

# 'Difficult' Words

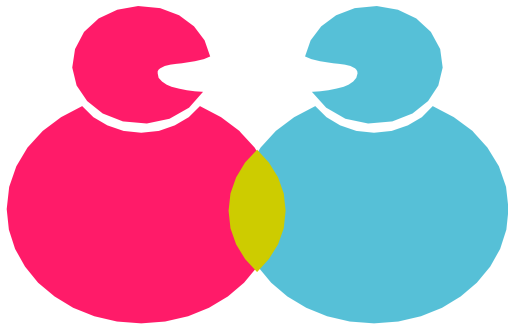
*Humans and NLP highlight different  
specialist terms in diverse academic fields  
and rank quality differently*

---

Helen McCombie

*Bureau de Traduction de l'Université*

*Université de Brest – Université de Bretagne Occidentale*



## Bureau de Traduction de l'Université (BTU)

Main activities: Translation, Interpreting, **Academic English Editing**

All types of document or media produced by the university and partner institutions:

Course materials, Websites, Brochures, Videos, Certificates,  
**Scientific papers...**

# Specialist terms

## How can we identify them in a text?

**Automatic strategies**

**Frequency (probability of occurrence)**

**Complexity (length)**

**Used only in a particular field**

**Already defined in reference works**

**Role/ Profile among the parts of speech**

**Own strategies**

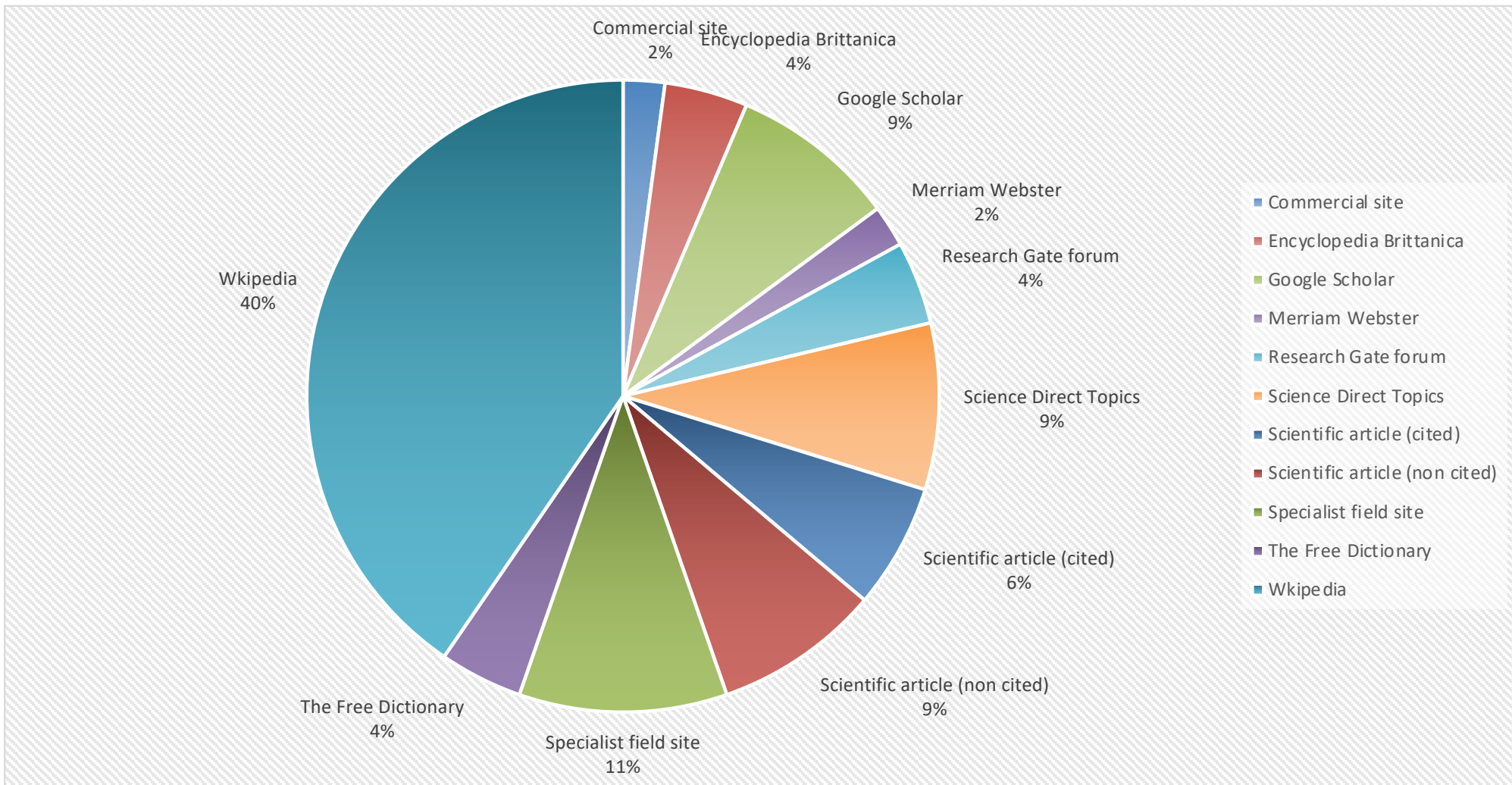
**Practically, in correction-revision they stand out as aids or hurdles to understanding**

**Need for understanding to edit out errors, improve communication and thus... understanding**

# Checking terms

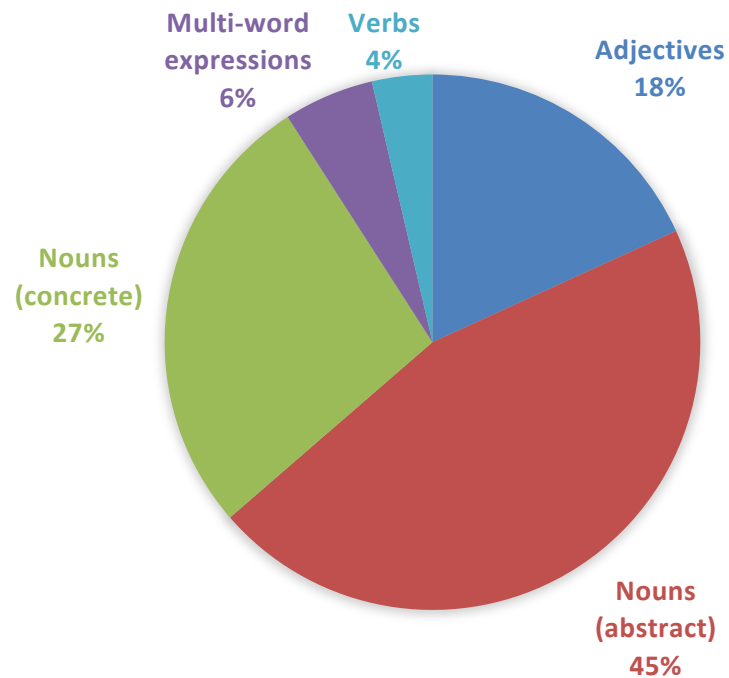
- Needs for checking
  - Spelling
  - Meaning
  - Usage (use in specific field, form, grammar etc.)

What tools do I use for this?

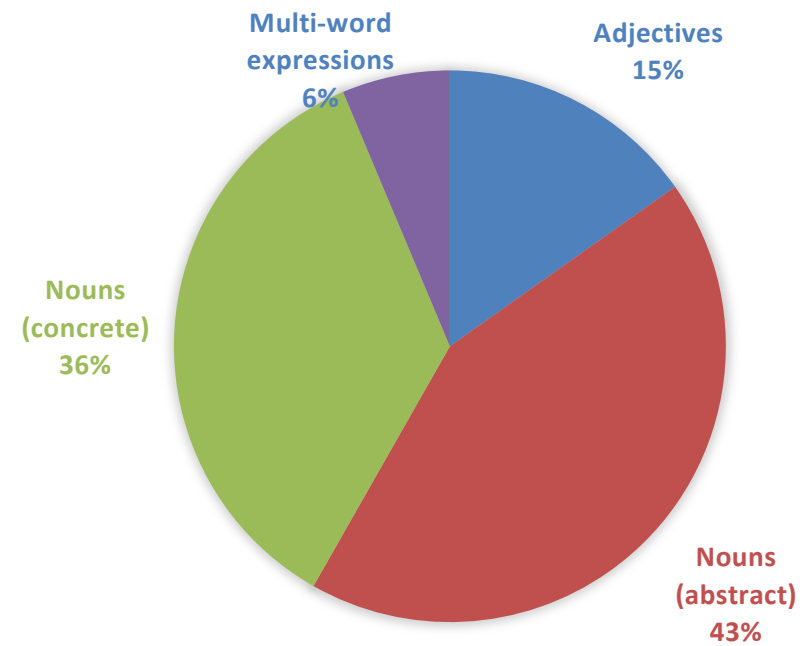


# What word types?

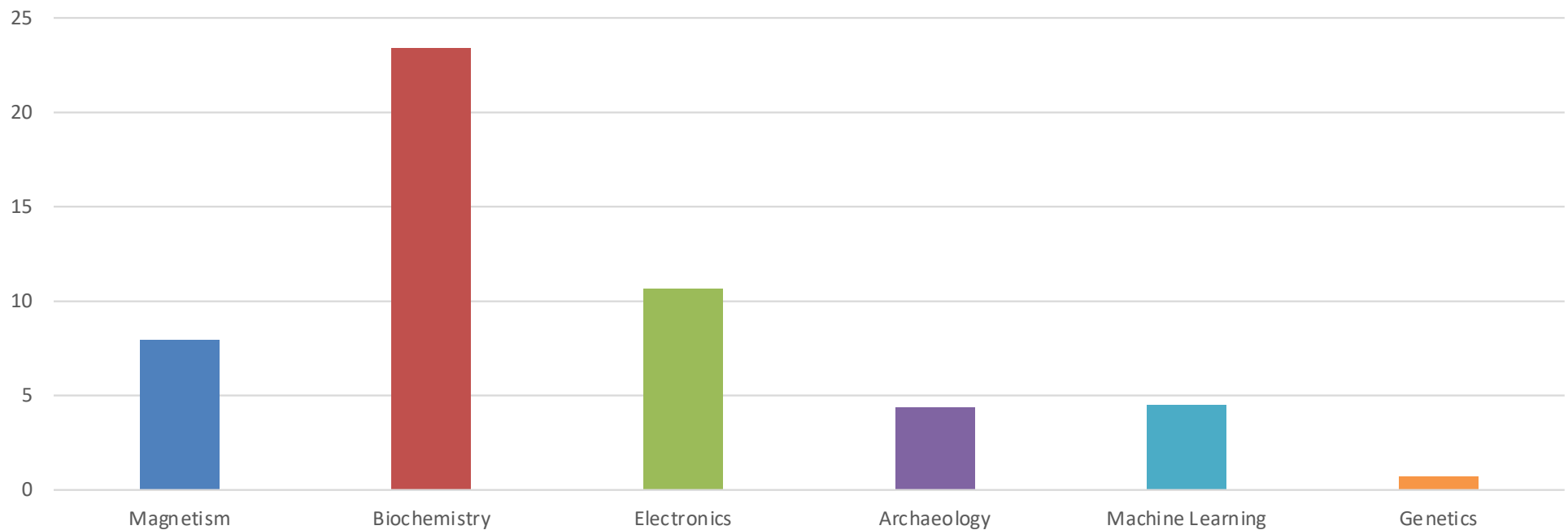
## Electronics



## Archaeology



# Number of words to be checked



A spellchecker (e.g. MS Word) underlines:

- Mistakes
- Rare terms (inc. those created by the author),
- Proper nouns (inc. citations, manufacturer names),
- Equations, acronyms, foreign language words or symbols

Add to MS Word's 'dictionary' (ignore) or define and add to own glossary

But what can other digital tools offer?



## Other help from digital tools

- ‘Term collectors’ e.g. TBXtools, TexSIS, ATR4S...  
for glossary building, term bases etc. Not relevant to my needs.
- Commercial tools: Trinka, Writeful, Paperpal  
trained on academic corpora to ignore or correct specialist terms
- ‘Search your own biblio’: Ref’n’Write (non AI)
- Chat GPT...

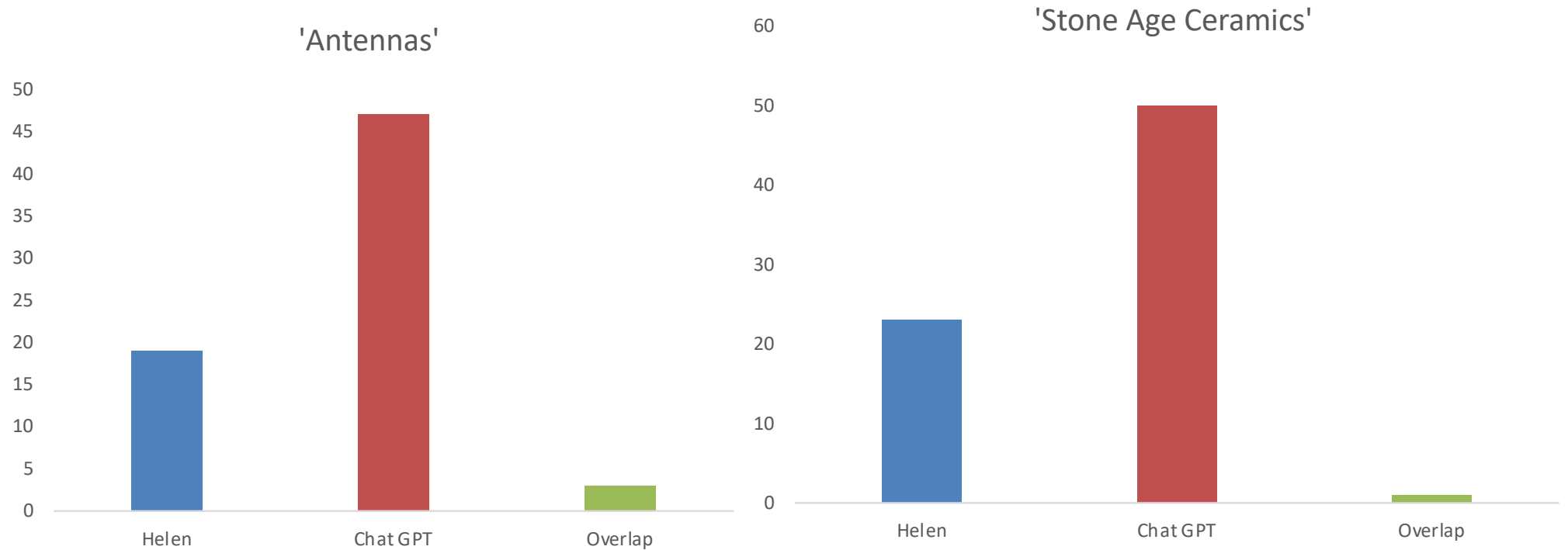
# Chat GPT

‘Please find the specialist terms in this text and give your criteria.’

- Chat GPT isolated more terms than I did
- When I asked it to find the same number or the ‘most specialised’ it took in order of appearance, i.e. no prioritisation

Specialist terms are technical terms that are specific to a particular field of study or subject area. The criteria used to identify specialist terms in the given text include:

1. **Technical language:** The use of language that is technical and specific to a particular field of study.
2. **Jargon:** The use of jargon, which is language that is specific to a particular profession or group of people.
3. **Complex concepts:** The use of complex concepts that are not commonly understood outside of a particular field of study.
4. **Uncommon words:** The use of uncommon words that are specific to a particular field of study.



- Repeat requests can give different results (e.g. sometimes chemical names, place names)
- Chat GPT found one adjective and the rest were nouns or nominal groups
- It did not detect 'common words used uncommonly': e.g. 'dig', 'occupation', 'fabrication'
- When I re-entered the list alone and asked for criteria, it provided definitions of each

- 1. Polycrystalline materials** - This term **is specialized because** it refers to materials that are composed of multiple small crystals. These materials have unique properties due to their grain boundaries and the interaction between the grains.
- 2. Anisotropy** - This term is specialized because it refers to the directional dependence of a material's physical properties. Anisotropic materials have different properties in different directions, which can be important in many applications.
- 3. Sintering treatment** - This term is specialized because it refers to a specific method of processing materials. Sintering involves heating a material to a high temperature to cause particles to fuse together, which can improve the material's properties.
- 4. Magnetostatic energy** - This term is specialized because it refers to the energy associated with the interaction between magnetic fields and materials. It is an important concept in the study of magnetic materials.
- 5. Magnetic moments** - This term is specialized because it refers to the magnetic dipole moment of an object. This concept is fundamental to the study of magnetism and magnetic materials.

## Specific AI tools for correcting academic writing

- Writefull, initially trained on Google Scholar content
- Paperpal, initially trained on Open access journals\*
- Trinka, initially trained on correction edition work on papers commercially edited by humans at Enago

All help mainly with style and grammar  
(rather than concepts).

Skip or correct specialist terms

# 'Preflight' by Paperpal

- Offered to publishers as a partnership to improve submission quality. Fee for author.
- Publishers specify criteria e.g. Abstract length, which the tool tests along with grammar etc.

# Your results

## Minor issues

Your manuscript can be submitted as it is. However, it is highly recommended that you address the issues listed before submitting. Your manuscript will have a higher chance of being accepted once these improvements have been made.

Get an edited Word document with language suggestions, for a one-time fee.

### Metadata

✓ Keyword list looks good

### Figures and tables

✓ All figures and tables have citations

✗ Figure legends are out of order

✓ Table legends are in order

### Counts

✗ Abstract is too long

### References

✓ Reference count is good

✓ All references have citations

✓ No use of self citing detected

### Language

✓ Issues found in Language quality, writing errors !

✓ Issues found in Language quality, mechanics and style !

✗ Issues found in Language quality, readability suggestions !

✓ Issues found in Language quality, vocabulary !



# When I gave published open access papers to Preflight and they were classed as poor. Why?

- Grammar assessed on a per sentence basis.  
e.g. misuse of articles 'the'
- Minor and major problems scored similarly.
- But, sometimes, there really are errors.

## Your results

### ■ Critical issues

Your manuscript has a very high risk of being rejected due to language quality or technical issues. We strongly suggest that you address critical issues before submitting your paper.

Get an edited Word document with language suggestions, for a one-time fee.

∨ See suggestions

#### Metadata

✓ Keyword list looks good

#### Figures and tables

✓ All figures and tables have citations

✓ Figure legends are in the proper order

✓ Table legends are in order

#### Counts

✓ Abstract length is good

#### References

✓ Reference count is good

✓ All references have citations

✓ No use of self citing detected

#### Language

✗ Issues found in Language quality, writing errors !

✗ Issues found in Language quality, mechanics and style !

✗ Issues found in Language quality, readability suggestions !

Readability suggestions covers issues like conciseness, redundancy, transition and flow, and suggests improved phrasing. We have identified 155 issues in this category.

#### ! CRITICAL CHECK

✓ Issues found in Language quality, vocabulary !

## Suggestions

We found issues to be resolved to make your manuscript fully submission-ready.

ABSTRACT Mobile instantaneous messaging (MIM) services significantly facilitate personal and business communications, inevitably consume substantial network resources, and potentially affect the network stability. In this study paper, we aimed to understand the traffic nature of MIM traffic in cellular networks. Specifically, in order to reach credible conclusions, our research considered takes the account of practical measurement records of MIM services from China Mobile at two different levels. First, a data-set of individual message level (IML) traffic is exploited and reveals power-law distributed message length and lognormal distributed interarrival time, the heavy-tailness of which completely diverts from the geometric model and the exponential models recommended by the 3rd generation partnership project (3GPP). Second, another data-set considers the statistical pattern of aggregated traffic within one whole base station, and demonstrates the accuracy of  $\alpha$ -stable models for the aggregated traffic. Furthermore, the results verify it verifies that the  $\alpha$ -stable models are suitable for characterizing the traffic in both the conventional fixed core networks and the cellular access networks. Finally At last, using with the aid of the generalized central limit theorem, we build up a theoretical relationship between the

Consistency: Different forms of this word have been used in the text. 'interarrival' no space [1 time] 'inter-arrival' with hyphen [12 times] Please pick one style and use it consistently throughout the text.

\$25

Download the full edited document with 322 more suggestions

ⓘ This document was generated automatically by AI services. It is not meant to replace human evaluation. Please be aware that language checks differ depending on the individual or tool performing the edits and must be interpreted carefully.

# Do these papers have errors and if so why were they accepted?

- Are errors exaggerated by the AI?
- Was the AI trained on the journal texts specifically? (One would expect more false negatives)
- Neither author nor reviewer are native English speakers?
- Reviewers read for content and meaning, not language use?

# Final thoughts

- Tools do not aim to understand, although they can aim to improve understanding through imitation
- If authors train themselves on their own biblio and this is flawed then the mistakes are perpetuated. Likewise for an AI with specialised training text – so, is it better to cast a larger net and
- Is the persistence of poor English really a problem if meaning is conveyed?