

## *Hypothesis, analysis and synthesis: it's all Greek to me!*

Ioannis Iliopoulos<sup>1</sup>, Sophia Ananiadou<sup>2</sup>, Antoine Danchin<sup>3,4</sup>, John P. A. Ioannidis<sup>5</sup>, Peter D. Katsikis<sup>6</sup>,  
Christos A. Ouzounis<sup>7</sup>, Vasilis J. Promponas<sup>8</sup>

...

<sup>1</sup>Division of Basic Sciences, School of Medicine, University of Crete, GR-71110 Heraklion, Greece

<sup>2</sup>School of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK

<sup>3</sup>Institute of Cardiometabolism & Nutrition, Hôpital de la Pitié-Salpêtrière, 47 blvd. de l'Hôpital, F-75013 Paris, France

<sup>4</sup>School of Biomedical Sciences, Li Kashing Faculty of Medicine, University of Hong Kong, 21 Sassoon Road, SAR Hong Kong, China

<sup>5</sup>Meta-Research Innovation Center at Stanford (METRICS), Stanford University, 1265 Welch Rd, MSOB X306, Stanford, CA 94305, USA

<sup>6</sup>Department of Immunology, Erasmus University Medical Center, 3015 CN Rotterdam, The Netherlands

<sup>7</sup>Biological Computation & Process Lab (BCPL), Chemical Process & Energy Resources Institute (CPERI), Centre for Research & Technology Hellas (CERTH), PO Box 361, GR-57001 Thessalonica, Greece

<sup>8</sup>Bioinformatics Research Laboratory, Department of Biological Sciences, University of Cyprus, PO Box 20537, CY-1678 Nicosia, Cyprus

...

e-mail: <sup>1</sup><[iliopj@med.uoc.gr](mailto:iliopj@med.uoc.gr)>, <sup>2</sup><[sophia.ananiadou@manchester.ac.uk](mailto:sophia.ananiadou@manchester.ac.uk)>,  
<sup>3,4</sup><[antoine.danchin@normalesup.org](mailto:antoine.danchin@normalesup.org)>, <sup>5</sup><[jioannid@stanford.edu](mailto:jioannid@stanford.edu)>, <sup>6</sup><[p.katsikis@erasmusmc.nl](mailto:p.katsikis@erasmusmc.nl)>,  
<sup>7</sup><[ouzounis@certh.gr](mailto:ouzounis@certh.gr)>, <sup>8</sup><[vprobbon@ucy.ac.cy](mailto:vprobbon@ucy.ac.cy)>

...

### **Abstract**

The linguistic foundations of science and technology have relied on a range of terms many of which are borrowed from ancient languages, a known but little researched fact from a statistical perspective. Precise definitions and novel concepts are often crafted with those – frequently used – terms, yet their etymology from Greek or Latin might not always be fully appreciated. Herein, we demonstrate that frequently used terms span almost the entire PubMed® database, while a handful of terms of Greek origin retrieve 80% of all entries. We argue that the etymology of those critical terms needs to be fully grasped to ensure correct use, in conjunction with other concepts. We further propose a number of terms for genomics, using prepositions that can accurately define subtle sub-disciplines of this ever-expanding field. Finally, we invite commentary by both the science community and the humanities, for possible adoption of suggested terms, not least to avoid inaccurate usage or inappropriate notions that may compromise clarity of meaning.

## Our etymological legacy

It is seldom in academic research that we, as scientists or engineers, explore or comprehend the deeper origins and subtle meanings of fundamental terms widely used in our everyday professional lives. These specialized vocabularies are primarily introduced during high school and university education to accurately describe concepts, phenomena, methodologies and techniques. At this stage, fledgling, aspiring young students usually focus on their subject matters often disregarding the linguistic dimension, as given. Later in life we get so accustomed to those terms, that it can be hard to dwell on their deeper meanings or derivatives (1), while their etymologies become overshadowed by intense usage. As a result, even entire disciplines may be wrongly named, thus creating terminological ambiguity (2).

This etymological deficit is alleviated by our linguistic backgrounds at various levels: each language contributes at different degrees to scientific and technological progress (3). It is widely accepted that the Greek language has provided the foundations of concepts to scientific research more than any other language (4) and, remarkably, continues to do so – as “the knowledge of names is a great part of knowledge” (5). We, as scientists and native speakers, carry this heavy linguistic legacy of Greek – a language with uninterrupted usage for thousands of years. Both science and technology are literally based on methods and ideas laid out and first described in Greek during the Golden Age and the Scientific Revolution of the Hellenistic period (6). Given that Greek provides the foundations of pretty much most technical, scientific, medical and philosophical terms with a rich meaning, we hereby attempt to raise a certain awareness of crucial aspects of semantics for the global community, in the age of instant information retrieval and rapid communication. Thus, we humbly urge colleagues to delve a little deeper into the etymology of standard terms of Greek origin and examine their meaning (7). Not only will non-Greek speakers be surprised by the richness as well as the structure of the Greek language, despite its often inept ‘naturalization’ in English or other languages, but they will most likely better understand their own thought processes and intellectual journey into their favorite subjects (8).

## Universal parallels unveiled

The subject matter of scientific and technical language goes beyond science, as it directly affects how our results are communicated to the wider public (9). Proper choice of terms and their exact meaning, often with Greek roots, deserve further explanation. Words such as ‘analysis’, ‘system’, or ‘dynamics’, may appear simple to grasp, yet they convey meanings that might have been transmuted over the past centuries to unrecognizable levels. Case in point is the word ‘machine’ which has a notion of an ‘engine’ and yet a very different meaning: ‘machine’ derives from ‘μηχανή’ (méchané; in Doric Greek: ‘μαχανά’) and it largely means ‘trick’. The word ‘diagnosis’ means ‘knowing through and through’, and so forth. All words ending with the suffix *-ics* were initially adjectives, not nouns: the term ‘mathematics’ is strictly speaking incomplete in ancient Greek, and was complemented by ‘philosophy’: ‘mathematical philosophy’ is a precise term, according to its original meaning<sup>1</sup> (10). Nouns with the same suffix are called ‘nominalized’ adjectives and also include physics, cybernetics, economics, or even the latin-derived informatics, and other *-ics*. The suffixes *-logy* and *-nomy* have been used interchangeably (e.g. ecology, astrology *versus* economy, astronomy) – why is there a ‘biology’ but not a ‘bionomy’? This series of arguments does not necessarily imply that we need to use those terms differently, but be aware of their roots and precise meaning, so that we exact control over the scientific lingua we use when we propose or invent new terms (11).

Our favourite example is of course ‘analysis’: everyone is using it, few fully understand it. ‘Lysis’ means ‘breaking up’, ‘ana-’ means ‘from bottom to top’ but also ‘again/repetitively’: the subtle yet ingenious latter meaning of the term implies that if you break up something once, you might not know how it works; if you break it twice, it means that you have implicitly reconstructed it so you understand the inner workings of the system. ‘System’ itself derives from *syn-* and *-stema*: ‘-stema’ meaning a ‘base’, a ‘component that stands up’; ‘syn-’ means ‘together’, ‘in conjunction with’, i.e. ‘syn-stema’, ‘systema’ means ‘a group of parts standing up together’. Now, you can almost speak Greek: ‘Anastema’ means ‘standing upwards’, ‘stature’ (in English). The etymology of the term ‘etymology’ points to ‘etymos’ meaning ‘real’, ‘genuine’ (11). In fact, we might

---

<sup>1</sup> ‘μανθάνω’: ‘learn’; mathematics: learnable, not logical or natural philosophy; philosophy: love for wisdom.

wonder how many PubMed® abstracts contain our title terms: hypothesis, analysis and synthesis disjointly return 11.7 million abstracts, an impressive fraction of a total of >28 million records (July 2018)<sup>2</sup>.

With this level of penetration into scientific vocabularies, established over the centuries by historical legacy and an indication of social status, the use of Greek is virtually unavoidable (8, 12). As an amusing illustration of this, one could not imagine reading the table of contents of a top journal and becoming excited about a breakthrough article titled “*Co-standing rebreaking of human firstin(?) cknits identifies colorbody segregation firstins(?)*”. In reality, this article does exist, and [in proper Greek!] has been published as “*Systematic analysis of human protein complexes identifies chromosome segregation proteins*”, a very interesting study indeed – no reference is provided, as this is essentially a random sample from a vast range of possible titles. Out of the ten words carefully selected by its authors in the title of this particular example, six have a Greek root. Yet, not everyone reflects on this amazing fact – including speakers of the Greek language. These words suddenly emerge as a gift from the remote past, hidden behind Latin characters and among other widely used terms (13).

### One sketchy analysis

To further substantiate our argument, we have carried out a systematic analysis of all PubMed® entries, to explore and quantify the presence of dominant terms that originate from Greek and convey vital meaning to the majority of the scientific literature in the biomedical domain. These most frequently used terms deserve special attention as they impact upon the majority of the written corpus of scientific research and training. This does not necessarily mean that less frequent terms are not critical, if ambiguous or misused – e.g. paediatrics (14), but that the most frequently used terms are of paramount importance (15). Evidently, proper term usage to ensure accuracy and careful dissemination of scientific results are also inextricably intertwined in our mission to keep pseudo-scientific claims at bay – e.g. physics *vs.* metaphysics, psychology *vs.* parapsychology, and others (16).

The exhaustive analysis of term frequencies across the PubMed® database, the first of its kind to the best of our knowledge, clearly demonstrates the preeminence of the Greek language in scientific literature, a fact that is well-known but not explicitly addressed in training curricula. For example, keeping terms that appear anywhere in more than one million entries each, we obtained a list of 243 distinct words which cover the database by almost 100% [DS01]. So-called ‘stop-words’ (words that are automatically excluded from any PubMed® query) are entirely represented by English terms. One would like to exclude articles (e.g. ‘the’), prepositions (e.g. ‘to’, ‘in’) and short verb forms (e.g. ‘is’, ‘was’), and retain mostly nouns and a few adjectives, to examine how many commonly used words deriving from Greek can cover (how much of) PubMed®. Note that no single gene or protein name appears in the list of most frequent terms (17), as those are represented by frequencies one order of magnitude smaller (10K-100K times for the most frequently used ones, e.g. lysozyme, collagen, globin, etc. – not shown). Using a part-of-speech tagger<sup>3</sup>, we have retained only words with rich meaning corresponding to nouns (NN\*), adjectives (JJ\*) and verb forms (VB\*), a total of 172 terms [DS02] (with >1M appearances) – subtracting 20 terms with length<4 characters<sup>4</sup>, we are left with 152 terms [DS03]. When combined into a disjoint query, using the OR operator [DS04], these 152 terms retrieve ~27.8M entries, the majority (~97%) of all entries in the PubMed® database (28,694,193 – 28 July 2018).

From this initial dataset, we have subsequently identified 15 terms with Greek origin using etymological dictionaries<sup>5</sup>: these terms might require a further qualification about etymology<sup>6</sup>. Importantly, the constructed query composed of those 15 terms – corresponding mainly to nouns<sup>7</sup> [DS05] – returns more

<sup>2</sup> [https://www.ncbi.nlm.nih.gov/pubmed/?term=\(hypothesis+or+analysis+or+synthesis\)](https://www.ncbi.nlm.nih.gov/pubmed/?term=(hypothesis+or+analysis+or+synthesis)) [accessed 28 July 2018]

<sup>3</sup> <http://cst.dk/tools/index.php>

<sup>4</sup> Most of these terms are single-letter codes, symbols (e.g. lt), abbreviations (e.g. DNA), and a few short English words (e.g. is, are).

<sup>5</sup> <https://www.etymonline.com/> (accessed, July 2018)

<sup>6</sup> It should be noted that there is a limit to which etymology can be traced back to roots, and at times, it can be erroneous: while some decisions about etymology may cause slight variations in statistical estimates, the overall trends remain the same.

<sup>7</sup> In fact, 13 out of the 15 terms of Greek origin are nouns.

than 23,000,000 distinct PubMed® entries i.e. 83% of the above mentioned 27.8M entries, or 80% of the entire database! An impressive percentage indeed.

(protein OR clinical OR analysis OR gene OR system OR method OR therapy OR type OR proteins OR based OR plasma OR diagnosis OR genes OR period OR surgery)

As a control, we use the 152 terms *minus* the above set of Greek words – with the NOT operator, the query being a composite of the total number of 137 frequent terms excluding the Greek ones, representing other words of different origins; note, however, that certain of these 137 query terms might be remotely connected by etymology to Greek words, for example ‘mice’/μύες. This composite query returns a mere 4,721,339 unique entries (28 July 2018) of which only 1,600,844 contain abstracts. In fact, most of the abstracts in this set also contain multiple Greek words with lower frequencies, such as ‘phenomenon’, or ‘hemorrhage’. Arguably, our estimates are on the conservative side, as low-frequency terms are not taken into account.

It should be noted that these numerical trends represent underestimates of the prevalence of Greek terms, as they do not necessarily refer to full-text documents. Another source of under-estimation, not impacting the ultimate conclusions, results from the dual presence of singular and plural forms in the query for a subset of terms; by further expanding the 15-term query to contain both forms for all terms, the reported coverage would possibly increase beyond the term expansion capabilities of Entrez®.

((were OR patients OR cells OR study OR have OR been OR treatment OR results OR cell OR using OR group OR used OR activity OR increased OR protein OR data OR levels OR other OR found OR disease OR compared OR significant OR expression OR clinical OR analysis OR effect OR associated OR effects OR showed OR different OR studies OR high OR control OR human OR years OR observed OR time OR cases OR higher OR blood OR gene OR groups OR rate OR such OR risk OR response OR present OR cancer OR increase OR well OR system OR changes OR normal OR patient OR first OR model OR mean OR role OR acid OR factors OR total OR function OR method OR health OR number OR level OR therapy OR development OR performed OR growth OR suggest OR important OR mice OR type OR days OR rats OR similar OR lower OR presence OR children OR months OR women OR specific OR proteins OR potential OR obtained OR studied OR tumor OR serum OR tissue OR reported OR binding OR including OR subjects OR reduced OR factor OR concentrations OR based OR plasma OR findings OR identified OR receptor OR differences OR care OR treated OR decreased OR test OR measured OR addition OR primary OR early OR concentration OR case OR diagnosis OR following OR genes OR infection OR conditions OR show OR period OR related OR induced OR examined OR evidence OR investigated OR positive OR developed OR range OR demonstrated OR determined OR revealed OR several OR activation OR samples OR liver OR surgery OR same OR major OR values OR greater OR survival OR species OR shown OR acute OR determine OR brain OR population OR complex OR pressure OR bone OR vitro OR drug) NOT (protein OR clinical OR analysis OR gene OR system OR method OR therapy OR type OR proteins OR based OR plasma OR diagnosis OR genes OR period OR surgery)) ;; explanation: terms marked by green include the detected 152 most frequent terms, while terms marked by blue include the 15 most frequent terms of Greek origin — the ‘NOT’ operator yields results for 137 most frequent terms of non-Greek origin

## Usage under control

The introduction of novel terms for new concepts and ideas will need to be carefully crafted, following best practices from the past so that our current insights do not play a role retrospectively but also prospectively (18). Hence, we should strive to direct our thinking towards a highly elaborate process of inventing or introducing new terms, following other scientific fields where – by tradition – there might be more attention for precise definitions or names (19). Our evolving nomenclatures should become tightly governed by greater control of language use and self-discipline. In the field of genomics, for instance, there have been creative usage patterns to qualify various genome biology sub-disciplines, for example “epigenomics”. Other -genomics / -omics terms might not be as precise, or charming – for more commentary on this particular topic, a well-known blog provides a treasure trove of such interesting definitions<sup>8</sup>. Case in point is the term “metagenomics”, which although difficult to revoke at this point, does not arguably carry the proper semantics with the *meta-* prefix (*meta*: meaning beyond). A better division of terms in this area might have been *endo*-genomics for host microbiomes (*endo*: meaning inside) and *exo*-genomics (*exo*: meaning outside) for environmental microbiomes (strictly speaking, these prefixes are not prepositions – see below). A similar situation might have occurred for the quest for life in the universe: the term *exo*-biology was gradually replaced by *astro*-biology, perhaps with limited linguistic (and possibly scientific) considerations (20) – one’s impression is that it had to rhyme with astro-nauts, and not exo-nauts. We conclude here with a classic

<sup>8</sup> [phylogenomics.blogspot.com/search/label/badomics](http://phylogenomics.blogspot.com/search/label/badomics)

recommendation: “Πάσαν γλώσσα βασάνιζε” (*tr.* “try every art of tongue” – Aristophanes, Σφήκες Sphekes; Latin: *Vespae*, 422 BC).

## A modest proposal

Evidently, all of research, training, education and knowledge transfer should take into account a much deeper appreciation and understanding of this linguistic treasure<sup>9</sup>, in order to satisfy the needs of future conceptual advances in science. For instance, a concept represented by the suffix ‘-some’ (from ‘soma’), meaning ‘body’ in Greek has already been used for proposed neologisms in the literature during recent times, capturing notions from molecular and cell biology to biochemistry and physiology (**Table 1**): while some of those have been successful, others have not. This situation emphasizes the urge to craft successful terminology as close to the conceptual roots as possible, to maintain accuracy, validity and perhaps community acceptance. One of us (AD) has proposed various modifications of existing terms ending in ‘-some’ with Greek roots that reflect concepts more accurately and even elegantly, for instance ‘schizosome’ instead of ‘divisome’, ‘coptosome’ instead of ‘spliceosome’ and ‘anxiosome’ instead of ‘stressosome’ (21).

Another group of overloaded terms in modern science has been, as we have already alluded to, those with the Greek suffix ‘-ome’ (from ‘-oma’), much used and abused in recent times – and relevant to our own academic activities. To exemplify the linguistic richness of the Greek language in this specific context, we thereby explore the use of Greek prepositions to qualify and accurately specify various aspects of genomics, with their respective subject matter (**Table 2**). Our intention is not only to introduce new concepts as an exercise, but seek a more precise definition of existing activities, currently undesignated. The usage of prepositions in general conveys additional, precise and logically consistent meaning, as in the case of *epi*-genomics. Other uses, not previously proposed, are herein offered for adoption by the global community: these terms can be explored as possible means to provide novel definitions with an accurate semantic representation of corresponding, ongoing research activities. This modest proposal will hopefully prevent the unfortunate use of alternative terms, which might not always articulate the deeper semantics of these sub-fields in the genomics domain. Once more, whether frequent or not, Greek terms might come to the rescue and act as safeguards, to ensure that we are able to express subtleties of evolving scientific concepts, without retreating to artificial or untested linguistic constructs.

## CONCLUSION

A number of historical factors over the past few centuries have kept apart the Greek-speaking world and turned the Greek language into a challenge, seen as ‘difficult’, maybe ‘foreign’ and definitely not very useful (12). It is often forgotten that some of the greatest achievements of science, engineering, literature, philosophy, arts and architecture, were communicated in Greek, not only in the remote past but throughout the recent centuries: the Greek language greatly contributed to the development of the Renaissance, the French and American Revolutions, and modern scientific exploration. Scholars of the Western world spoke and wrote in Greek, from Newton and Leibniz to Goethe and Wittgenstein. The same situation holds for scholars who influenced the development of the life sciences, as exemplified by taxonomic names, a tradition that has survived alongside Latin since Linnaeus and Darwin. Perhaps today we can all learn a thing or two from them, now that we have the capacity to extract meaning and sense in our electronic universe with automated translation, dictionary and etymology tools, in order to achieve a precise scientific discourse and maintain accuracy of our ‘logos’ (as in -logy).

## Epi-logos: Φοβού τούς Δαναούς...

The proper usage of terms and composites requires careful assessment, open dialogue, scientific reasoning and deep knowledge of each specific natural science domain, as well as constructive feedback from linguists and scholars from the humanities (8, 22). As an opportunity to bridge the sciences, from natural to social, we hope that this contribution will stimulate the fundamental discourse requisite to elicit consistent and accurate terminology usage in modern science and technology.

---

<sup>9</sup> ‘treasure’ arising from: *thesaurus* «θησαυρός»

## **ACKNOWLEDGMENTS**

We thank Ben Blencowe (Univ. of Toronto), Georgios Floros (Univ. of Cyprus), Peter Karp (SRI International), Nikos Kyrpides (JGI Berkeley), Nikolas Papanikolaou (Univ. of Crete), Nikos Sarantakos (European Parliament) and Spyros Sfenthourakis (Univ. of Cyprus) for insightful comments and constructive criticisms.

## **SUPPLEMENTARY INFORMATION**

- ▶ Data supplement (and invited commentary): <https://figshare.com/s/33fbde421cdfb5cc0245>
- ▶ Etymology of the select terms of Greek origin: <http://troodos.biol.ucy.ac.cy/Etymology.html>

## **TABLES**

### **Table 1**

Terms that appear in the literature from different disciplines, ending in -some.

Terms with suffix -some, (column one) with a short definition (column two) and the general context in the literature (column three); number of PubMed® entries (column four) and year of first appearance in PubMed® (column five) are also listed.

<b>term</b>	<b>definition</b>	<b>context</b>	<b>PubMed®</b>	<b>year</b>
catansome	catanionic vesicle	synthetic biochemistry, surfactants	2	2008
ejectosome	extrusive organelle	cell biology and physiology	16	1984
histosome	a type of ejectosome	cell biology and physiology	1	2015
hyposome	cellular structure	Dinoflagellate biology	7	2010
prostosome	prostate gland vesicle	sperm mobility and physiology	218	1982
remosome	remodeled nucleosome	non-canonical chromatin structure	2	2010

### **Table 2**

Potential term definitions using prepositions and the suffix -genome.

Prepositions of Greek origin (column one) that can be associated with the suffix -genome (currently undesignated in italics, column two); interpretation of terms (column three) defines the suggested use. It can be seen that certain prepositional prefixes have been successfully used in genomics, while others have not. Certain combinations offer extremely interesting opportunities to summarize vast swathes of genome biology, qualifying specific sub-fields or aspects of the particular area of research (23).

<b>preposition</b>	<b>-genome</b>	<b>interpretation of terms</b>
ana	<i>anagenome</i>	possible term for usage over time, to monitor population variation
amphi	<i>amphigenome</i>	potential use for polyploid genomes, and sex differences
anti	antigenome	used for antigen identification and immune system response (24)
apo	<i>apogenome</i>	could be used to convey non-DNA large scale analysis
dia	<i>diagenome</i>	useful concept for comparative genomics
eis	<i>eisgenome</i>	nutrigenomics, including substances (alcohol, tobacco, etc.)
ek	<i>ecgenome</i>	alternative for exo-genomics (see text)
en	<i>engenome</i>	alternative for endo-genomics (see text)
epi	epigenome	exemplary use in existing field, correct meaning (25)
kata	<i>katagenome</i>	great term to describe developmental processes over time
meta	metagenome	beyond' genome, for environmental sequencing (26)
para	<i>paragenome</i>	genomics of paralogs, not satisfactory, but conveys meaning
peri	<i>perigenome</i>	great term to describe developmental processes over space
pro	<i>progenome</i>	meaning 'previous', used for paleogenomics (limited use) (27)
pros	<i>prosgenome</i>	meaning 'after', could be used for synthetic & future, and similar
syn	<i>syngenome</i>	could be used to add meaning for genomics of symbioses
hyper	<i>hypergenome</i>	synthetic genome with added functions (limited, other use) (28)
hypo	<i>hypogenome</i>	synthetic genome with depleted functions
(pan)	pangenome	not strictly speaking a preposition, but conveys meaning nicely (29)

## REFERENCES

1. H. Askitopoulou, I. A. Ramoutsaki, E. Konsolaki, Analgesia and anesthesia: etymology and literary history of related Greek words. *Anesth Analg* **91**, 486 (Aug, 2000).
2. I. Ramoutsaki, I. Ramoutsakis, D. Bouros, Pneumology or pneumology?: An etymologic approach. *Chest* **121**, 1385 (May, 2002).
3. A. Danchin, Les langues de la découverte scientifique. *Archicube, le magazine de l'Association des Anciens Élèves, Élèves et Amis de l'École Normale Supérieure* **9**, 48 (2010).
4. W. E. Flood, *Scientific words: their structure and meaning*. (Oldbourne, London, 1960).
5. A. Silverman, Plato's Cratylus: The Naming of Nature and the Nature of Naming, in *Oxford Studies in Ancient Philosophy: Volume X: 1992*, J. Annas, Ed. (Clarendon Press, 1992).
6. L. Russo, *The forgotten revolution: how science was born in 300 BC and why it had to be reborn*. (Springer, Berlin/New York, 2004).
7. A. Tamis, The Importance of Learning the Greek Language, *Greek Reporter Australia*, 2016.
8. B. Snell, The Forging of a Language for Science in Ancient Greece. *The Classical Journal* **56**, 50 (1960).
9. S. Montgomery, Of towers, walls, and fields: perspectives on language in science. *Science* **303**, 1333 (Feb 27, 2004).
10. P. H. Nidditch, The First Stage of the Idea of Mathematics: Pythagoreans, Plato, Aristotle. *Midwest Studies In Philosophy* **8**, 3 (1983).
11. G. R. Welch, Physiology, physiomics, and biophysics: a matter of words. *Prog Biophys Mol Biol* **100**, 4 (Sep-Oct, 2009).
12. K. N. Lewis, The language of modern medicine: it's all Greek to me. *Am Surg* **70**, 91 (Jan, 2004).
13. G. F. Steffanides, The Role of Greek and Latin in Science. *The American Biology Teacher* **27**, 785 (1965).
14. G. F. Still, Paediatrics or Pediatrics. *Can Med Assoc J* **20**, 84 (Jan, 1929).
15. D. Newman, S. Karimi, L. Cavedon, Using topic models to interpret MEDLINE's medical subject headings, in *AI 2009: Advances in Artificial Intelligence*, A. Nicholson, X. Li, Eds. (Springer, Melbourne, Australia, 2009), pp. 270-279.
16. W. T. Jarvis, Misuse of the Term 'Allopathy' in <https://www.ncahf.org/articles/a-b/allopathy.html> National Council Against Health Fraud, (1996).
17. M. R. Seringhaus, P. D. Cayting, M. B. Gerstein, Uncovering trends in gene naming. *Genome Biol* **9**, 401 (Jan 31, 2008).
18. D. M. Jackson, Hellenomania-hellenophobia. *Lancet* **2**, 541 (Sep 02, 1961).
19. H. G. Trüper, How to name a prokaryote? *FEMS Microbiology Reviews* **23**, 231 (1999).
20. M. Gargaud, P. López-García, H. Martin, *Origins and evolution of life: an astrobiological perspective*. (Cambridge University Press, Cambridge, UK ; New York, 2011).
21. A. Danchin, The significance of split genes to developmental genetics. *Adv Genet* **24**, 243 (1987).
22. H. R. Wulff, The language of medicine. *J R Soc Med* **97**, 187 (Apr, 2004).
23. G. L. Banay, An Introduction to Medical Terminology I. Greek and Latin Derivations. *Bull Med Libr Assoc* **36**, 1 (Jan, 1948).
24. A. Sette, T. R. Schenkelberg, W. C. Koff, Deciphering the human antigenome. *Expert Rev Vaccines* **15**, 167 (2016).
25. A. Murrell, V. K. Rakyant, S. Beck, From genome to epigenome. *Hum Mol Genet* **14 Spec No 1**, R3 (Apr 15, 2005).
26. W. R. Streit, R. A. Schmitz, Metagenomics--the key to the uncultured microbes. *Curr Opin Microbiol* **7**, 492 (Oct, 2004).
27. R. Ferreira, F. J. Pontes, B. de Barros Neto, P. M. Farias, Do older taxa have older proteins? *Z Naturforsch C* **59**, 454 (May-Jun, 2004).
28. V. Sgaramella, The hypergenome in inheritance and development. *Cytogenet Genome Res* **139**, 215 (2013).
29. H. Tettelin *et al.*, Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* **102**, 13950 (Sep 27, 2005).