

Non-indexed citedness

Endre Száva-Kováts

Two recent fact-finding investigations demonstrate that eponymal and indirect-collective citedness are very frequent and long-standing phenomena in the journal literature of physics, at present not indexed by Citation Indexes. In consequence of non-indexed literature citedness, the Indexes are not suitable for the measurement of real literature citedness. The meaning and value of quantified citation data of the Citation Indexes must be reduced, especially in the cases of the (mis)use of the data for the purpose of evaluating scientists as scientists.

COMMUNICATING scientists as authors of journal articles, right from the start and out of various motives have made reference to the works of other scientists bearing relation to their own work. These reference acts form bibliographic connections which have been shown for decades by the Citation Indexes in ever-wider fields of current science.

For a scientist, the Indexes show which references were made in the observed and processed literature to a work by himself or another scientist working in the same field, thus giving information on the progress of the related literature.

However, the grandiose (originally) bibliographic instrument published by the Institute for Scientific Information (ISI) views these bibliographic connections from the angle of the cited author and the cited work and regards and names the reference act as a citation 'received' by the cited work. However, following the publication of the Indexes, in this approach primitive quantification and measurement appeared (because it had become possible): research to determine how many citations were 'obtained' by the communication of a given scientist, by the articles of a journal, by the research work of a scientific institute, and so on.

Naturally, citation-counting measurement was followed by numerical comparison, and this in turn was followed by quantitative evaluation. The fashionably modern and current research trend which has had a big influence on science administration, at first known as 'citation analysis' but now as 'scientometrics' or 'informetrics', essentially still deals with these aspects. It is not unusual now for its indices calculated for individuals – the 'impact factor' and the 'citation index' – to determine the career of a scientist¹.

The raw material of this research is the data of the Indexes, that is, the reference stock printed in the text of articles separated from the main body of the text, and indexed by the ISI in the processed part of the literature

as 'citations'. The concept of indexed citations, however, is narrower than that of formal references, because the ISI indexes as 'citations' only the formal references of direct and itemized nature.

Although the ISI now indexes as citations not only the *literature* references (i.e., published works), but also *verbal* ones ('personal communications'), it does not take stock of and show the *total real* reference acts contained in the articles of the processed journals.

Some of the non-indexed but real reference acts cannot be taken into account, but the majority of them can. It is not possible to take into account those reference acts (even if of a 'formal' nature), the objects of which are not given in a concrete or itemized form; e.g. in a single issue of an elite science journal we find 'and prior publications' 'and later papers', 'and previous papers', 'and related papers', 'and earlier papers'². The non-indexed but real reference acts which can be taken into account are of two kinds: eponymal references and indirect – and often collective – references. Eponymal references (the so-called eponyma, e.g. 'Pauli exclusion principle') are of non-formal nature, existing in the main text of an article. The indirect-collective references (e.g. '... and the references cited therein') can be of a formal nature, existing in the text of a formal (and indexed!) reference act, or of non-formal nature, existing in the article's main text (possibly in the context of a formal reference).

It can be stated that the stock shown by ISI represents only the indexed-formal references out of the total-real reference stock. The two stocks are generally not identical, and where this is the case, the indexed-formal reference stock is always the smaller. It is this stock, smaller than the total-real reference stock, which the Indexes show and which is the basis of all the results and conclusions of the citation studies.

Right from the start, Citation Indexes and 'citation analysis' have been the target of severe and many-sided but ineffectual criticism, recently reviewed and summarized by MacRoberts and MacRoberts³. The existence of non-indexed but real references as a defect does not fig-

Endre Száva-Kováts is at Magyar István utca 2, Budapest II-1114, Hungary.

ure in this criticism, nor is it found in Sandison's famous paper 'thinking about citation analysis'⁴, or in the more recent article of Liu⁵ reviewing the citation studies, or in the most recent corresponding discussion about citation analysis in *Nature*⁶⁻¹². Nevertheless, the existence of formally non-cited eponyms in scientific literature had already been raised in 1974 in a critical letter of the physicist Goudsmit¹³ in the famous literature case of the so-called Ortega Hypothesis in *Science*, launched by Cole and Cole with their citation analysis study¹⁴. Despite this, the literature phenomenon of non-indexed citedness did not appear as a point of criticism in any of the numerous communications dealing later with this citation analysis case.

The silence of the related literature to date concerning this point has obviously been due to the fact that no investigation had been made either on the nature or the frequency of this literature phenomenon, and in the absence of factual research data, the bibliometric and other '-metric' significance of its existence were not known. No investigation was made either of the non-indexed eponymal or of the non-indexed indirect-collective citedness in the literature.

In the last few years, two investigations of these phenomena have been made to determine their nature. The first fact-finding investigation, directed at non-indexed eponymal citedness, was made in the material of two elite physics journals: the general *Physical Journal (PHJ)* and the special *Journal of the Optical Society of America (JOSA)*. The research covered two years (1939, 1969) which are representative of the last, mature period of Little Science and of early Big Science, and in the case of *JOSA* was extended longitudinally to the period from 1934 to 1974, in five-year steps. The processed volumes were: *PHJ* 55, 56, 177; *JOSA* 24, 29, 34, 39, 44, 49, 54, 59, 64; the number of processed articles: *PHJ* 922, *JOSA* 1131, giving a total of 2053. In this way the full material of the individual years was processed, with the exception of the *PHJ* in 1969: only the January material of vol. 177 was examined for this journal which has grown to enormous proportions.

According to the numerical results¹⁵ of this first investigation, which was of both synchronous and diachronous nature, non-indexed eponymal citedness (NIEC) is a very frequent and long-standing phenomenon in the journal literature of physics, with permanent and growing importance in respect of both the number of scientists in question and the number of all eponymal citations not indexed by the Citation Indexes¹⁶. To mention only the most important numerical result: the non-indexed eponymal reference stock of studied articles of both *PHJ* and *JOSA* amounts to a third of the quantity of the stock of indexed references, for both 1939 and 1969. Furthermore, it was proved that because of the NIEC phenomenon, the producers of really important and long-lasting scientific achievements – continuously

many hundreds of prominent earlier scientists – although textually really cited, do not obtain indexed citations of their outstanding achievements. Numerical data of the investigation also indicate that the number of non-indexed eponymal citations in the case of the majority of these prominent eponymous authors not only equals, but often exceeds the number of indexed formal citations of prominent scientists of the present period.

The non-numerical finding of the investigation was that the nature and phenomenology of the NIEC phenomenon, as a conventional form of naming and of referring to long-existing scientific knowledge, were revealed and outlined, showing the origin and development of NIEC. This development already began in the second, declining stage of indexed-formal citedness of the knowledge, when the knowledge was still formally referenced; after this NIEC becomes increasingly frequent, then predominant and finally exclusive. However, by this time the practice of naming and mentioning the knowledge in the anonymous form in the text of articles also generally begins. The anonymous form in reality represents the practice of non-referencing; e.g. 'the second law of thermodynamics' (in the case of a law), or 'as it is well known' (in place of a formula in a mathematical demonstration). Moreover, the final textual form in which long-existing knowledge is mentioned in the physics literature – and also in the literature of other sciences – is now increasingly becoming the acronym, containing either the eponymal or the anonymial designation; e.g. DWBA (= 'distorted-wave Born approximation'), and RI (= 'retention index'). It was found and stated that the emergence of the NIEC phenomenon marks the end of the first stage of the famous Merton 'obliteration by incorporation' as a process. From the viewpoint of indexed-formal citedness, the literature phenomenon of NIEC is a special case of Garfield's well-known 'uncitedness'.

The second fact-finding investigation, which was directed at non-indexed indirect-collective citedness, and the publication of its results is in progress¹⁷ but due to insuperable problems of method, this dealt with only the 1969 part of the first investigation. The number of processed articles was: *PHJ* 331, *JOSA* 290, giving a total of 621; however these were studied with the required philological methods and precision. It was established that non-indexed indirect-collective reference(s) can be found in 7.2% of the articles of *JOSA* – which contains a large proportion (26.6%) of 'Letters' – while the corresponding figure was 14.8% for the articles of *PHJ*, which contains only 2.1% of short communications. However, these incidence ratios are far from expressing the weight and significance of indirect-collective references as non-indexed citations. To determine these, a few normal – non-review – articles containing indirect-collective reference(s) were examined for the number of indirectly cited works and then their ratio to indexed

citations was calculated. The result was surprising: the works cited in a non-indexed indirect-collective way in the articles was at least ten times greater than the number of works cited in the indexed-formal way. This means that both the number of scientists who did not receive their citations in the Indexes and the number of non-indexed but real citations was *many times greater* than the number of indexed scientists and the number of their indexed citations. Naturally, the number of all non-indexed – eponymal and indirect – references found in these articles exceeded the number of indexed references to an even greater extent. This surprising result was checked by tests in the material examined for 1939. These checks confirmed that the phenomenon (not only of eponymal, but also) of indirect-collective reference acts existed already in the last, mature period of Little Science. The checks confirmed that in the articles where they existed, the number of objects of such reference acts was also many times greater than the number of formal references which would have been indexed by ISI and would have figured in the Indexes – if they had existed at the time.

As the numerical result of the second investigation it can be stated that in the 1969 material studied, the quantity of non-indexed indirectly-collectively cited works in itself also reaches that of the indexed formally cited works, while the quantity of eponymally and indirectly cited works, i.e. the total non-indexed citations far exceeds that of the indexed ones. This means that in the first decade of the period when both Big Science and the Citation Indexes existed, in the exemplary elite journal literature of the leading science – physics – non-indexed citedness as a literature phenomenon exceeded indexed citedness in frequency.

Until a new, similar investigation covering the current years produces contrary results, the following conclusive statements are justified. The ISI does not take into account and therefore the Citation Indexes do not show the total-real reference stock of processed articles as citations, and so the Indexes do not give a real picture of the work of the scientists who are contributing to the progress of science. Citation data of the Indexes do not fully indicate the actual contribution – recognized

really-textually by the publishing scientists – and the quantified data do not quantify the individual extent of this contribution. The failure to take indirect-collective references into account neglects innumerable cases of real recognition of the work of countless contemporary scientists, while the failure to consider eponymal references neglects the impact of the most outstanding earlier creators in the field of science and their genuinely important results which have stood the test of time.

In consequence of the phenomenon of non-indexed citedness, it must (should) be taken into consideration that the stock of indexed references in the journal literature of natural sciences is only *a part* of the real stock of references; that indexed citedness is *not* identical with real citedness; that the Citation Indexes are *not* suitable for the measurement of real literature citedness. There is a need to amend, limit and reduce the now commonly accepted meaning, value and validity of the quantified data of the Citation Indexes and of 'citation analysis' and '-metrics' studies based on them. This need especially applies where these studies wish to (mis)use the citation data of the Citation Indexes for the evaluation of scientists as scientists or other similar, non-bibliographic purposes.

1. Calza, L. and Garbisa, S., *Nature*, 1995, 374, 492.
2. See *J. Org. Chem.*, 1969, 34, 1803, 1845, 1857, 1863, 1962.
3. MacRoberts, M. H. and MacRoberts, B. R., *J. Am. Soc. Inf. Sci.*, 1989, 40, 342–349.
4. Sandison, A., *J. Doc.*, 1989, 45, 59–64.
5. Liu, M., *J. Doc.*, 1993, 49, 370–408.
6. Strata, P., *Nature*, 1995, 375, 624.
7. Motta, G., *Nature*, 1995, 376, 720.
8. Metcalfe, N. B., *Nature*, 1995, 376, 720.
9. Lewison, G., *et al.*, *Nature*, 1995, 377, 671.
10. Yaalon, D. H., *Nature*, 1995, 378, 760.
11. Maffulli, N., *Nature*, 1995, 378, 760.
12. Persson, O., *Nature*, 1996, 380, 100.
13. Goudsmit, S. A., *Science*, 1974, 183, 28.
14. Cole, J. R. and Cole, S., *Science*, 1972, 178, 368–375.
15. Száva-Kováts, E., *Tud. Műsz. Táj.*, 1987, 34, 523–542; 1988, 35, 195–219; 1989, 36, 291–319, 515–535; 1991, 38, 83–101; 1992, 39, 178–192.
16. Száva-Kováts, E., *J. Inf. Sci.*, 1994, 20, 55–70.
17. Száva-Kováts, E., *Tud. Műsz. Táj.*, 1995, 42, 259–273.