Observations from the essay drafts

Wolfgang Menzel

Department of Informatics
Hamburg University

Overview

- Content and structure of the presentation
  - Why do we need a research question?
  - Why should we avoid repetitions
- Quality of Content
- Linguistic Quality

Why do we need a research question?

Why do we need a research question?

- It helps to structure the essay

Why do we need a research question?

- It helps to structure the essay
  - the question and its motivation $\rightarrow$ the introduction
  - the background information $\rightarrow$ foundations
  - previous findings (lessons learned) $\rightarrow$ previous work
  - own findings $\rightarrow$ experiments
  - pro and counter arguments $\rightarrow$ discussion
  - the answer $\rightarrow$ the conclusions

Why do we need a research question?

- It helps to structure the essay
  - the question and its motivation $\rightarrow$ the introduction
  - the background information $\rightarrow$ foundations
  - previous findings (lessons learned) $\rightarrow$ previous work
  - own findings $\rightarrow$ experiments
  - pro and counter arguments $\rightarrow$ discussion
  - the answer $\rightarrow$ the conclusions
- It helps to streamline the essay

Why do we need a research question?

- It helps to structure the essay
  - the question and its motivation $\rightarrow$ the introduction
  - the background information $\rightarrow$ foundations
  - previous findings (lessons learned) $\rightarrow$ previous work
  - own findings $\rightarrow$ experiments
  - pro and counter arguments $\rightarrow$ discussion
  - the answer $\rightarrow$ the conclusions
- It helps to streamline the essay
  - only issues relevant for answering the question should be included
- It helps to identify open issues

Why should we avoid repetitions?

Why should we avoid repetitions?

- They are wasting precious space.

Why should we avoid repetitions?

- They are wasting precious space.
- They bore the reader.

Why should we avoid repetitions?

- They are wasting precious space.
- They bore the reader.
- They distract they reader, forcing her to check and evaluate differences.

Quality of Content

What are core requirements for a good presentation in CS?

Quality of Content

What are core requirements for a good presentation in CS?

- precision

Quality of Content

What are core requirements for a good presentation in CS?

- precision
- intelligibility

Quality of Content

What are core requirements for a good presentation in CS?

- precision
- intelligibility

Both are equally important.

Quality of Content

What are core requirements for a good presentation in CS?

- precision
- intelligibility

Both are equally important.

How to improve precision?

How to improve precision?

- use precise terminology

How to improve precision?

- use precise terminology
- use the terminology consistently: do not switch it without reason

How to improve precision?

- use precise terminology
- use the terminology consistently: do not switch it without reason
- properly introduce new terms
    - definition
    - explanation
    - illustration
    - discussion

How to improve precision?

- use precise terminology
- use the terminology consistently: do not switch it without reason
- properly introduce new terms
    - definition
    - explanation
    - illustration
    - discussion
- chose terms on the proper level of granularity
    - neither too general, nor too specific

How to improve precision?

- use precise terminology
- use the terminology consistently: do not switch it without reason
- properly introduce new terms
    - definition
    - explanation
    - illustration
    - discussion
- chose terms on the proper level of granularity
    - neither too general, nor too specific
- use formal notation
    - but it needs to be explained to make it intelligible

How to improve intelligibility?

How to improve intelligibility?

- prefer concise explanations

How to improve intelligibility?

- prefer concise explanations
- use pictures
    - they need to be integrated into the presentation
    - by explaning them

How to improve intelligibility?

- prefer concise explanations
- use pictures
  - they need to be integrated into the presentation
  - by explaning them
- do not distract the reader
  - include only relevant content

Linguistic quality

The most frequently annotated problem:

Linguistic quality

The most frequently annotated problem:

the proper use of determiners

Linguistic quality

The most frequently annotated problem:

the proper use of determiners

- (All) Ethiopian languages lack the distinction between definite and indefinite noun phrases.

Linguistic quality

The most frequently annotated problem:

the proper use of determiners

- (All) Ethiopian languages lack the distinction between definite and indefinite noun phrases.
- But the rules are not so complicated

Linguistic quality

The most frequently annotated problem:

the proper use of determiners

- (All) Ethiopian languages lack the distinction between definite and indefinite noun phrases.
- But the rules are not so complicated
- Determiners are not arbitrary, they provide important pragmatic information to the reader

Linguistic quality

The most frequently annotated problem:

the proper use of determiners

- (All) Ethiopian languages lack the distinction between definite and indefinite noun phrases.
- But the rules are not so complicated
- Determiners are not arbitrary, they provide important pragmatic information to the reader
  - namely the distinction between new and old entities in the discourse
  - old ones hev been introduced before or are already well known to (almost) every potential reader

The proper use of determiners (without warranty)

- the default case: a determiner is almost always needed, either a definite or an indefinite one

  The house has been built in 1973.
  The new models are out of order.

  A substantial improvement will be necessary.
  Some members do not support this view.

The proper use of determiners (without warranty)

- the default case: a determiner is almost always needed, either a definite or an indefinite one

    The house has been built in 1973.
    The new models are out of order.

    A substantial improvement will be necessary.
    Some members do not support this view.

- but not in case of names or generic terms (in singular or plural):

    *The John sleeps.

    *The neural networks are powerful.
    but: *The neural network approach performed better.*

    *The machine learning can be used for almost everything.
    but: *The machine learning approach ...*

    ?An artificial intelligence has been developed.
    ?Artificial intelligences are developed by many companies.

The proper use of determiners (without warranty)

- a definite determiner can only be used if the entity has been introduced before or is already familiar to the intended audience

  (We trained a neural network, ...) The network consists of three layers.

  Yesterday, the Ethiopian government announced ...

The proper use of determiners (without warranty)

- a definite determiner can only be used if the entity has been introduced before or is already familiar to the intended audience

    (We trained a neural network, ...) The network consists of three layers.
    Yesterday, the Ethiopian government announced ...

- in all other cases an indefinite determiner has to be used.
    Some improvements can be suggested.
    A neural network might be a better solution.

The proper use of determiners (without warranty)

- a definite determiner can only be used if the entity has been introduced before or is already familiar to the intended audience

  (We trained a neural network, ...) The network consists of three layers.
  Yesterday, the Ethiopian government announced ...

- in all other cases an indefinite determiner has to be used.
  Some improvements can be suggested.
  A neural network might be a better solution.

- Only after a new concept has been introduced, it can be referenced by means of a definite noun phrase or a personal pronoun.

The proper use of determiners (without warranty)

- a definite determiner can only be used if the entity has been introduced before or is already familiar to the intended audience

  (We trained a neural network, ...) The network consists of three layers.
  Yesterday, the Ethiopian government announced ...

- in all other cases an indefinite determiner has to be used.
  Some improvements can be suggested.
  A neural network might be a better solution.

- Only after a new concept has been introduced, it can be referenced by means of a definite noun phrase or a personal pronoun.

- exceptions to the exceptions, only for extreme rare cases, e.g.
  It was not Jane Miller but a Cindy Miller.

Linguistic quality

The second most frequently annotated problem:

Linguistic quality

The second most frequently annotated problem:

the proper use of number

Linguistic quality

The second most frequently annotated problem:

the proper use of number

- Number is not an arbitrary choice
  - it conveys important semantic information

The proper use of number (without warranty)

- the rules are even more simple.

The proper use of number (without warranty)

- the rules are even more simple.
- and they are the same as for Amharic etc.
  - if there is a single entity to be mentioned, use singular.

    I trained a single network

The proper use of number (without warranty)

- the rules are even more simple.
- and they are the same as for Amharic etc.
  - if there is a single entity to be mentioned, use singular.
    I trained a single network
  - if there are several entities to be mentioned, use plural.
    I trained several networks

The proper use of number (without warranty)

- the rules are even more simple.
- and they are the same as for Amharic etc.
  - if there is a single entity to be mentioned, use singular.
    I trained a single network
  - if there are several entities to be mentioned, use plural.
    I trained several networks

The proper use of number (without warranty)

- Some words have no plural form: *information*, *research*, *?input*, *?output*, ...

The proper use of number (without warranty)

- Some words have no plural form: *information*, *research*, *?input*, *?output*, ...
- Some words have no plural form: *scissors*, *pants*, *binoculars*, ...

The proper use of number (without warranty)

- Some words have no plural form: *information*, *research*, *?input*, *?output*, ...
- Some words have no plural form: *scissors*, *pants*, *binoculars*, ...
- a borderline case: *data*
    - etymologically the plural of datum
    - but nowadays also used in singular for collections of data items

        The data is/are noisy.
        The data is insufficient.

The proper use of number (without warranty)

- Some words have no plural form: *information*, *research*, *?input*, *?output*, ...
- Some words have no plural form: *scissors*, *pants*, *binoculars*, ...
- a borderline case: *data*
  - etymologically the plural of datum
  - but nowadays also used in singular for collections of data items

    The data is/are noisy.
    The data is insufficient.

- The number of the subject NP and the predicate have to agree with respect to number.

    The method is better suited for low-resourced languages.
    The methods have been applied to low-resourced languages.