86

Knowl. Org. 33(2006)No.2
J. Gabel. Improving Information Retrieval of Subjects Through Citation-Analysis

Improving Information Retrieval of Subjects Through Citation-Analysis*

Jeff Gabel

Brooklyn Law School Library, 250 Joralemon Street,
Brooklyn, New York 11201, jeff.gabel@brooklaw.edu

Jeff Gabel is currently a Cataloging Librarian at Brooklyn Law School Library. He received his M.L.I.S. from the Palmer School of Long Island University. He also holds an M.F.A. from Pratt Institute and a B.F.A. from Kansas State University. Gabel's publications include "Grammatical Noun Cases for Non-Linguists: Bibliometrics Applied to a Subset of the Literature on Finnish Linguistics," published in Collection Management in 2005, and a presentation at the Ninth International ISKO Conference, Vienna, July 5-7, 2006 under the title "Improving Information Retrieval of Subjects through Citation Analysis: a Study."

ABSTRACT: Citation-chasing is proposed as a method of discovering additional terms to enhance subject-search retrieval by broadening and prioritizing the results. Subjects attached to records representing cited works are compared to subjects attached to records representing the original citing sources, and to the subjects yielded by chasing see-also references from the latter group of headings. Original citing sources were yielded via a subject-list search in a library catalog using the subject heading "Language and languages – Origin." A subject-search was employed to avoid subjectivity in choosing sources. References from the sources were searched in OCLC where applicable, and the subject headings were retrieved. The subjects were ranked first by number of citations from original sources, then by total citation-frequency. The results were tiered into four groups in a Bradford-like distribution. A similar rank and division was performed on the subjects representing the original citing sources, and those yielded by chasing see-also references. Both in terms of subject frequency and topic type, positive comparisons between citation-chasing and see-also references show a confirmation of different methods of yielding alternative subjects. Exclusive results suggest potential mutual complementary value among these different methods.

1. Introduction

The purpose of this study is to broaden and prioritize subject-search retrieval results by using citation-chasing. Citation-chasing, simply put, is the process of retrieving the references cited by a work. In the context of topical search-retrieval, citation-chasing is a potential source of subject terms that can: expand the set of terms that apply to a given topical search, reveal groupings and relative prioritization among the terms in such a set; as well as patterns and prominence among topical types suggested by the terms in such a set.

Research has shown that subject searching consistently produces low recall, and remains the most problematic aspect of OPAC searching, even though it is the most common type of search (Bates, 2003).

Graham (2004) notes the disappointment "that the subject-searching capabilities of Web catalogs appear to be much the same as those of pre-Web, second-generation systems," despite the "significant expansions to the accessibility and content of library catalogs" with the emergence of Web interfaces in the mid-1990s.

Bates makes a case for end-user entry vocabulary with expanded terms, whose basic designs could consist of human-made clusters of terms with computer support. She notes that "the range of vocabulary used by information system users is extremely wide and varied … with the total number of different terms used among a group of people found to be almost always high," and that people can recognize information much easier than they can recall it (Bates, 2003). Bates notes incidentally that information systems

should also support users in other information searching behaviors, including citation-chasing, which is useful for its ability to directly point to potentially relevant works. But in the context of terminology expansion, what comes to mind is the process's ability to yield large amounts of data, and to present these data in clusters according to relevance. Therefore, an obvious extension of this process would be the use of subject terms found on records yielded through citation-chasing. The identification of these sets through citation-chasing creates the potential for discovering previously unknown relationships among works, or in this case, subjects. Larsen (2002) has shown that following citations from a subject-search can improve recall.

Starting with a known work entails the problem of starting with what Birger Larsen (2002) calls the "good seed document." This problem can be eliminated by starting with one or more subjects, rather than pre-chosen titles, as entry points. But in both cases, the user or researcher is ultimately relying on subject-assignment decisions that are made on a title-by-title basis as they apply to potential target documents. This type of assignment does not necessarily reveal all relationships, or even the best relationships, between relevant titles and subjects. Many users employ citation-chasing to bypass the subject search process altogether. This provides direct access to potential target materials, and in the process, it bypasses the problems of assignment subjectivity. But because citation-chasing is mentioned in cursory fashion in subject retrieval literature, it appears that it is not seen as a complete substitute for subject searching.

An added benefit of citation-chasing is the ability to retrieve large amounts of bibliographic data for analysis, which can be subjected to clustering and ranking. But chasing citations in many environments does not retrieve subject terminology data the way it does author, title, title-word or journal data. This is particularly true with respect to books. However, books frequently are repositories of quite large numbers of citations. If one can use the citations in a book to evaluate and measure author-name, journal-name, and title-word data attached to those citations, then one should be able to do the same with a subjects or subject-words attached to the same citations. Though subject information is generally not directly attached to citations in books, it might be appropriate by analogy to consider subject-headings found on bibliographic records for cited works as having been "cited" by the sources containing those cita-

tions, just as the authors, source-titles, etc., of those citations are considered to be cited by the sources.

One returns to the problem of starting with a good seed document. Therefore, this study begins with a subject-heading, rather than a chosen work, as the starting point. But rather than treating the sets of cited works as an end, we attempt to investigate them as means to an end, by taking advantage of the cited works' functions as citation data to yield expanded subject terminology.

1.1 Research Questions

Within the limits of a given search inquiry, how does the set of terms yielded by a common terminology-retrieval strategy compare to the corresponding set yielded by citation-chasing, and how do the occurrence-frequencies of these terms compare across the two strategies? In this case, the common retrieval strategy is the library catalog using the *Library of Congress Subject Headings* (*LCSH*) as topical retrieval language. Furthermore, we are interested in learning whether topic-type patterns can be identified as properties of these sets of terms, so that topic-type production and patterns can be compared across the different methods of yielding subjects.

1.2 Scope

This study represents an attempt to investigate correlations between co-occurrence, frequency-rankings, and topical types resulting from different methods of yielding subjects. Observations or claims regarding the objectivity of citation-chasing compared to that of an OPAC or any *LCSH*-driven inquiry are meant to refer to comparisons of the different means of retrieving subject headings. There is no attempt in this study to address the subjectivity involved in *LCSH* creation or assignment practices. It should also be noted that the methodology for this study was time consuming, and yielded a large amount of data. This raises issues of practicality with regard to citation-chasing as a retrieval strategy. The problem of physically implementing citation-chasing for live end-user retrieval is outside the scope of this study.

2. Methodology

A subject-list search was performed using the subject LANGUAGE AND LANGUAGES – ORIGIN in the online catalog at Long Island University (LIU). The titles linked to this subject heading were

88

Knowl. Org. 33(2006)No.2
J. Gabel. Improving Information Retrieval of Subjects Through Citation-Analysis

retrieved, as well as those linked to all instances of this subject heading followed by subheadings. Thirty-seven monograph titles were located in this manner. For a manageable study size and reasonable access, and for currency, the results were restricted to works located at the local Brooklyn Campus, and to works that were published during the 10-year period from 1995 through 2004. Thirteen titles remained, which will be referred to as 'original sources' or 'citing sources.' The works are listed in Table 1-1 below. A 'cited work-subject' refers to a subject heading found on the OCLC record that represents a work cited in an original source (see Table 1-2). For example, the subject-heading HUMAN EVOLUTION is considered to be cited in the original source work by Armstrong in the context of this article.

The references found in the original sources were searched in the OCLC WorldCat. Varying *LCSH* assignments from different records for the same works were included. For example, different editions are often cataloged with different choices of subject headings by different catalogers. Care was given to avoid misleading variations (for example, when a record shows the work bound with another work in one record). The subject headings were not checked for accuracy or obsolescence. To keep the retrieval set manageable, non-*LCSH* headings attached to records were not used.

The cited-work-subjects were ranked according to the number of original citing sources that yielded them, from 13 to 1, and divided into Bradford-like tiers. The subjects were sorted first by number of citing sources, then by number of total citations. The results were divided into four tiers in which the number of total citations was roughly equal among the tiers, while the number of subject headings in each tier increased at a rate between one-third and one-fifth as the frequencies decreased.

3. Results

Figure 1 shows the Bradford-like distribution of the 2525 subject-headings that were associated with cited works 6853 times. Only 745 (29.5%) of the total 2525 subject-headings were associated with works cited by more than one of the original 13 sources.

Title	Author	Date
From hand to mouth: the origins of language	(CO) Corballis, Michael C.	2002
The evolutionary emergence of language: social function and the origins of linguistic form	(KN) Knight, Chris Studdert-Kennedy, Michael Hurford, James R. (editors)	2000
Original signs: gesture, sign, and the sources of language	(AR) Armstrong, David F.	1999
The origins of language: what non-human primates can tell us	(KI) King, Barbara J. (editor)	1999
The development of language: acquisition, change, and evolution	(LIG) Lightfoot, David	1999
How the brain evolved language	(LO) Loritz, Donald	1999
Eve spoke: human language and human evolution	(LIE) Lieberman, Philip	1998
The seeds of speech: language origin and evolution	(AI) Aitchison, Jean	1996
Grooming, gossip, and the evolution of language	(DU) Dunbar, R.I.M.	1996
The prehistory of the mind: a search for the origins of art, religion, and science	(MI) Mithen, Steven J	1996
Human evolution, language, and mind: a psychological and archaeological inquiry	(NO) Noble, William Davidson, Iain	1996
The biology of language	(PU) Puppel, Stanisław (editor)	1995
Music and the origins of language: theories from the French Enlightenment	(TH) Thomas, Downing A.	1995

Table 1-1. Titles retrieved by searching "Language and languages – Origin" in the Subject-List index in Long Island University's OPAC (restricted to the Brooklyn Campus location and to publication dates 1995-2004).

Knowl. Org. 33(2006)No.2
J. Gabel. Improving Information Retrieval of Subjects Through Citation-Analysis

89

Subject	Total	AI	AR	CO	DU	KI	KN	LIE	LIG	LO	MI	NO	PU	TH
Language and languages – Origin	12	1	1	1	1	1	1	1	1	1	1	1		1
Human evolution	5				1		1	1		1		1		
Linguistic change	3	1								1		1		
Grammar, Comparative and general	2								1	1				
Historical linguistics	2	1			1									
Language acquisition	2	1							1					
Social evolution	2				1						1			
Animal communication	1					1								
Anthropological linguistics	1						1							
Art, Prehistoric	1										1			
Behavior evolution	1							1						
Biolinguistics	1								1					
Biolinguistics - Congresses	1												1	
Brain - Evolution	1										1			
Cognition	1										1			
Communication - Social aspects	1				1									
France - Intellectual life - 18th century	1													1
Genetic psychology	1										1			
Gesture	1		1											
Gossip - History	1				1									
Group identity	1				1									
Historical linguistics - Congresses	1												1	
Human behavior	1				1									
Human evolution - Congresses	1												1	
Human evolution - Philosophy	1										1			
Interpresonal relations	1				1									
Language and languages - Origin - Congresses	1												1	
Language and languages - Origin - History - 18th century	1													1
Language and languages - Sex differences	1				1									
Music and language	1													1
Primates**	1					1								
Psycholinguistics	1											1		
Reason	1										1			
Sign language	1		1											
Signs and symbols	1											1		
Sociolinguistics	1				1									
Thought and thinking	1										1			
TOTALS	58	4	3	1	11	3	3	3	4	4	9	5	4	4

Table 1-2. Subject headings attached to the titles in Table 1-1.

Only 376 (14.9%) were associated with works cited by more than 2. The 1st tier consists of the subject-headings associated with works cited by 10 to 13 original sources, where 1.4% of the subjects account for 22.3% of all subject-citation association (also see Table 2). Furthermore, within the group of 35 sub-
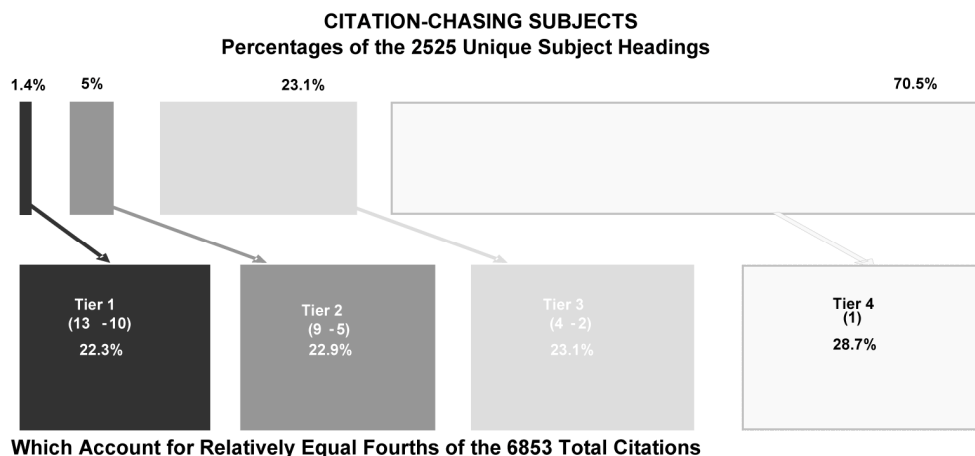
90

Knowl. Org. 33(2006)No.2

J. Gabel. Improving Information Retrieval of Subjects Through Citation-Analysis

Figure 1. Distribution-pattern of impact of the subjects cited by the source titles

jects in Tier 1, only 3 of the subjects were associated with works cited by all 13 sources. These data suggest that ranking subjects by citing sources creates a rigorous and highly selective system. Each citing source yielded hundreds of bibliographic records for subject data, so in theory, each subject had many chances to be yielded at least once by all of the citing sources. Similarly, analysis of cited authors and journal titles resulting from the same original sources showed that none of the authors or journal titles were cited by all 13 sources – and only a couple each were cited by 12 or 11 sources.

Note that only roughly the top half of the 1st tier of subject headings is displayed, and the other tiers are not shown at all, due to the sizes of the lists. The chart shows a predominance of general linguistic, evolution, biology and psychology topics, sometimes intermixed within a heading. The linguistic topics are the most frequently used, after which they mingle with the latter three topic-types in the frequency lists. In general, granularity increases as the frequencies decline. An outstanding exception is the heading PSYCHOLOGY near the bottom. A look at the remainder of tier 1 and the upper part of the 2nd tier, neither of which is displayed here, would show a sharper continuation of this tendency. More general headings might be more frequent because they are easier to determine and assign than specific headings. It is more difficult to comment on the influence of the nature of the topic 'language origin' on this tendency, since journal-article sources are not accounted for in the study. However one chooses to perceive granularity, Table 2 suggests a pattern of relevance of topic types yielded by citation-chasing, which will be discussed in Section 3.4 below.

Subject	Total citations	Source citations
Language and languages	109	13
Language and languages – Philosophy	63	13
Linguistics	53	13
Human evolution	129	12
Psycholinguistics	101	12
Language acquisition	97	12
Evolution	75	12
Language and languages – Origin	59	12
Evolution (Biology)	48	12
Natural selection	46	12
Biolinguistics	33	12
Neuropsychology	32	12
Human beings - Origin	44	11
Psychology, Comparative	39	11
Sign language	39	11
Brain - Evolution	36	11
Psychology	24	11
Behavior evolution	23	11
Cognition and culture	19	11

Table 2. Top half of Tier 1. Subject-Headings Associated with Works Cited by 11 to 13.

The column on the right shows the number of citing sources associated with each subject. The column immediately to the right of the subject headings shows the total number of bibliographic records on which this subject heading was found. The headings were sorted in the following order: by number of citing

sources, then by total number of bibliographic records on which the subject was found, then alphabetically.

There was one subject that appeared 15 times in total, but was only associated with a work cited by one source: "Music – Philosophy and aesthetics." In addition, there were 12 others associated with works cited by one source that received between 9 and 4 citations. Of these 13 subject headings, all except one, a 4-cite subject, were associated with works cited by Thomas. All of these subject headings are related to music, aesthetics, or French philosophy, anomalous subjects given Thomas's divergent treatment of language origin. Of the 13 original sources, the work by Thomas was responsible for the most anomalies in subject frequencies. It raises the question of the influence that anomalous titles and their topics have on the results. These numbers are such a small portion of the 4th tier, all of which could be considered to be anomalous, that their removal would hardly affect the results. However, if Thomas's work were to be removed as an anomaly, many of the distinctions among the most frequently cited works would collapse. In a further study, this anomaly could possibly be addressed by widening the net for original citing sources and noting any proportional increase, or lack thereof, in the topics in question. On the other hand, the anomaly strengthens the notion that citation-chasing creates a rigorous selective system, but theo-

retically making the chances more difficult for any given subject to be in the 13-cite group rather than the 12-cite group.

3.2 Comparisons of methods of yielding subjects

For comparative results, similar frequency rank charts and tiers were created for the subjects found directly on all versions of the bibliographic records for the original sources. The resulting subjects will be called original source subjects. These 13 titles were retrieved by searching "Language and languages – Origin." By searching the titles in the local catalog as well as OCLC, and checking all records for each title, anywhere from 1 to 11 subject headings were gathered. 37 headings account for 58 citations, and the average is 4.5 headings per title. These subject-headings can be divided into two tiers along the same concept that Table 2 was divided. 18.9% of the headings (cited 2 to 12 times) account for about half (48.3%) of the citations. 81.1% of the headings (cited once) account for the other half (51.7%).

Figure 2 shows a side by side comparison of the top tiers of the original source subjects and the subjects yielded by citation-chasing. The rank number beside each subject represents that particular subject's placement in the opposite chart. The color bars basically reinforce the numbers next to them. For

Frequency-Rank Comparison
Original Citing Source Subjects and Citation Chasing Subjects

ORIGINAL SOURCE SUBJECTS Top 7 (Tier 1)	Rank in CITATION CHASING CHART
Language and languages--Origin	8th
Human evolution	4th
Linguistic change	69th
Grammar, Comparative and general	24th
Historical linguistics	54th
Language acquisition	6th
Social evolution	38th

CITATION CHASING CITED SUBJECTS Top 12 (Tier 1 - top half)	Rank in ORIGINAL SOURCE SUBJ CHART
Language and languages	--
Language and languages -- Philosophy	--
Linguistics	--
Human evolution	2nd
Psycholinguistics	once
Language acquisition	4th
Evolution	--
Language and languages -- Origin	1st
Evolution (Biology)	--
Natural selection	--
Biolinguistics	once
Neuropsychology	--

Figure 2. Frequency-Rank Comparison between Original Citing Source Subjects and Citation-Chasing Subjects

92

Knowl. Org. 33(2006)No.2
J. Gabel. Improving Information Retrieval of Subjects Through Citation-Analysis

example, the chart on the left shows that the most frequent original source subject is ranked 8th in the citation-chasing chart.

Conversely, the top three spots on the citation-chasing frequency chart, in the right column, are blank, indicating that they did not appear at all on the original citing source records. A major generalization in charts is that all of the subjects in the top tier of the citing source records were yielded by citation-chasing, and many of them were ranked among the top 10 of the 2500 citation-chasing subjects. The lowest is ranked 69th, which is still within the top 3% of the citation-chasing charts.

The chart on the right shows that 7 of the 12 most frequent citation-chasing subjects were not revealed at all by looking at subjects on the citing source bibliographic records. The original source term, "Language and languages – Origin," placed behind 7 other subjects in the citation-chasing chart.

3.3 Subjects yielded by chasing see-also references

LCSH facilitates further methods of yielding subjects. Figure 3 displays similar types of distribution for the citing source subjects, and then the see-also subjects that they yielded. These will be called see-also subjects. Figure 3 shows that when see-also references support the original source subjects, they pick up some of the discrepancy between the two charts in Figure 2.

All of the see-also subjects were yielded by citation-chasing, but the chart on the left shows some high numbers, indicating citation-chasing subjects that appeared relatively low in the 3rd Tier. In the context of clusters of terms to facilitate recognition, the numbers of subjects in the 3rd Tier become too big to be practical. However, when a few such 3rd Tier subjects, for example, are ranked among the top 5 or 10 of another method of yielding subjects, as is the case here, this suggests the possibility of selective evaluation of subjects in the larger tiers through outside corroboration.

The chart on the right shows that some of the highest-frequency citation-chasing subjects were confirmed by see-also references, although some are unique. The rest of the 23 subjects in the 1st Tier of the Citation-chasing Chart on the right show a greater proportion of subjects that are unique altogether, both

Frequency-Rank Comparison
See-Also Subjects with Citing Source Subjects
and Citation Chasing Subjects

SEE-ALSO SUBJECTS Top 7 (Tier 1)	Rank in CITATION CHASING CHART
Linguistics	3rd
Language and languages	1st
Evolutionary psychology	177th
Physical anthropology	207th
Evolution (Biology)	9th
Human beings -- Origin	13th
Psychology	17th
Historical linguistics	54th
Psycholinguistics	5th

CITATION CHASING CITED SUBJECTS Top 12 (Tier 1 – top half)	Rank in SEE-ALSO SUBJ CHART	Rank in ORIGINAL SOURCE SUBJ CHART
Language and languages	2nd	--
Language and languages -- Philosophy	--	--
Linguistics	1st	--
Human evolution	--	2nd
Psycholinguistics	8th	once
Language acquisition	--	4th
Evolution	twice	--
Language and languages -- Origin	--	1st
Evolution (Biology)	5th	--
Natural selection	--	--
Biolinguistics	once	once
Neuropsychology	--	--

Figure 3. Frequency-Rank Comparison between See-Also Source Subjects with Citing Source Subjects and Citation-chasing Subjects

to the citing source subjects and see-also subjects. The top tier of the citation-chasing subjects shows a relatively high amount of unique subject terms.

3.4 Topic types

The chart in Figure 4 shows the topic types that appeared in the top tiers of the charts we just saw, starting with citation-chasing subjects. Each bar represents a subject heading. The most notable observation here is the exclusive appearance of general language topics in the 13 citing source areas. Looking at the next few levels of bars, the chart indicates that the general language core is supported by a secondary core of evolution-related topics, as well as the topics that represent a specific context of language investigation. Just after this, the chart indicates a large supporting area of other biology and behavior topics, and topics related to psychology, intellect and the brain.

The three charts in Figure 5 compare these results with the topic types suggested by the other 2 methods of yielding subjects. The chart on the left is a repeat of the previous one, but it is cut off at 11 citing sources for a manageable size. A few properties of this chart are obvious right away. One is that the see-also-subject chart on the right yielded results similar to the citation-chasing chart, both of which diverge from the data in the original-source-subjects chart.

The most obvious similarity is the exclusive highest-frequency area of General Language. The other notable feature here is the appearance of anthropology, with a relatively high frequency in the see-also reference chart. The citation-chasing method yielded a fairly high amount of anthropology terms in the lower tiers, and in lower parts of the 2nd tier, but not in the 1st Tier.

Figure 6 compares the different results by topic types. This is perhaps a better overview of the results. In short, the results of the two methods of yielding alternative subjects diverge from the results of original source subjects in similar ways. The original-source-subjects chart shows a lack of general language topics, but the specific language context topics are its major core, which is also the category to which the original search topic belongs.

The original-citing-source chart includes the topic "language structural domain." This topic also appeared a few times on the citation-chasing chart. It was not yielded by see-also references. Unlike the two methods of yielding alternative subjects, the original citing sources yielded a lack of non-language related topics in general, except for evolution, which is a major core in all three charts. Biology and behavior topics were unique to citation-chasing in the high frequency tiers, while anthropology was unique to the see-also subjects, though it appeared in lower tiers of the citation-chasing charts, including the 2nd Tier.

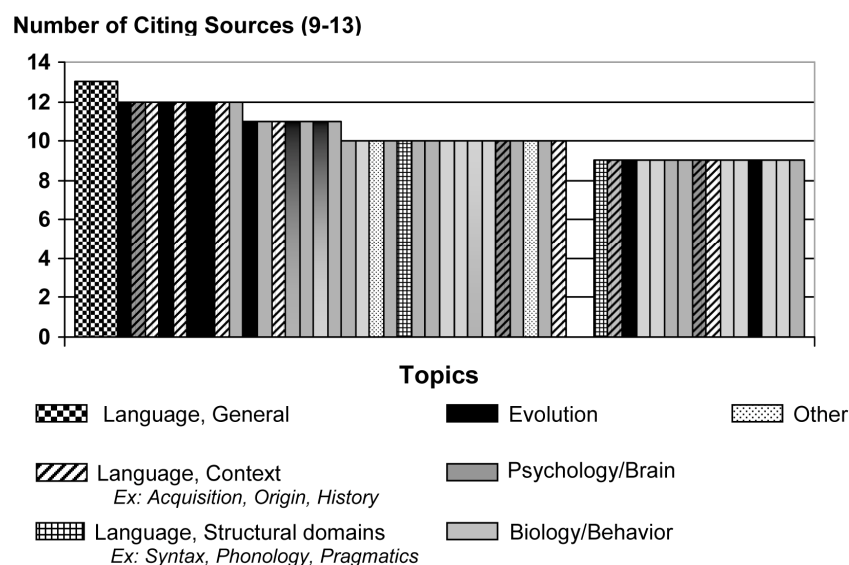## Topic Types Suggested by Citation Frequencies of Subjects Produced by Citation-chasing

Figure 4. Topic Types Suggested by Citation Frequencies of Subjects Yielded by Citation Chasing

94

Knowl. Org. 33(2006)No.2
J. Gabel. Improving Information Retrieval of Subjects Through Citation-Analysis

Comparison of Topic Types Suggested by Citation Frequencies of Subjects Produced by 3 Different Methods
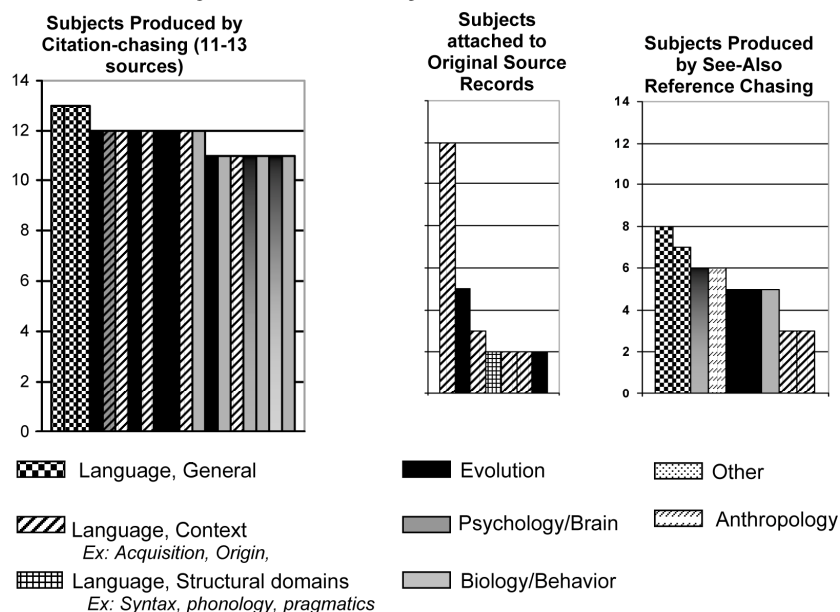

Figure 5. Comparison of Topic Types Suggested by Citation Frequencies of Subjects Yielded By 3 Different Methods

Topic Types

■ General Language
Citation Chasing - **the small major primary core**
See-Also - **the small major primary core**.
Citing Sources Records - **do not show prominently**.

▬ Lang, Treatment specific
Citation Chasing - **secondary core**
See-Also – **secondary core, to a smaller degree**
Citing Source Records - **the major core**

▭ Language, Natural Properties
Citation Chasing - **low frequencies, low occurrence**
See-Also - **low frequencies, low occurrence**
Citing Source Records - **secondary core**

■ Evolution   **Topic Types**
Citation Chasing - **major secondary core**   *Continued*
See-Also - **major secondary core**
Citing Source Records - **major secondary core**

▬ Psychology/Brain
Citation Chasing - **large Supporting Core**
See-Also - **supporting core**
Citing Source Records - **do not show prominently**.

▬ Biology/Behavior
Citation Chasing - **large supporting core**
See-Also - **do not show prominently**
Citing Source Records - **do not show prominently**

▭ Anthropology
Citation Chasing - **low frequencies, low occurrence**
See-Also - **major secondary core**
Citing Source Records - **low frequency, low occurrence**

Figure 6.

4. Conclusions

Citation-chasing yielded subjects unique to the original source subjects, and to a lesser degree, unique to see-also subjects. Citation-chasing and see-also chasing yielded a mix of concurring and unique topic types, both diverging from the original source subjects. Both in terms of subject frequency and topic type, positive comparisons between citation-chasing and see-also references show a confir-

mation of different methods of yielding subjects. Exclusive results suggest potential mutual complementary value among these different methods.

It is important to view these results in the context of some of the points made about the benefits of citation-chasing. The process provides an objective layer, in the sense that the retrieval of bibliographic records containing subjects is driven by citations, which represent the actual research, rather than the subjective assignment of headings. Also, following

citations yields a large number of subject headings. The relatively small amounts retrieved by viewing original source records and see-also references exhibit small frequency ranges, which don't promote the ranking of subjects, and make priority among distribution-groups less evident. Finally, citation-chasing enhances ranking and distribution by creating a rigorous selection system, where each subject in theory has many chances to be yielded at least once by the hundreds of citations in each one of the original citing sources.

5. Suggestions for further study

The methodology is this study could be applied to different subject-term systems, with less restrictive or unbounded terminology. It could also be applied to environments that are less restrictive in the format types they include. Anomalous titles and their subjects could be compensated for by investigating the titles and topics of a larger body of citing sources. A further study could make use of a more critical investigation of the naming and divisions of the topic types that were suggested by this study. Perhaps the most obvious potential for the further use of the data in this study is for the investigation of the idiosyncrasies and subjectivity in *LSCH* assignment.

Regarding the topic "Language and languages – Origin," a more involved study might analyze the fields of linguistics and complimentary or related subjects, like evolution. Detailed treatments of the nature of human language through specific traditional linguistic domains, like phonology, are not among the most prevalent in the citation patterns of any of the three methods of yielding subjects in this study. The more applied components of linguistics, or those in which more generalized aspects or functions are inherent, like history or acquisition of language, are given a different weight in each of the three charts. Both methods of yielding alternative subjects show an exclusive group of very general topics within linguists that is not revealed by the original sources records. When linguistics forms part of a co-topic with another discipline like evolution, can it generally be expected that the nature of the linguistic material drawn upon tends to be exceptionally general, and does this say something about the nature of linguistic science? There is a fair

amount of literature that discusses the scientific status of the field of linguistics, which is often placed in the humanities, the social sciences, or the sciences, depending on the source (Georgas and Cullars, 2005). There is also considerable discussion about the questionable adequacy of methods and results in linguistics, where explanatory adequacy often does not match the scientific mode of investigation (Yngve, 1985).

* This paper is an expanded report of Gabel (2006).

References

Bates, Marcia J. 2003. *Task force recommendation 2.3 research and design review: improving user access to library catalog and portal information: final report (version 3)*, report resulting from the Bicentennial Conference On Bibliographic Control For The New Millennium, Library Of Congress, Washington, D.C. http://www.loc.gov/catdir/bibcontrol/ 2.3batesreport6-03.doc.pdf.

Ding, Ying, Chowdhury, Gobinda G., Foo, Schubert. 2001. Bibliometric cartography of information retrieval research by using co-word analysis, *Information processing and management* 37: 817-42.

Gabel, Jeff. 2006. Improving information retrieval of subjects through citation analysis: a study. In Budin, Gerhard, Swertz, Christian and Mitgutsch, Konstantin, eds., *Knowledge organization for a global learning society: Proceedings of the Ninth International ISKO Conference (Vienna, Austria, July 4-7, 2006)*, pp. 19-26.

Georgas, Helen, Cullars, John. 2005. A citation study of the characteristics of linguistics literature. *College & research libraries* 66: 496-515.

Graham, Rumi Y. 2004. Subject no-hits in an academic library online catalog: an exploration of two potential ameliorations. *College & research libraries* 65: 36-54.

Larsen, Birger. 2002. Exploiting citation overlap for information retrieval: generating a boomerang effect from the network of scientific papers. *Scientometrics* 54: 155-78.

Stegmann, Johannes, Grohmann, Guenter. 2003. Hypothesis generation guided by co-word clustering. *Scientometrics* 56: 111-35.

Yngve, Victor H. 1986. *Linguistics as a science.* Bloomington: Indiana University Press.