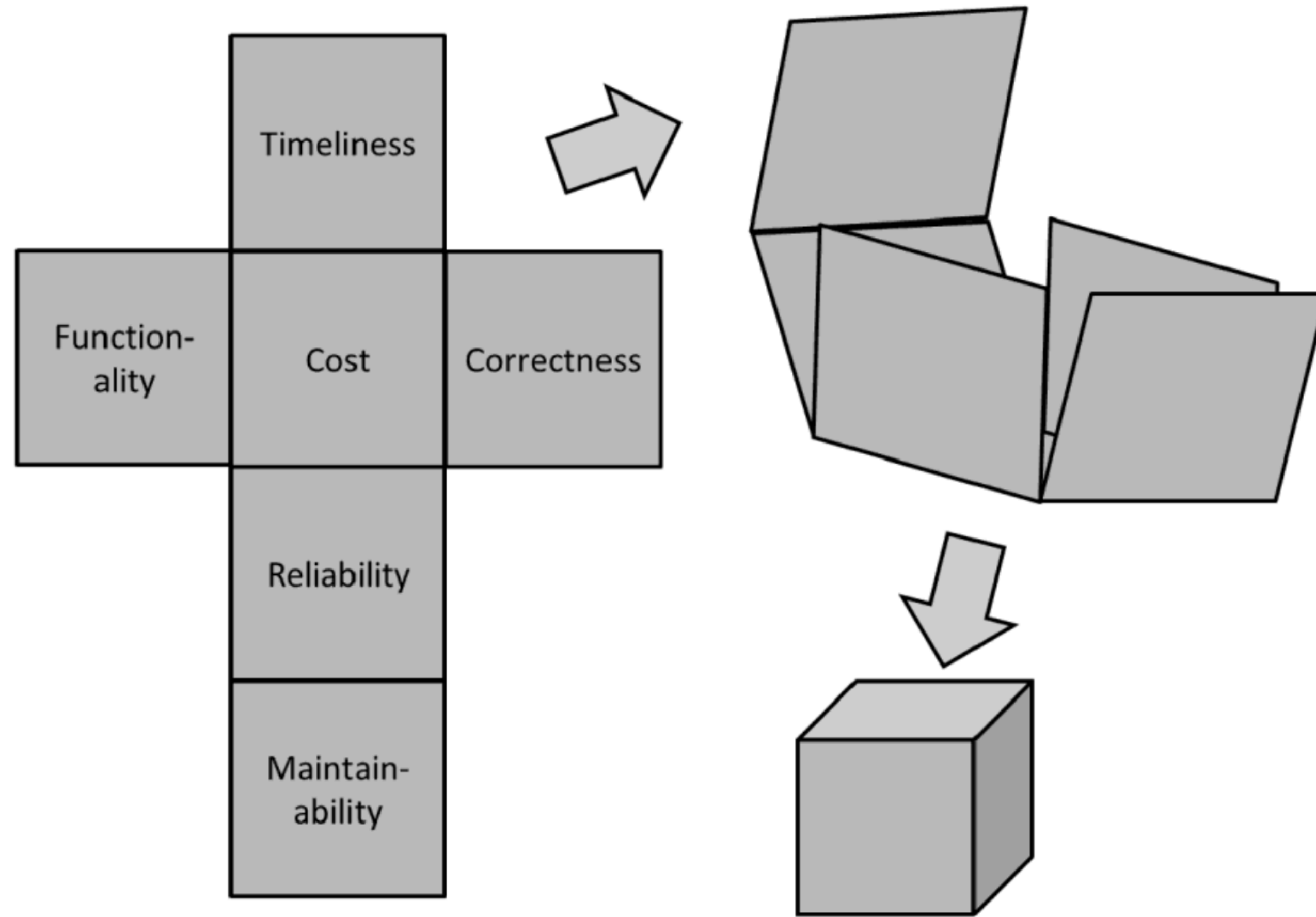
Gillies et al. 2011

How do researchers measure comment quality?

How do researchers measure comment quality?

Systematic literature review

10 years timeline **(2011-2020)**

195 software engineering venues

332 proceedings

2353 candidate papers

47 relevant papers

Quality attributes

21 quality attributes

Quality attributes

21 quality attributes

Quality attributes

23 | **Consistency**

12 | **Completeness**

Accuracy

Readability

Up-to-date-ness

Content relevance

Maintainability

Spelling and grammar
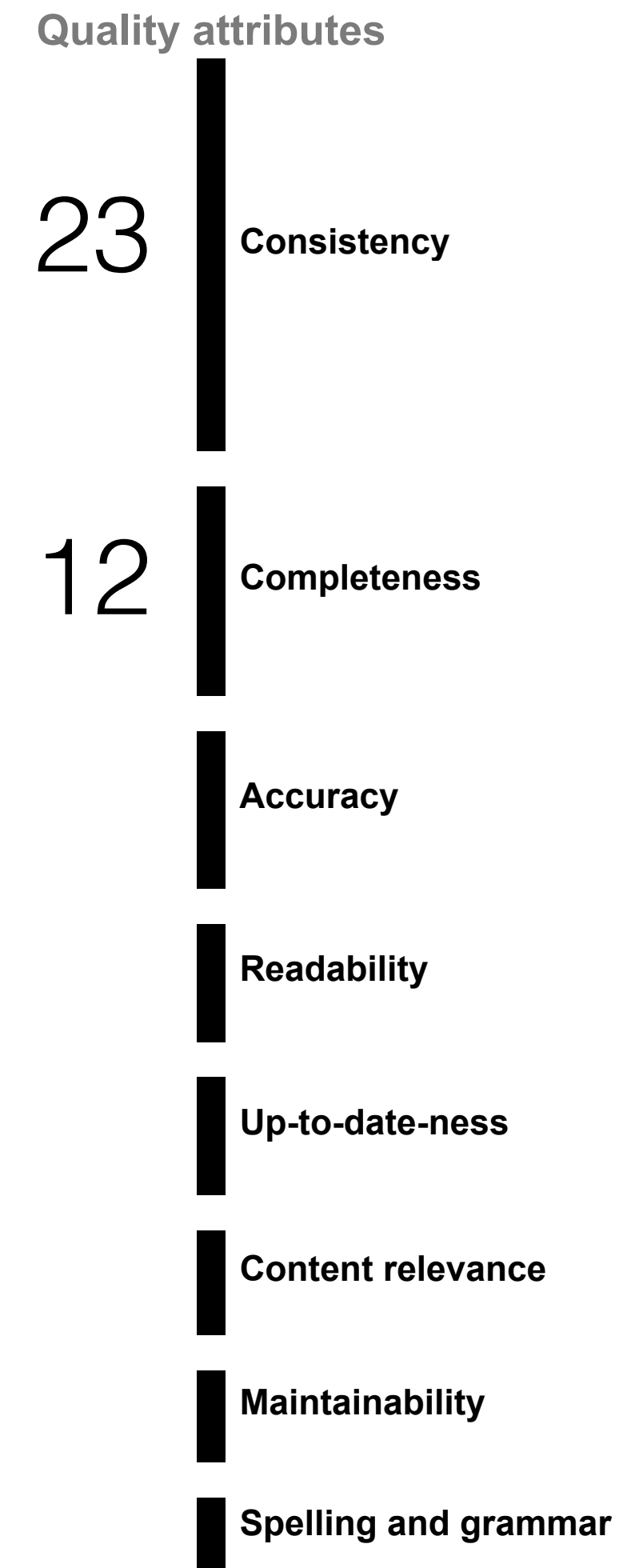
Quality attributes

Some are frequently considered

Quality attributes

- Consistency
- Completeness
- Accuracy
- Readability
- Up-to-date-ness
- Content relevance
- Maintainability
- Spelling and grammar

Quality attributes

Consistency

Completeness

Accuracy

Readability

Up-to-date-ness

Content relevance

Maintainability

Spelling and grammar

Conciseness

Usability

Correctness

Traceability

Accessibility

Coherence

Format

Information organization

Understandability

Documentation technoogy

Internatioalization

Author-related

Some are rarely considered

15

Techniques

Consistency

Completeness

Accuracy

Readability

Up-to-date-ness

Content relevance

Maintainability

Spelling and grammar

Conciseness

Usability

Correctness

Traceability

Accessibility

Coherence

Format

Information organization

Understandability

Documentation technoogy

Internatioalization

Author-related

16

Techniques

Consistency

Completeness

Accuracy

Readability

Up-to-date-ness

Content relevance

Maintainability

Spelling and grammar

Conciseness

Usability

Correctness

Traceability

Accessibility

Coherence

Format

Information organization

Understandability

Documentation technoogy

Internatioalization

Author-related

Manual assessment

ML-based

Heuristic-based

Metric-based

Static analysis

NLP

DNN-based

First order logic

Empirically

fNIRS

17

Techniques

Quality attributes:
- Consistency
- Completeness
- Accuracy
- Readability
- Up-to-date-ness
- Content relevance
- Maintainability
- Spelling and grammar
- Conciseness
- Usability
- Correctness
- Traceability
- Accessibility
- Coherence
- Format
- Information organization
- Understandability
- Documentation technoogy
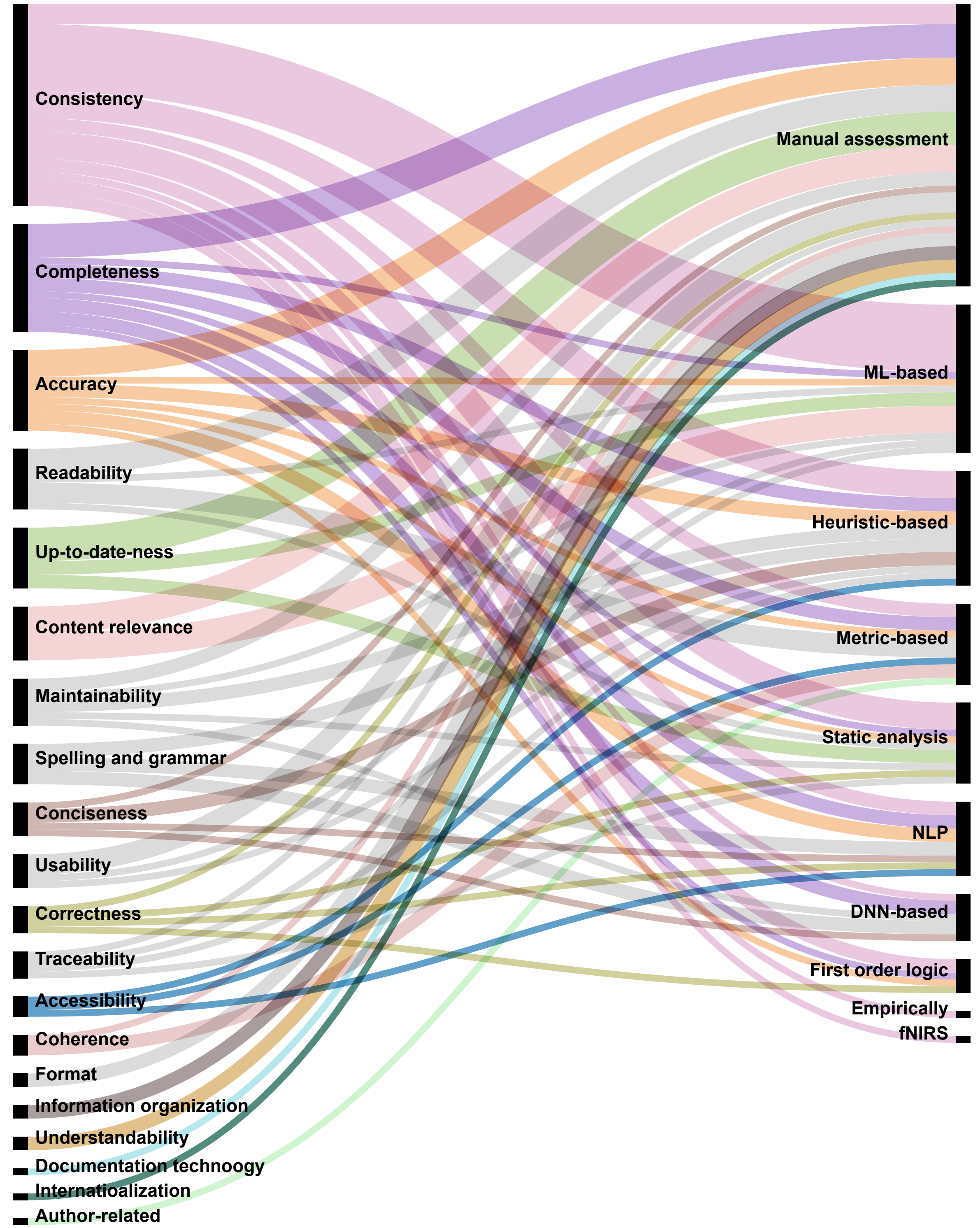- Internatioalization
- Author-related

Techniques:
- Manual assessment
- ML-based
- Heuristic-based
- Metric-based
- Static analysis
- NLP
- DNN-based
- First order logic
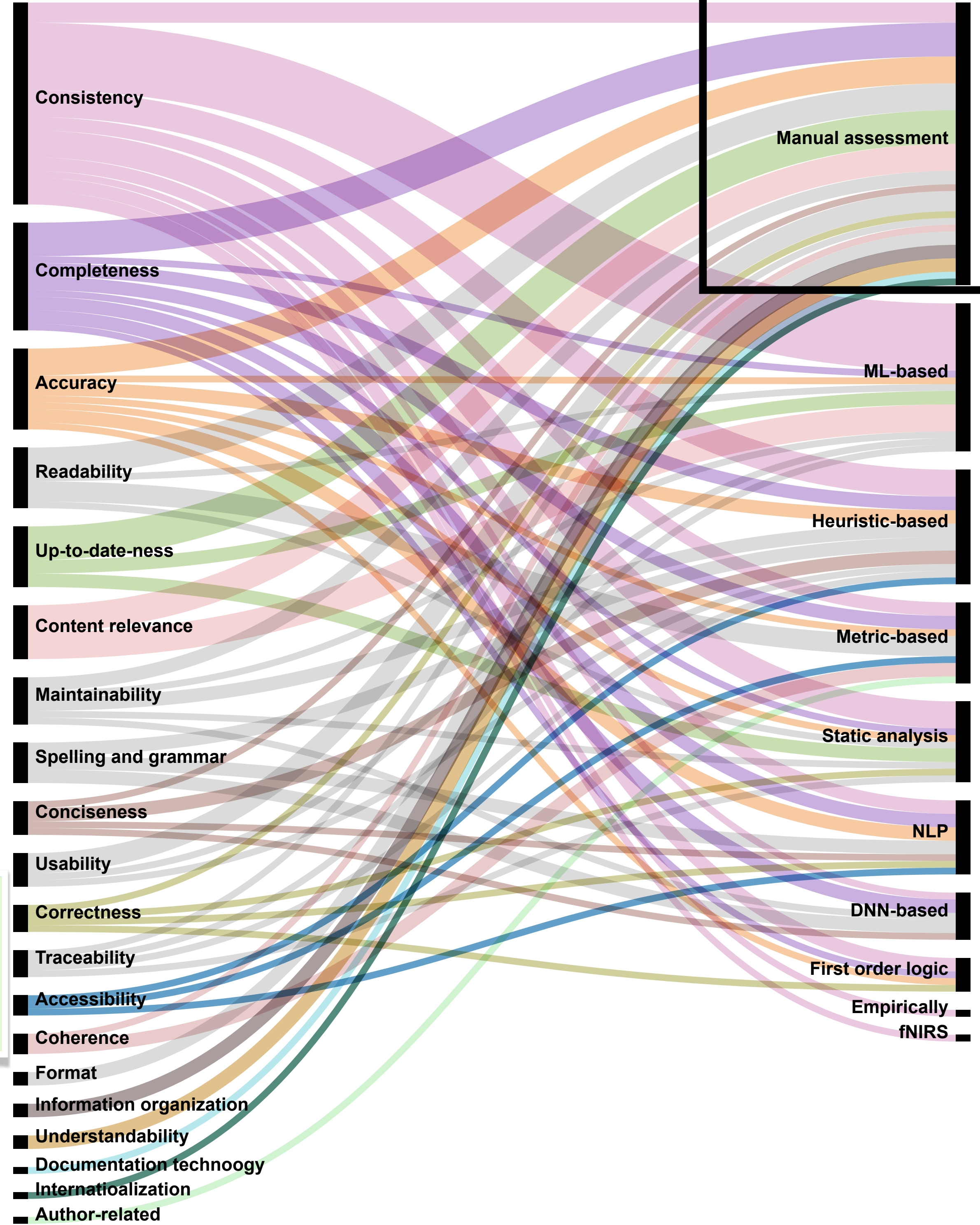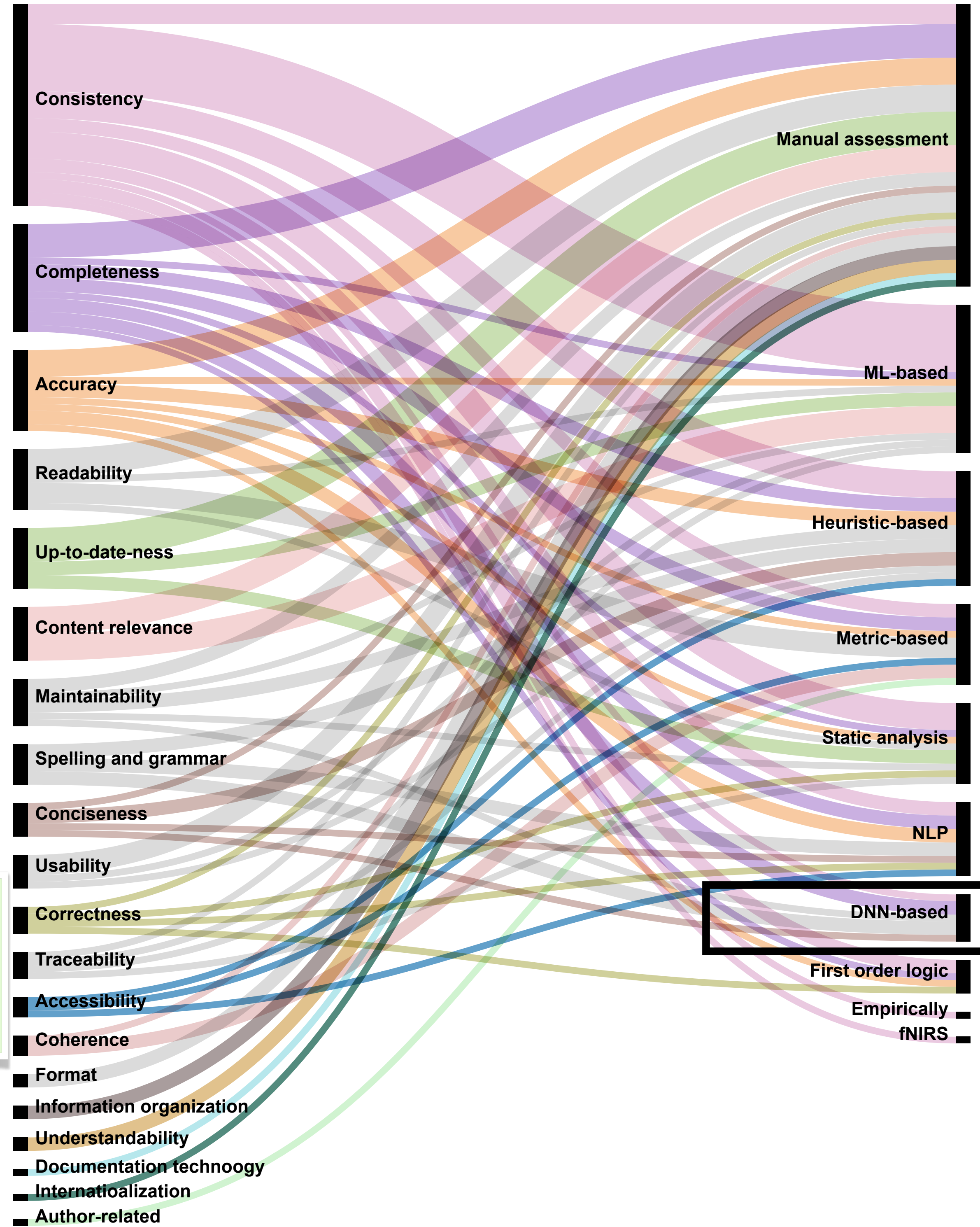- Empirically
- fNIRS

18

Techniques

Manual assessment is still the most frequent technique to measure quality attributes
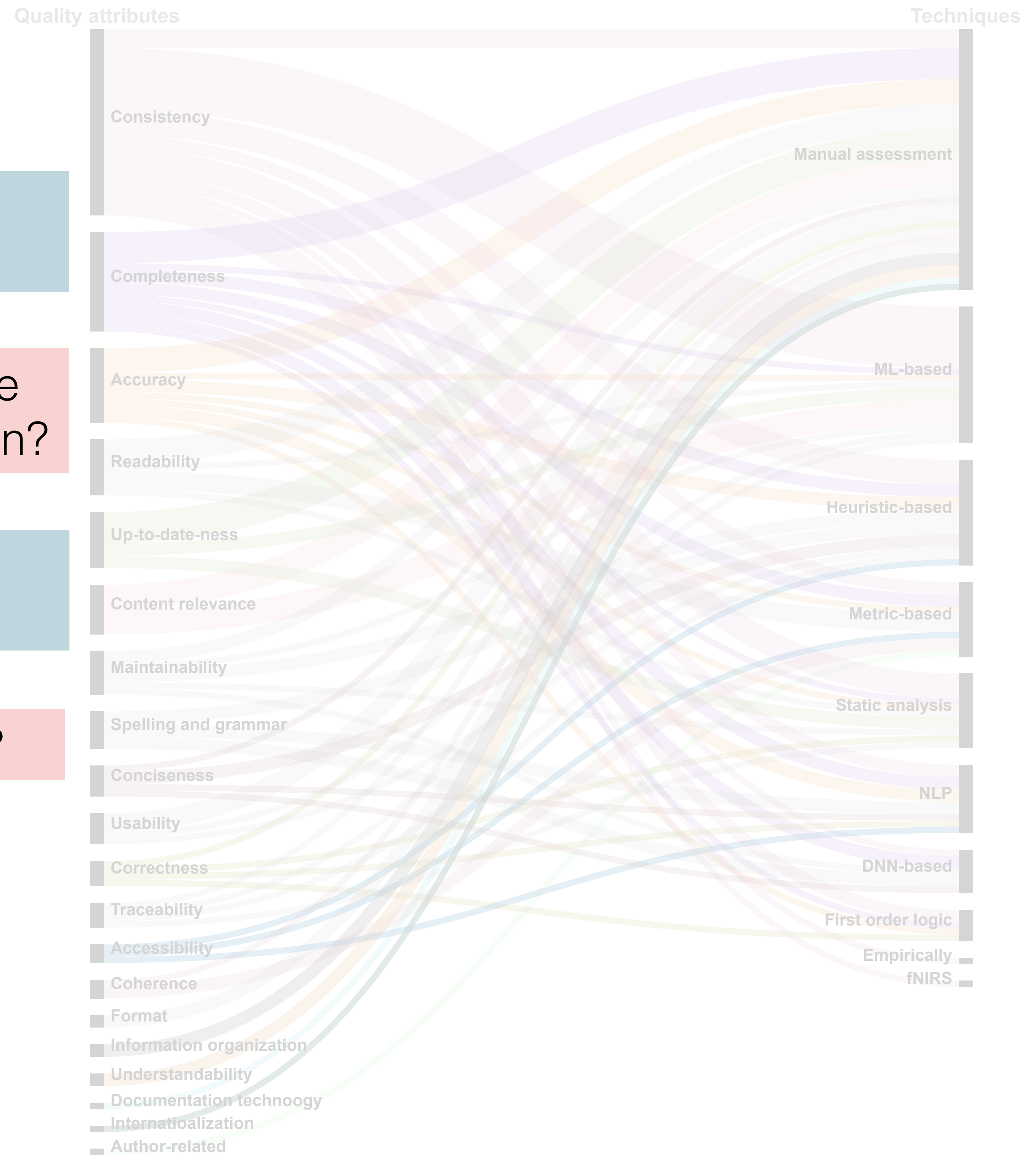
Techniques

Deep learning-based techniques have not been extensively explored for comment analysis

20

Which of these quality attributes are important to which developers?

Do code reviewers care more for accurate comments than complete comments? When?

How to automatically measure these quality attributes?

Can LLMs help writing better comments?

Quality attributes

Consistency

Completeness

Accuracy

Readability

Up-to-date-ness

Content relevance

Maintainability

Spelling and grammar

Conciseness

Usability

Correctness

Traceability

Accessibility

Coherence

Format

Information organization

Understandability

Documentation technoogy

Internatioalization

Author-related

Techniques

Manual assessment

ML-based

Heuristic-based

Metric-based

Static analysis

NLP

DNN-based

First order logic

Empirically

fNIRS

21

"Documentation: castor oil of programming"
Weinberg 1975
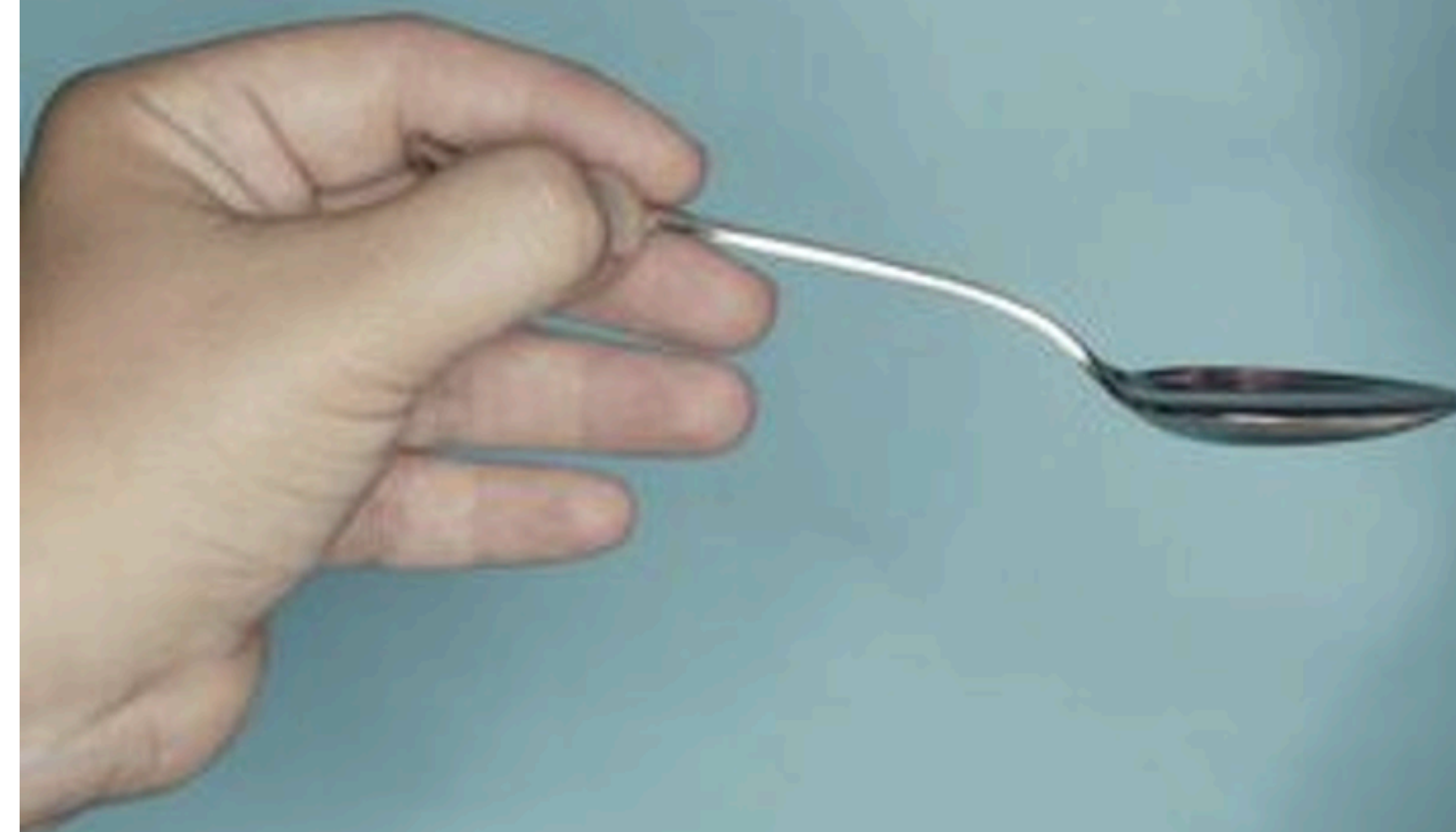
"Incomplete or confusing documentation is the top pain point of programmers", GitHub Open Source Developer Survey 2017

Is comments worth the efforts?

Future work

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {
    …
}
```

Information

Summary

Usage

Pointer

Quality attributes

Accessibility

Readability

Author-related

Correctness

Completeness

Consistency

Task   Design   Implementation   Maintenance

Audience   User   Developer

23

A Decade of Comment Quality Assessment: A Systematic Literature Review

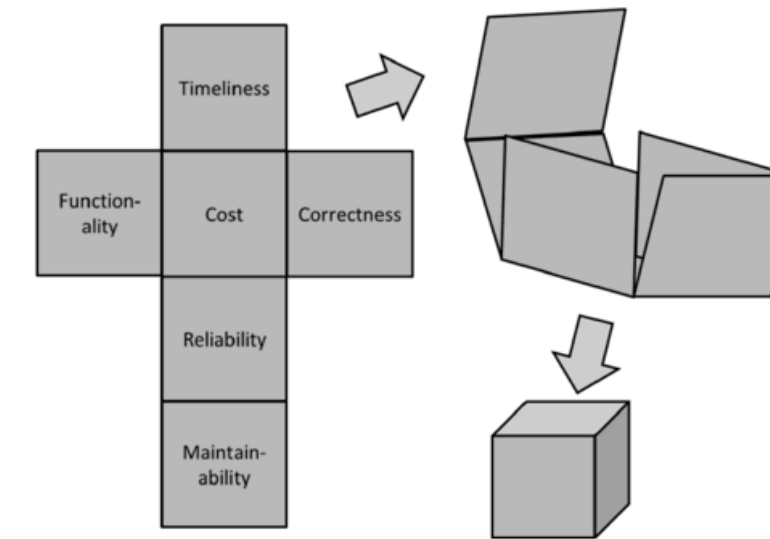Code comments

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {
    …
}
```

Trustworthy form of documentation

High-quality comments support developers
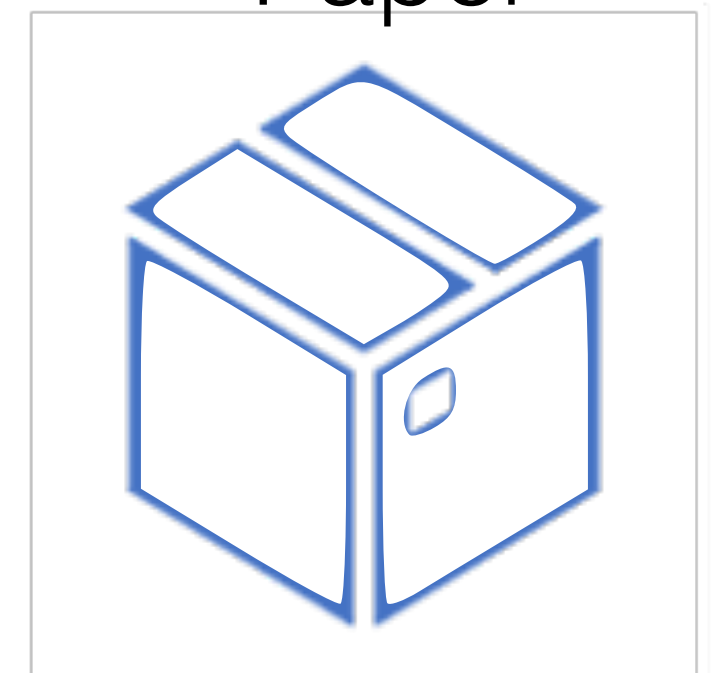
7

Quality is a multi-dimensional concept

Gillies et al. 2011

7

Techniques
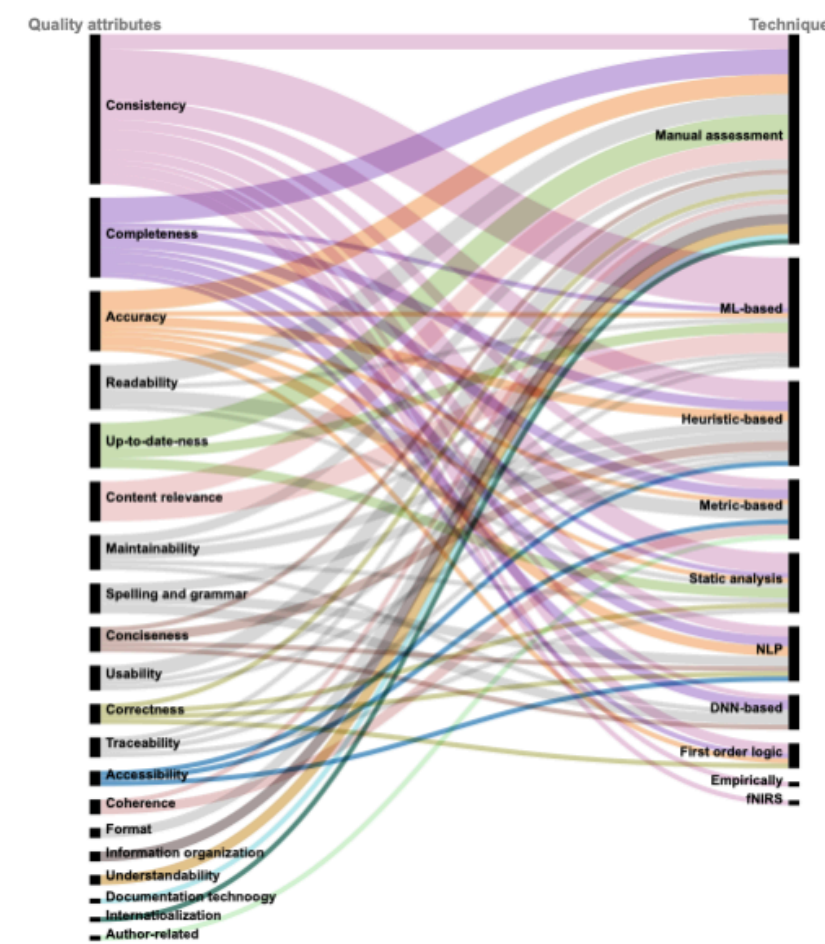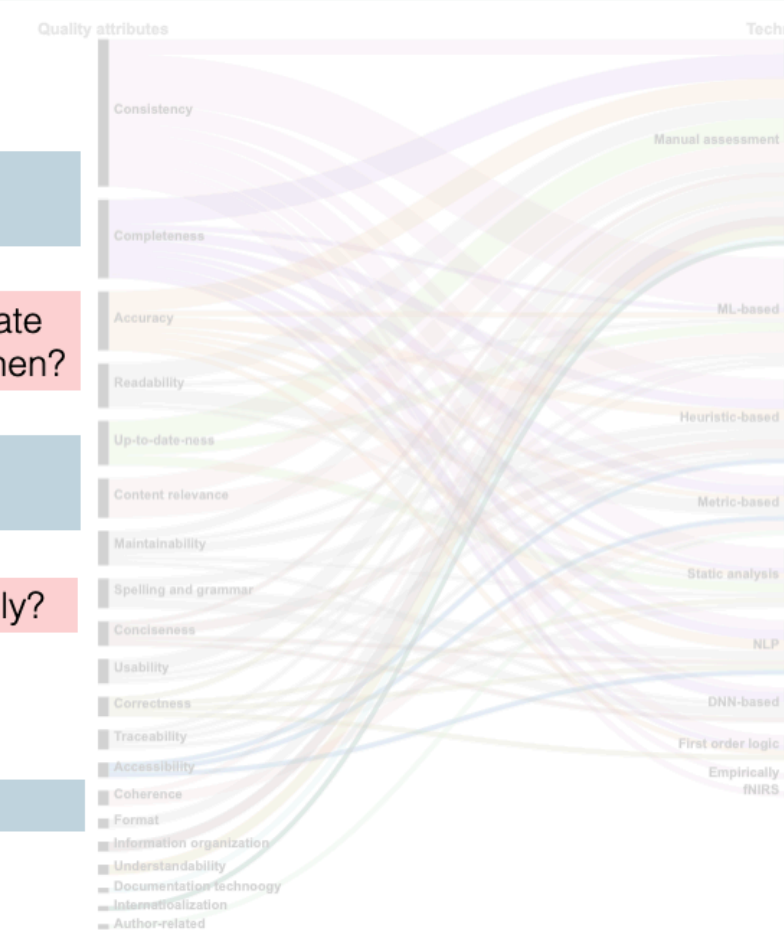
39

Which of these quality attributes are important to which developers?

Do code reviewers care more for accurate comments than complete comments? When?

How to automatically measure these quality attributes?

Can we measure readability automatically?

Is comments worth the efforts?

21

Paper

Replication: https://zenodo.org/record/4729054

pooja-rani-01985b68/

24

More Curious?

Comments of **multi-languages**

```
/**
* A class representing a window on the screen.
*
* For example:
* <pre>
* Window win = new Window(parent);
* win.show();
* </pre>
*
* @author Sami Shaio
* @version 1.13, 06/0
* @see java.awt.BaseW
*/

class Window extends
..
}
```

```
class OneHotCategorical(Distribution):
    r"""
    Creates a one-hot categorical distribution parameterized by :attr:`probs` or
    :attr:`logits`.

    Samples are one-hot coded vectors of size ``probs.size(-1)``.

    .. note:: The `probs` argument must be non-negative, finite and have a non-zero sum.

    See also:
    :attr:`pr
    Example::

        >>> m
        >>> m
        tenso
    Args:
        probs
        logit
    """
```

? Comment

```
I represent a message to be scheduled by the WorldState.

For example, you can see me in action with the following example which print 'alarm test' on Transcript
one second after evaluating the code:

Transcript open.
MorphicUIManager currentWorld
        addAlarm: #show:
        withArguments: #('alarm test')
        for: Transcript
        at: (Time millisecondClockValue + 1000).


* Note *
Compared to doing:
[(Delay forMilliseconds: 1000) wait. Transcript show: 'alarm test'] forkAt: Processor activeProcess
priority +1.


the alarm system has several distinctions:
- Runs with the step refresh rate resolution.
- Alarms only run for the active world. (Unless a non-standard scheduler is in use)
- Alarms with the same scheduled time are guaranteed to be executed in the order they were added
```

Smalltalk class comment

Methodology

CORE 2020

1 *195 venues*

2 *26 relevant venues*

3 *332 proceedings*

keyword-based publication filtering

4 *2,043 potentail papers*

filtering by title and abstract

5 *71 candidate papers*

filtering by full text

6 **30 initial papers**

Methodology

CORE 2020

1 *195 venues*

2 *26 relevant venues*

3 *332 proceedings*

keyword-based publication filtering

4 *2,043 potentail papers*

filtering by title and abstract

5 *71 candidate papers*

filtering by full text

6 **30 initial papers**

forward and backward snowball

Methodology

CORE 2020

1 — 195 venues

2 — 26 relevant venues

3 — 332 proceedings

keyword-based publication filtering

4 — 2,043 potentail papers

filtering by title and abstract

5 — 71 candidate papers

filtering by full text

6 — **30 initial papers**

forward and backward snowball

7 — *3,704 unique papers*

keyword-based publication filtering

8 — *311 potentially relevant papers*

filtering by title and abstract

9 — *39 new candidate papers*

filtering by full text

10 — **18 snowballed papers**

11 — **47 papers**
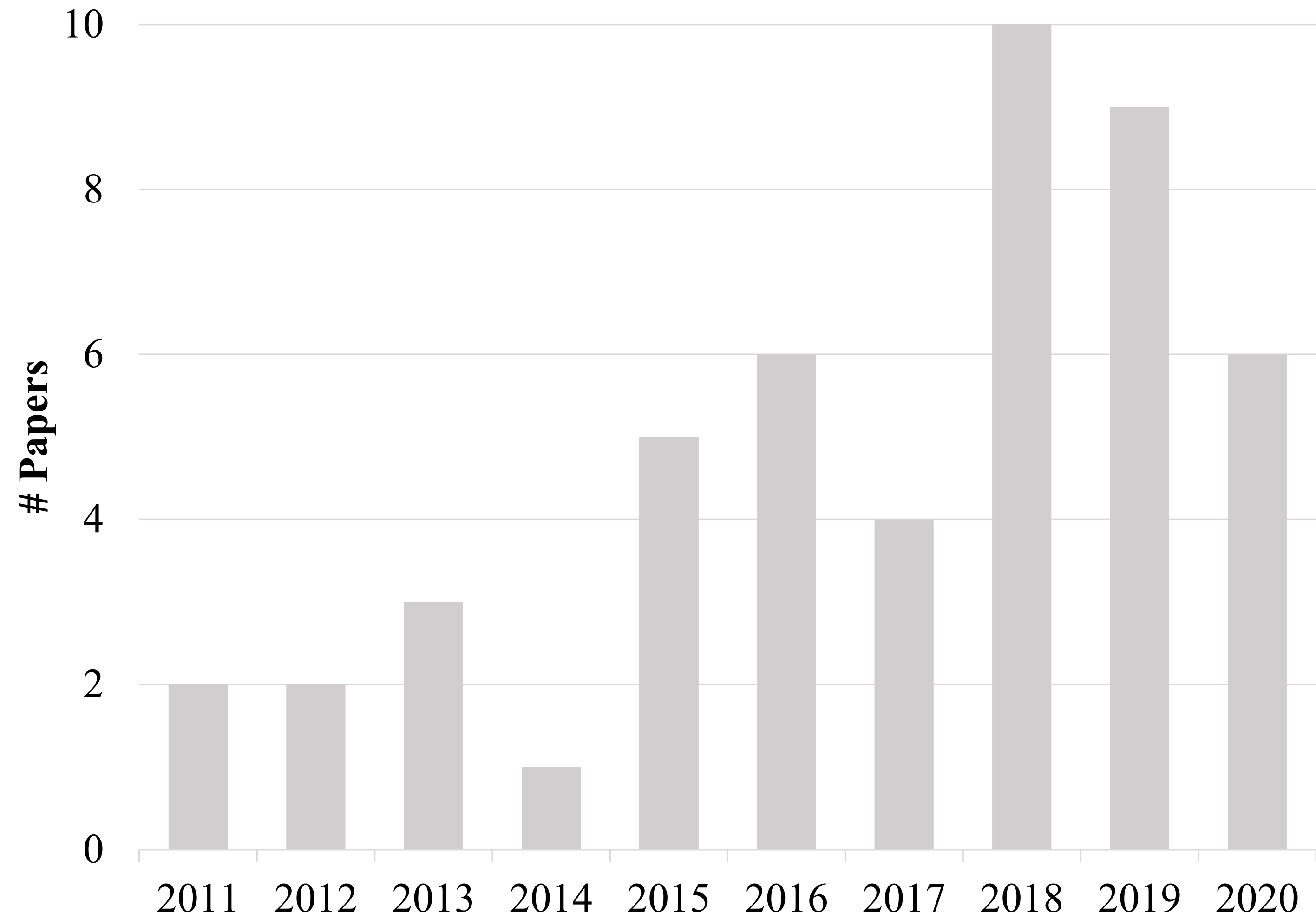
47 papers over years

Dimensions analyzed

Comment types

method comments, inline comments

Quality attributes

consistency, completeness

Techniques

heuristic-based, machine learning-based

Data availability

tool, dataset

# Code comment metrics

```
/**
 * A class representing a window on the screen.
 *
 * For example:
 * <pre>
 *    Window win = new Window(parent);
 *    win.show();
 * </pre>
 *
 * @author Sami Shaio
 * @version 1.13, 06/08/06
 * @see java.awt.BaseWindow
 * @see java.awt.Button
 */
class Window extends BaseWindow {
    …
}
```

Consistency:

Overlap between the terms used in comment and code. Higher value of CIC with higher readability of that code.

Readability:

Flesch reading-ease test
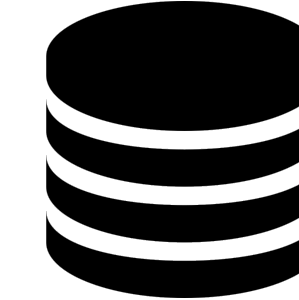
Nearly 50% of the studies still lack in the replicability dimension, as their respective dataset or tool is often not publicly accessible.

Comment types

Code comments

Method comments

API documentation

License comments

Inline comments

TODO Comments

Software documentation

Deprecation Comments

Comment types

Code comments

Method comments

API documentation

License comments

Inline comments

TODO Comments

Software documentation

Deprecation Comments

52% of the studies

Comment types

Code comments

Method comments

API documentation

License comments

Inline comments

TODO Comments

Software documentation

Deprecation Comments

48% of the studies

Comment types