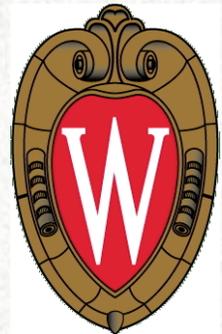# Achieving a Readable Style
Part 2: Sentence Structure

A writing workshop presented by BACTER
(Bringing Advanced Computational Techniques to Energy Research)
University of Wisconsin-Madison
October 2009

Last time we looked at…

Part One: Word Choice

- Acronyms
- Plain language
- Wordiness

This time we'll examine…

Part Two: Sentence Structure

- Subject-verb agreement

- Maintaining linkage

BACTER

Can you spot the problem here?

Original: Although the method yielded a relatively high prediction performance (more than 80% accuracy) in cross-validation and usefulness in the comprehensive prediction of target proteins for a given chemical compound with tens of thousands of prediction targets, it suffered from the problem of predicting many false positives when comprehensive predictions were conducted.

The subject is separated from its verb by 34 words

Original: Although the **method** yielded a relatively high prediction performance (more than 80% accuracy) in cross-validation and usefulness in the comprehensive prediction of target proteins for a given chemical compound with tens of thousands of prediction targets, it **suffered** from the problem of predicting many false positives when comprehensive predictions were conducted.

Subject-verb separation

**When people read, they expect a subject to be followed quickly by a verb. When this doesn't happen, they tend to be distracted by the lack of syntactical closure and skim through everything in between.*

The problem is easily remedied, however…

Here's a solution

Revised: ~~Although~~ The method yielded a relatively high prediction performance (more than 80% accuracy) in cross-validation, and was useful in the comprehensive prediction of target proteins for a given chemical compound with tens of thousands of prediction targets; **however, it suffered** from the problem of predicting many false positives when comprehensive predictions were conducted.

Another example

Original: "**Master switches**", a central set of regulatory genes whose states (on/off or activated/deactivated) determine specific developmental fate or cell-fate specification, **play** a pivotal role for whole developmental processes.

BACTEЯ

And one possible solution…

<u>Revised</u>: Pivotal in whole developmental processes are "master switches," a central set of regulatory genes whose states (on/off or activated/deactivated) determine specific developmental fate or cell-fate specification.

BACTEЯ

What noun and verb are separated in this example?

<u>Original</u>: Addition of an Escherichia coli 50S subunit containing a Cy5-labeled L11 N-terminal domain within the GTPase-associated center to an E. coli 30S initiation complex containing Cy3-labeled initiation factor 2 complexed with GTP leads to rapid development of a FRET signal during formation of the 70S initiation complex.

How can this be fixed?

Original: **Addition** of an Escherichia coli 50S subunit containing a Cy5-labeled L11 N-terminal domain within the GTPase-associated center to an E. coli 30S initiation complex containing Cy3-labeled initiation factor 2 complexed with GTP **leads** to rapid development of a FRET signal during formation of the 70S initiation complex.

One solution…

<u>Revised</u>: Rapid development of a FRET signal occurs during formation of the 70S initiation complex upon adding an Escherichia coli 50S subunit containing a Cy5-labeled L11 N-terminal domain within the GTPase-associated center to an E. coli 30S initiation complex containing Cy3-labeled initiation factor 2 complexed with GTP.

**This sentence has other issues, but it's a little better

BACTER

This paragraph illustrates another common problem

<u>Original</u>: Ruminants form the cotyledonary placenta at the feto-maternal interface. Two specific types of trophoblast cells, trophoblast giant binucleate cells (BNCs) and trophoblast mononucleate cells (TMCs), play a crucial role in ruminant placentation. The properties of BNC-specific genes...have been investigated, and TMC-expressed interferon-tau is the molecule for maternal recognition of pregnancy. BNC and TMC individually produce numerous proteins of unknown function. It is important to identify the genes that are specifically expressed in each cell type in order to systematically decipher the function of the trophoblast cells.

Maintaining Linkage

**The problem is a lack of linkage between sentences. That is, they tend to read like stand-alone statements strung together; there's little coherence or transition between ideas. It's as if the authors took a set of bullet points and condensed them into a paragraph.

For example…

BACTEЯ

What's the connection between these ideas?

Ruminants form the cotyledonary placenta at the feto-maternal interface. Two specific types of trophoblast cells, trophoblast giant binucleate cells (BNCs) and trophoblast mononucleate cells (TMCs), play a crucial role in ruminant placentation. The properties of BNC-specific genes...have been investigated, and TMC-expressed interferon-tau is the molecule for maternal recognition of pregnancy. **???** BNC and TMC individually produce numerous proteins of unknown function. **???** It is important to identify the genes that are specifically expressed in each cell type in order to systematically decipher the function of the trophoblast cells.

BACTEЯ

Here are my best guesses, which I've indicated by adding connecting words

Revised: Ruminants form the cotyledonary placenta at the feto-maternal interface. Two specific types of trophoblast cells, trophoblast giant binucleate cells (BNCs) and trophoblast mononucleate cells (TMCs), play a crucial role in ruminant placentation. The properties of BNC-specific genes... have been investigated, and TMC-expressed interferon-tau is the molecule for maternal recognition of pregnancy. **However,** BNC and TMC **also** individually produce numerous proteins of unknown function. **Thus, as a first step toward (?)** systematically deciphering the function of **these** trophoblast cells, we sought to identify the genes that are specifically expressed in each cell type.

BACTEЯ

Maintaining linkage

**I say "guesses," because the author's failure to use connecting words leaves me unsure of the true meaning.

A good use of connecting words

A challenge of using genomics databases is the automated retrieval of relevant data. **For example**, naive approaches to retrieve gene expression studies about 'brain' will fail to find datasets that only mention 'cerebrum' in their descriptions, because free text-based retrieval algorithms generally cannot make the inference that 'cerebrum' is part of 'brain'. **In addition**, using free text for information retrieval can produce false positives due to ambiguity, and false negatives due to synonyms. **For these reasons**, it is valuable to use formal ontologies to describe genomics studies… **However**, tagging studies with terms from ontologies is currently done by human curators.

BACTEЯ

And how hard this is to understand once they're removed

A challenge of using genomics databases is the automated retrieval of relevant data. Naive approaches to retrieve gene expression studies about 'brain' will fail to find datasets that only mention 'cerebrum' in their descriptions, because free text-based retrieval algorithms generally cannot make the inference that 'cerebrum' is part of 'brain'. Using free text for information retrieval can produce false positives due to ambiguity, and false negatives due to synonyms. It is valuable to use formal ontologies to describe genomics studies… Tagging studies with terms from ontologies is currently done by human curators.

BACTEЯ

Some examples. Now, use them!

Transition **ideas** and the connecting words that serve them:

Result: therefore, thus, hence

Example: for example, specifically, an illustration

Contrast: but, yet, however, on the other hand

Time: meanwhile, after, later

Sequence: first, then, finally

--from the UW-Madison Writing Center Sourcebook

BACTEЯ

Next…

Sentence Structure, continued

- More on maintaining linkage

- Noun clusters

- Passive voice

BACTEЯ