

# Scientific Writing: Clarity, Conciseness, and Cohesion

Prepared by  
Nathan Sheffield,  
Institute for Genome Sciences and Policy, Duke University  
In collaboration with  
Duke Writing Studio

Updated September 8, 2011

Let's start with an example...

# Example 1

## What makes this sentence unclear?

*“The assumptions that all sites evolve at one of two evolutionary rates (conserved and nonconserved), that these rates are uniform across the genome, that sites evolve independently conditional on whether they are in conserved or nonconserved regions, and that the phylogenetic models for conserved and nonconserved regions have the same branch-length proportions, base compositions, and substitution patterns, all represent oversimplifications of the complex process of sequence evolution in eukaryotic genomes.”*

# Example 1

## Distance between subject and verb

*“The assumptions that all sites evolve at one of two evolutionary rates (conserved and nonconserved), that these rates are uniform across the genome, that sites evolve independently conditional on whether they are in conserved or nonconserved regions, and that the phylogenetic models for conserved and nonconserved regions have the same branch-length proportions, base compositions, and substitution patterns, all represent oversimplifications of the complex process of sequence evolution in eukaryotic genomes.”*

# Example 1

## Complex subject

*“The assumptions that all sites evolve at one of two evolutionary rates (conserved and nonconserved), that these rates are uniform across the genome, that sites evolve independently conditional on whether they are in conserved or nonconserved regions, and that the phylogenetic models for conserved and nonconserved regions have the same branch-length proportions, base compositions, and substitution patterns, all represent oversimplifications of the complex process of sequence evolution in eukaryotic genomes.”*

# Example 1

## Implied actions (nominalizations) vs. verbs

*“The **assumptions** that all sites **evolve** at one of two **evolutionary** rates (**conserved** and **nonconserved**), that these rates **are** uniform across the genome, that sites **evolve** independently conditional on whether they **are** in **conserved** or **nonconserved** regions, and that the phylogenetic **models** for **conserved** and **nonconserved** regions **have** the same branch-length proportions, base **compositions**, and **substitution patterns**, all **represent** oversimplifications of the complex **process** of **sequence evolution** in eukaryotic genomes.”*

# Example 1

## **Context** comes after the **list**

*“The assumptions that all sites evolve at one of two evolutionary rates (conserved and nonconserved), that these rates are uniform across the genome, that sites evolve independently conditional on whether they are in conserved or nonconserved regions, and that the phylogenetic models for conserved and nonconserved regions have the same branch-length proportions, base compositions, and substitution patterns, **all represent oversimplifications of the complex process of sequence evolution in eukaryotic genomes.**”*

**Clarity**

```
graph TD; Clarity[Clarity] <--> Conciseness[Conciseness]; Clarity <--> Cohesion[Cohesion]; Conciseness <--> Cohesion;
```

**Conciseness**

**Cohesion**



# Four Problems

# Four Problems

## Things that make science writing unclear

1. Subjects and verbs too far apart
  2. Overabundance of nominalizations
  3. Poor flow (misplacement of old and new information)
  4. Excessive/unnecessary use of passive voice
- NOT the complexity of the topic!

# Four Problems

## 1. Subjects and verbs too far apart

- English readers expect *doers* to be near their *actions*.
- Complex subjects (subjects modified with essential clauses) can violate this expectation.

# Four Problems

## 2. Overabundance of Nominalizations

- English readers expect *actions* to be in *verbs*.
- Nominalizations are actions that appear in parts of a sentence *other than* a verb (e.g. in nouns or adjectives).
  - The word “nominalization” is a nominalization of the verb “to nominalize.”
- Nominalizations aren’t all created equal: some are clear, others reduce clarity.

# Four Problems

## 2. Overabundance of Nominalizations

### **Actions in Nominalizations:**

*The **assumption** that all RNAs are polyadenylated **is** an **oversimplification** of the transcription process.*

### **Actions in Verbs:**

*The model **oversimplifies** the transcription process because it **assumes** that all RNAs are polyadenylated.*

# Four Problems

## 3. Poor flow (lack of cohesion)

- A cohesive sentence **links** with neighboring sentences by **starting** with familiar (“old”) ideas and **ending** with new ideas.
- Disrupt flow by:
  - **Starting** with unfamiliar ideas
  - **Ending** with backwards-linking ideas
- Also evident at the paragraph-level.

# Four Problems

## **3. Poor flow (lack of cohesion)**

(in a paper about farmers...)

Farmers try to provide optimal growing conditions for crops by using soil additives to adjust soil pH. Garden lime, or agricultural limestone, is made from pulverized chalk, and can be used to raise the pH of the soil. Clay, which is a naturally acidic soil type, often requires addition of agricultural lime.

# Four Problems

## 3. Poor flow (lack of cohesion)

Old Information vs. New information

Farmers try to provide optimal growing conditions for crops by using soil additives to adjust soil pH. Garden lime, or agricultural limestone, is made from pulverized chalk, and can be used to raise the pH of the soil. Clay, which is a naturally acidic soil type, often requires addition of agricultural lime.



# Four Problems

## 3. Poor flow (lack of cohesion)

Old Information vs. New information

Farmers try to provide optimal growing conditions for crops by using soil additives to adjust soil pH. One way to raise the pH of the soil is an additive made from pulverized chalk called garden lime or agricultural limestone. Agricultural limestone is often added to naturally acidic soils, such as clay.

# Four Problems

## 4. Excessive/unnecessary use of passive voice

- Passive voice is sometimes useful, but...
- Passive voice has several effects:
  - Reverses the order of the sentence (A-B vs. B-A)
    - I stole the money
    - The money was stolen by me
  - It can eliminate the actor (causing ambiguity)
  - It often increases length

# Four Problems

## 4. Excessive/unnecessary use of passive voice

- Passive voice is NOT inherently scientific!
- Consider cohesion: **Don't** choose passive voice simply out of habit. **Do** choose passive voice when it improves cohesion by putting familiar ideas first.
- Most scientific journals encourage authors to **use active voice** for the sake of *clarity*, *conciseness*, and *cohesion*!

# Passive Voice

What do the journals say?

## Science

“Use active voice when suitable, particularly when necessary for correct syntax (e.g., ‘To address this possibility, we constructed a  $\lambda$ Zap library...,’ not ‘To address this possibility, a  $\lambda$ Zap library **was constructed**...’).”

# Passive Voice

## What do the journals say?

### Nature

“Active voice has been Nature policy for as long as I can remember; it is enshrined in our style manual and is specifically recommended to all authors as part of our standard acceptance procedure. However, if an author insists on the passive, we would probably allow it...So you will see papers in Nature in the passive voice, but you can be assured that this is at the author's insistence rather than Nature policy.”

–Maxine Clark, editor

# Passive Voice

**What do the journals say?**

## [Astronomical Society of the Pacific](#)

“Use active voice as much as possible, and avoid passive voice as you would avoid the Ebola virus. This means writing ‘Astronomers discovered a new planet’ (active voice) rather than ‘A new planet was discovered by astronomers’ (passive voice). You should write less than 10 percent of your sentences in passive voice.”

# Passive Voice

## The standard objection

“But using active voice means using first person pronouns, and first person pronouns aren’t allowed in my field!”

But compare:

- The substrate surface **was mapped** using an Atomic Force Microscope.
- We mapped the substrate surface using an AFM.
- The AFM mapped the substrate surface.

# Revision Techniques



# Revision Techniques

**Four ways to improve clarity, conciseness, and cohesion**

1. Omit unnecessary words
2. Put actions in verbs (avoid nominalizations)
3. Place verbs near subjects
4. Put familiar information first

# Revision Techniques

## 1. Omit unnecessary words

- It is absolutely vital that...
  - ✓ We must...
- At the same time...
  - ✓ Simultaneously/furthermore...
- There were five mice receiving antibiotics...
  - ✓ Five mice received antibiotics.

# Revision Techniques

## 2. Put actions in verbs

- We performed an **analysis**...
  - ✓ We **analyzed**...
- The **quantification** of the atoms was done.
  - ✓ The atoms were **quantified**.
- The MS managed the **measurement** and **identification** of the proteins.
  - ✓ The MS **measured** and **identified** the proteins.

# Revision Techniques

## 2. Put actions in verbs

- Nominalizations are useful when they summarize the action of the previous sentence:
- Our analysis using regression and k-means clustering revealed that...
  - ✓ We **analyzed** the data with regression and k-means clustering. This **analysis** revealed that...

# Revision Techniques

## 3. Place verbs near subjects

- **DNA** in repeat regions or with long stretches of the same base or small microsatellites **causes** problems for next-gen sequencers.
  - ✓ **DNA causes** problems for next-gen sequencers when it is in repeat regions or has long stretches of the same base or small microsatellites .

# Revision Techniques

## 4. Put familiar information first

- We **searched** the database of sequences to look for similar structures. A **protein** involved in the regulation of the BRCA1 gene in humans was found by **the search**.
- ✓ We **searched** the database of sequences to look for similar structures. **This search** found a **protein** involved in the regulation of the BRCA1 gene in humans.

Now for some practice...

# Example 2

**What would you do?**

*This component will chiefly involve a description and quantitative analysis of the study's data collection process.*



# Example 2

**We suggest: Put actions in verbs**

*This component will chiefly involve a description and quantitative analysis of the study's data collection process.*

***This component describes and quantitatively analyzes the data collection process.***

# Example 2

## Positive Consequences

*This component will chiefly involve a description and quantitative analysis of the study's data collection process.*

***This component describes and quantitatively analyzes the data collection process.***

- ✓ The sentence is more concise (10 vs 16 words).
- ✓ The meaning is clearer.

# Example 3

**What would you do?**

*Detailed analyses of the evolutionary features of different types of regulatory elements are an important area for future research.*

# Example 3

**We suggest: Put actions in verbs**

*Detailed analyses of the evolutionary features of different types of regulatory elements are an important area for future research.*

Consider **implied actions** vs. **verb**:

*Detailed **analyses** of the **evolutionary features** of **different** types of **regulatory** elements **are** an important area for future **research**.*

# Example 3

## Positive Consequences

*Detailed analyses of the evolutionary features of different types of regulatory elements are an important area for future research.*

***Future research should analyze the evolutionary features of different types of regulatory elements.***

- ✓ The sentence is more concise (13 vs 19 words).
- ✓ The subject is clearer.
- ✓ The subject and verb are closer together.

# Example 4

**What would you do?**

*Improvements are expected in the predictive power of all the scores being computed on multispecies alignments.*

## Example 4

**We suggest: Use active voice, omit unnecessary words**

*Improvements **are** expected in the predictive power of all the scores **being** computed on multispecies alignments.*

***[We expect to/Model changes should] improve the predictive power of our multispecies alignment scores.***

# Example 4

## Positive Consequences

*Improvements **are** expected in the predictive power of all the scores **being** computed on multispecies alignments.*

***[We expect to] improve the predictive power of our multispecies alignment scores.***

- ✓ The sentence is more concise (12 vs 16 words).
- ✓ Prepositions no longer disrupt flow.
- ✓ Sentence is more direct.



# Example 5

## What would you do?

*Some astonishing questions about the nature of the universe have been raised by scientists studying the nature of black holes in space. The collapse of a dead star into a point perhaps no larger than a marble creates a black hole.*

# Example 5

**We suggest: Put familiar information first, omit needless words**

*Scientists studying ~~the nature of~~ black holes ~~in space~~ have raised some astonishing questions about ~~the nature of~~ the universe. ~~The collapse of a dead star into a point perhaps no larger than a marble creates a black hole.~~ **A black hole is created by the collapse of a dead star into a point perhaps no larger than a marble.***

# Example 5

## Positive Consequences

*Scientists studying black holes have raised some astonishing questions about the universe. **A black hole is created by the collapse of a dead star into a point perhaps no larger than a marble.***

✓ The **link** is clearer; these sentences are more cohesive.

✓ ***But what if the paragraph were about different types of collapses, rather than about black holes?***

# Example 6

**What would you do?**

*The second reaction is really the end result of a very large number of reactions. It is also worth noting that these two reactions form a simple linear chain whereby the product of the first reaction is the reactant for the second.*

# Example 6

**We suggest: Eliminate unnecessary words**

*The second reaction is really the end result of **a very large number of** reactions.*

The second reaction is [really] the [end] result of **numerous** reactions.

***It is also worth noting that** these two reactions form a simple linear chain, whereby the product of the first reaction is the reactant for the second.*

***[Moreover,]** these two reactions form a simple linear chain, whereby the product of the first reaction is the reactant for the second.*

# Example 6

## Positive Consequences

***The second reaction is the result of numerous reactions. Moreover, these two reactions form a simple linear chain whereby the product of the first reaction is the reactant for the second.***

✓ More concise (32 vs. 42 words)

# Example 7

**What would you do?**

*Significant positive correlations were evident between the substitution rate and a nucleosome score from resting human T-cells.*

# Example 7

**We suggest: Put actions in verbs**

*Significant positive correlations were evident between the substitution rate and a nucleosome score from resting human T-cells.*

***In resting human T-cells, the substitution rate correlated with a nucleosome score.***



# Example 7

## Positive Consequences

*Significant positive correlations were evident between the substitution rate and a nucleosome score from resting human T-cells.*

***In resting human T-cells, the substitution rate correlated with a nucleosome score.***

- ✓ More concise (12 vs. 17)
- ✓ The verb is now “correlate” rather than the nebulous “were evident”

# Example 8

## What would you do?

*We identified genes that are differentially expressed between species. A phylogenetic tree based on the number of differentially expressed genes between species recapitulates their known phylogeny.*

# Example 8

**We suggest: Put actions in verbs**

*We identified genes that are differentially expressed between species. A phylogenetic tree based on the number of differentially expressed genes between species recapitulates their known phylogeny.*

***We identified genes that are differentially expressed between species. The number of differentially expressed genes can be used to build a phylogenetic tree that recapitulates the known phylogeny.***

# Example 8

## Positive Consequences

*We identified genes that are differentially expressed between species. ~~A phylogenetic tree based on the number of differentially expressed genes between species recapitulates their known phylogeny.~~ **The number of differentially expressed genes can be used to build a phylogenetic tree that recapitulates the known phylogeny.***

- ✓ The second sentence now links back at the beginning
- ✓ Subject-verb are now closer in the second sentence.

# Example 1 (again)

## What would you do? (Now with more context)

*The model used by the software is a fairly rich probabilistic model, but it is clearly not realistic in several respects. The assumptions that all sites evolve at one of two evolutionary rates (conserved and nonconserved), that these rates are uniform across the genome, that sites evolve independently conditional on whether they are in conserved or nonconserved regions, and that the phylogenetic models for conserved and nonconserved regions have the same branch-length proportions, base compositions, and substitution patterns, all represent oversimplifications of the complex process of sequence evolution in eukaryotic genomes.*

# Example 1 (again)

**We suggest: Place verbs near subjects**

**The gist of the sentence:**

*Certain assumptions oversimplify the complex process of sequence evolution in eukaryotic genomes.*

Should the gist of the sentence go first or last?  
Before the list of assumptions or after it?

# Example 1 (again)

## A possible revision

*[Our model admittedly] oversimplifies the complex process of sequence evolution in eukaryotic genomes by assuming that: (1) all sites evolve at one of two evolutionary rates (conserved and nonconserved), (2) these rates are uniform across the genome, (3) sites evolve independent of whether they are in conserved or nonconserved regions, and (4) the phylogenetic models for conserved and nonconserved regions have the same branch-length proportions, base compositions, and substitution patterns.*

# Example 1 (again)

## Positive Consequences

- ✓ The most important action (oversimplify) is now a verb
- ✓ The verbs follow closely after the subjects
- ✓ The sentence is more **cohesive**: familiar information links to the previous sentence at the beginning
- ✓ The sentence contains cues for parsing information (*by, [1, 2, 3, 4], however, etc.*)



# Summary of revision techniques

**Four ways to improve clarity, conciseness, and cohesion**

1. **Omit unnecessary words**: use find.
2. **Put actions in verbs**: find nominalizations and see if they should be converted to verbs.
3. **Place verbs near subjects**: check for subject-verb distance
4. **Put familiar information first**: check each sentence for old and new information.

# References and further reading

- Please check out the [Duke Scientific Writing Resource](#)
- *How to write consistently boring scientific literature* (2007), Kaj Sand-Jensen
- *The infectiousness of pompous prose* (1992), Martin W. Gregory
- *How we write about biology* (1991), Randy Moore
- *Writing intelligible English prose for biomedical journals* (2007), John Ludbrook
- *Whose literature is science?* (2003), Judith A. Swan
- *Dancing with professors: the trouble with academic prose* (1993), Patricia Nelson Limerick
- *What is the scientific literature?* (1986), John Maddox
- *Scientific literature: Clear as mud* (2003), Jonathan Knight
- *Style: Toward clarity and grace* (1990), Joseph Williams
- *The science of scientific writing* (1990), George Gopen, Judith Swan
- *The readability of marketing journals: are award-winning articles better written?* (2008), Sawyer, Laran, & Xu

# Duke Scientific Writing Resource

<http://cgi.duke.edu/web/sciwriting/index.php>

Duke

GRADUATE SCHOOL  
SCIENTIFIC WRITING RESOURCE

## INTRODUCTION

Introduction

## LESSONS

Disclaimer

Lesson I

Subjects/Actions

Principles

Examples

Worksheet

Lesson II

Cohesion/Emphasis

Principles

Examples

Worksheet

Lesson III

Simplicity

Principles

Examples

Worksheet

Example

## INTRODUCTION

There is no form of prose more difficult to understand and more tedious to read than the average scientific paper (Francis Collins in *The Astonishing Hypothesis*).

### Why is writing important in science?

Writing is the most common form of scientific communication, yet scientists have a reputation for being poor writers. Why? One reason could be that writing is never really taught to scientists. Better writing will benefit your science career in several ways. Within the scientific community, improved communication leads to improved collaboration, easier access to cross-disciplinary knowledge, and faster, less painful training. Besides this, you will be able to communicate better not only with other researchers, but with the public, who funds your research. If scientists were better writers, the gap between the public and academy would shrink.

### How can I use this resource?

The Scientific Writing Resource is course material that teaches how to write effectively. The material is not about correctness (grammar, punctuation, etc), but about *communicating what you intend to the reader*. It can be used

# Duke Scientific Writing Resource

<http://cgi.duke.edu/web/sciwriting/index.php>

You can understand the gist of the sentence just by reading the succession of subjects. The point of this example is to illustrate that you don't need every paragraph to have exactly 1 topic and subject. Instead, just be aware of what your subjects are, and if they match the structure of the idea you intend to communicate.

## REVISION TECHNIQUE

Highlight the subject of each sentence. Does the structure of your subjects match the information you intend to convey? In other words, are the subjects of the sentences jumping from one thing to another, or do they shift only when you intend to shift the topic under discussion?

**Note:** One problem that frequently makes scientific writing confusing is a sentence without a character; such sentences can be caused by passive voice, which can leave a reader to guess the actor (that's a Bad Thing). More on this in the section on passive voice.

## Principle 3: Keep subjects near verbs

Recall the two primary pieces of information a reader looks for:

1. *who* is the sentence about?
2. *what* are they doing?

When these two pieces of information are far apart, that usually means one of them isn't arriving until the end of the sentence. This confuses readers, because they can't piece together the whole picture without answers to these questions. In science writing, this is often caused by long, complex subjects. I find many sentences that go on and on and finally provide the verb at the end of the sentence. When this happens, readers must re-read the sentence, now that they know the action.

---

# Duke Scientific Writing Resource

<http://cgi.duke.edu/web/sciwriting/index.php>

## Examples:

### Example 1:

✘ The TRANSFAC database has been subject to different improvements, modifications, and extensions in structure and content over the years.

This sentence relies on nominalizations to convey action. The awkward verb of the sentence ("has been subject to") is basically meaningless; the authors likely intended to convey action in the words *improvement*, *modification*, and *extension*. But these are all nominalizations. By converting these into verbs, we get a much clearer sentence, and eliminate "has been subject to":

The TRANSFAC database has been improved, modified, and extended in both structure and content over the years.

To clarify even further: doesn't *improved* imply *modified*? Possibly it even implies *extended*. To strip it to exactly what you mean, what about this?

✔ The curators have improved the structure and content of the TRANSFAC database.