Psientometrics, Vol. 60 (1992) 19-24

THE ORGANIZATION OF THE SEMANTIC SPACE OF AN AUTHOR

L. LEYDESDORFF

University of Amsterdam, Science Dynamics (The Netherlands)

Scientometric mappings reconstruct cognitive multidimensional spaces of social units of analysis by using textual indicators. Small (personal communication) has raised the question of whether similar techniques could be used to trace the cognitive space of individuals, and also to map their careers. In this study, I explore the possibility of using words and their co-occurrences for mapping the semantic space at this lower level of aggregation of a single author.

Introduction

The author under study is Professor Tibor Braun. Braun has been the Editor of the journal *Scientometrics* since its appearance in 1978. During his career Braun functioned most of the time as a professor in analytical chemistry at the Loránd Eötvös University in Budapest, Hungary. Additionally, he is director of the Information Science and Scientometric Research Unit (ISSRU) at the Library of the Hungarian Academy of Sciences.

Callon et al. (1986) have analyzed co-occurrences of words as indicators of semantic space in document sets [1]. Citations have been proposed as indicators of concept symbols [2]. Elsewhere [3], I have shown that words can be used for the analysis of structure in a single text, since authors are consistent in word usage within the framework of a single text, and distributions are skewed. However, when these structures are compared among texts, both the word frequencies and the underlying word structures exhibit variances. Thus, there is both conceptual and semantic change, and it is impossible to sort the two dimensions of cognition apart on the basis of their phenomenological appearance in terms of words [4].

At the lower level of aggregation of the individual, the processing of symbols and the attribution of meaning to words will be differently organized than in a scientific communication system. Phenomenological variances may be caused also in this case by conceptual or semantic developments. However, in this study I shall only explore the possibilities to use word occurrences for mapping these processes. Thereafter,

the analysis can straightforwardly be combined with that of other textual indicators, and also be extended with the dynamic perspective. [5]

Methods

For the data I used the on-line *SciSearch* edition of the *Science Citation Index* at *DIALOG* which covers publications in the natural and biomedical sciences from 1974 onwards. Braun is also a prolific author in the journal *Scientometrics* which is not included in this database, but in the equivalent for the social sciences. However, since he is Editor of this journal, I decided to focus on articles in the *Science Citation Index* only.

The search with ((author=Braun T) and (corporate country=Hungary)) retrieved 81 publications. These publications contain 327 different title words. The distribution of the words is skewed (as expected) with 196 words only occurring once, and one word ("of") occurring 55 times. For the further analysis, I used only title words which occurred more than three times, and which contain more than one character, and no prepositions, articles, numerals or conjunctions. Plurals and singulars are equated.

This leaves me with 45 words which occur in 78 of the 81 retrieved titles. However, in 10 texts there is only one word occurrence, and one word ("bibliography") occurs only once in this sample. Since a singular occurrence leads to a vanishing variance, these cases and variables are discarded. Thus, the further analysis is based on 68 titles and 44 words.

The analysis is pursued on this matrix with factor analysis (SPSS), and multidimensional scaling on the correlation matrix by using MINISSA.

Results

The factor analysis does not exhibit a clear factor structure at this level of aggregation. Instead, 14 factors have an eigenvalue of more than 1.0, and thus the reduction of complexity is only in the order of one third (14 factors for 44 variables). The scree-plot (Figure 1) suggests a four factor solution. Four factors explain only 43.8% of the variance.

The four factor solution is provided in Table 1. The multidimensional scaling of the correlation matrix is in Figure 2. The factor solution is penciled into the picture.

Table 1

Four Factor Solution for 44 words in 68 titles of Professor T. Braun

Rotated Factor Matrix:	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
reagent	.85586	-.02340	-.04583	-.04164
cell	.83553	-.02858	.09172	.01416
open	.83553	-.02858	.09172	.01416
preconcentration	.79674	-.03053	-.01982	-.00649
organic	.79529	-.03750	.00359	.10979
plasticized	.75212	-.01565	-.04233	-.09542
matrix	.69826	-.02364	-.15777	-.09362
polyurethane	.69063	-.15856	.21786	.20516
trace	.56491	-.07087	-.19690	.13495
diethyldithio carbamate	.52247	-.02980	-.12550	-.07596
foam	.43519	-.16051	.15033	.21478
loaded	.39327	-.01256	.04615	.02254
resilient	.26877	-.06568	.13145	.26376
flash	-.15574	.95594	-.10700	-.03013
cold	-.15574	.95594	-.10700	-.03013
annotated	-.14787	.95100	-.10313	-.02687
selective	-.14787	.95100	-.10313	-.02687
fusion	-.14534	.91774	-.10333	-.02888
world	-.17195	.90659	-.12177	-.08086
analytical	-.34127	-.36392	-.03253	-.35378
phase	-.04357	-.03758	.72224	.01444
separation	.17839	-.06497	.70265	.15808
solid	-.17686	-.11600	.69727	.09113
new	-.16478	-.08979	.60347	-.07344
extraction	-.07151	-.04743	.50833	.05667
liquid	-.16898	-.09660	.50426	-.09217
column	.25925	-.01218	.42883	.10652
media	-.08571	-.05673	.32780	-.00934
some	-.19497	-.22510	-.29482	.25863
exchange	.05434	-.02637	.29172	.00516
evaluation	-.16764	-.17589	-.22809	.02513
literature	-.17173	-.16108	-.22117	.10536
growth	-.15583	-.15682	-.19654	.00576
scientometric	-.09200	-.08586	-.11558	-.05468
trends	-.17325	-.21223	-.28674	.69441
recent	-.13412	-.14288	-.17117	.69242
instrumental	-.15259	-.14532	-.28137	.63488
chemistry	-.29160	-.33786	-.15571	-.50909
national	-.15201	-.21266	-.19609	-.43269
research	-.19568	-.21084	-.22440	-.40637
sorbents	.00230	-.11386	.11871	.34982
publication	-.10998	-.14758	-.12131	-.34918
analysis	-.05528	-.11520	-.16319	.22533
journals	-.16098	-.18299	-.14165	-.22014

Varimax Rotation 1, Extraction 1, Analysis 1 - Kaiser Normalization.
Varimax converged in 11 iterations.

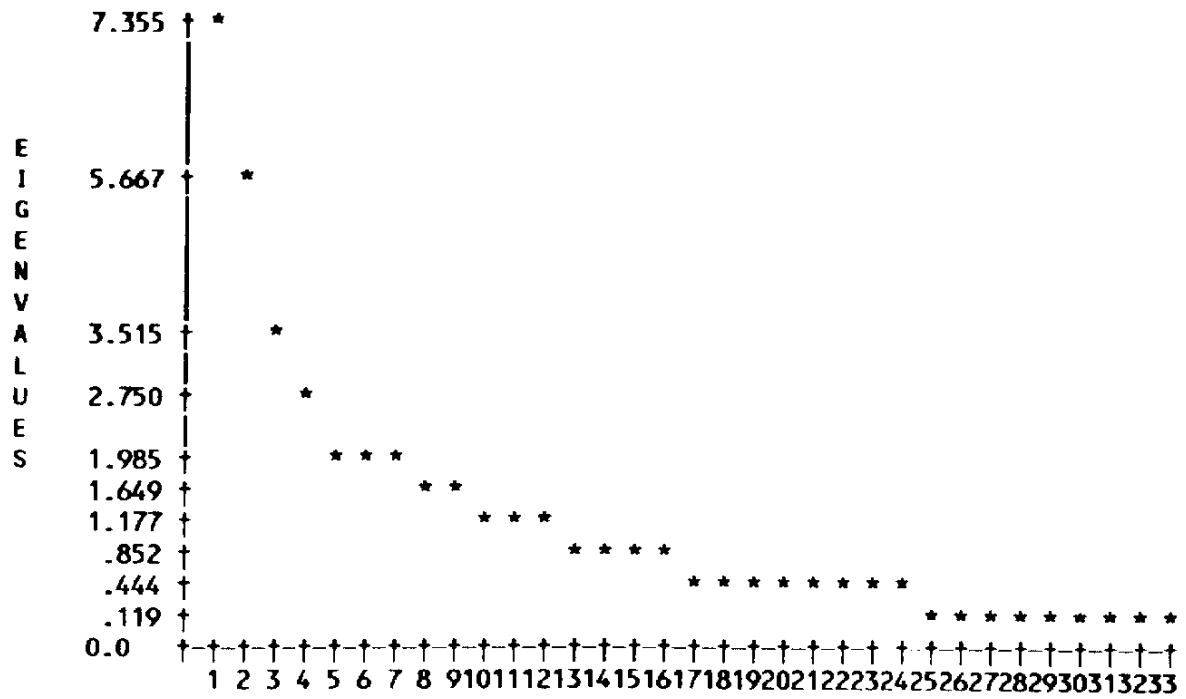

Fig.1 Scree Plot Factor Analysis.
44 words in 68 titles of Prof. T. Braun

Discussion and Conclusions

Three factors represent areas which cover different interests of Professor Braun in chemistry. Only Factor Two is different. This factor is well known in the community as the "world flash," which one seems to be able to turn on and off in Budapest. A number of articles relate this "world flash" strongly to "cold fusion." A similar focus on news related items is exhibited in Factor Four. The words with highest factor loading on this factor are "recent," "trends," "instrumental," and "sorbent."

Several important words do not load significantly on one of the factors in a four dimensional solution. Among them is such a crucial word as "scientometrics." The interpretation is that this word remains very much in the origin of Braun's semantic space.

A group of words which is technically categorized under Factor Four exhibits negative loading on all factors. These words are "chemistry," "national," "research," and "publication." We may wish to call these words "meta-scientific," and thus they might exhibit a reflection of Braun's social interests in community building, and national science policies.

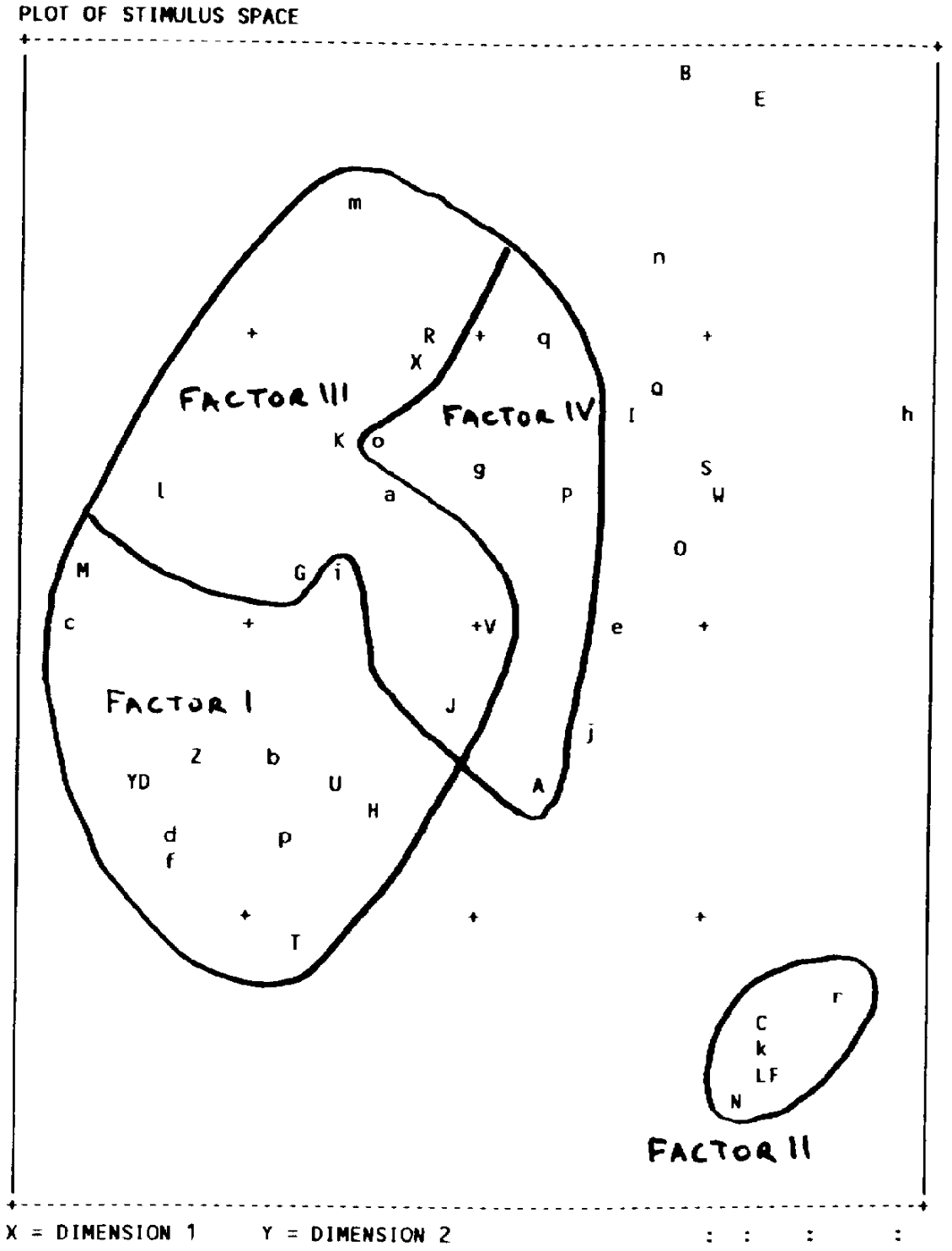

Fig. 2 Multidimensional Scaling of 44 words in 68 titles by Professor T. Braun
(Legends on next page)

In summary, the instrument of using words and their co-occurrences in the titles of documents has provided us with a means to map the semantic space of Professor Braun. In comparison with the high degree of organization among title-words and

their co-occurrence in scientific document sets [6], personal semantics seems less specific in its spatial organization.

Legends for Fig. 2

		DIM 1	DIM 2
A	analysis	0.138	-0.261
B	analytical	0.443	0.993
C	annotated	0.654	-0.724
D	cell	-0.732	-0.277
E	chemistry	0.608	0.942
F	cold	0.674	-0.811
G	column	-0.402	0.068
H	diethyldithiocarbamate	-0.223	-0.341
I	evaluation	0.329	0.367
J	exchange	-0.046	-0.131
K	extraction	-0.323	0.321
L	flash	0.674	-0.812
M	foam	-0.870	0.111
N	fusion	0.580	-0.853
O	growth	0.451	0.136
P	instrumental	0.201	0.217
Q	journals	0.392	0.405
R	liquid	-0.098	0.487
S	literature	0.499	0.276
T	loaded	-0.403	-0.553
U	matrix	-0.317	0.272
V	media	0.042	0.001
W	national	0.539	0.210
X	new	-0.137	0.453
Y	open	-0.732	0.277
Z	organic	-0.631	-0.210
a	phase	-0.202	0.235
b	plasticized	-0.447	-0.207
c	polyurethane	-0.891	-0.001
d	preconcentration	-0.680	-0.377
e	publication	0.304	-0.007
f	reagent	-0.668	-0.395
g	recent	0.001	0.258
h	research	0.957	0.371
i	resilient	-0.316	0.080
j	scientometric	0.260	-0.183
k	selective	0.653	-0.725
l	separation	-0.707	0.251
m	solid	-0.271	0.757
n	some	0.405	0.654
o	sorbents	-0.228	0.342
p	trace	0.418	-0.364
q	trends	0.129	0.511
r	world	0.807	-0.664

References

1. M. Callon, J. Law, A. Rip (Eds), *Mapping the Dynamics of Science and Technology*, MacMillan, London, 1986.
2. H. G. Small, *Social Studies of Science*, 8 (1978) 327-340
3. L. Leydesdorff, *Social Studies of Science*, 21 (1991) 75-110
4. L. Leydesdorff, From "Words" to "Information", (forthcoming)
5. L. Leydesdorff, *Scientometrics* 18 (1990) 281-307
   L. Leydesdorff, *Scientometrics*, 19 (1990) 271-96
6. L. Leydesdorff, *Research Policy*, 19 (1989) 209-223