Punctuation crash course

Maybe 80% of students struggle with punctuation. Punctuation is part of the English grammar, and it's used to improve readability.

InEngl,ishwriTingWith...theApp'ropriat"e"punctuation—is*ImPoR* tAnT;

So, I think it's time for a crash course in punctuation. I can't go through everything; just the common problems.

The most common problem is incorrect spacing!

Citations

What's wrong with this punctuation?

> tion. In the previous work we employed the global OTSU binarization for video caption extraction and achieved good performance[8, 5]. However, by observing more practical cir-
>
> — Xu et al., *Automatic Text Discovering ...*, Proc. ACM Multimedia, 2008.

1. There's no space between "performance" and the citation "[...]".
2. The numbers in the brackets are out of order. (The package `\usepackage{cite}` puts them in order.)
3. (There's also grammar errors, but we'll ignore those.)

We change this to:

> In the previous work we employed the global OTSU binarization for video caption extraction and achieved good performance [5, 8].

Many journals and conferences have their own style. This is the most common style in computer science.

Spaces and brackets

In English, we (sometimes) use brackets within sentences. (We can also use them to enclose whole sentences.) In this case, there's space before the bracket "(" and after the close bracket ")".

No space:

- We write "$f(x)$" and "$(x+y)/2$".
- We write "Figure 1(a)" or "Fig. 1(a)" to refer to subfigure (a) of Figure 1.
- We write "server(s)" when it's unclear if there's one or more server(s). It's slightly informal. (Importantly, "(s)" does not indicate the unit is "seconds".)
- Some authors write "(s)he" (or "s/he") to refer to a person with unknown gender. I recommend "they" (known as *singular they*[1]) or something non-gendered like "the reader", "the user", etc.

[1] `https://en.wikipedia.org/wiki/Singular_they`

Spaces and brackets (cont.)

Space:

- When introducing acronyms, e.g. "... the time in advance (TIA) is measured ...".
- For listed items, e.g. "We experiment with (a) 1 GPU, and (b) 2 GPUs".
- For an equation reference, e.g. "Thus, (1) implies ...".
- In tables, where we want to be succinct, e.g.:
  - Units, e.g. "time (s)" or "time (sec.)"; "number of documents ($\times 10^6$)"; or "FDR (%)" (where FDR means "failure detection rate").
  - We might write "CPU (1 thread)" and "CPU (4 threads)".

What's wrong with the punctuation?

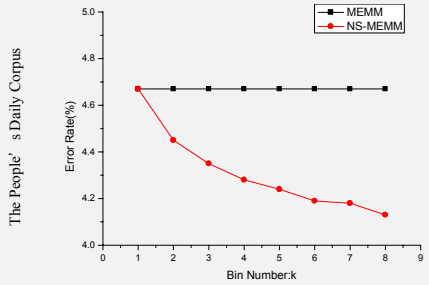
There is a problem with this sentence (which contains a bracket.)

The sentence starts outside the brackets, and ends inside the brackets.

(There is a problem with this sentence too).

The sentence starts inside the brackets, and ends outside the brackets.

What's wrong with the punctuation?



— Xiao et al., *The Study of* ..., ACM Trans. Asian Lang. Inf. Proc., 2007.

- It should be "Error Rate (%)" with a space.
- It's "People's Daily" not "People' s Daily".
- There should be no colon ":" before k.

Another very common error: bin number does not mean the same thing as number of bins.
(Compare: phone number with number of phones.)

Periods

A sentence ends with a *period*, i.e., ".", also called a *full stop*—this is the usual usage. There's no space before a full stop, and there is space after the full stop. (LaTeX determines the size of the space.)

We also use periods to denote abbreviations such as:

- "no." (meaning "number" or "number of"),
- "sec." (meaning "seconds"),
- "IEEE Trans. Big Data" (meaning the journal "IEEE Transactions on Big Data"), and
- "Proc. Web Search and Data Mining" or "Proc. WSDM" (meaning the conference proceedings for the conference Web Search and Data Mining). Not "Proc. of ...", just "Proc. ...".

By the way:

- *Abbreviations* look like: IEEE Trans. Comput., Discrete Math., etc.
- *Acronyms* look like: SIGMOD, VLDB, etc.

We also use periods in Latin abbreviations, such as:

- "etc." (meaning "and so on"),
- "i.e." (meaning "that is"),
- "e.g." (meaning "for example"),
- "et al." (meaning "and others"),
- "op. cit." (meaning "already cited"), and
- "p." (meaning "page") and "pp." (meaning "pages"), and so on.

We can write Latin abbreviations as either "et al." or "*et al.*" (in italics). But we must be consistent—choose one and stick to it. (I suggest not using italics; it's simpler.)

We can also just write the words in English. Doug West writes:

> *In formal mathematical writing, abbreviations (except as notation) are like contractions; it is better to avoid "i.e." and "e.g." altogether.*

Abbreviations to avoid

Do not write these:

- "w.r.t." meaning "with respect to";
- "s.t." meaning "such that" (or sometimes "subject to"); and
- "w/" and "w/o" meaning "with" and "without", respectively.

There is no advantage in writing these; there is a strong disadvantage: the reader might not know what they mean!

Just write it in words.

The space at the end of a sentence is different to the space after an abbreviation. This is a common problem when writing "et al.":

LaTeX input	compiles to
`Wang et al. \cite{Wang09}`	Wang et al. [1]
`Wang et al.\ \cite{Wang09}`	Wang et al. [1]
`Wang et al.~\cite{Wang09}`	Wang et al. [1]
`Xu et al. proved ...`	Xu et al. proved ...
`Xu et al.\ proved ...`	Xu et al. proved ...
`Xu et al.~proved ...`	Xu et al. proved ...

Two different spaces: "\ " is a breaking space (permits LaTeX to break the line in this place) "~" is a non-breaking space. Without one of these, LaTeX interprets the period as the end of the sentence, and adds a bigger space.

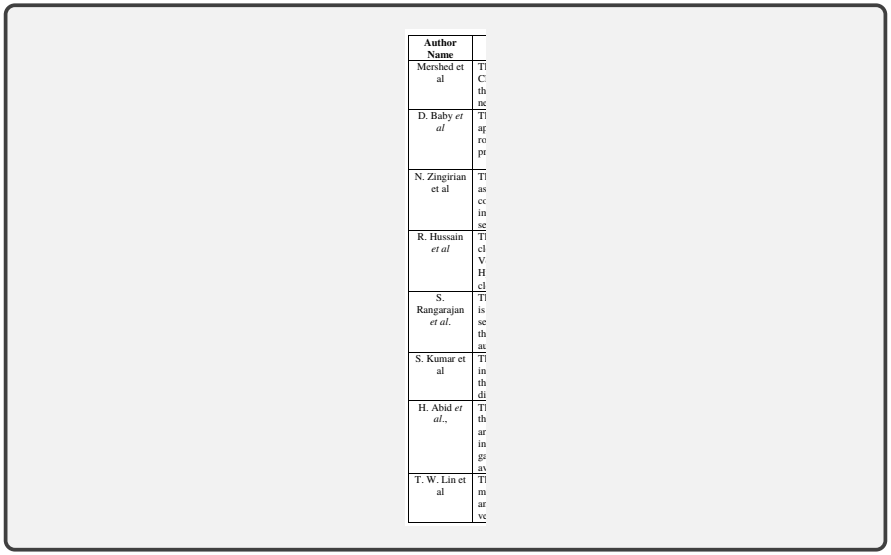
What's wrong with the punctuation?

> environments. Atsushi Manabe et al [23] discuss how to integrate the HPSS into a Grid architecture to handle petabytes of data produced by the Atlas experiment and to share such data among the regional collaborators. DataCutter is a middleware
>
> — Deng, Wang, *A Heterogeneous ...*, ACM SIGOPS Op. Syst. Review, 2007.

It should be "et al." not "et al". (It is not "et. al." either.)

(There's other minor problems with this snippet, but I just want to point this one out.)

What's wrong with the punctuation?



Author Name	
Mershed et al	Th Co th ne
D. Baby *et al*	Th ap ro pr
N. Zingirian et al	Th as co in se
R. Hussain *et al*	Th cl V H el
S. Rangarajan *et al.*	Th is se th au
S. Kumar et al	Th in th di
H. Abid *et al.*,	Th th ar in ga av
T. W. Lin et al	Th m ar ve

(a) The use of italics is inconsistent; and (b) the period is sometimes missing. (And it should just say "Authors".)

What's wrong with the punctuation?

> [3] W.Peter, E. G.Robert, G.Henry, et al. IP SAN – From iSCSI to IP-addressable Ethernet Disks. In *Proc. Of 20th IEEE/11 th NASA Goddard Conference on Mass Storage Systems and Technologies (MSS'03)*, San Diego, California, 2003, 189-194.
>
> — Deng, Wang, 2007 (op. cit.).

Here's the paper they're citing:

IP SAN – From iSCSI to IP-addressable Ethernet Disks

Peter Wang	Robert E. Gilligan	Henry Green	Jeff Raubitschek
Intransa, Inc.	*Intransa, Inc.*	*Intransa, Inc.*	*Intransa, Inc.*
Peter.wang@intransa.com	robert.Gilligan@intransa.com	henry.green@intransa.com	Jeff.raubitschek@intransa.com

Punctuation problems (there's *many* more errors):

- "W.Peter" should be "W. Peter" (with a space), and so on for the other authors. (Actually, the name is wrong: it should be "P. Wang".)
- The abbreviation "et al." means "and others", but there is only one other author (not "others").
- Don't need "Proc. Of" (or "Proc. of"), simply "Proc." is enough.

Commas

Commas are used to separate parts of a sentence, and indicate where a reader should pause.

There should never be a space before a comma, and ordinarily there's a space after a comma.

There's a lot to say about commas[2], and every rule I've seen gets broken repeatedly. Consequently, I'm unsure how to explain how to use commas. (I'll just list some good and some bad examples.)

[2] https://en.wikipedia.org/wiki/Comma

Comma usage: list separators

We use them to separate lists; there's two styles:

- planes, trains and automobiles; and also
- planes, trains, and automobiles.

The latter comma is called the *Oxford comma*[3] (more common in American English). Either is fine, as long as you're consistent throughout the paper.

> It keeps enough replicas of the data across different nodes in different fault domains to keep data durable within the stamp in the face of disk, node, and rack failures.
>
> — Calder, et al., *Windows Azure Storage ...*, Proc. SOSP, 2011.

[3] https://en.wikipedia.org/wiki/Serial_comma

Comma usage: separating two clauses

Examples:

> Algorithms developed under this assumption may not be applicable for analyzing data collected from heterogeneous settings, where the set of genes being monitored may be different and expression levels may be not directly comparable even for the same gene.
>
> — Zhang, Wang, *Mining Coherent Patterns ...*, Proc. CIKM, 2006.

> The rating matrix is usually extremely sparse with most entries undefined, since a user may only rate a few POIs.
>
> — Zhang, et al., *Trip Recommendation ...*, ACM Trans. Info. Syst., 2016.

Usually the structure is "Clause 1, Clause 2" where Clause 2 depends on Clause 1, but Clause 1 does not depend on Clause 2.

In most sentences, using commas like this is complicated by many things.

Comma usage: sentences starting with "Therefore", etc.

> Therefore, to increase the amount of storage, we deploy one or more storage stamps in the desired location's data center and add them to the LS.
>
> — Calder, et al., 2011 (op. cit.).

> However, the energy consumed at different time contributes differently to the cooling cost because the electricity price in the market may fluctuate significantly.
>
> — Wang, et al., *Leveraging Thermal Storage ...*, Proc. HotPower, 2011.

> In other words, data are transmitted while the sender and receiver run in parallel.
>
> — Wu, et al., *Whispers in the Hyper-Space ...*, IEEE/ACM Trans. Netw., 2015.

These are called "sentence adverbs" or "sentence adverbials"[4].

[4] `https://en.wikipedia.org/wiki/Disjunct_(linguistics)`

Comma usage: alternative to brackets

> The fundamental assumption is that the randomness and uniqueness of keys and IVs, relative to encrypted content, can be largely achieved, but that, as we will demonstrate, can be very difficult in modern storage systems.
>
> — Diesburg, et al., *When Cryptography Meets Storage*, Proc. StorageSS, 2008.

This is equivalent to:

> The fundamental assumption is that the randomness and uniqueness of keys and IVs (relative to encrypted content) can be largely achieved, but that (as we will demonstrate) can be very difficult in modern storage systems.

> *Enclose parenthetic expressions between commas.*
>
> — *Strunk & White, Section II.3*

I disagree with Strunk & White, and I think using brackets is better—it's easy to confuse the distinct comma usage.

Comma usage: mathematical expressions

Most of the time, LaTeX handles commas in mathematical expressions appropriately.

Side note: Don't write e.g.

> ... the number of nodes, n, satisfies ...
> ... the number of nodes (n) satisfies ...

and just write

> ... the number of nodes n satisfies ...

There's too much punctuation nearby mathematical expressions otherwise.

Colons

We use a colon ":" before we explain something in more detail. A good example is the following:

> First, the formulation is uniform: all memory models are defined via serial views, and memory models only differ in the serial views that they require.
>
> — Furbach et al., *Memory-Model-Aware Testing: A Unified Complexity Analysis*, ACM Trans. Embed. Comp. Sys., 2014.

We also use a colon before lists, such as

> We experiment with two iPhone models: iPhone X and iPhone 5s.

There's no space before the colon, and there's space after the colon.

We avoid colons if they're unnecessary. E.g., we shouldn't use a colon in "The parameter α satisfies: $\alpha \leq t^2$."

Colons (cont.)

A colon is commonly used in paper titles, such as

> *When encryption is not enough: privacy attacks in content-centric networking*

and

> *TrueErase: Leveraging an Auxiliary Data Path for Per-File Secure Deletion.*

Here it separates the eye-catching part of the title, or the software name. We might think of this as "title: subtitle".

In mathematical expressions, we might write $\{n^2 : n \equiv 2 \pmod 4\}$ and $f : \mathbb{R} \to \mathbb{R}$. The spacing is better if we write `\colon` instead of `:`, but probably nobody (except me) would notice.

LaTeX input	compiles to
`$\{n^2: n \equiv 2 \pmod 4\}$`	$\{n^2 : n \equiv 2 \pmod 4\}$
`$\{n^2 \colon n \equiv 2 \pmod 4\}$`	$\{n^2 \colon n \equiv 2 \pmod 4\}$

Colons are used in ratios, like 1 : 3. But ratios should not be used in papers.

Semicolons

A common use for semicolons are when we have two consecutive sentences, Sentence 1 and Sentence 2, and the context of Sentence 1 is used in Sentence 2. This is a good example:

> In BS, each data item is stored as a collection of blocks that are kept in pseudorandom locations; the server sees only a super-set of the locations where the data item is stored, rather the exact set of locations.
>
> — Wang, Zhao, *Secure Dynamic SSE ...*, Proc. ASIA CCS, 2016.

If we made this into two sentences, the second sentence might begin "In this way, the server sees only ...". However, using a semicolon

- we reduce the number of words (i.e., it's more succinct) and
- we avoid vagueness with "this" (referring to the previous sentence).

It's possible to make this into one sentence, e.g., "... are kept in pseudorandom locations, so the server sees only ...". (It's often possible to glue the two sentences together like this.)

Semicolons (cont.)

In complicated lists (e.g., A, B, and C), we may use semicolons instead of commas (e.g., A; B; and C). A good example is:

> Some aspects of the user model examined include
>
> – the number of times a person must enter an encryption key per session;
> – the ease with which the method is invoked; and
> – the number of encryption keys or passwords a person or a system must remember.
>
> — Diesburg, Wang, *A Survey of Confidential ...*, ACM Comput. Surv., 2010.

In complicated lists, a comma would be ambiguous: does it separate two list items, or is it part of the sentence?

What's wrong with the punctuation?

Putting aside other grammar problems, what's wrong with the punctuation here?

> If the integrity of the overwrited data is correct after each data updating operation, the original data is overwrited by the randomly generated data blocks and become unrecoverable forever; Otherwise, the cloud misbehaves during the deletion process.
>
> — Luo, et al., *Enabling Assured Deletion ...*, Proc. SCC, 2016.

The word "otherwise" should start with a lowercase "o".

> This demo is to show a prototype of AI3. The demo has four parts: 1) ASP-independent data management in AI3; 2) ASP-independent management of users' social relations in AI3; 3) inter-domain data transport and user roaming; 4) real-time communications by using AI3.
>
> — Zhang, et al., *Enabling Assured Deletion ...*, Proc. SIGCOMM, 2014.

We need an "and", since we're writing a list "A, B, C, and D."

I also recommend avoiding "1)" because of the missing bracket, and "(1)" because it can be confused with equations numbers. In my opinion, the best choice is "(a)", etc.

What's wrong with the punctuation? (cont.)

> The improvement of wide-area connectivity is constrained by factors such as how quickly we can dig ditches to bury fibers in the ground; and the cost of furnishing "last-mile" wiring can be prohibitively high.
>
> — Wang, et al., *Turning the Postal System ...*, Proc. SIGCOMM, 2004.

The sentence is just a normal sentence with an "and" in it—we should use a comma or nothing, rather than a semicolon.

If we really wanted to use a semicolon, we could write "... in the ground; the cost of ...", but there's no point.

What's wrong with the punctuation? (cont.)

> **ALGORITHM 1:** Task Scheduling of Partial Garbage Collection Tasks
>
> **Input**: A set of data request task ($V_T = \{T_1, T_2, \ldots, T_n\}$), an empty set of garbage collection tasks ($V_G = \emptyset$), garbage collection threshold (ρ_{th}), and the number of free pages (Φ).
> **Output**: A task schedule.
>
> 1 **for** *each* $T_i \in V_T$ *is executed* **do**
> 2 **if** $T_i \in T_w$ **then**
> 3 **if** $V_G \neq \emptyset$ **then**
> 4 $r_{G_j} \leftarrow r_{T_i} + e_{T_i}$;
> 5 *remove_gc_task*(G_j);
> 6 **end**
> 7 **if** $\Phi < \rho_{th}$ **then**
> 8 $PBN_{victim} \leftarrow$ *get_victim_block*() ;
> 9 $V_G \leftarrow$ *generate_gc_tasks*(PBN_{victim}) ;
> 10 $r_{G_j} \leftarrow r_{T_i} + e_{T_i}$;
> 11 *remove_gc_task*(G_j) ;
> 12 **end**
> 13 **end**
> 14 **end**
>
> — Zhang, et al., *Lazy-RTGC: ...*, ACM Trans. Design Auto. Elec. Syst., 2015.

We don't use semicolons to end lines in pseudocode—we're communicating with a human, not a computer.

Likewise, we don't use other computer notation, e.g., (a) `2.90E-05` to denote 2.90×10^5, and (b) "==" for equals.

Dashes, hyphens, subtraction

These are all different:

- *Hyphen*: A single dash, e.g., "left-hand side" (typeset `left-hand side`), such as for compound adjectives.
- *En dash*: We use a double dash for page numbers, e.g.:
  - pp. 11–20 (typeset `pp.~11--20`) and
  - lists of equations (1)–(5) (typeset `\eqref{..}--\eqref{..}`).
- *Em dash*: Triple dash "---" is a way separating parts of a sentence (an alternative to brackets and commas):

  Most authors—including the present authors—use θ to denote an angle.

Notice: no spaces for any of these.

Subtraction and negatives in mathematics are different from the above, e.g.:

$$1 - 1 = 0 \quad \text{and}$$
$$-1 \neq 0.$$

Same length, but different spacing (LaTeX does this for you).