The Craft of Scientific Writing

Revised Edition, 2013-2014

Gershom (J.M.L.) Martin,
gershom@weizmann.ac.il
Kimmelman 361

Unit 5b

- Punctuation (part 2)

(22) One, hand, clapping

- **The "opening and closing" rule:**
  A parenthetical phrase occurring in the middle of a sentence need not *if it is simple* be separated by punctuation at all. If it is separated, as in this example, by commas or (as may happen in the case of lengthier phrases) by brackets then these symbols — and this also applies to dashes — should always *occur in pairs*.

For want of a comma…

- For want of a nail the shoe was lost.
- For want of a shoe the horse was lost.
- For want of a horse the rider was lost.
- For want of a rider the battle was lost.
- For want of a battle the kingdom was lost.
- And all for the want of a horseshoe nail.
  (Traditional English nursery rhyme)

An extreme example: 2nd Amendment to the US Constitution (1791)

- How many decades of heated debate have there been about commas in the following "simple" English sentence:

- A well regulated militia, being necessary to the security of a free state, the right of the people to keep and bear arms, shall not be infringed.
  - Clearly neither Schoenfeld nor yours truly, wrote this sentence ;-)

Unpaired commas or dashes

- Parenthetical clause at the beginning or end of a sentence
  - "Under forcing conditions, dehydration occurs."
  - "All our science, measured against reality, is primitive and childlike — and yet it is the most precious thing we have." (Albert Einstein)
- Commas preceding conjunctions (such as *if*, *and*, and *but*)
  - No comma if both clauses are equally important
    - The isomers were formed in high yield and proved easy to separate.
  - Subtle emphasis on 2nd clause possible by comma
    - The isomers were formed in high yield, but were difficult to separate.

Commas in enumeration

- Many newspapers and weeklies prefer not to use the serial comma
  - "Liberty, equality and brotherhood."
- Almost all American scholarly journals (and many British ones) favor the serial comma (a.k.a. "Harvard comma", a.k.a. "Oxford comma")
  - "Liberty, equality, and brotherhood."

Anti-ambiguity commas

- Schoenfeld: “Anti-garden-path commas”, from English expression “to lead somebody up the garden path” (to mislead somebody)
  - Unfortunately diazotized anthranilic acid decomposed…
  - Unfortunately, diazotized anthranilic acid decomposed…
  - Unfortunately diazotized, anthranilic acid decomposed…
  - However examplitol was treated… [No matter how]
  - However, examplitol was treated

Dixit *Lynne Truss*:

A panda walks into a café. He orders a sandwich, eats it, then draws a gun and fires two shots in the air.

"Why?" asks the confused waiter, as the panda makes towards the exit. The panda produces a badly punctuated wildlife manual and tosses it over his shoulder.

"I'm a panda," he says at the door. "Look it up."

The waiter turns to the relevant entry and, sure enough, finds an explanation.

"Panda. Large black-and-white bear-like mammal, native to China. Eats, shoots and leaves."

- So punctuation really does matter, even if it is only *occasionally* a matter of life and death.

(Schoenfeld, Chapter 6) English scientists secretly practice German vice!

- German (and its sister language, Dutch) are notorious for long "sausage words"
  - Hauptwortkombinationenzusammenstellungs-bedürfnis (compound noun assembly mania)
  - Wapenstilstandsonderhandelingen (armistice negotiations): longest entry in Dutch dictionary
- But what about "proton magnetic resonance spectroscopy literature survey" or "cyclic ligand planar nitrogen array"?
  - Only *seem* less unwieldy because of spaces
  - At least in Dutch and German, the "links" in the sausage all have to be nouns

Linguistic "peptide bond"

- Common and indispensable in Chemical English as a device for expressing a subtle relationship ("protein crystallography", "trial run", "hydrolysis experiment", …)
  - Even in everyday English: "Vice squad investigates call girl racket."
- As a rule: "hydrolyze" overly long "peptide chains" like "ring junction carbon environment differences" to, e.g., "differences in the environment of the ring junction carbons" or "differences in the environment of the carbons at the ring junctions"
  - Exception 1: names of chemical compounds are to be regarded as single "amino acids"
  - Exception 2: overly long words or chemical names should not be used in this manner at all

(7) Of nuts, muttons, and shotguns

- The hyphen (-) has three “big brothers” in typography
  - The en-dash (–, a.k.a. “nut”)
    - In LaTeX, type --; in M$ [fourletter]Word, Alt -
  - The em-dash (—, a.k.a. “mutton”)
    - In LaTeX, type ---; in M$ […], Alt Shift -
  - The minus sign (in formulas, same width as an en-dash but does not fill space entirely).
    - In LaTeX, just - inside math mode; in M$ Word, use Equation Editor

En-dash ("nut")

- Width of letter "n" (hence the name)
- For numerical ranges (20–30 kcal/mol, pp. 184–185)
- For single bonds in in-text chemical formulas
- Between *different* proper names
  - Diels–Alder reaction (after Diels & Alder)
  - *But*: Lennard-Jones potential (after J. W. Lennard-Jones)
  - *And*: the Feynman–Gell-Mann theory (after Richard P. Feynman and Murray Gell-Mann)

Em-dash ("mutton")

- Width of a letter "m" (hence the name)
  - should be surrounded by spaces
- Used to draw attention to a parenthetical clause
  - German: "Gedankenstreifen"
  - Dutch: "aandachtstreepje" (attention dash)
- In my personal opinion, the most unfairly neglected punctuation mark by inexperienced writers
  - Yours truly may be overly fond of it ;-)
- If no reason for emphasis, "slaughter" the muttons and replace by simple commas

"Shotgun marriage" of binary compound nouns

- "In the marriage of two nouns, the hyphen is the shotgun"
  - Hyphen is used if pair seems ill-matched and/or unfamiliar at first sight
  - Once the link between the two words has become familiar, the hyphen can be dispensed with: the words then coalesce (e.g., "bookcase") or stay apart ("filter paper"), largely dependent on which looks better/is more widely adopted ("website" or "web site"? "wavefunction" or "wave function"?)
    - Rule of thumb: British English retains hyphen in many binary compound nouns where American English would either coalesce or leave a space.

Ternary compound nouns

- If even a remote ambiguity exists, add a hyphen where necessary
  - Near-ultraviolet spectrum
  - “Complex ion mechanism” means one thing
    - A complicated mechanism involving ion
  - “Complex-ion mechanism” means another
    - A mechanism involving complex ions
- … there is no substitute for human common sense (אין תחליף לשכל הישר)!

Quaternary and higher compound nouns

- Usually best avoided in the first place (cf. “German vice”)
- Sometimes number of “links” in sausage can be greatly reduced by judicious use of acronyms (e.g., “DFT literature survey” instead of “density functional theory literature survey”)
  - Obviously, acronym must have been properly defined previously
- Speaking of acronyms: define *all* of them on first use, except for extremely common ones like IR, NMR,…
  - “DFT” may be “density functional theory” to everybody we know, but it means “discrete Fourier transform” to some people and everybody *they* know

(8) Tetravalency of carbon disproved!

- Many scientific terms derive from Latin or Greek
- Common sense suggests: do not mix them up in a single word: no “sha`attnez” (שעטנז)
  - WRONG: polylingual OR multiglot
  - RIGHT: multilingual OR polyglot
  - “unimolecular reaction” and not “monomolecular reaction”
    - although “monomolecular film” seems to be somewhat accepted

(8) Tetravalency of carbon disproved!

- Normally, Latin and Greek should not be mixed in a single word (cf.איסור שעטנז Lev. 19:19 and Deut. 22:9-11): one is multilingual or polyglot, but not "polylingual" or "multiglot" [sic]
- Exceptions in Chemical English:
  - In chemical nomenclature, Greek numerical prefixes (mono-, di-, tri-, tetra-, penta-, hexa-, …,poly-) are *always* preferred over Latin ones (uni-, bi-, tri-, quadri-, …, multi-)
  - Cf. also bis-, tris-, tetrakis-, pentakis-, …
  - bivalent, trivalent, tetravalent, pentavalent, … are accepted Chemical English
  - "unimolecular reaction", not "monomolecular reaction"
    - although "monomolecular film" seems to be somewhat accepted

(23) Alphabetical disorder

- **Roman type= regular type**
  - Origin of name quite clear from examples
  - Upper case: like inscriptional capitals
  - Lower case: bit like Roman "bookhand"

Italics

- Originally invented by pioneering Italian printer Aldo Manuzio (Latin: Aldus Manutius, 1449-1515)
  - Original purpose: allowing more compact printing for pocket books
  - Modern purposes:
    - (A) for *emphasis*
    - (B) scalar variables in mathematical expressions
      - $P V = n R T$
    - (C) qualifiers in chemical structure names (*exo*, *endo*, *cis*, *trans*)
    - (D) for foreign words and expressions that have *not yet* become "naturalized citizens" in English
      - *Ab initio*, *in situ*, *force majeure*, *allegro con brio*
      - but: Per capita, laissez-faire, vice versa

Use of foreign-language words in English writing

- On the one hand, strikes many readers (especially Americans) as conceited, snobbish, *je ne sais quoi*
  - Common "abusage" by Europeans (especially with Latin phrases) and especially by native speakers of Dutch or other "boutique" languages (who are generally, of necessity, partly or wholly fluent in several other languages and assume that the same is true of their readership)
    - Cfr. Israelis saying (while otherwise speaking Hebrew) "as a matter of fact" instead of "*le-ma`ase*". Ehud Barak: "Ke'inyan shel uvda"
  - Schoenfeld's recommendation: avoid if a perfectly good English alternative exists
- On the other hand, sometimes nothing conveys the meaning better than *davka* the foreign word
  - In fact, Latin/Greek derived scientific lingo is often still more "internationally" understood than "native" English equivalent
  - Examples: *schadenfreude*, *pièce de résistance*, *apparatchik*, *ars*
  - Sometimes the word subtly changes meaning
    - *Verboten* in German = exact synonym of prohibited/forbidden
    - *Verboten* when used in English = prohibited by narrow-minded, dictatorial, bureaucratic *fiat*.

A comment: Calques (loan translations) in formal English writing

- Calque/loan translation=literal translation of foreign term or phrase. Many have become part of the standard vocabulary.
- Hebrew examples:
  - names of elements (Wasserstoff->מימן, Stickstoff->חנקן), Sauerstoff->חמצן
    - גן ילדים from Kindergarten, obsolete שח-רחוק from Fernsprecher (itself a German calque of telephone)
    - תפוח אדמה(Fr.: *pomme de terre*), כדורגל
- English examples: marriage of convenience, New Wave (*nouvelle vague*), point of view (*point de vue*), [learn] by heart (*par cœur*), masterpiece (*Meisterstück*, *meesterstuk*), Devil's advocate (*advocatus diaboli*), cookie (<= Dutch *koekje*, "little cake") Gospel ("Good spell" [=news], Old English calque of the Greek *evangelion*), pineapple (<= Dutch *pijnappel*), antibody (*Antikörper*), homesickness (Ger. *Heimweh*, D. *hijmwee*), loanword (*Lehnwort*), "the straw that broke the camel's back" (often: "the last straw")
- Nonstandard calques of Hebrew/Dutch/French/… expressions: can be very colorful *when no corresponding "native" English expression exists* (otherwise just looks goofy). Use *very* sparingly in formal writing.

To serif or not to serif

- Serif fonts like Times, Baskerville, Garamond
  - On paper:
    - remain easily readable even at small point sizes
    - At large point sizes, look “newspaper-like” and old-fashioned
    - Many journals typeset body text in serif
    - On computer displays (including LCD projectors):
    - Comparatively hard to read due to “quantization error”
      - note typical resolutions 72–96 dpi, compared to 600 dpi on laser printers, or 1200 to 2400 dpi in professionally printed books or journals
      - My recommendation: use only at large point sizes
      - Update 2014: as “retina displays” become more common, serif fonts maybe again the more readable choice on computer monitors.

To serif or ***not to serif***

- Sans-serif fonts like Helvetica, Arial, Futura, Geneva
  - On paper:
    - Much less readable at small point sizes
    - At large point sizes, look more “modern”
    - Note: many journals typeset titles, captions, etc. in sans-serif fonts at larger point sizes and/or boldfaced
  - On computer displays (including LCD projectors):
    - More readable, because lower resolution will still render all features correctly
    - Generally preferable for presentations
  - General features:
    - Italics are generally *oblique* fonts, rather than true italics
    - Can be confusing for symbols (I vs. l) — mathematical equations generally typeset in serif font for that reason

A compromise: "humanist" fonts

- Sans-serif fonts with some calligraphic features, such as Optima, Gill sans, Verdana, Trebuchet…
  - In print, almost as readable as Times, but looks more "modern"
  - On screen at small point sizes: as hard to read as Times (or harder)
  - On screen at large point sizes: work really well (the present slide is in Optima, even though most of the other slides were done in Trebuchet)

Special-purpose fonts

- Monospaced fonts like Courier and Monaco: often used for examples of computer input and output
- **Dingbats: printer slang for fonts made up of special symbols, ideograms, or printer's ornaments. Example (Zapf Dingbats):**
  ❑◑❄▼❒▼❙◆◆❒❑❇◆○✳✳✳❄❖✪○➪➽➽✓✖✕➽✓✔✕
  ✖✗✘✗✙✐✁✠✁❀☎☛✆✿❀✉
- Symbol font (obsolete): !∀∃ΑΒΧΔΕΦΓΗΙϑΚΛΜΝΟΠΘΡΣΤΥςΩΞΨΖ[∴]⊥_ ‾αβχδεφγηιφκλμνοπθρστϖϖωυξψζ←↔↑→↓etc.

What the PC is Unicode? (1)

- Computer representation of text: de facto standard was 7-bit ASCII (American Standard Code for Information Interchange)
  - Codes 0-31 reserved for control characters
    - 9=TAB, 10=LF, 13=CR, 14=hard page break, etc.
    - 32=space, 127=DEL
    - 33–126= !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
    - For standard English, nothing else is needed
- 8-bit ASCII: codes 128-255 used for foreign characters and symbols
  - Every country developed its own => Chaos Manor

Unicode and UTF-8

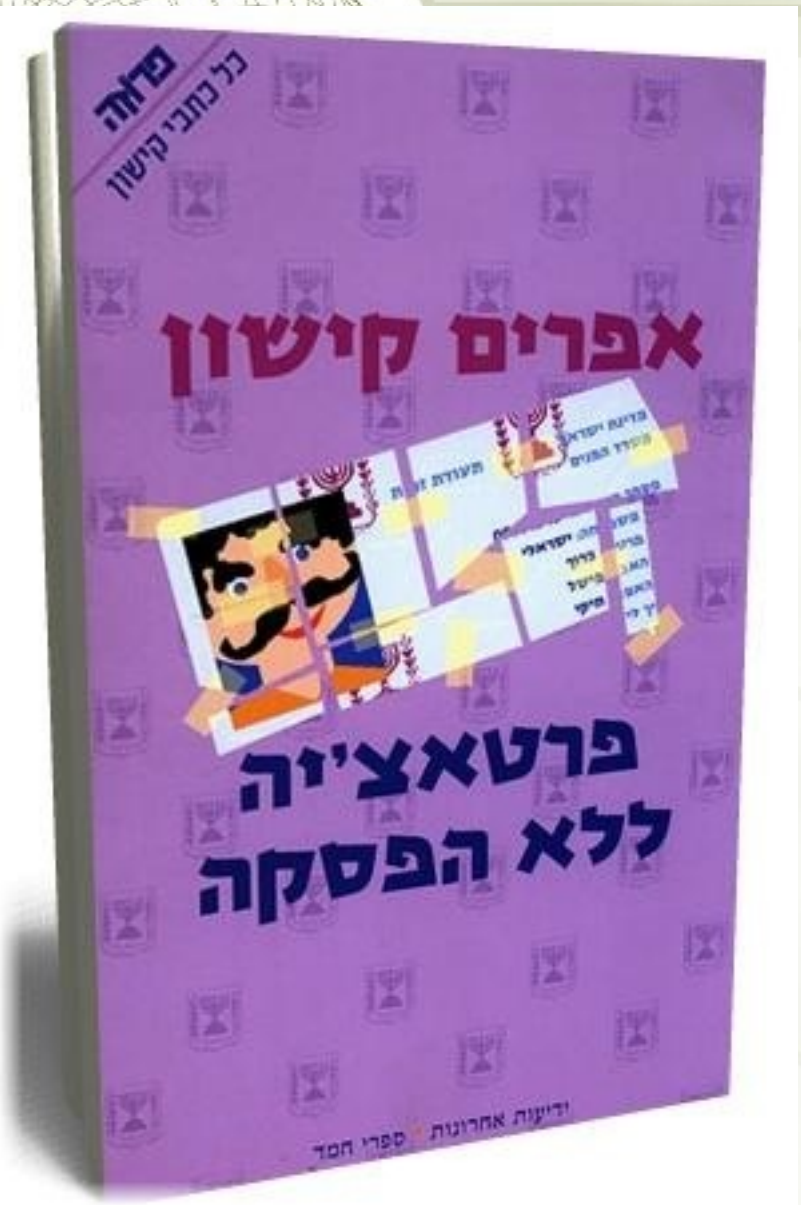
- Seeks to include all letters, ideograms, symbols,… in present or past human use
- Obviously 1-byte codes out of the question, even 2-byte codes will eventually run out
- Most common representation: UTF-8
  - Codes 0–127 (00–FF): 1-byte, identical to ASCII
  - Codes 128–255: 1st byte of multibyte Unicode sequences
  - More details at:
    http://en.wikipedia.org/wiki/UTF-8
    http://unicode-table.com
    http://www.utf8-chartable.de

(24) Is you is or is you ain't my data?

- Foreign words imported into English: plural generally follows the original language

Singular	Plural	Source
lingua franca	lingue franche	Italian
kibbutz	kibbutzim	Hebrew
festschrift [note:l.c.]	festschriften	German
Beduin	Beduin	Arabic
spectrum	spectra	Latin
datum	data	Latin
appendix	appendices	Latin
vertex	vertices	Latin
matrix	matrices	Latin
phenomenon	phenomena	Greek
criterion	criteria	Greek
medium	media	Latin
phenomenon	phenomena	Greek

"Partatch"/slapdash



- Many slapdash writers do not seem to know that criteria, data, media, and phenomena are plurals
  - And may indeed invent bogus plurals like "criterias", "phenomenas",…
  - Creates *unbelievably* sloppy impression
  - The only more irritating mistake I know may be confusing "its" with "it's".
- Consider this: If your research isn't important enough to take the trouble to write up properly, why should it be important enough for people to read it?

In English, no rules without exceptions?

- Some borrowings have two plurals
  - **Indices** (mathematical), but **indexes** (all other senses)
  - **Formulae** and **formulas** are both acceptable (although only "formulae" is etymologically correct)
- Other borrowings are petrified plurals turned singular
  - **Agenda** is originally the plural of "agendum" (something to act upon). Now singular: plural=**agendas** (hidden or other)
- Yet other borrowings (generally more recent ones) always use the English plural
  - Protons, neutrons, electrons,… not prota, neutra, electra,…

Foreign words with English endings

- Some borrowings (notably from German) acquired English endings, and thus also English plurals
  - Eigenvalues (not *Eigenwerte*)
  - Eigenfunctions (not *Eigenfunktionen*)
  - Eigenvectors (not *Eigenvektoren*)
- A curious example: ansatz (German: original plural *Ansätze* or *Ansaetze* — nouns are capitalized in German) could be ansaetze or ansatzes
  - The exponential *ansatz* $\Psi = \exp(-\hat{T})\psi_0$