Methoden Wetenschappelijk Onderzoek

Scientific Writing
This lecture

• Useful information for writing scientifically
  – Organization
  – References (repeat)
  – Illustrations

• Nothing new, but stuff that bears repeating
  – From now: no excuse for not getting it right!
Lecture Plan

- Things to do before you begin
- Organization
- Writing
- A repeat about references
- Diagrams, figures, tables, code etc.
Before you begin (1)

- Identify your goal(s)
  - Report of a project?
  - Project proposal?
  - Presentation of a new idea?
  - Presentation of results?
  - Contribution to a discussion?
  - A review article?
  - …

- There may be multiple goals
Before you begin (2)

• Who is your audience?
  – Your professor?
  – Close colleagues?
  – Other researchers?
    • Friendly
    • Unfriendly
  – A wider audience?
    • Scientists
    • Lay audience
Before you begin (3)

• Where/how will you publish?
  – Journal?
    • Finished work, ~8000 words
  – Conference?
    • Latest results, ~3000–4000 words
  – Book chapter?
    • Like a journal article, but less strong peer review
  – Thesis?
  – Report?
    • More room for detail
Before you begin (4)

• Answer, with some detail
  – Don’t start anything before you have answered these questions

• What is your message?
  – What is your conclusion?
  – What is the impact of your work?
  – What do you want to convince other researchers of?

• What is your contribution?
  – What is new?
  – What are your results?
  – Which results are relevant to your conclusion?
Before you begin (5)

• What can I assume my audience knows?
• How much room do I have to elaborate?
• What background do I need to present?
  – Such that the target audience understands
    • Your questions
    • Your methods
    • Your conclusions
  – How much detail do you need?
  – What can I use references for?
• Whose work did you use?
  – And whom do you need to impress?
  – Avoid plagiarism!!!
Before you begin (6)

• Make a **schematic** of how parts of the content depend on each other
  – Helps to **avoid confusion**
  – Helps to avoid explaining things **twice**

```
New speech features
Hidden Markov Model
Robustness
Speech recognition
Feature vectors
```
Before you begin (6)

• Make a **schematic** of how parts of the content depend on each other
  – Helps to **avoid confusion**
  – Helps to avoid explaining things **twice**
Organization
The canonical form

• Abstract
• Introduction
• Method
• Results
• Conclusion
• Discussion
Organization (1)

• Prepare a division in chapters
  – **Before** you start writing!
  – Describe (on the level of sections < chapters) what you **want/need** to write
Organization (2)

• Ideally, chapters are about equally long
  – Just long enough to read in one sitting
  – Split them up if they become too long
  – Use headings and subheadings if sections become too long (> 1–2 pages)
  – Avoid too many levels of subsections

5.1.3.2 The parity bit

This is a sign your chapters are too long, or contain things that should not be together in one chapter
Preface

- Only occurs in larger texts (books, theses)

- Background/history of the project
  - Not the scientific, but the organizational background
  - Acknowledgments of contributions, friends, family etc.
The abstract

• Is often all someone reads of your paper
• Among other reasons because it may be the only thing they can get for free
• Should be self-contained
  – No references
  – It is not a teaser!
• Contains
  – Your research question
  – The essence of your method
  – Your main results
Introduction

• **Scientific background** of the project
  – Why was this research done?

• Research **question(s)**

• A short **summary of methods**
  – Just reading the introduction and conclusion should give you a good idea what was done

• (In shorter works) review of **background material**
  – In larger works this would have chapters of its own
Background

• Larger works have separate chapters that review background material
  – If you have a multi-disciplinary audience, you can divide material from different disciplines over different chapters

• In all cases: keep your target audience in mind!
Your contribution

• Define good research questions
• Creative work to create your methods
• Your results

• For multiple models/ experiments
  – [methods; results]$^n$ or methods$^n$; results$^n$?
Methods

• A detailed description of your system
  – Either experimental or computational
  – Anything special about the analysis of your data

• Experiments need to be replicable
  – Give all details and parameters

• Models need to be re-implementable
  – Give a detailed description

• If you base yourself on well-published stuff, you can always use references
• Sometimes you can use appendices for details
Results

• Contains the factual results of your system
  – And nothing but the results
  – The statistical tests & descriptions go here

• Do not be tempted to give descriptions of your methods or interpretations of your results in the results section!
Conclusion

• **Summary** of results
  – Again: Introduction and conclusion should together give a complete picture of your work

• Help in **interpreting** the results in light of the **research questions**
Discussion

• Comparison of your work with existing work
• What is the broader impact/ significance of your results?
• Future work
  – But only if it is serious!
Acknowledgment

• Usually an acknowledgment and a preface are mutually exclusive
  – Preface for books etc.
  – Acknowledgment for articles
  – Often funding agencies and contributors are thanked in the acknowledgment

Example

ACKNOWLEDGMENTS

This work is part of the NWO vidi project “Modeling the evolution of speech” Grant No. 016.074.324. The author thanks Didier Demolin for discussion on the function and anatomy of air sacs, Rob van Son for comments on the manuscript, and Wendy van Bohemen for letting the author use the anatomical collection of the Amsterdam Zoological Museum.

de Boer, 2009
Appendices / Supplementary material

• Appendices (or on line supplements) contain
  – Extra detailed results (not really necessary for your conclusion, or raw data if you mention descriptive statistics)
  – Technical details, not necessary for understanding the gist of your methods
  – Extra background for a less specialized audience (mostly only in books etc.)
Bibliography/references

• At the back of your work
  – This way it is easiest to find

• More about this in a minute
Writing
When to start writing

• Ideally, after everything (programming, experiments, analysis) is done
• In practice, a parallel effort

• Sometimes, the effort of writing things down shows you that you need more data
  – Ideally, this should not happen
  – Can only be done if experiments are simple
  – But should not be a problem, if this happens rarely
  – But be careful not to get completely sidetracked
  – This is why you make a research plan
Writing (1)

• Writing is an iterative process
  – Don’t be afraid to modify stuff you wrote
  – But be careful to maintain consistency
  – An outline and a research plan help
Writing (2)

- Write, mature, reread
  - Time creates the necessary distance
  - If you think something isn’t good, you can be sure it isn’t good

- Let your supervisor (or your colleagues) read your stuff regularly
  - But don’t hope your supervisor will re-read your thesis twenty times

- Take criticisms seriously, but not personally
Advice (1)

• **Scientific writing has its own style**
  – Read a lot to learn this
  – But don’t forget to read *popular science* as well (you want your stuff to be fun to read)
Advice (2)

• Use one **perspective** from which to write
  – First person singular (I) is almost **never** used
  – First person plural (we) is **often** used
    • weird if there is only one author “tutorial we”
  – Neutral, **passive voice** is often used
    • But it does not always result in beautiful style
  – Richard Feynman successfully uses “you”

• **Don’t change** your perspective halfway
Advice(3)

• Optimize your **paragraphs**
  – First sentence **announces** content
  – Material in between **elaborates**
  – Don’t make them **long**
  – Last sentence introduces **transition** to next paragraph

*Most of the world’s 7,000 present-day languages fall into language families — such as the Indo-European family, to which this journal’s English, along with 140 other languages, belongs. Languages of the same family resemble each other because they arose from a common ancestor and then diverged. The difficulty in detecting ancient linguistic relationships — those beyond the 5,000–10,000-year barrier — is that all languages share the same pool of consonants, vowels and grammatical structures. Therefore, some*

Diamond, 2011
Advice (4)

• Try to write **simple** texts
  – No “*officialese*”
  – Avoid *jargon* and *abbreviations* if you can
  – Avoid the temptation to use *synonyms*

• Don’t write **overlong** sentences
  – But don’t use *three-word* sentences either

• **Read a style manual!**
  – Links will be added to the website
More about references
Plagiarism

• When using others’ work, include references
• Never copy anything directly without mentioning the source: plagiarism
  – Not even if you translate
  – Not if they say it a hundred times better than you

• Don’t paraphrase either
  – Copy a sentence, but rephrase it slightly
  – Understand before you write!
References (1)

• Short text: use quotes, and perhaps italics:
  – As the bard of Avon said: “A rose by any other name would smell as sweet.” (Shakespeare 1597)

• Longer texts: make it a separate paragraph, clearly separated from the main text (e.g. through indentation and white space)
References (2)

• Always mention the source
  – Better too often than forget once

• Include the source in your reference list

• Also do this for figures, tables and other graphical materials
  – Often you need to clear copyright for this!
References (3)

• You may need to give more information:
  – …rete algorithm (Russell & Norvig 1995, section 10.5)…

• Referring to sections is better than referring to page numbers if there are multiple editions (e. g. paperback, hardback)
Bibliography

• You ought to know by now how to make a bibliography (reference list)
• Hard cases:
  – **Unpublished** stuff
    • ...(Smith, *unpublished manuscript*)
    • ...(Jones, *in press*)
    • ...(White, *in preparation*)
    • ...(Brown, *personal communication*)
  – Technical reports
    • Often have a year and an internal number, always mention the institution and the city
  – PhD theses, MA theses etc.
    • More difficult to locate, but specify university, department and any internal codes
Illustrations
Diagrams (1)

- **Diagrams** are often useful
  - Especially for algorithms and architectures
  - But also for explaining experimental methods
- They have to be **clear and spacious**
  - Never scan and copy
  - Always use vector graphics
Diagrams (2)

Diagrams allow you to focus on what you think is important

Nishimura et al. 2007

de Boer, 2012
Diagrams (3)

- Often, a drawing is **better** than a photo
Graphs

• Graphs are an important way to present results

• Choose the right graph for your results
  – What does the graph need to express?

• Don’t forget titles, labels, axes, units etc.

• Make it readable!
Graph (example 1)

Development of communicative success in a population of 25 agents

Axes: Generation, Communicative success

Line labels: Average over 10 runs, 90% confidence interval

Axis labels: Error bars, Use of a line to suggest continuity

Title: Development of communicative success in a population of 25 agents
Graph (example 1)

Development of communicative success in a population of 25 agents

This is bad practice, because it is an impossible value!

Bad

Good

Axes

Error bars

Title

Line labels

Axis labels

Use of a line to suggest continuity

Average over 10 runs

90% confidence interval
Graph (example 2)

Vergelijking experiment 1 en 2

Bad

Good

Vergelijking witheid van tanden direkt na poetsen met 95% betrouwbaarheidsinterval

Zonder ingrediënt X

Mèt ingrediënt X
Color

• If color use is possible, then it is advisable
  – Printing color is expensive
  – But many articles also appear as pdf (and this is used by most people anyway)
  – But make sure your graphs also make sense in black and white!
    (Print in B&W to make sure)
  – Make sure most color blind people can use them, too (avoid red/green contrasts)
Tables?

- Only put tables in the main text if they can be understood at a glance
  - They can be long, but they must be simple
  - What does the reader really need?
- Otherwise, use a graph and put the table in an appendix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.250443</td>
<td>0.73488</td>
<td>0.948836</td>
</tr>
<tr>
<td>2</td>
<td>0.807682</td>
<td>0.467691</td>
<td>0.728936</td>
</tr>
<tr>
<td>3</td>
<td>0.451348</td>
<td>0.076974</td>
<td>0.093809</td>
</tr>
<tr>
<td>4</td>
<td>0.896178</td>
<td>0.812178</td>
<td>0.512827</td>
</tr>
<tr>
<td>5</td>
<td>0.217635</td>
<td>0.615393</td>
<td>0.682811</td>
</tr>
<tr>
<td>6</td>
<td>0.571587</td>
<td>0.990073</td>
<td>0.084405</td>
</tr>
<tr>
<td>7</td>
<td>0.771654</td>
<td>0.721151</td>
<td>0.152322</td>
</tr>
<tr>
<td>8</td>
<td>0.627987</td>
<td>0.633807</td>
<td>0.168597</td>
</tr>
<tr>
<td>9</td>
<td>0.668921</td>
<td>0.829536</td>
<td>0.002979</td>
</tr>
<tr>
<td>10</td>
<td>0.684140</td>
<td>0.087422</td>
<td>0.464807</td>
</tr>
<tr>
<td>11</td>
<td>0.150617</td>
<td>0.787560</td>
<td>0.077427</td>
</tr>
<tr>
<td>12</td>
<td>0.421771</td>
<td>0.842447</td>
<td>0.619870</td>
</tr>
<tr>
<td>13</td>
<td>0.421567</td>
<td>0.896218</td>
<td>0.951763</td>
</tr>
<tr>
<td>14</td>
<td>0.418193</td>
<td>0.753472</td>
<td>0.471015</td>
</tr>
<tr>
<td>15</td>
<td>0.288770</td>
<td>0.061868</td>
<td>0.505044</td>
</tr>
<tr>
<td>16</td>
<td>0.225232</td>
<td>0.615185</td>
<td>0.643771</td>
</tr>
<tr>
<td>17</td>
<td>0.214955</td>
<td>0.493694</td>
<td>0.881810</td>
</tr>
<tr>
<td>18</td>
<td>0.289220</td>
<td>0.468677</td>
<td>0.653768</td>
</tr>
<tr>
<td>19</td>
<td>0.726684</td>
<td>0.091652</td>
<td>0.870418</td>
</tr>
<tr>
<td>20</td>
<td>0.580616</td>
<td>0.329993</td>
<td>0.585588</td>
</tr>
<tr>
<td></td>
<td>9.689611</td>
<td>11.504710</td>
<td>9.456472</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Aantal A</th>
<th>Aantal B</th>
<th>Aantal C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groep 1</td>
<td>0.82</td>
<td>0.05</td>
<td>0.77</td>
</tr>
<tr>
<td>Groep 2</td>
<td>0.05</td>
<td>0.69</td>
<td>0.91</td>
</tr>
<tr>
<td>Groep 3</td>
<td>0.61</td>
<td>0.09</td>
<td>1.00</td>
</tr>
<tr>
<td>Groep 4</td>
<td>0.46</td>
<td>0.57</td>
<td>0.31</td>
</tr>
<tr>
<td>Totaal:</td>
<td>1.95</td>
<td>1.40</td>
<td>2.98</td>
</tr>
</tbody>
</table>
Algorithms

- **Avoid** putting code in the text
  - Depends on *programming language*
  - Too many *details*
  - No clear distinction between *essence* and *detail*

- **Pseudocode** is better
  - Keep it *short* (split up in functions)
  - Use notation and formalisms that anyone who knows programming and maths can *understand*
Algorithm (example)

Bad

Good

\begin{verbatim}
void DetermineClust( int Size, int **Neighbours, int *NewClusts )
{
    const int MaxIts = 128;
    int index, indey;
    int change = true;
    int Steps = 0;
    int MaxNeighs;
    int* Clusts = new int [ Size ];
    // find maximal number of neighbours...
    MaxNeighs = Neighbours[ 0 ][ 0 ];
    for( index = 1; index < Size; index++ )
        if( Neighbours[ index ][ 0 ] > MaxNeighs )
            MaxNeighs = Neighbours[ index ][ 0 ];
    // And reserve memory for sorted neighbours list
    int* Sorted = new int[ MaxNeighs-1 ];
    while( change && (++Steps <= MaxIts) )
    {
        for( index = 0; index < Size; index++ )
            Clusts[ index ] = NewClusts[ index ];
        change = false;
        for( index = 0; index < Size; index++ )
        {
            if( Neighbours[ index ][ 0 ] == 1 )    // No neighbours
                BestClass = Clusts[ index ];
            else
            {
                // Find class that occurs most frequently in nearest points
                for( indey = 1; indey < Neighbours[ index ][ 0 ]; indey++ )
                    Sorted[ indey-1 ] = Clusts[ Neighbours[ index ][ indey ] ];
                // qsort( void* Sorted, Neighbours[ index ][ 0 ]-1, sizeof( int ), CompareInt );
                qsortint( Sorted, Neighbours[ index ][ 0 ] -1 );
                FindBest( Sorted, Neighbours[ index ][ 0 ]-1, BestClass );
                if( BestClass != Clusts[ index ] )
                    change = true;
            }
            NewClusts[ index ] = BestClass;
        }
    }
    delete [] Sorted;
    delete [] Clusts;
}
\end{verbatim}

\begin{verbatim}
C ← Unsupervised Classify ( S, R ) [ S is set of points, R is radius of neighbourhood ]
C ← Initial Classification( S, D )
for ∀ c_i ∈ C:
    N_i ← ∅ [N_i is neighbourhood of c_i]
    for ∀ c_j ∈ C, c_j ≠ c_i:
        if distance( c_i, c_j ) ≤ R
            N_i ← N_i ∪ c_j
    end if
end for

do
    for ∀ c_i ∈ C simultaneously:
        set label of c_i to
        most frequent class label in N_i
    end for
while( C changed ∧ iterations < 128 )
\end{verbatim}
Conclusion

• This is a huge subject
  – Read books (or internet stuff) about this

• Read other people’s work
  – Note what works and what doesn’t

• Everybody has their own preferences
  – You can’t argue about taste
  – But some things really work better than others

• Quality comes with experience

• But after this lecture: no excuse for sloppiness!