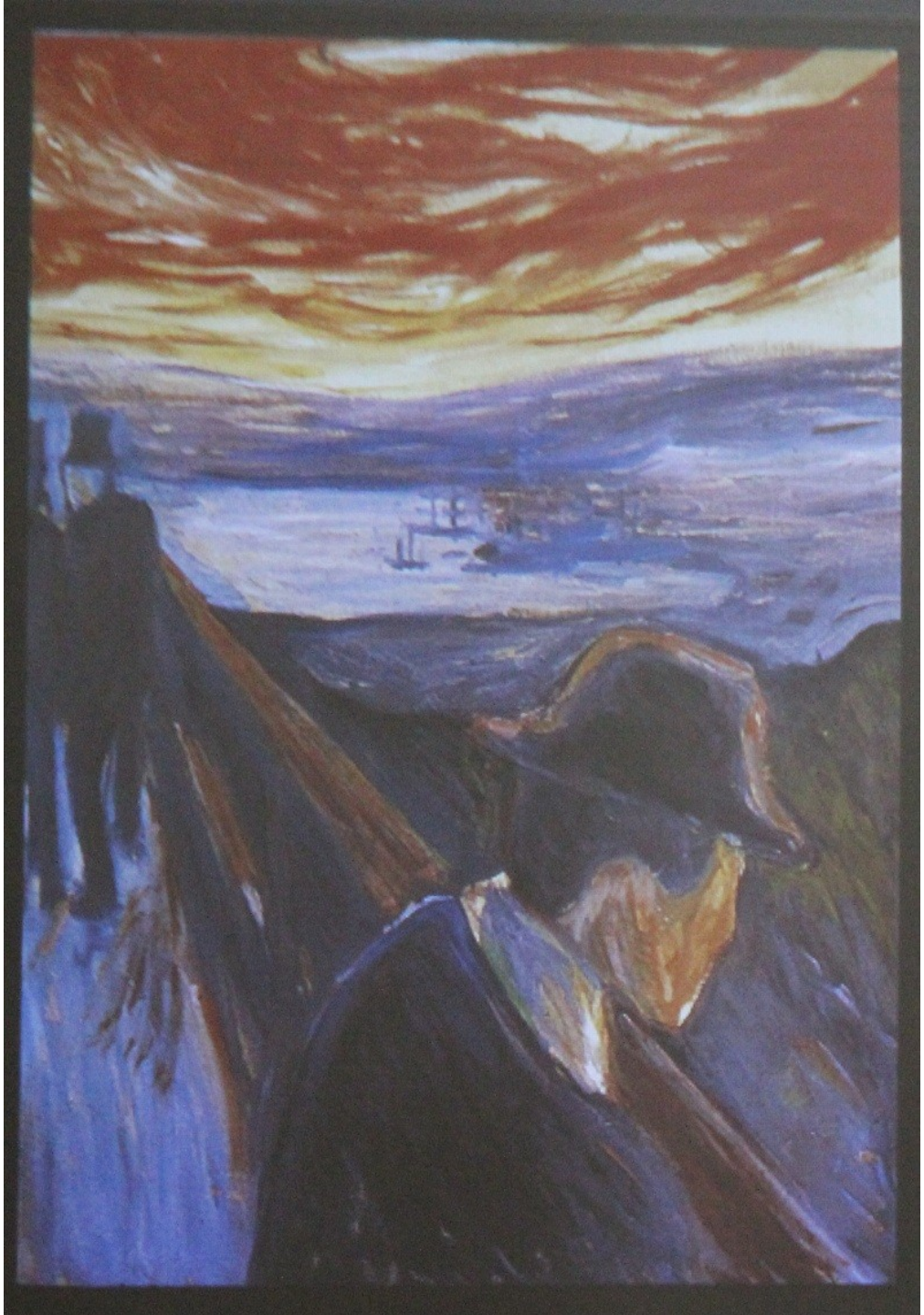
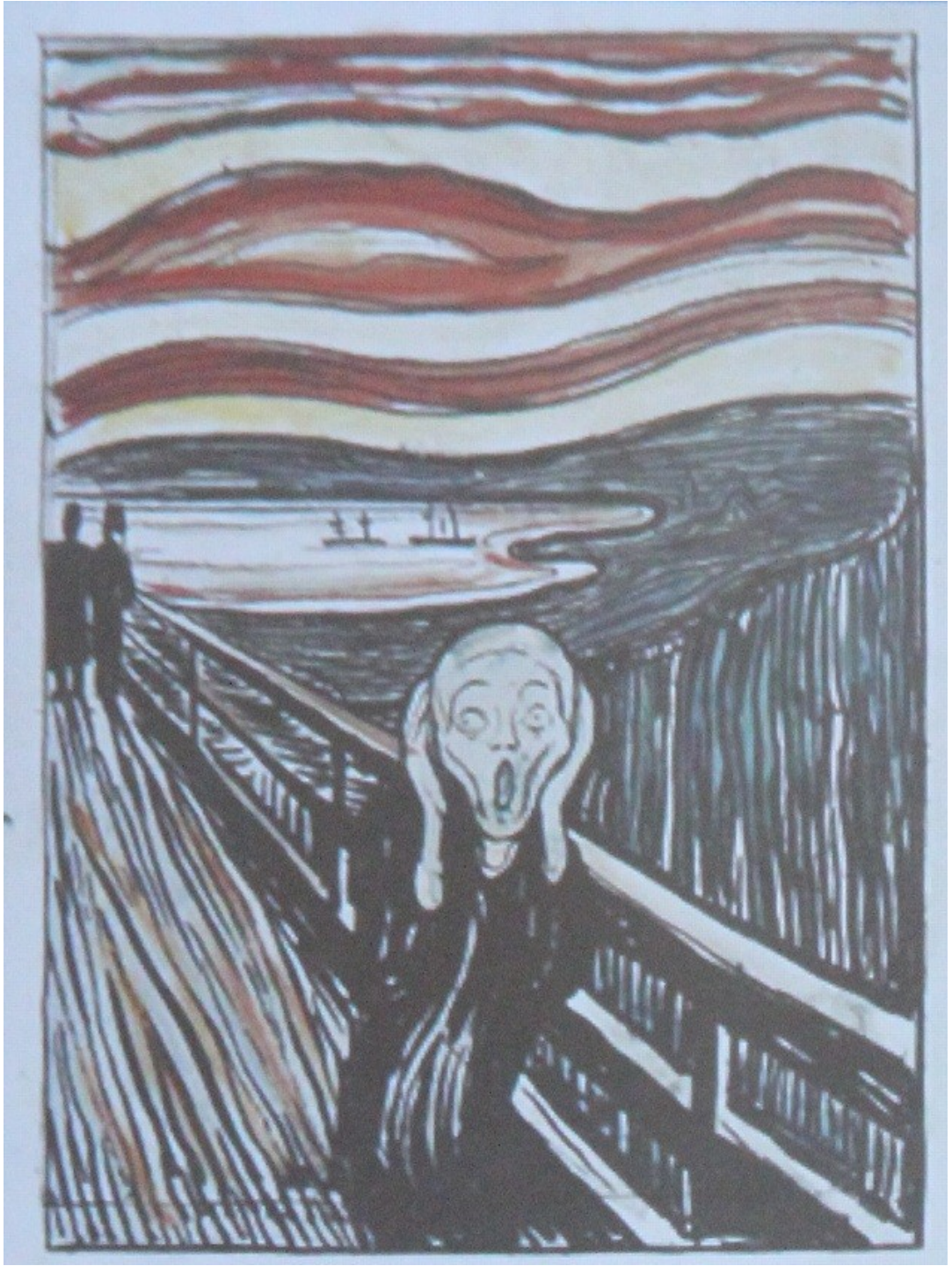


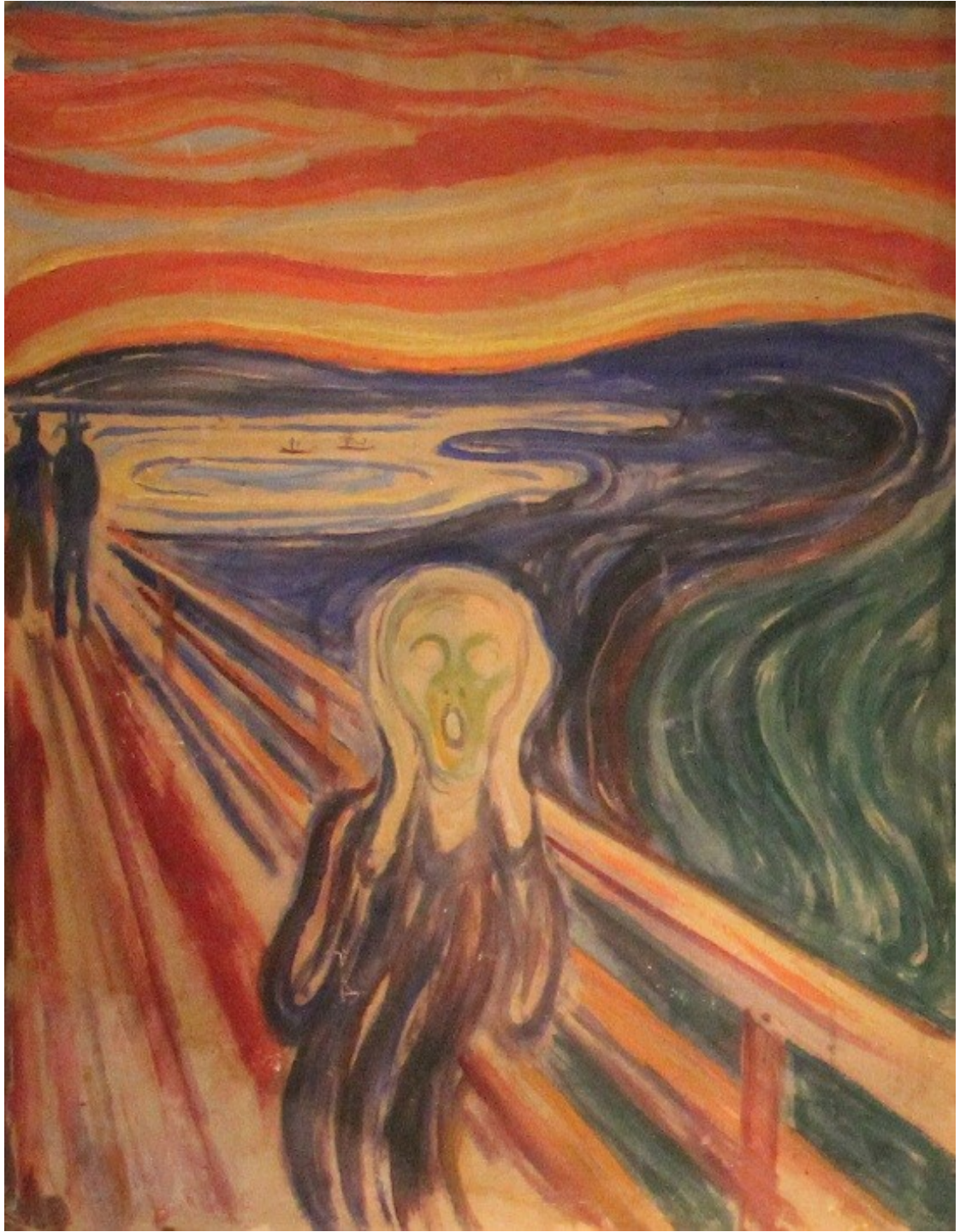
Despair, 1892,
Thielska Galleriet,
Stockholm

The Scream, 1st painted version, 1893, National Gallery



The Scream,
lithography, 1895

The Scream,
pastel, 1895,
$119,922,500 (2 May 2012)



The Scream, 2nd painted version, c. 1910, Munch Museum

Angst,
woodcut, 1896,
Munch Museum

Pilestredet 30

Lexicography in the Crystal Ball: Facts, Trends and Outlook

Gilles-Maurice de Schryver

Dept of Languages and Cultures, Ghent University (Belgium) &

Xhosa Dept, University of the Western Cape, Cape Town (South Africa)

1. THE EURALEX CONGRESSES CRYSTALLIZE

From milestone to milestone

- Half a century ago ‘a small group of linguists and lexicographers met at Indiana University to discuss a variety of problems related to the making of dictionaries’ (Householder 1962: v).
- The proceedings of that conference (Householder & Saporta 1962) set in motion the emergence of lexicography as a modern scientific discipline.

From milestone to milestone

- A decade later, Ladislav Zgusta's *Manual of Lexicography* (1971) gave every aspiring lexicographer something solid to hold onto.
- And have we held onto it: Zgusta's magnum opus remains one of the most cited works of our field.

From milestone to milestone

- Another decade later, another milestone.
- The year is 1983, when Reinhard Hartmann organizes a major international conference on lexicography in Exeter — baptized LEXeter ’83 — where the basis is inter alia laid for:
- the international encyclopaedia of lexicography *Wörterbücher / Dictionaries / Dictionnaires* (published a decade later, in three massive volumes),

From milestone to milestone

- the book series *Lexicographica. Series Maior* (which started appearing in 1984) as well as
- the journal *Lexicographica. International Annual for Lexicography* (as of 1985), and last but not least,
- where the *European Association for Lexicography* itself — EURALEX — was established (cf. Hartmann 2008).

From milestone to milestone

- The LEXeter ’83 proceedings (Hartmann 1984) thereby automatically became the proceedings of the first EURALEX congress.
- The second EURALEX congress was organized in 1986, with the proceedings appearing two years later (Snell-Hornby 1988).
- From then on, EURALEX has gathered biennially, with proceedings appearing two years after the event for the third and fourth congress, and simultaneously with the event as of the fifth congress onwards.

EURALEX CONGRESS					EURALEX PROCEEDINGS		
No.	**Year**	**City**	**Country**	**Acronym**	**Editor(s)**	**Year**	**Publisher**
1	1983	Exeter	UK	LEXeter '83	Hartmann	1984	Max Niemeyer Verlag
2	1986	Zurich	Switzerland	ZüriLEX '86	Snell-Hornby	1988	A. Francke Verlag
3	1988	Budapest	Hungary	BudaLEX '88	Magay & Zigány	1990	Akadémiai Kiadó
4	1990	Málaga	Spain	EURALEX '90	Alvar Ezquerra	1992	Biblograf
5	1992	Tampere	Finland	EURALEX '92	Tommola et al.	1992	Tampereen Yliopisto
6	1994	Amsterdam	Netherlands	Euralex '94	Martin et al.	1994	Vrije Universiteit Amsterdam
7	1996	Gothenburg	Sweden	Euralex '96	Gellerstam et al.	1996	Göteborgs Universitet
8	1998	Liège	Belgium	EURALEX'98	Fontenelle et al.	1998	Université de Liège
9	2000	Stuttgart	Germany	EURALEX 2000	Heid et al.	2000	Universität Stuttgart
10	2002	Copenhagen	Denmark	EURALEX 2002	Braasch & Povlsen	2002	Københavns Universitet
11	2004	Lorient	France	EURALEX 2004	Williams & Vessier	2004	Université de Bretagne Sud
12	2006	Turin	Italy	XII EURALEX	Corino et al.	2006	Edizioni dell'Orso
13	2008	Barcelona	Spain	XIII EURALEX	Bernal & DeCesaris	2008	Universitat Pompeu Fabra
14	2010	Leeuwarden	Netherlands	XIV Euralex	Dykstra & Schoonheim	2010	Fryske Akademy
15	2012	Oslo	Norway	EURALEX OSLO 2012	Fjeld & Torjusen	2012	Universitetet i Oslo

Aim of the present paper

- Although the EURALEX board went on to launch the quarterly *International Journal of Lexicography* in 1988, the material published in the biennial EURALEX proceedings held its own over the years.
- The body of research reported on in the EURALEX proceedings is now so substantial that an in-depth analysis is in order. This is exactly the aim of the present paper.
- The study is *driven* by the data in the proceedings.

2. THE EURALEX METADATA DATABASE

Setting up for meta-analysis

- A corpus was built containing *all* the material found in all fourteen proceedings published so far, as well as *all* the material (bar the current paper) accepted for presentation at the fifteenth congress.
- In the corpus each paper (and each piece of editorial material) is a separate file with a unique identifier.
- A separate database contains all the metadata for each file.

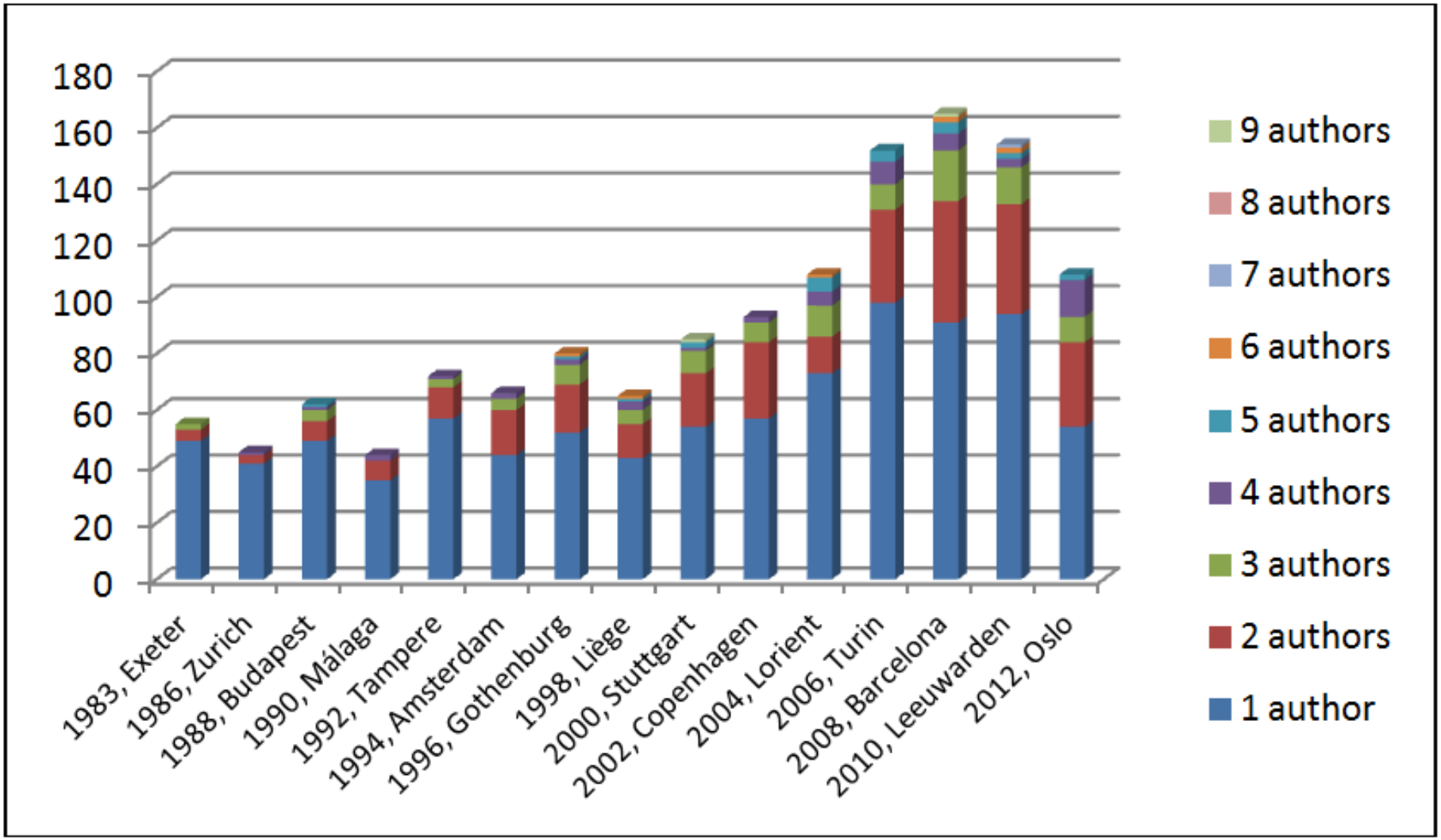
Setting up for meta-analysis

- The first four proceedings having been produced after the congresses took place, they do not necessarily contain all that was presented.
- Conversely, the proceedings of the next eleven congresses — the so-called *pre*ceedings — do contain a few papers which were not presented in the end.

Setting up for meta-analysis

- Overall, however, the proceedings represent the congresses well, even though one should keep in mind that more activities are typically taking place at the congresses themselves, which may include:
  - workshops,
  - symposia,
  - round tables,
  - structured debates,
  - poster and demo session, etc.
- Not to forget the publisher booths and the social programme.

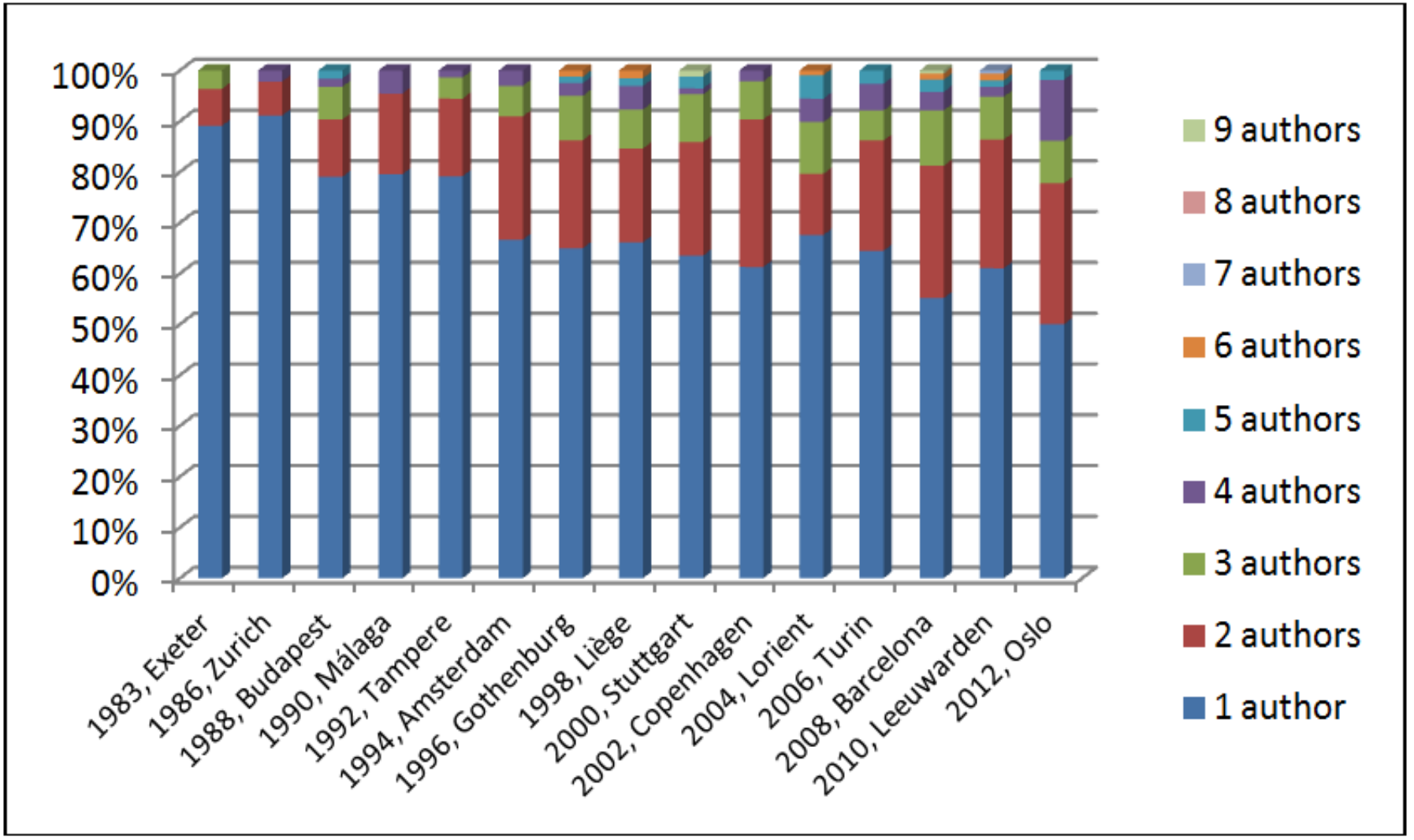
Papers per congress, showing number of authors per paper



Discussion …

- The number of papers grew exponentially over the years, up to and including the 2008 congress, after which the number went down again, likely to a more manageable number.
- In total, a massive 1 354 papers have been written so far.

Papers per congress, with number of authors per paper in %



Discussion …

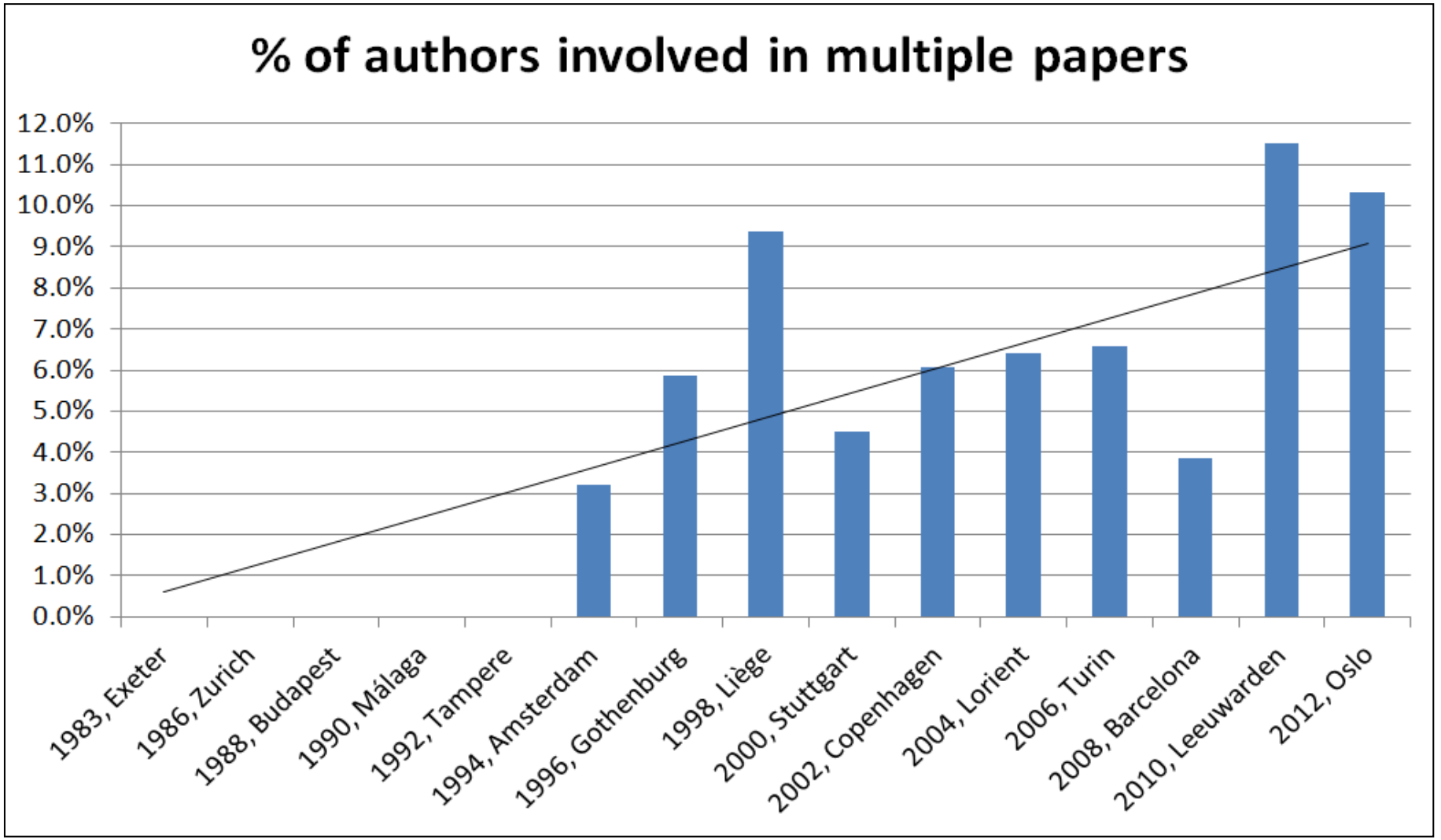
- The number of single-authored papers is steadily declining; in 2012 descending below the 50% level for the first time.
- The number of co-authors per paper indeed tends to grow with each new congress.
- One dares suggest that lexicography is becoming ever more complex, needing the input of more than one scholar, and especially the input from multiple disciplines.

Average number of authors per paper at each congress

Discussion …

- Overall, there are 2 130 authors for the 1 354 papers written so far.
- The number of authors per paper rose — nearly linearly — from an average of about 1.1 three decades ago, to about 1.9 today.
- The average number of authors per paper nearly doubled.

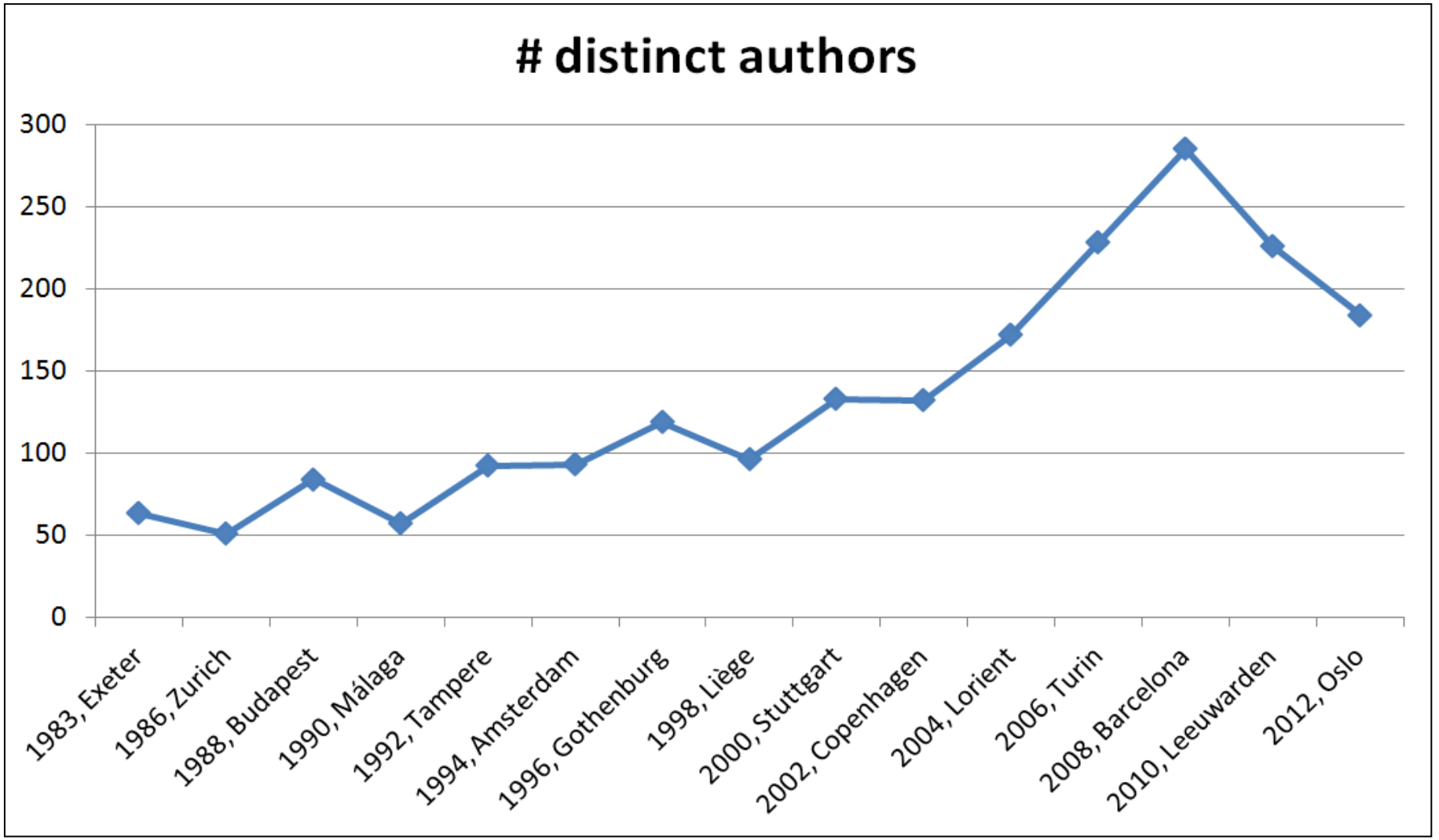
% of authors involved in multiple papers at each congress



Discussion …

- The number of scholars who are involved in multiple papers at the same congress is on the rise.
- This phenomenon that started in 1994, where about 3% of the presenters were involved in multiple papers, a figure which has risen to over 10% today.

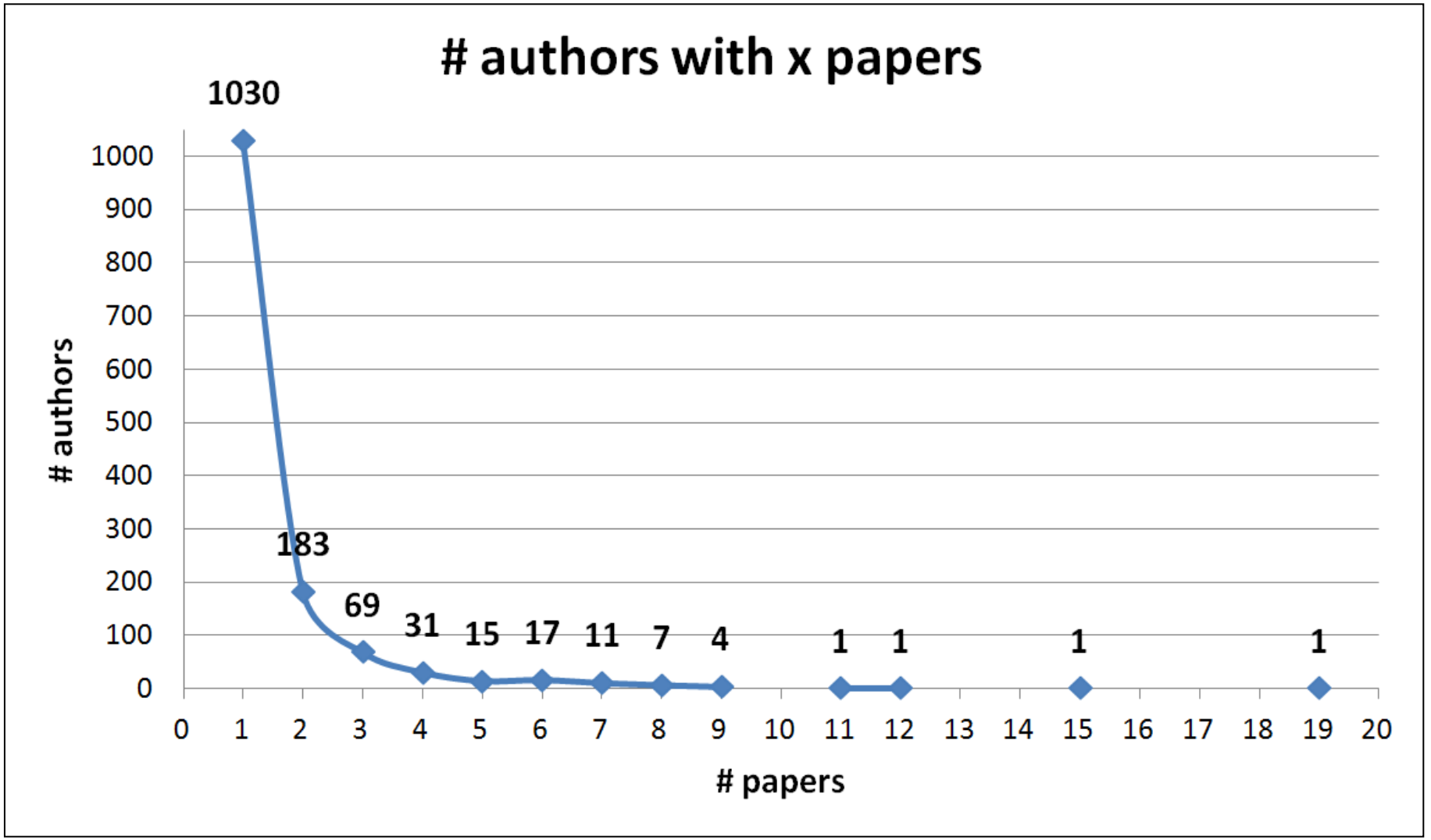
Number of distinct (i.e. unique) authors at each congress



Discussion …

- Given that ever more scholars co-write (and co-present) papers, the actual number of distinct (i.e. unique) authors is thus lower than 2 130.
- Over the years, this metric went from about 50 (in 1983) to nearly 300 (in 2008), and is now back at about 200 authors (in 2012).
- Still an impressive number.

Number of authors with x papers, across all congresses



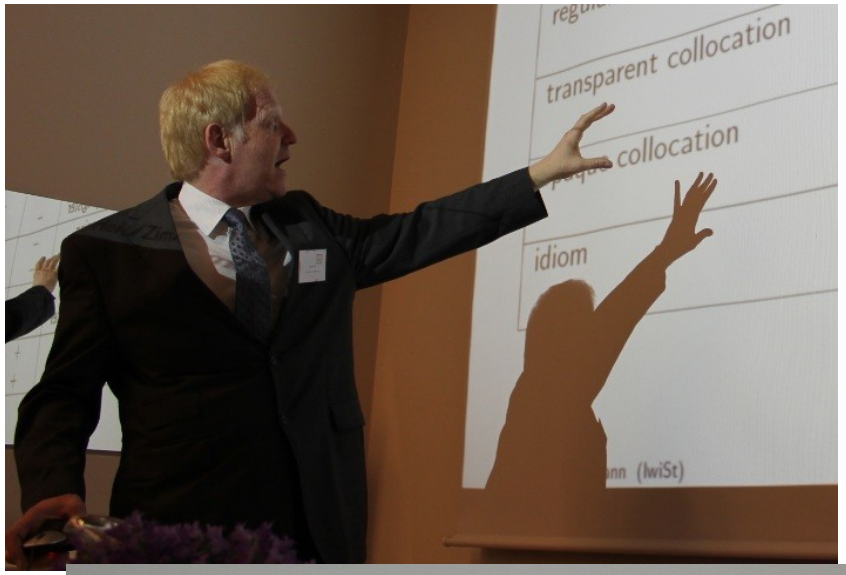
Discussion …

- Loyal (and even very loyal) colleagues do join in with papers time and again.
- A study of all authors, across all fifteen congresses, reveals that a grand total of 1 371 distinct scholars have written papers for EURALEX over the the past three decades.
- The distribution is clearly Zipfian.

Author returns across the various congresses (with > 7 papers)

Author	Papers	1983	1986	1988	1990	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012
Heid, Ulrich	19			1	1		2	2	2	1	1	2	1	1	2	3
Kilgarriff, Adam	15						1	1	1	1	1	1	2	2	3	2
Hanks, Patrick	12		1	1					1	2		1	1	2	2	1
Fontenelle, Thierry	11				1	1	1	2	1	1	1	1	1	1		
Calzolari, Nicoletta	9	1	1	1		1	1	2	2							
de Schryver, Gilles-Maurice	9									1	1	3	1	2		1
DeCesaris, Janet	9								1	1	1	1	2	1	1	1
Verlinde, Serge	9					1	1		2		2	1	1		1	
Abel, Andrea	8									1	1	1	1	2	1	1
Atkins, B. T. Sue	8			1	1			2			1		1		2	
Binon, Jean	8					1	1		2		1	1	1		1	
Picchi, Eugenio	8		1	1		1	1	2	1				1			
Prinsloo, Daan J.	8									1	1	1	1	2	1	1
Rundell, Michael	8		1				1				1		1	1	2	1
ten Hacken, Pius	8					1				1	2		1	1	1	1
Bogaards, Paul	7					1	1		1		1	1		1	1	
Braasch, Anna	7						1			1	1	1	1	1	1	
Čermák, František	7						1		1	1	1		1	1	1	
Dobrovol'skij, Dmitrij O.	7						1			1	1	1	1	1	1	
Gouws, Rufus H.	7		1							1		1	2	1	1	
Lew, Robert	7										2	1	1	1	1	1
Martin, Willy	7			1	1	1		1			2		1			
Moon, Rosamund	7		1			1		1	1	1	1	1				
Swanepoel, Piet H.	7				1	1	1		1	1			1		1	
van der Meer, Geart	7							1	1	1	1	1	1	1		
Varantola, Krista	7					1	1		2	1		2				

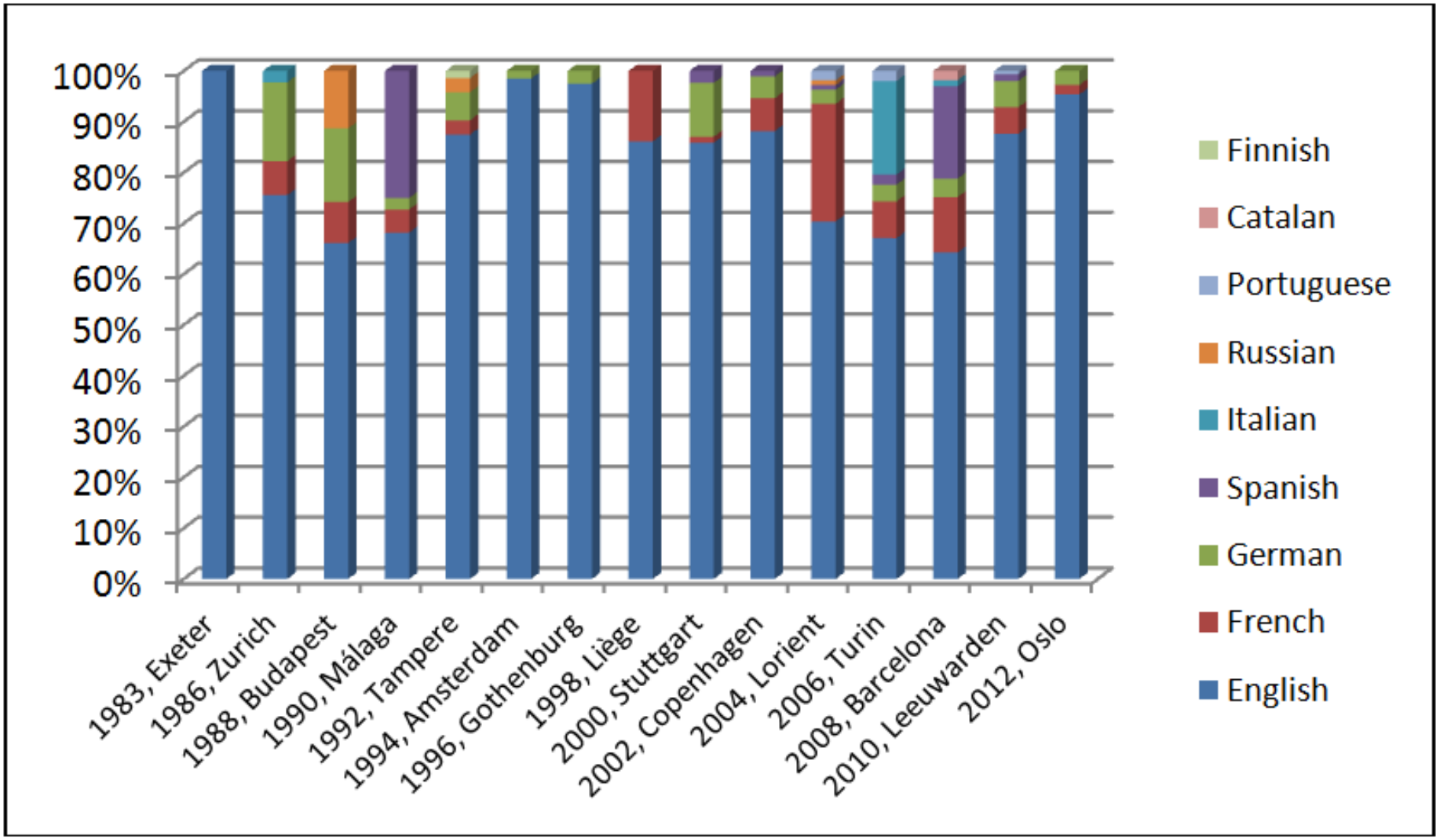
Top 4 author returns across the various congresses



Discussion …

- To the insider, it will of course not really come as a surprise to see that Ulrich Heid, Adam Kilgarriff, Patrick Hanks and Thierry Fontenelle top this list.
- Each of them has become synonymous with major developments in the field at large, and it is gratifying to see their devotion to EURALEX.
- All other scholars listed in Table 2 are most certainly 'must-reads' as well.

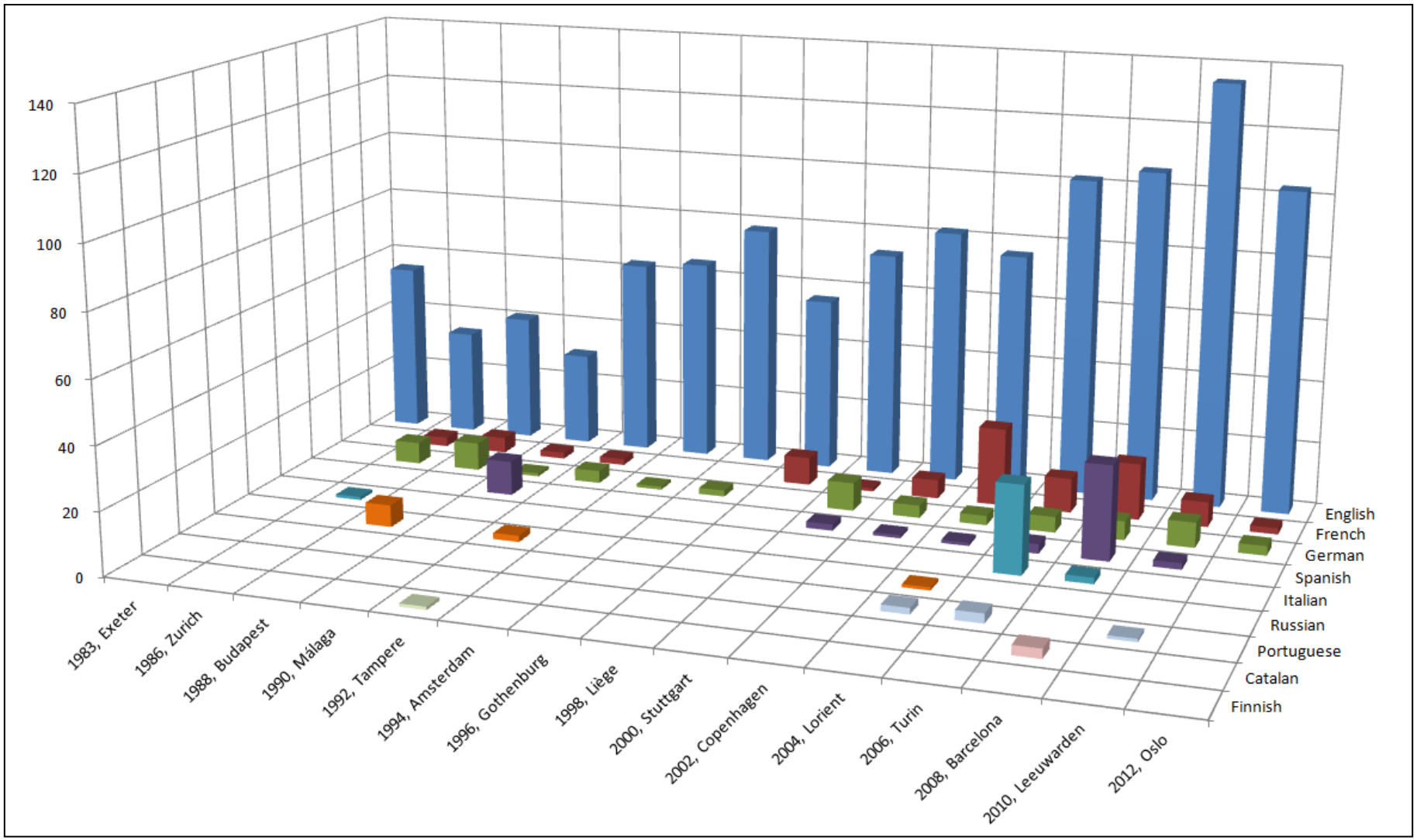
Languages of papers, in % per congress



Discussion …

- Nine languages have been used for the 1 354 papers to date:
  - 1 099 in English (81.2%),
  - 92 in French (6.8%),
  - 62 in German (4.6%),
  - 50 in Spanish (3.7%),
  - 31 in Italian (2.3%),
  - 10 in Russian (0.7%),
  - 6 in Portuguese (0.4%),
  - 3 in Catalan (0.2%), and
  - a single one in Finnish (0.1%).

Languages of papers, actual number per congress



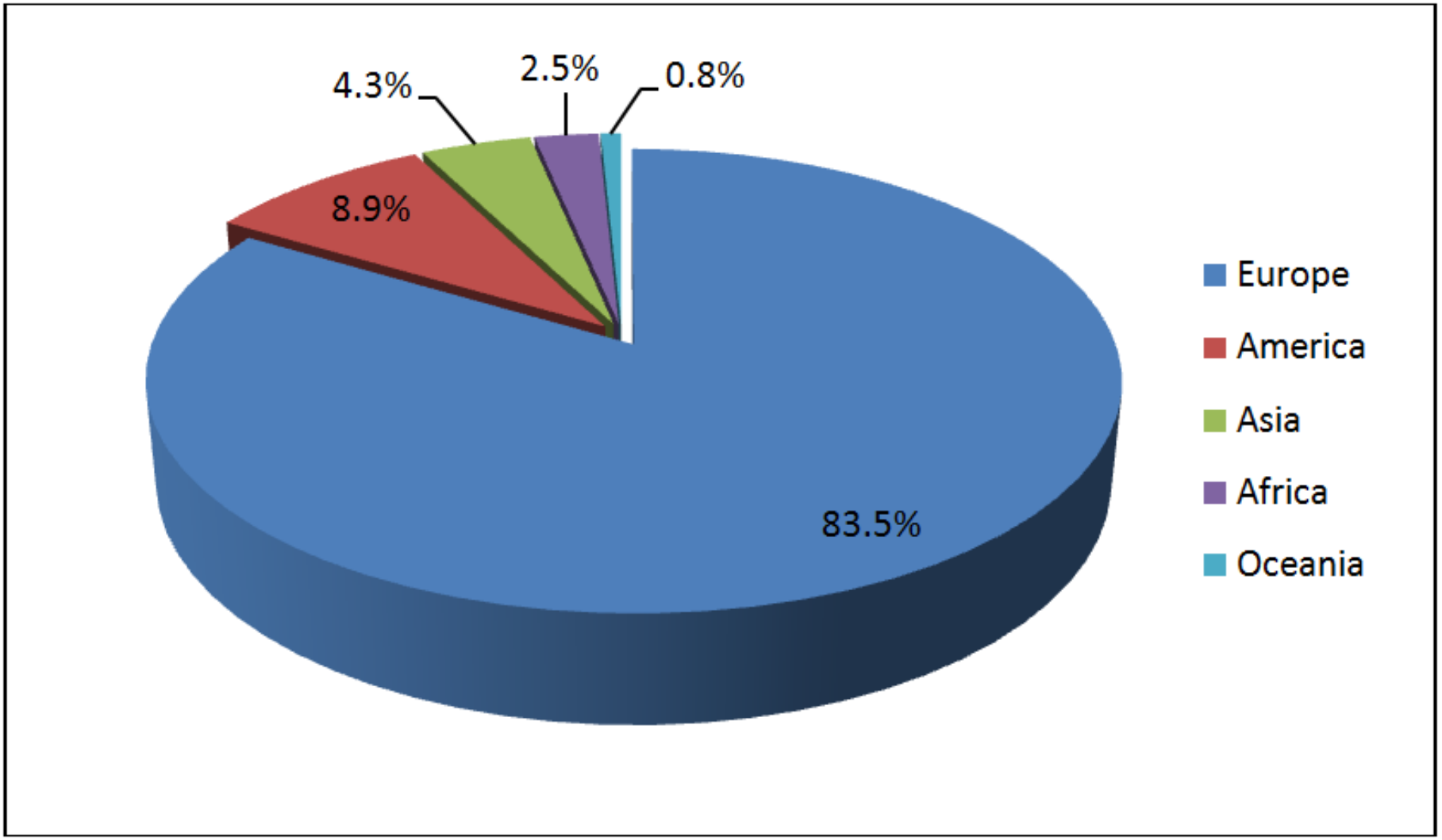
Discussion …

- EURALEX congresses clearly seem to act as a magnet for local researchers, turning EURALEX congresses in combined international and national gatherings as they move around the continent.

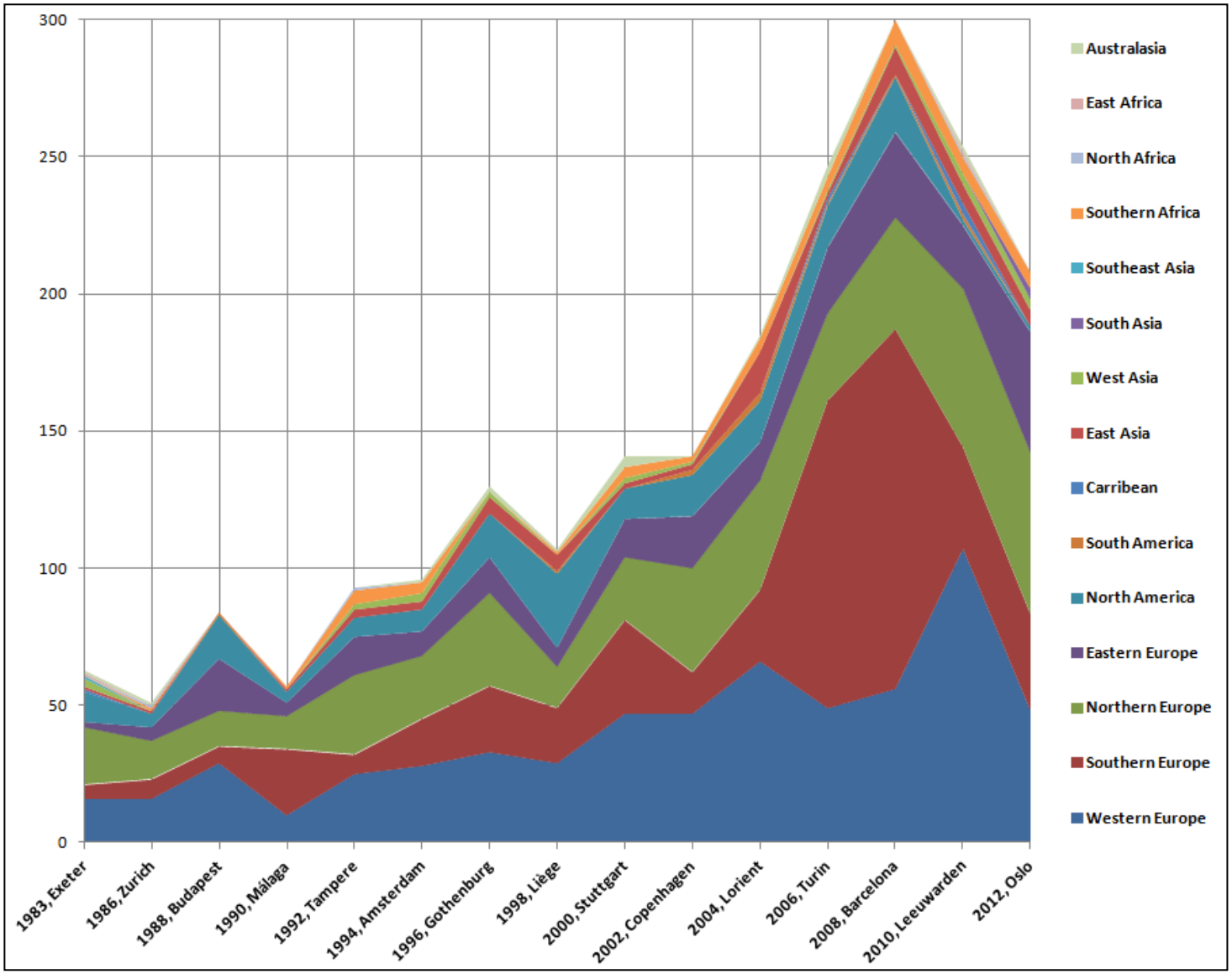
Country distribution of the affiliations for all authors

Region	Sub-region	Country	Papers	%
Europe	Southern Europe	Spain	222	10.29%
Europe	Southern Europe	Italy	199	9.23%
Europe	Northern Europe	United Kingdom	192	8.90%
Europe	Western Europe	Germany	179	8.30%
Europe	Western Europe	The Netherlands	141	6.54%
Europe	Western Europe	France	127	5.89%
America	North America	USA	100	4.64%
Europe	Western Europe	Belgium	86	3.99%
Europe	Eastern Europe	Russia	75	3.48%
Europe	Northern Europe	Denmark	71	3.29%
America	North America	Canada	63	2.92%
Europe	Northern Europe	Sweden	62	2.87%
Europe	Eastern Europe	Czech Republic	57	2.64%
Africa	Southern Africa	South Africa	49	2.27%
Europe	Eastern Europe	Poland	49	2.27%
Asia	East Asia	Japan	37	1.72%
Europe	Western Europe	Switzerland	36	1.67%
Europe	Western Europe	Austria	35	1.62%
Europe	Eastern Europe	Hungary	32	1.48%
Europe	Northern Europe	Ireland	30	1.39%
Europe	Southern Europe	Slovenia	27	1.25%
Europe	Northern Europe	Norway	26	1.21%
Europe	Northern Europe	Finland	25	1.16%
Europe	Northern Europe	Estonia	24	1.11%

Region distribution of the affiliations for all authors



Sub-region contribution per congress



3. THE EURALEX CITATION DATABASE

Setting up for citation analysis

- Not all papers make a lasting impact.
- Those that do, typically attract a number of citations over the years.
- Although this is not a substitute for inherent quality — after all, one can theoretically also and only refer to a paper merely to point out its infelicities — high citation counts typically correspond to satisfaction.

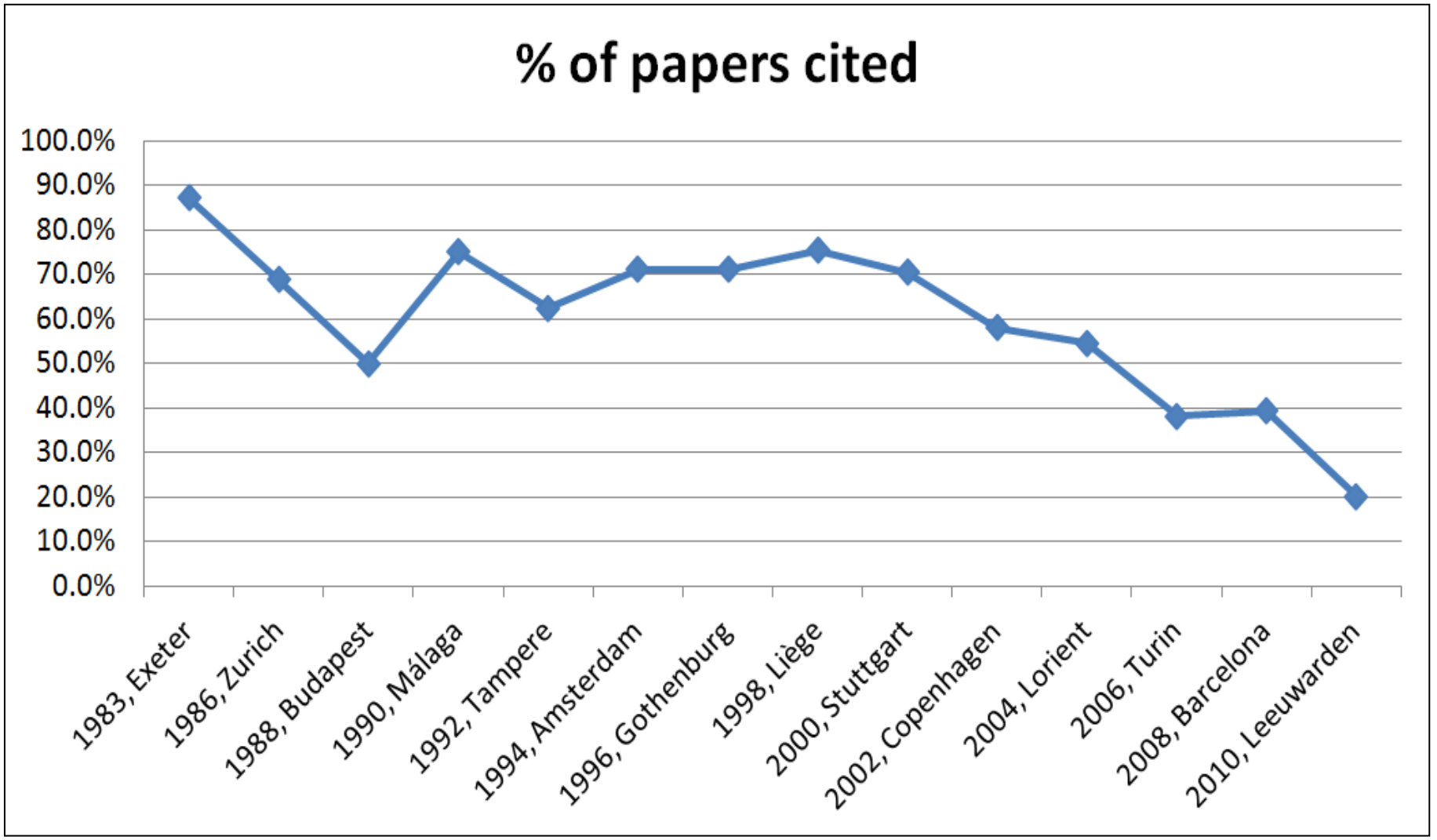
Setting up for citation analysis

- Today, the most convenient way to determine a paper's citations is simply to query Google Scholar.
- In what follows, the citation status in Google Scholar as reflected on 24 July 2012 is used.
- The Google Scholar database does *not* see everything (yet), so all values are minimum values.

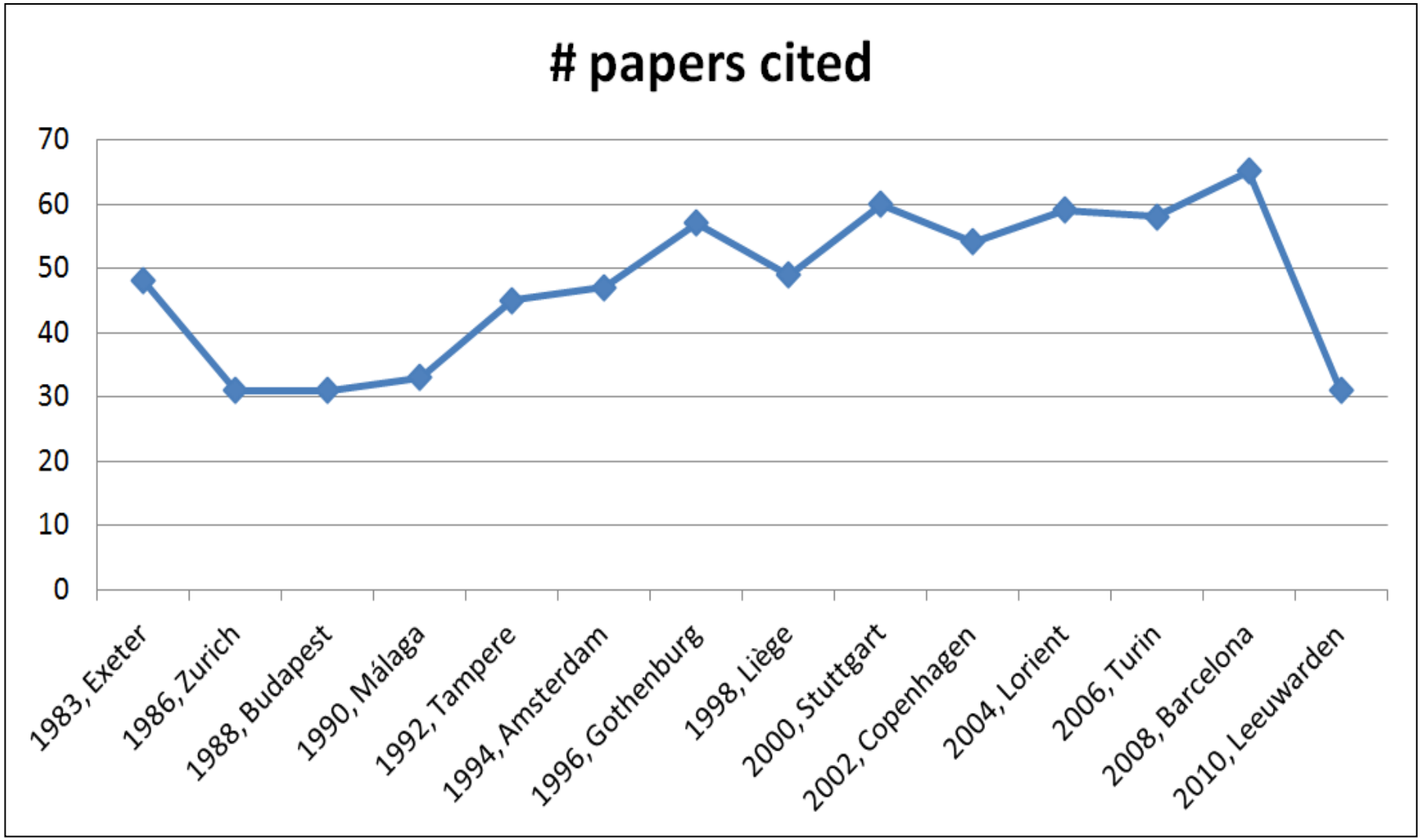
Setting up for citation analysis

- Given a congress paper first has to be published this section of the study looks at all the papers from the first fourteen congresses only.
- In all, there are 1 246 papers for this period.
- 668 (or thus 53.6%) have been cited at least once.

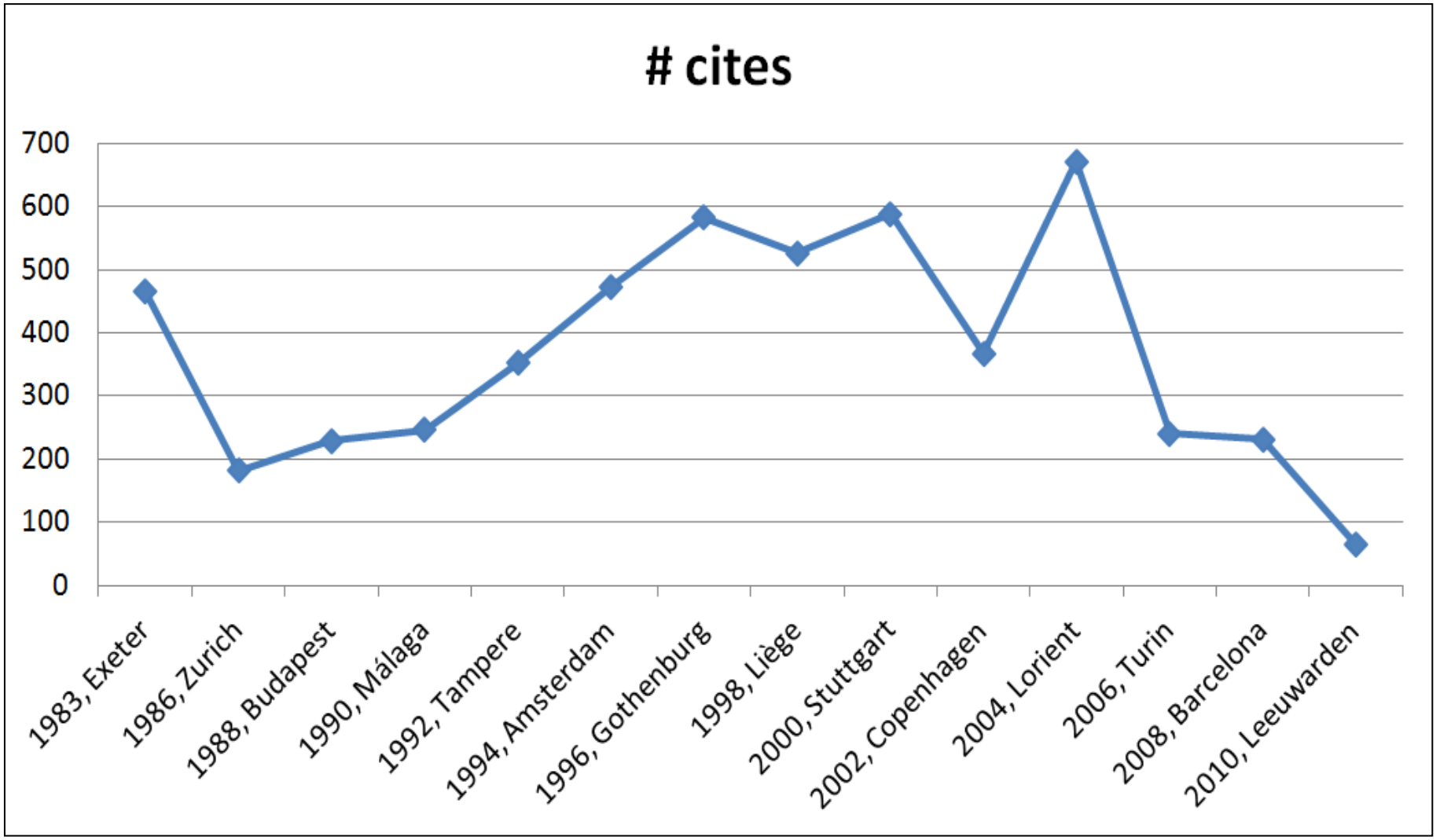
Percent of papers cited, per congress



Number of papers cited, per congress



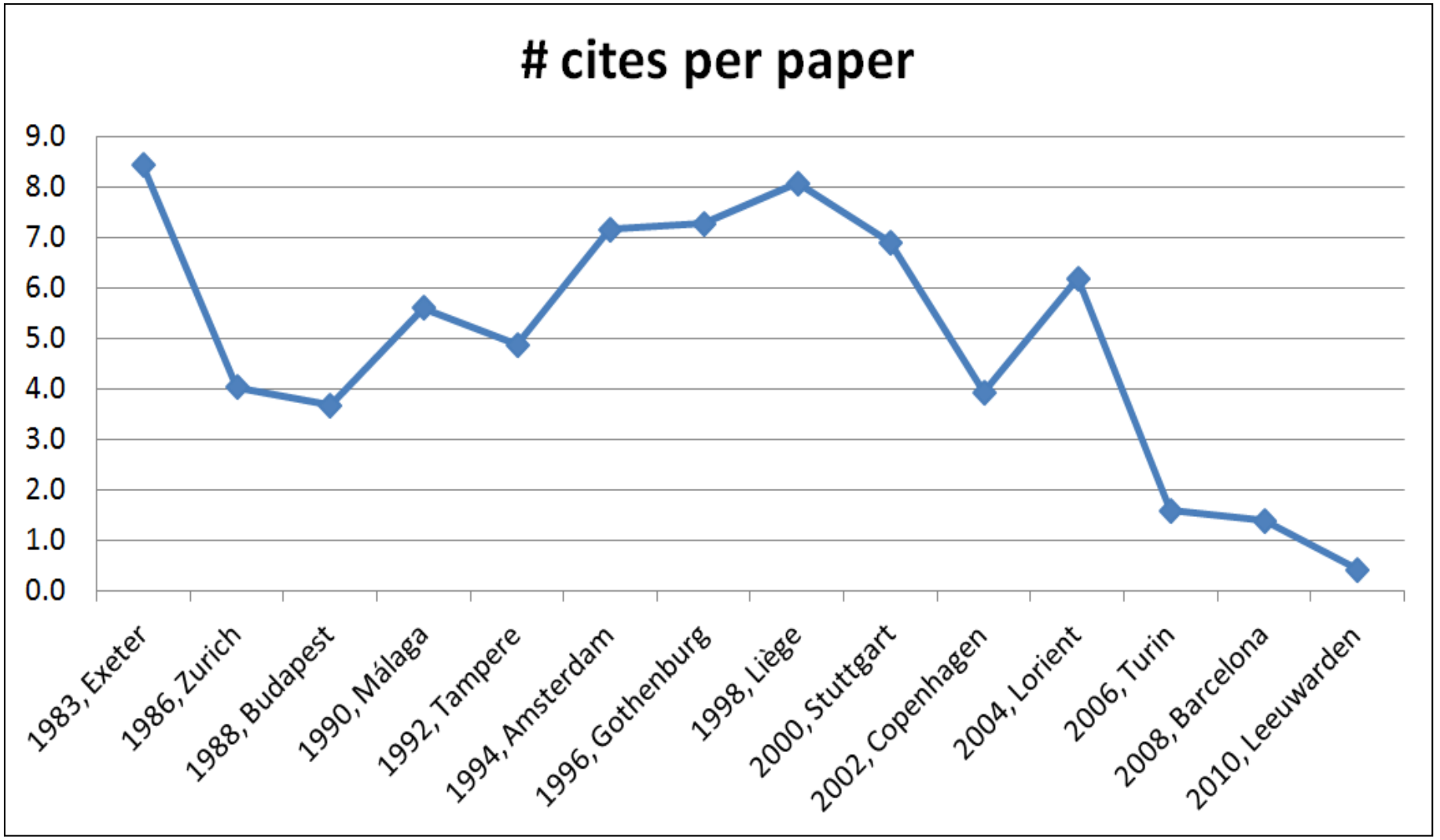
Number of cites, per congress



Discussion …

- The actual number of references to the first fourteen congresses adds up to 5 220 cites.
- In terms of overall cites, the 2004 (Lorient) congress made the biggest impact so far.
- Of course this may be (and *is*) the result of just a single very-high-impact paper.
- A better way to study the data is therefore to look at the average number of references per paper presented at each congress.

Average number of cites per paper at each congress



Putting a face on the bleak statistics

- The sub-discipline of computational lexicography easily elbows out the more traditional aspects of the discipline.
- Especially NLP topics top the more recent lists.
- This is a trend set in motion at the 1992 (Tampere) congress, gaining strength at the 1994 (Amsterdam) and 1996 (Gothenburg) congresses, and unleashed in full as of the 1998 (Liège) congress.

Cites	Title	Author(s)	Year
349	The Sketch Engine	Kilgarriff, Adam; Rychlý, Pavel; Smrz, Pavel; Tugwell, David	2004
130	SENSEVAL: An Exercise in Evaluating Word Sense Disambiguation Programs	Kilgarriff, Adam	1998
105	Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French	Polguère, Alain	2000
102	COMLEX Syntax: An On-Line Dictionary for Natural Language Processing	Macleod, Catherine; Grishman, Ralph; Meyers, Adam	1996
102	NOMLEX: a lexicon of nominalizations	Macleod, Catherine; Grishman, Ralph; Meyers, Adam; Barrett, Leslie; Reeves, Ruth	1998
82	Corpus-Derived First, Second and Third-Order Word Affinities	Grefenstette, Gregory	1994
73	Systematic polysemy in lexicology and lexicography	Nunberg, Geoffrey; Zaenen, Annie	1992
71	On Ways Words Work Together – Topics in Lexical Combinatorics	Heid, Ulrich	1994
65	Bilingual Dictionaries: Past, Present and Future	Atkins, B. T. Sue	1996
59	Interim Report on the EURALEX/AILA Research Project Into Dictionary Use	Atkins, B. T. Sue; Knowles, Frank E.	1990
56	Translational equivalence in the bilingual dictionary	Zgusta, Ladislav	1984
53	Studying dictionary use: some findings and proposals	Hatherall, Glyn	1984
50	Lexical Profiling Software and its Lexicographic Applications - a Case Study	Kilgarriff, Adam; Rundell, Michael	2002
48	GDEX: Automatically Finding Good Dictionary Examples in a Corpus	Kilgarriff, Adam; Husák, Miloš; McAdam, Katy; Rundell, Michael; Rychlý, Pavel	2008
45	On the structure and contents of a general theory of lexicography	Wiegand, Herbert Ernst	1984
41	The FrameNet Database and Software Tools	Ruppenhofer, Josef; Baker, Collin F.; Fillmore, Charles J.	2002

4. THE EURALEX PROCEEDINGS CORPUS

Setting up for corpus analysis

- The EURALEX proceedings corpus =
  - the full-text corpus of all the papers and editorial material of the fifteen EURALEX congresses to date;
  - close to five million running words;
  - the English part is about 4 million words strong, with 146 thousand distinct words.
- Why singling out the English component?
  - the idea is to study trends based on keywords;
  - not enough data with a good distribution for the other languages.

Setting up for corpus analysis

- Determining the keywords:
  - the EURALEX proceedings corpus was compared to the 100-million-word BNC;
  - overall 'keyness values' were calculated;
  - using the log-likelihood statistic;
  - with minimum frequency set at 3; and
  - maximum probability at 0.000001.
- About 15 thousand types were found to be 'key' (i.e. positively outstanding).
- The first 1 000 were studied in detail.

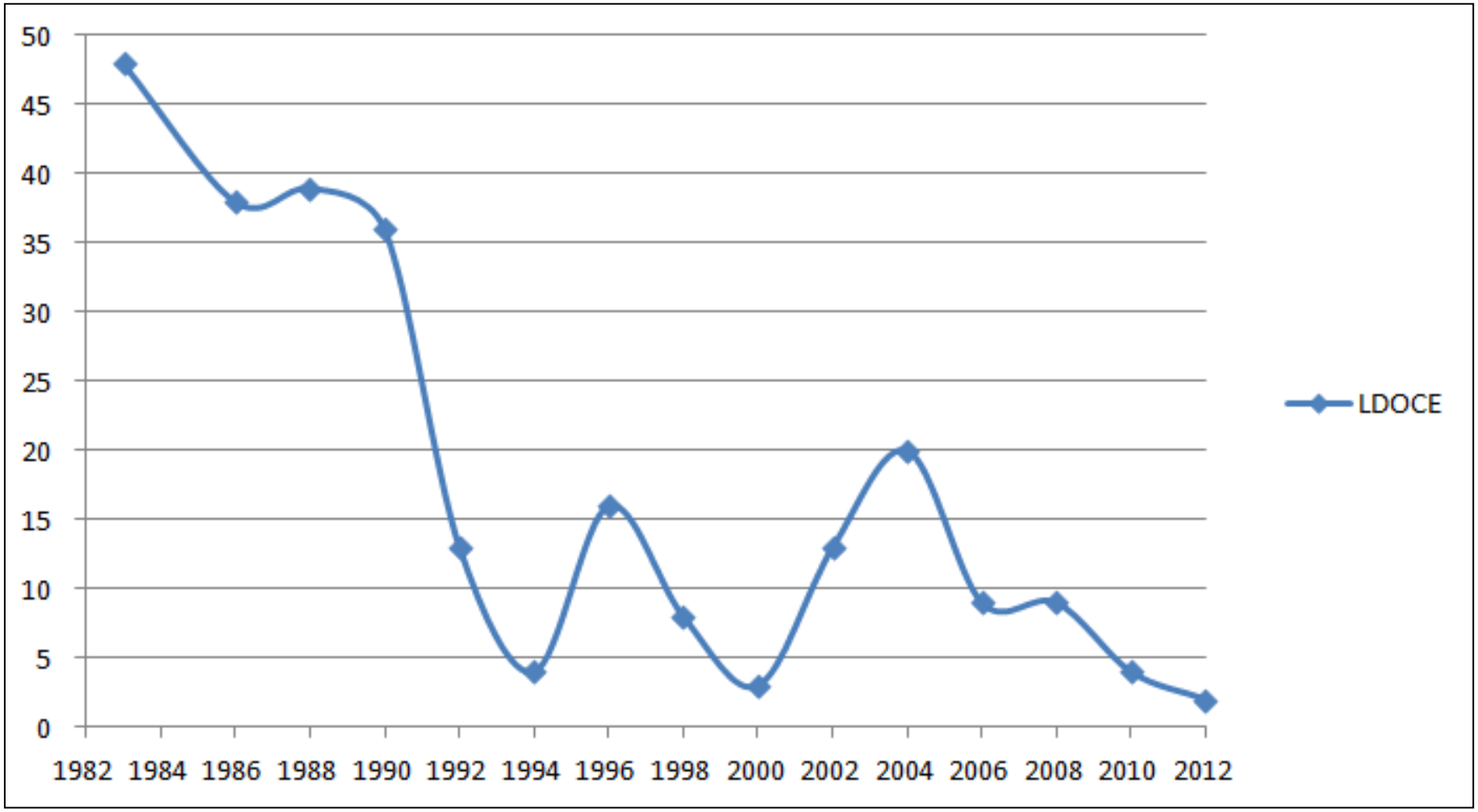
Setting up for corpus analysis

- For each of these 1 000 keywords, the frequency in each of the fifteen congress sub-corpora was determined.
- In order to be able to compare the frequencies across the congress sub-corpora the frequencies were normalised to show number of occurrences per 100 thousand words.
- The result of this analysis is shown in the Addendum of the printed version.

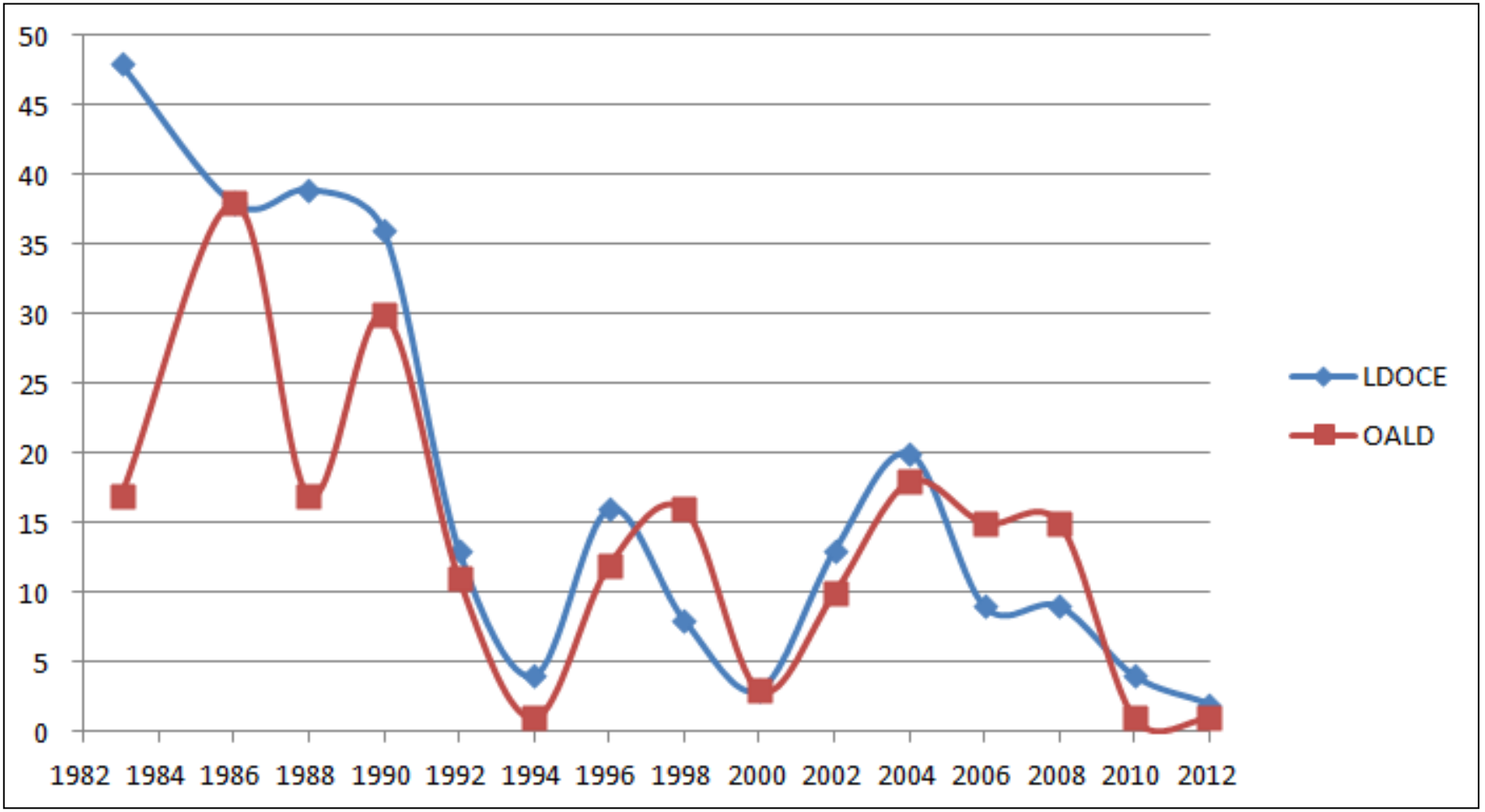
Setting up for corpus analysis

- The possible uses of the data shown in the Addendum are many and varied, and only a small selection will be presented here.
- The interested reader is invited to look at the Euralex keywords not covered, guided by their interest in certain topics.
- To begin with, however, a true EURALEX classic …

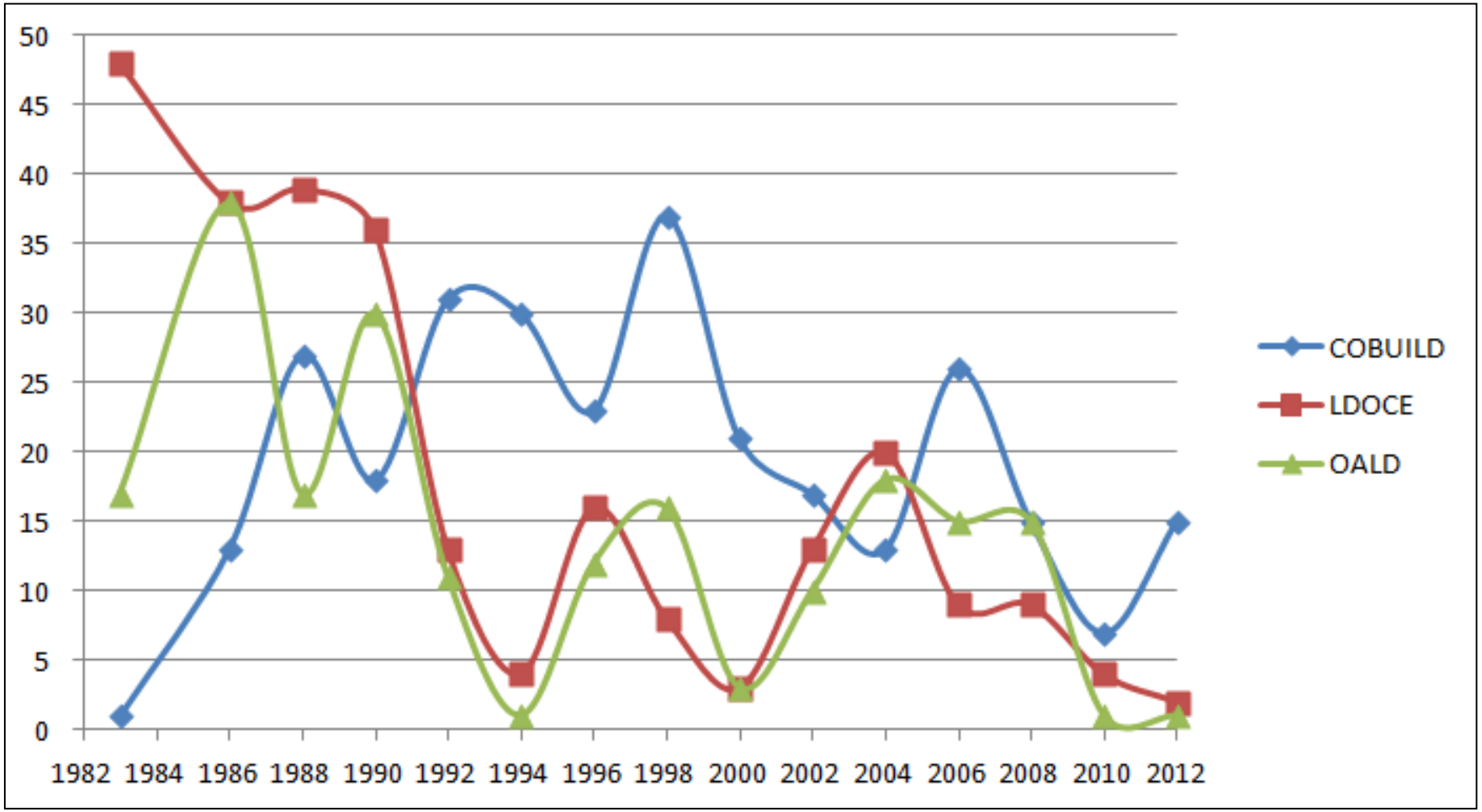
Trend for the Big Five monolingual learners' dictionaries



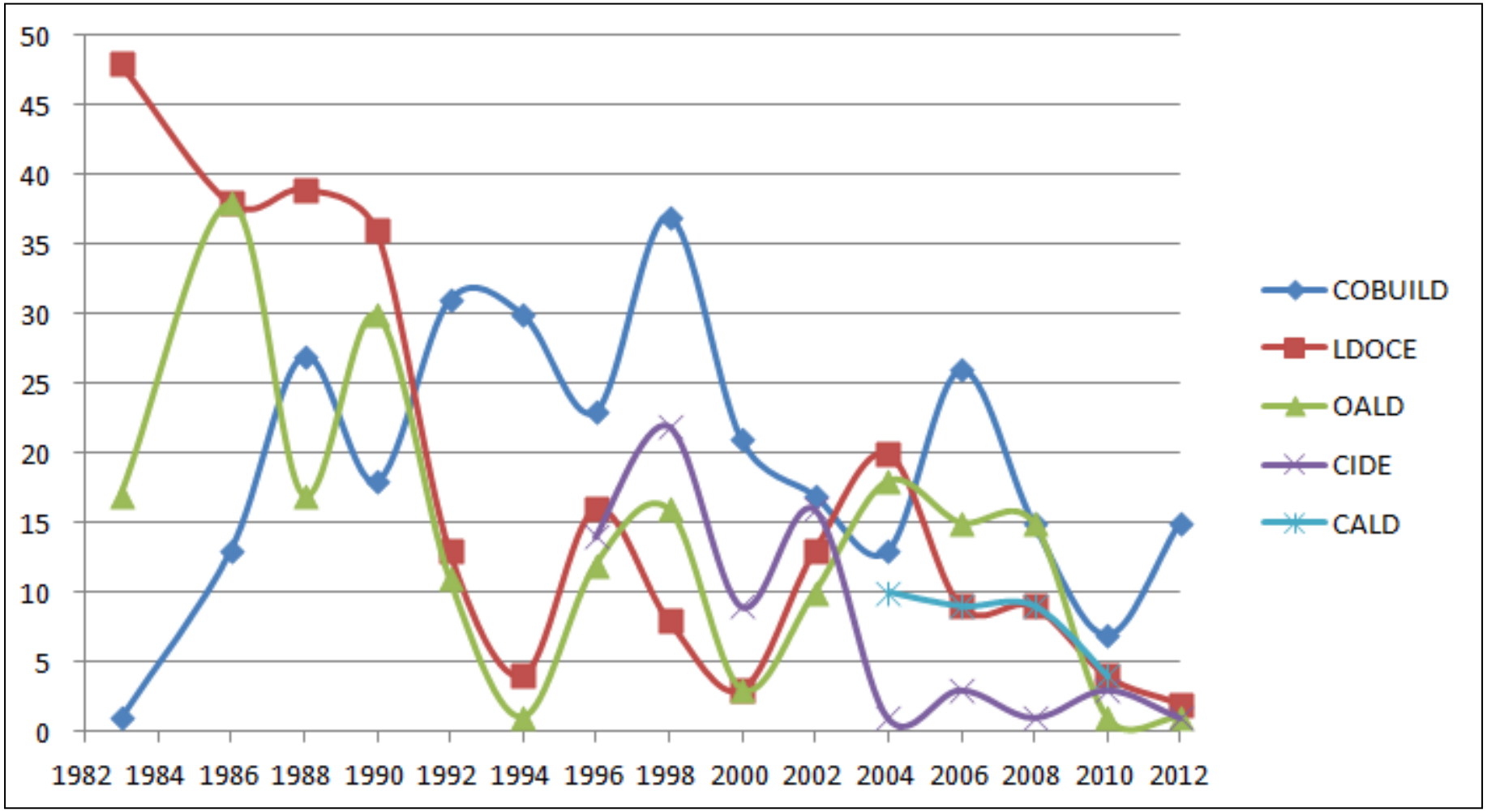
Trend for the Big Five monolingual learners' dictionaries



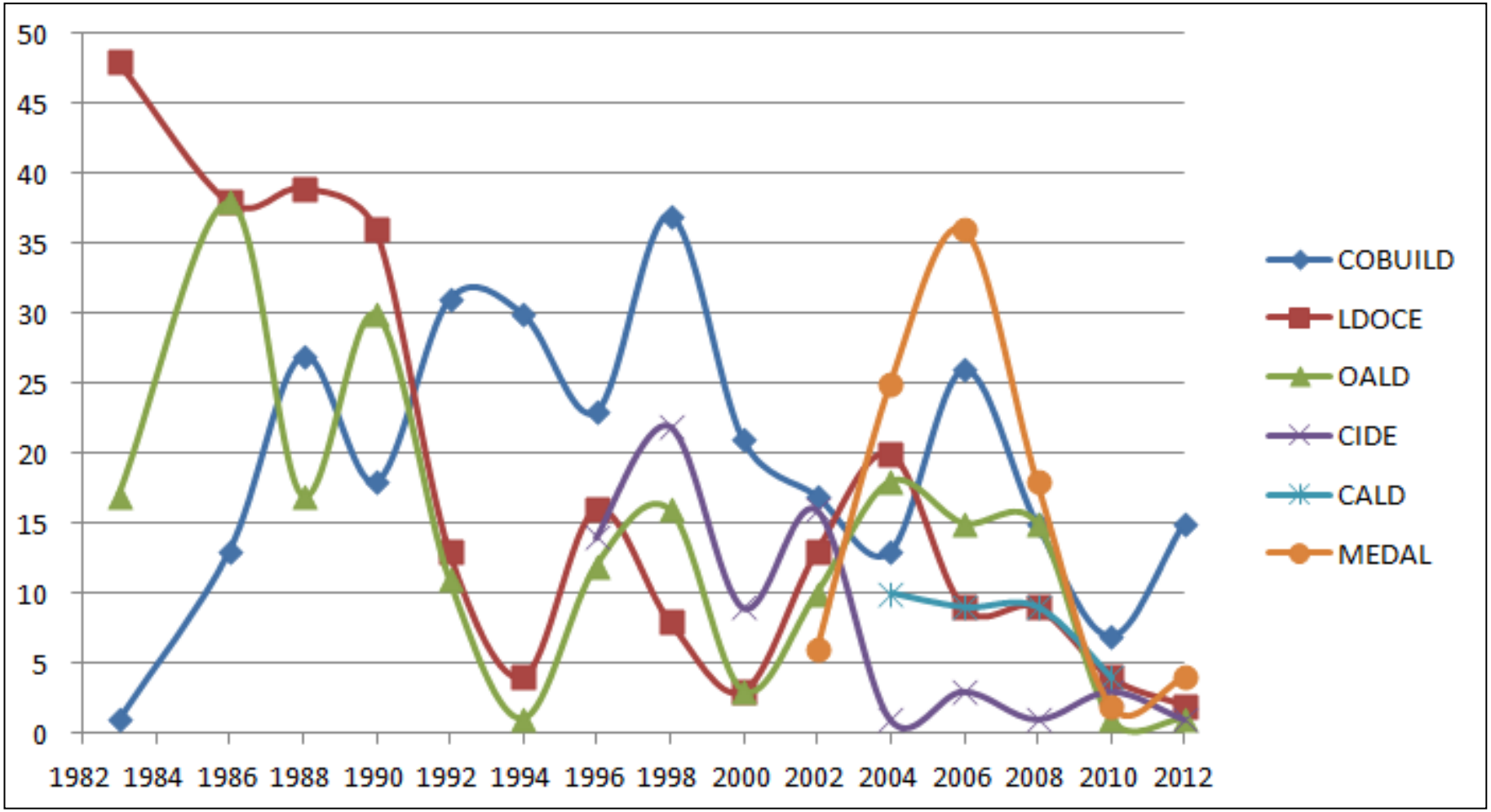
Trend for the Big Five monolingual learners' dictionaries



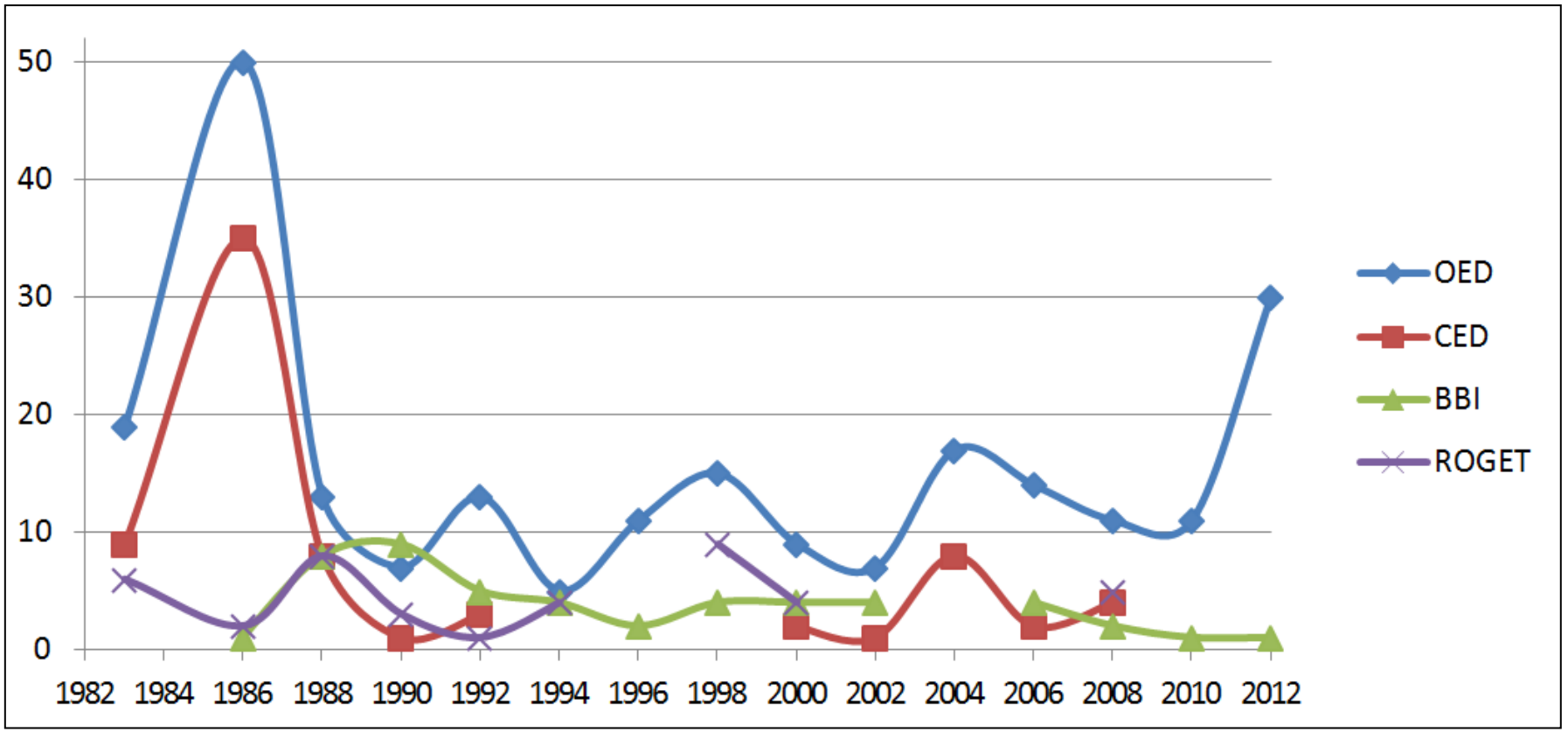
Trend for the Big Five monolingual learners' dictionaries



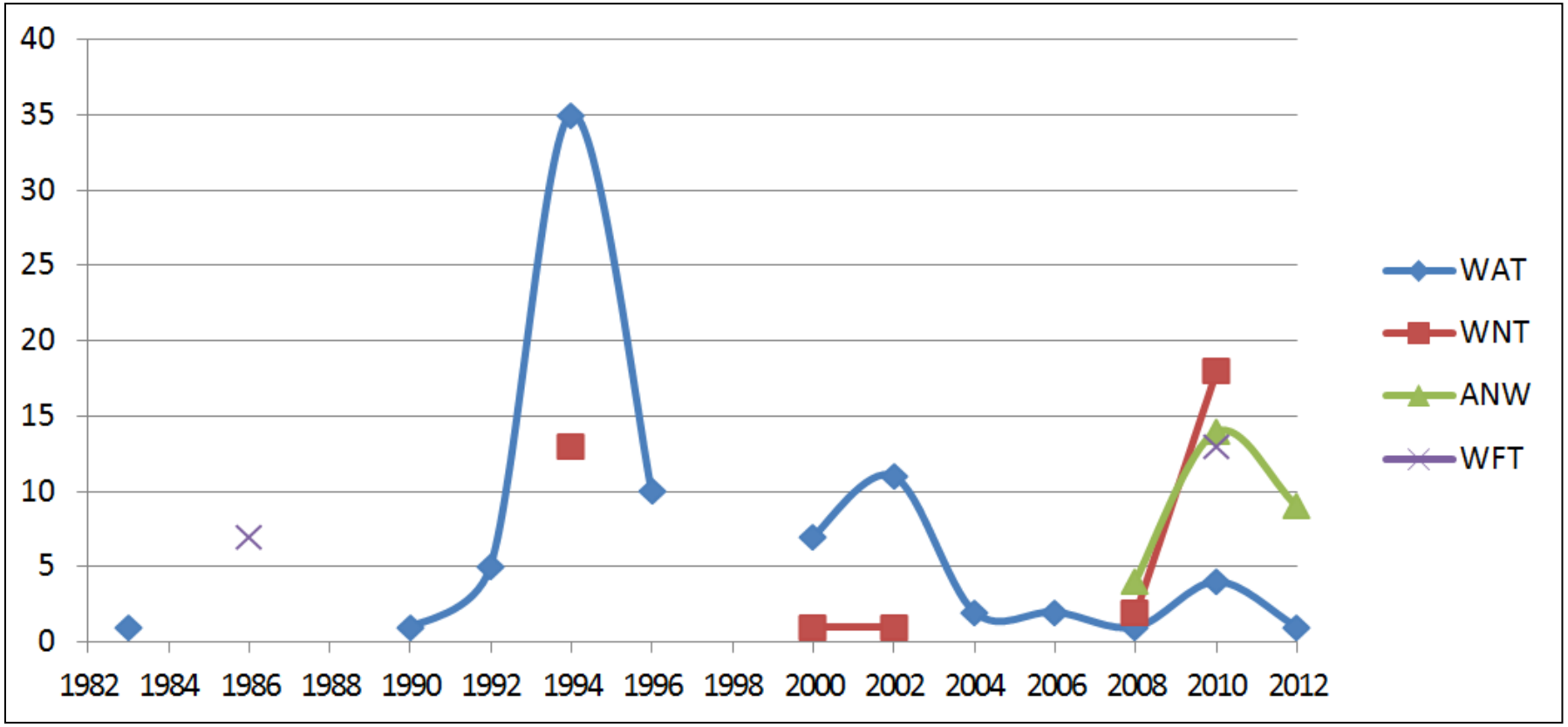
Trend for the Big Five monolingual learners' dictionaries



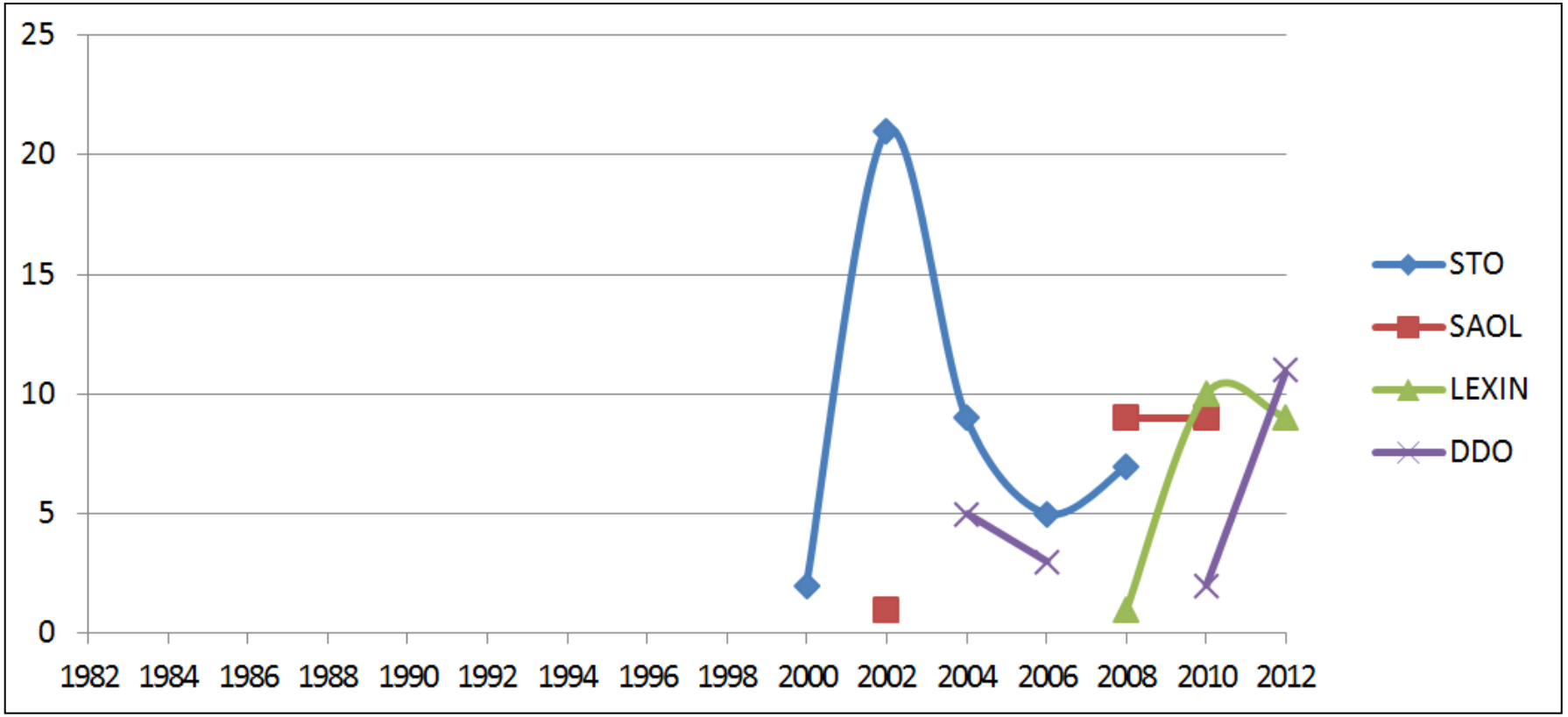
Trend for four types of English dictionaries



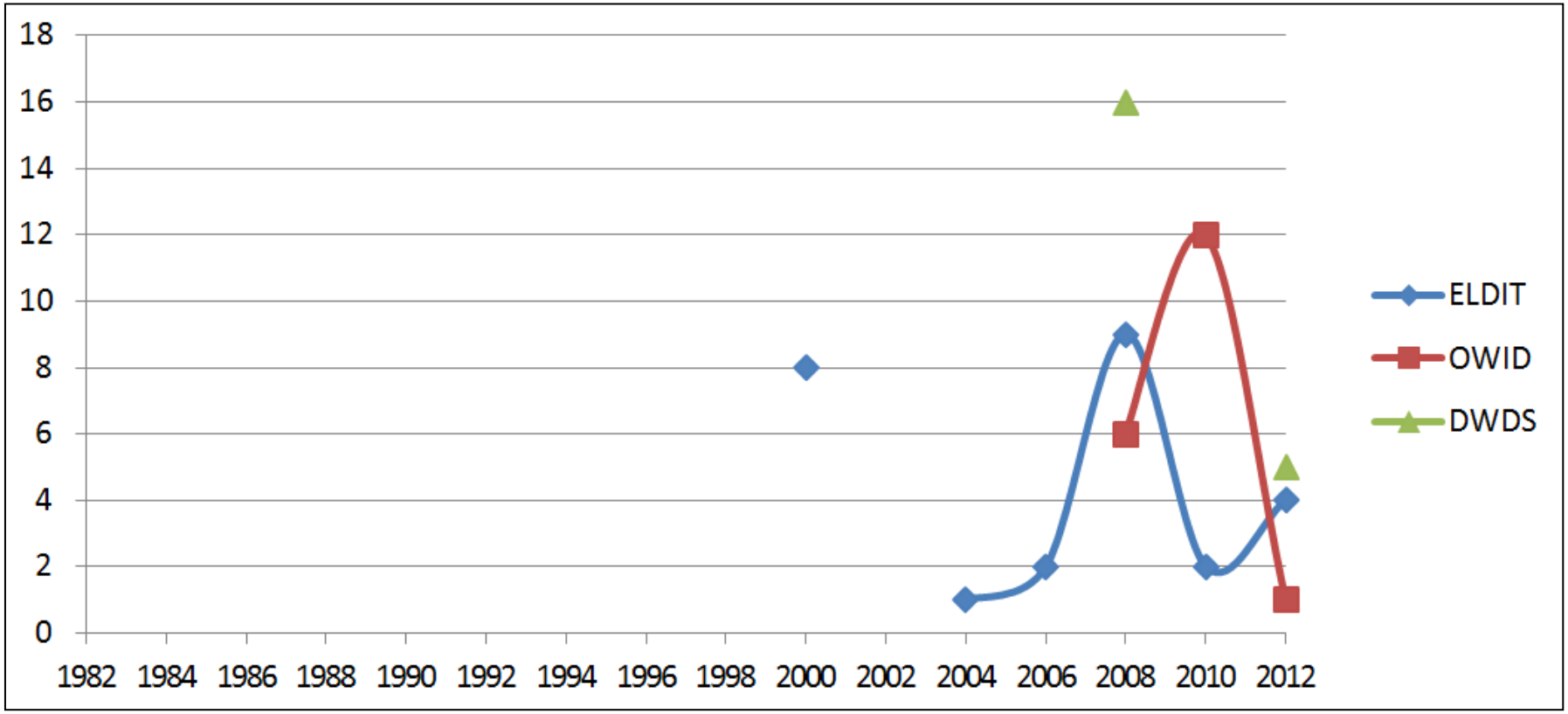
Trend for four dictionaries in Afrikaans, Dutch and Frisian



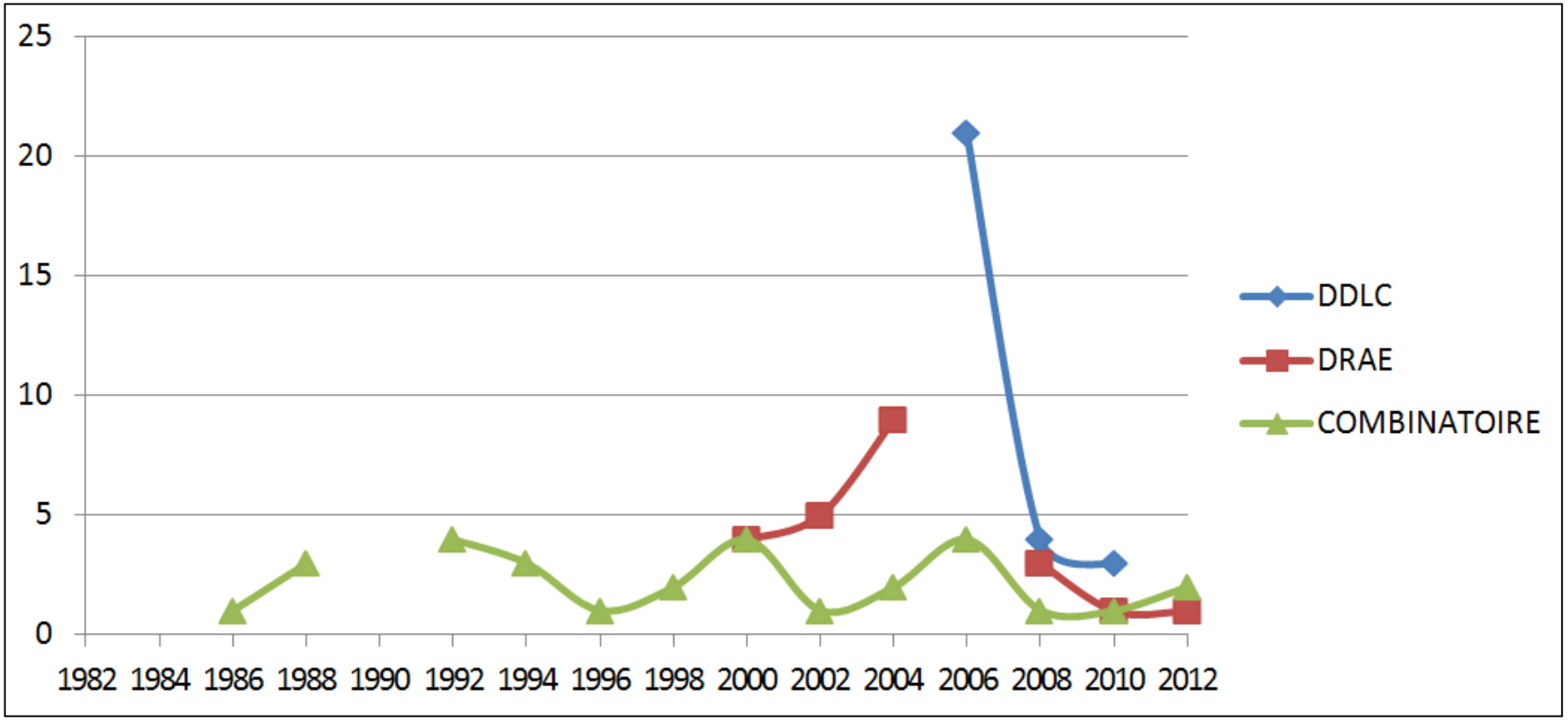
Trend for two Danish and two Swedish dictionaries



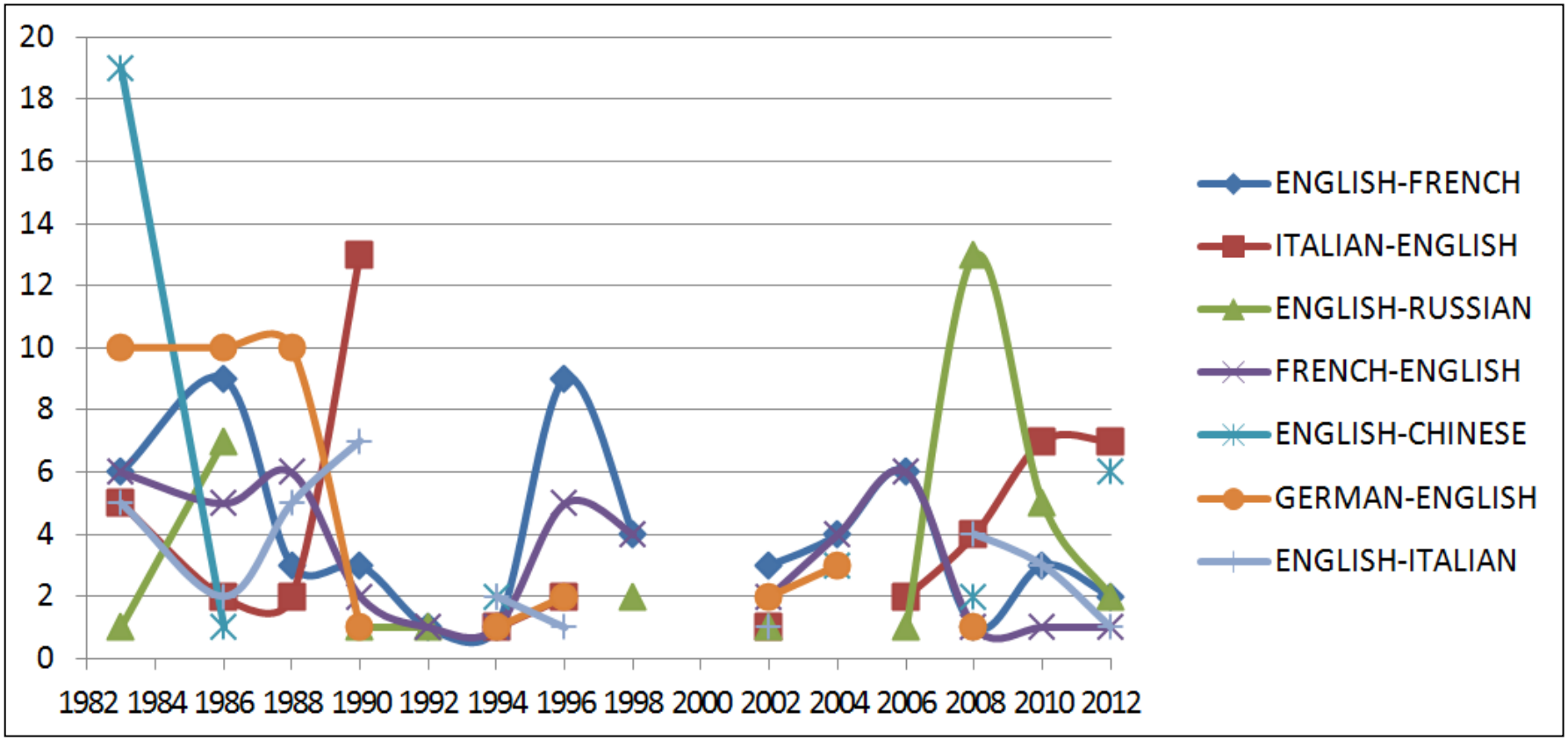
Trend for dictionaries involving German



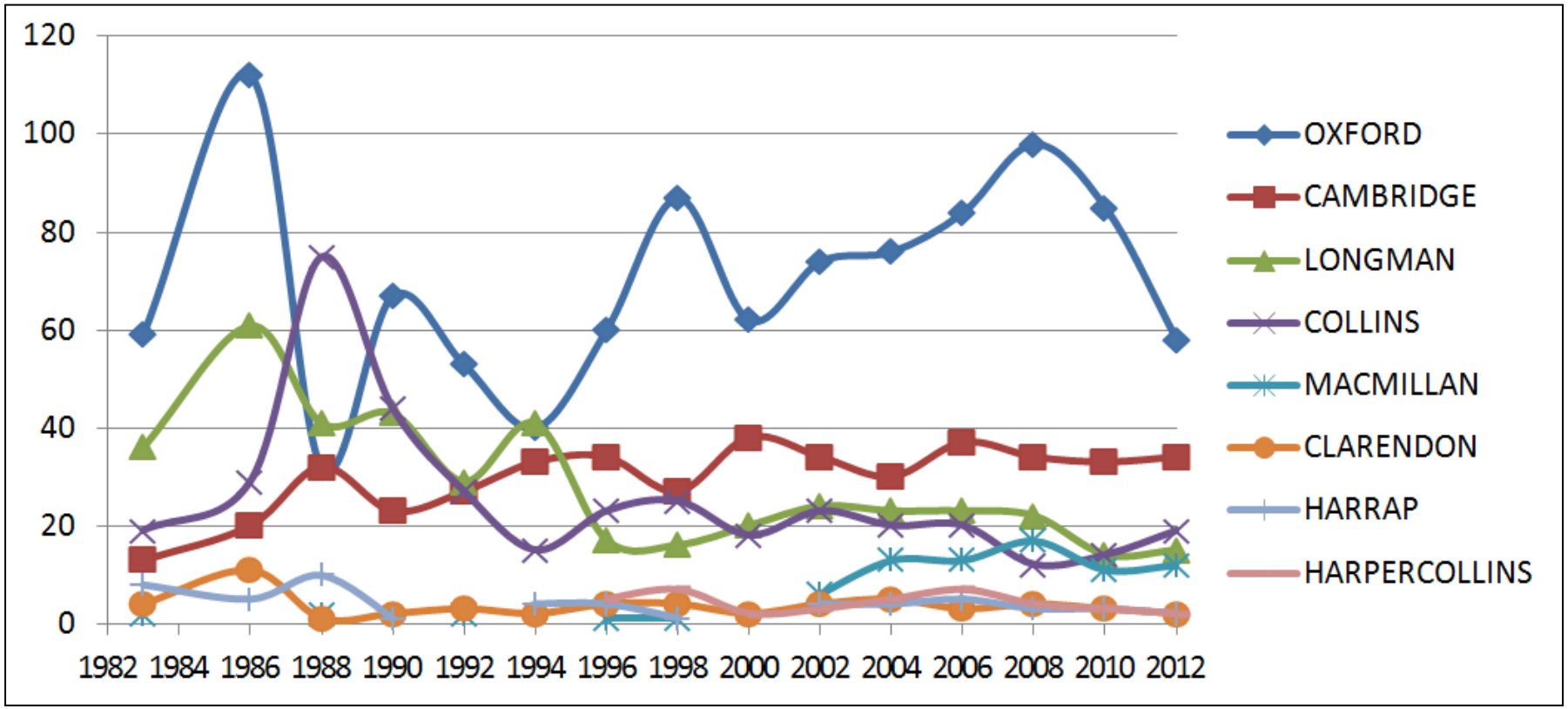
Trend for dictionaries of Romance languages



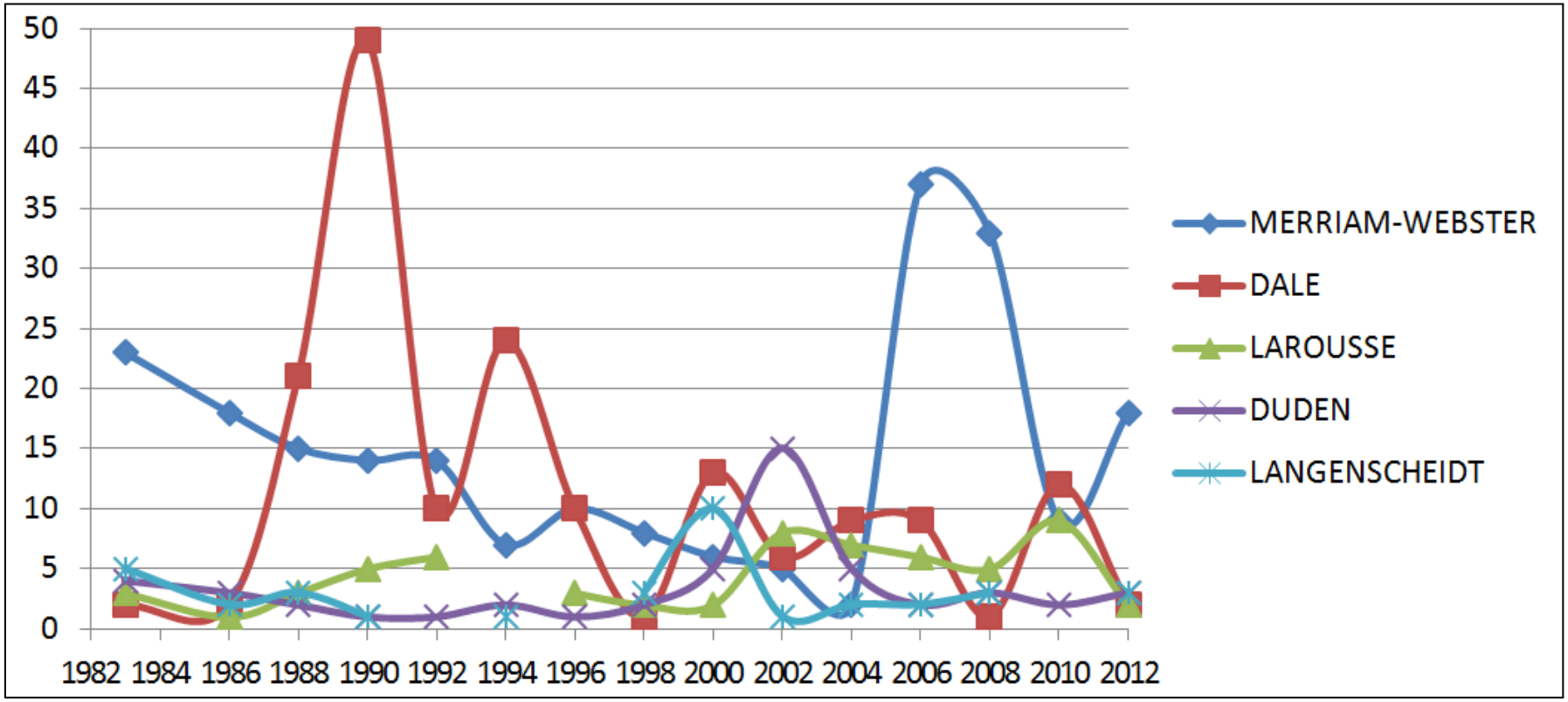
Trend for popular dictionary language pairs



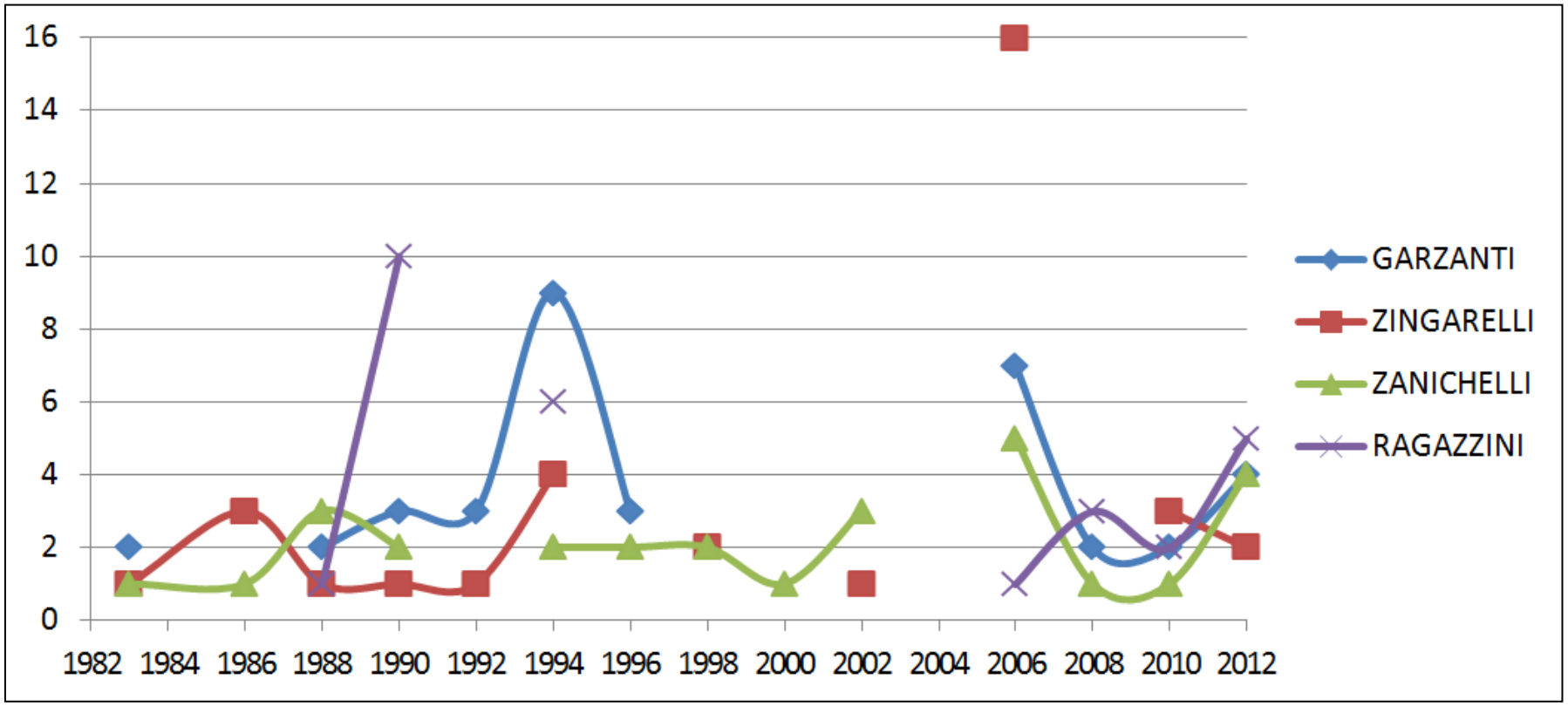
Trend for a number of British dictionary publishers



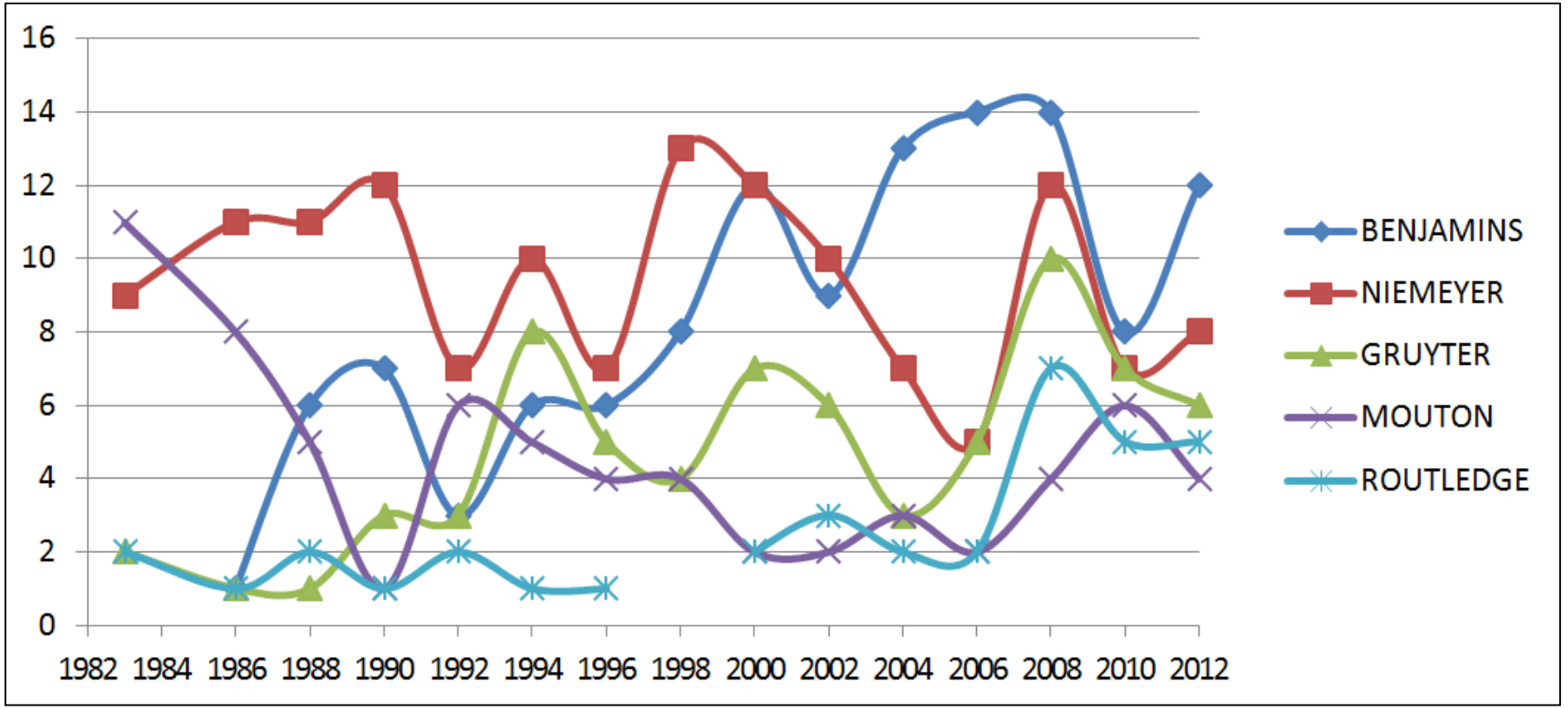
Trend for a number of US / continental dictionary publishers



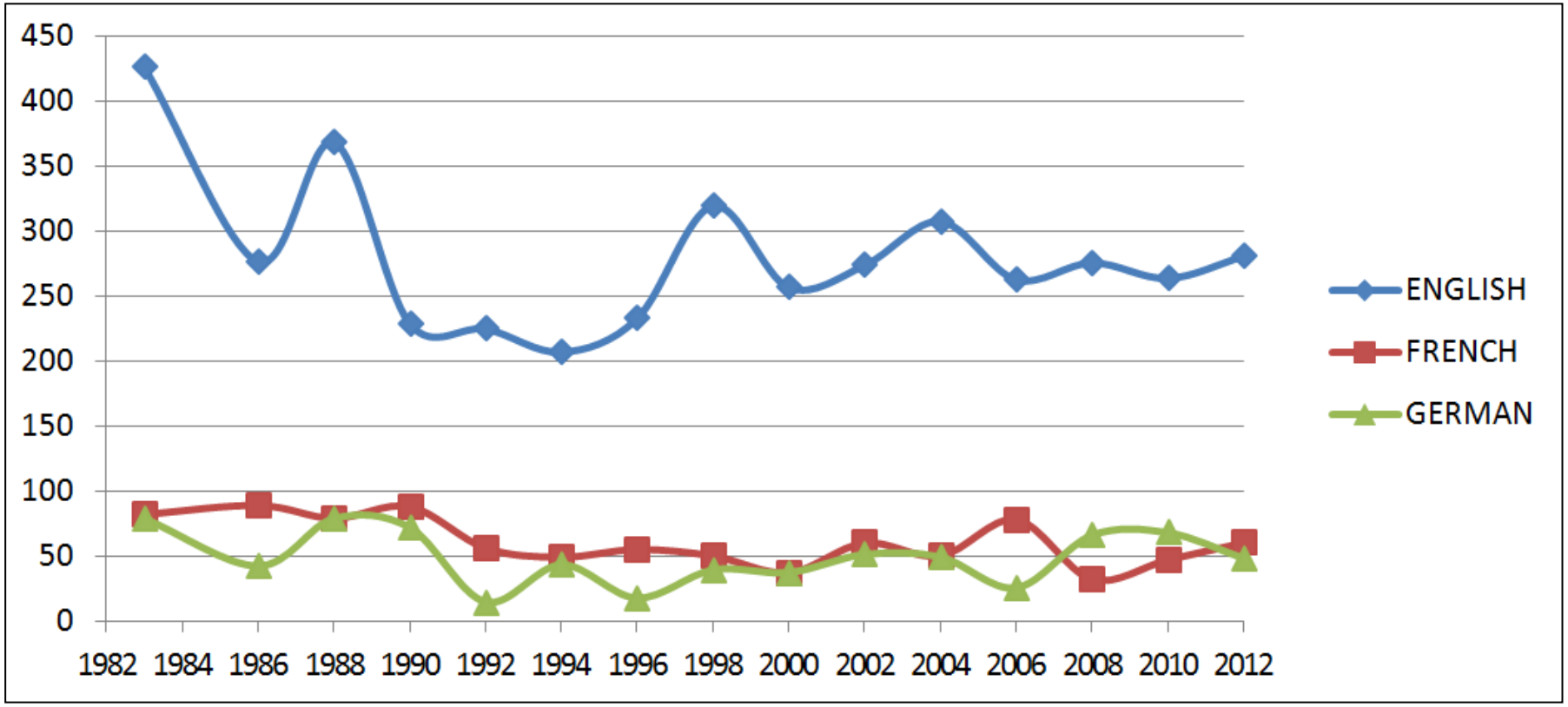
Trend for a number of Italian dictionary publishers

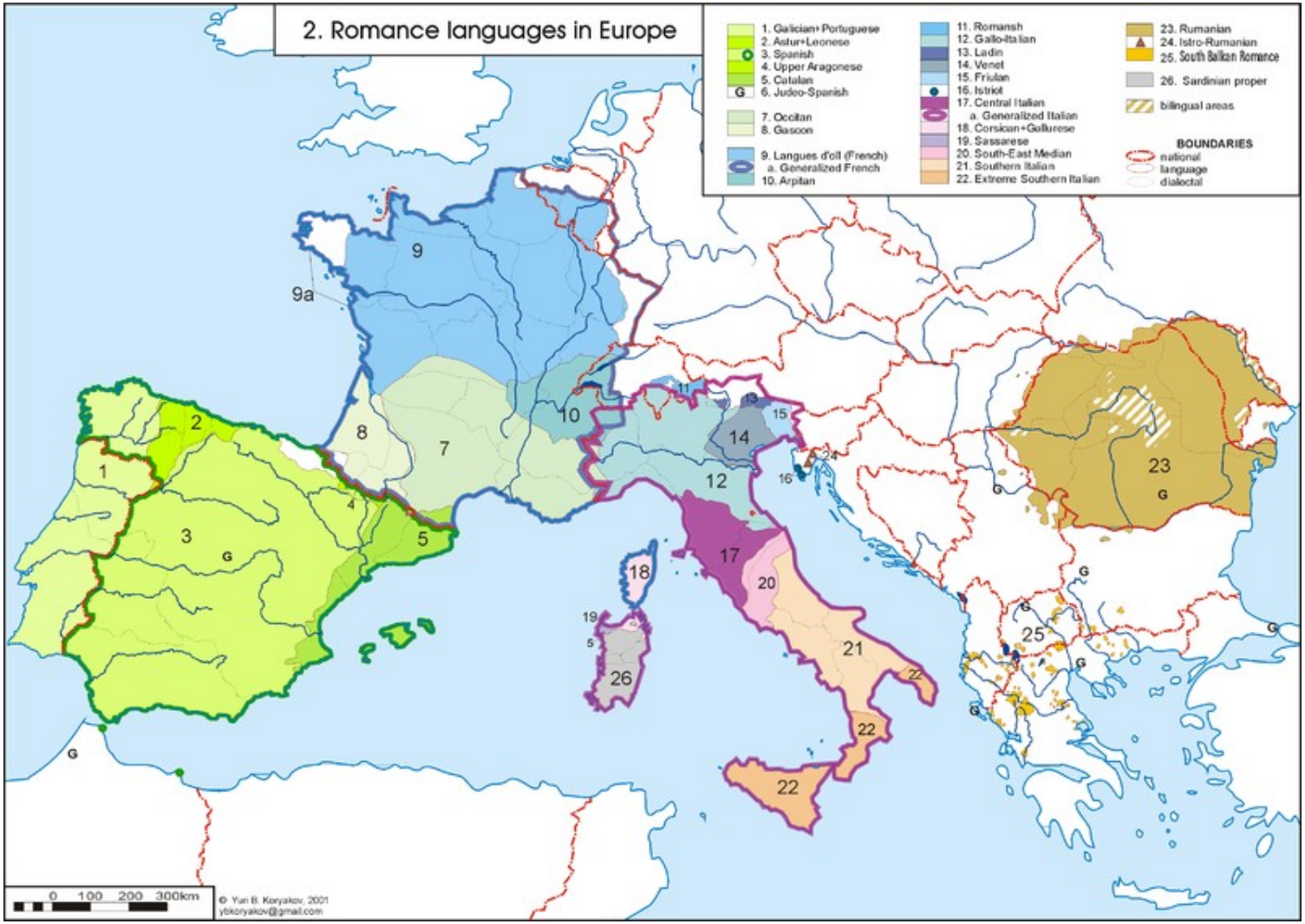


Trend for publishers of general linguistic works



Trend for English vs. French and German

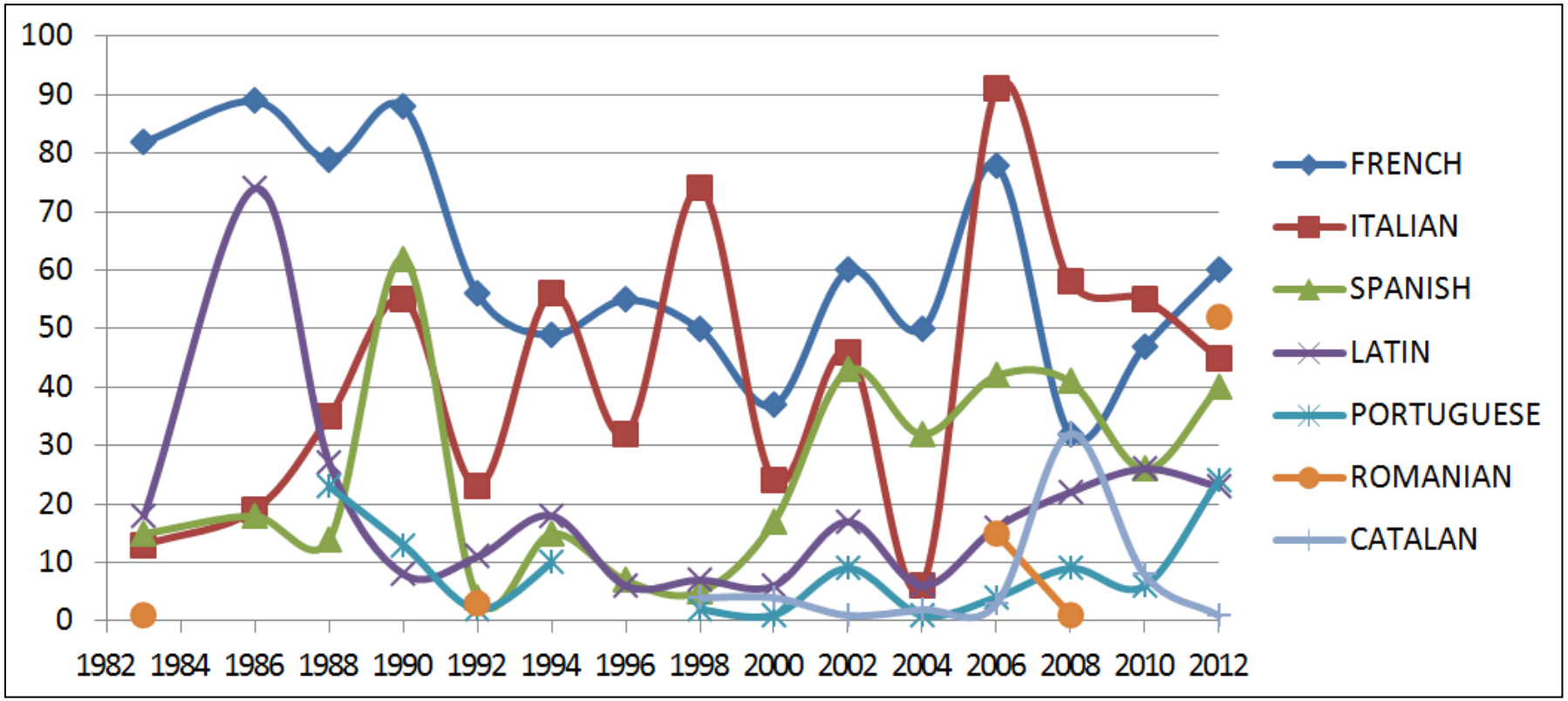


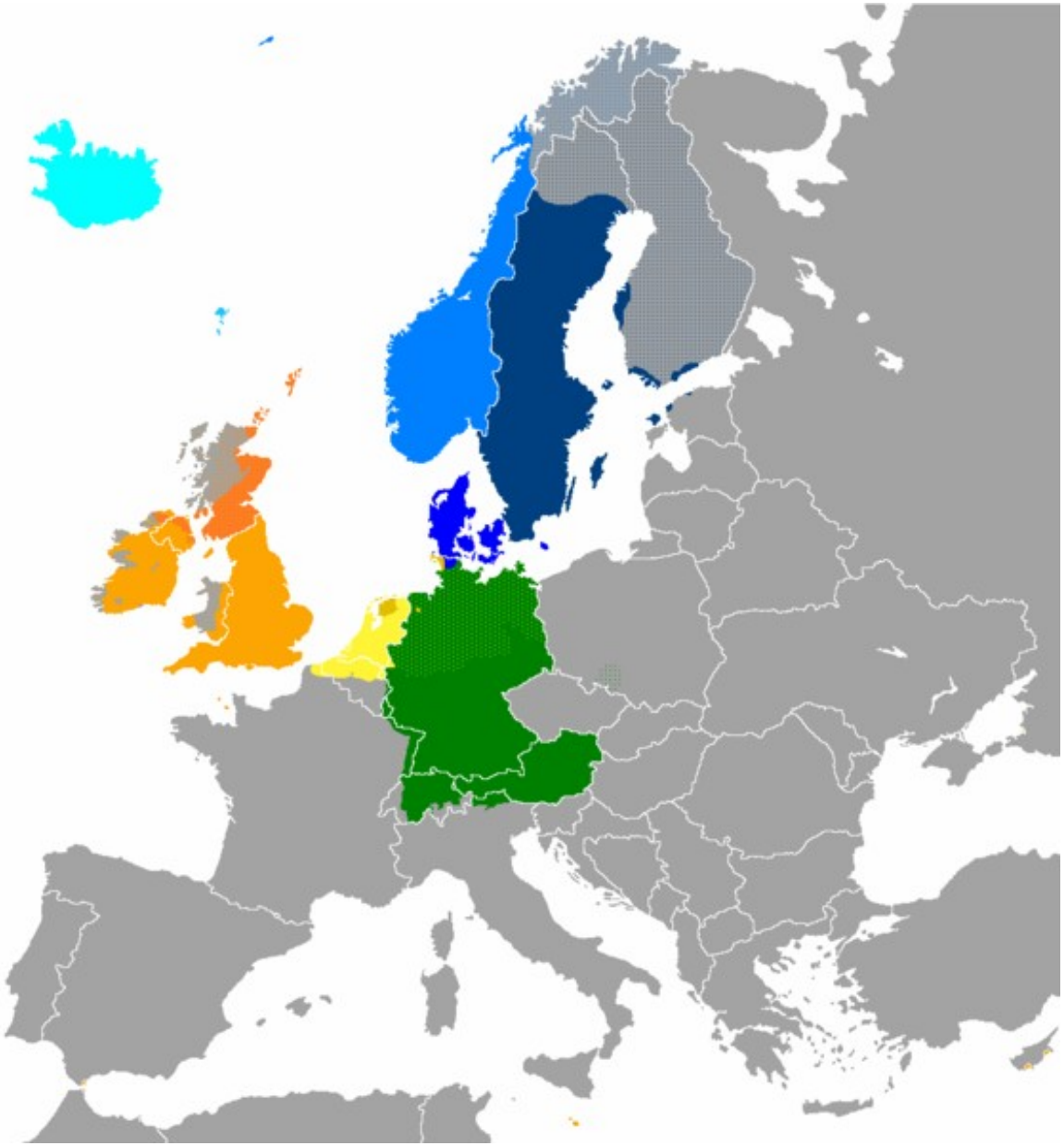


Romance languages

http://en.wikipedia.org/wiki/File:Romance_20c_en.png

Trend for Romance languages

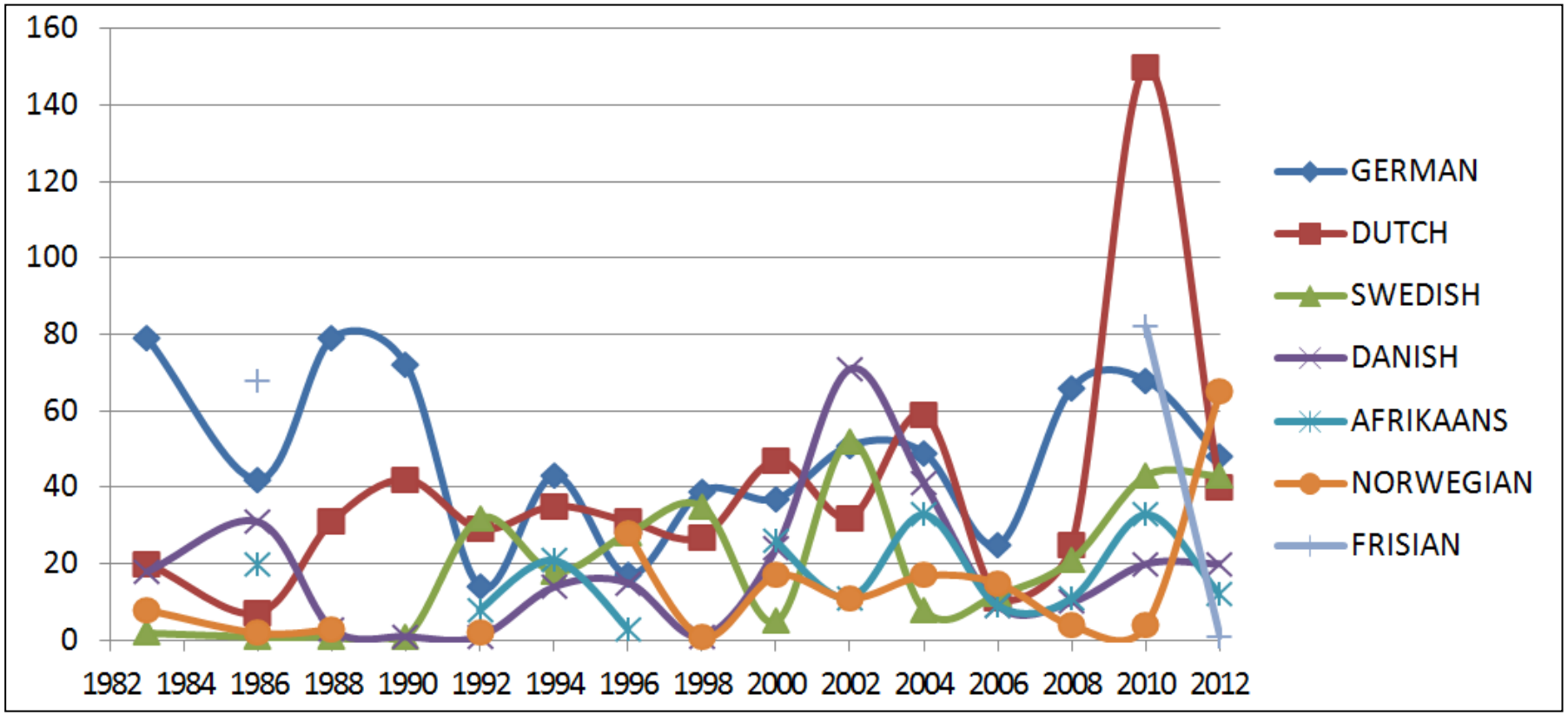




Germanic languages

http://en.wikipedia.org/wiki/File:Germanic_languages_in_Europe.png

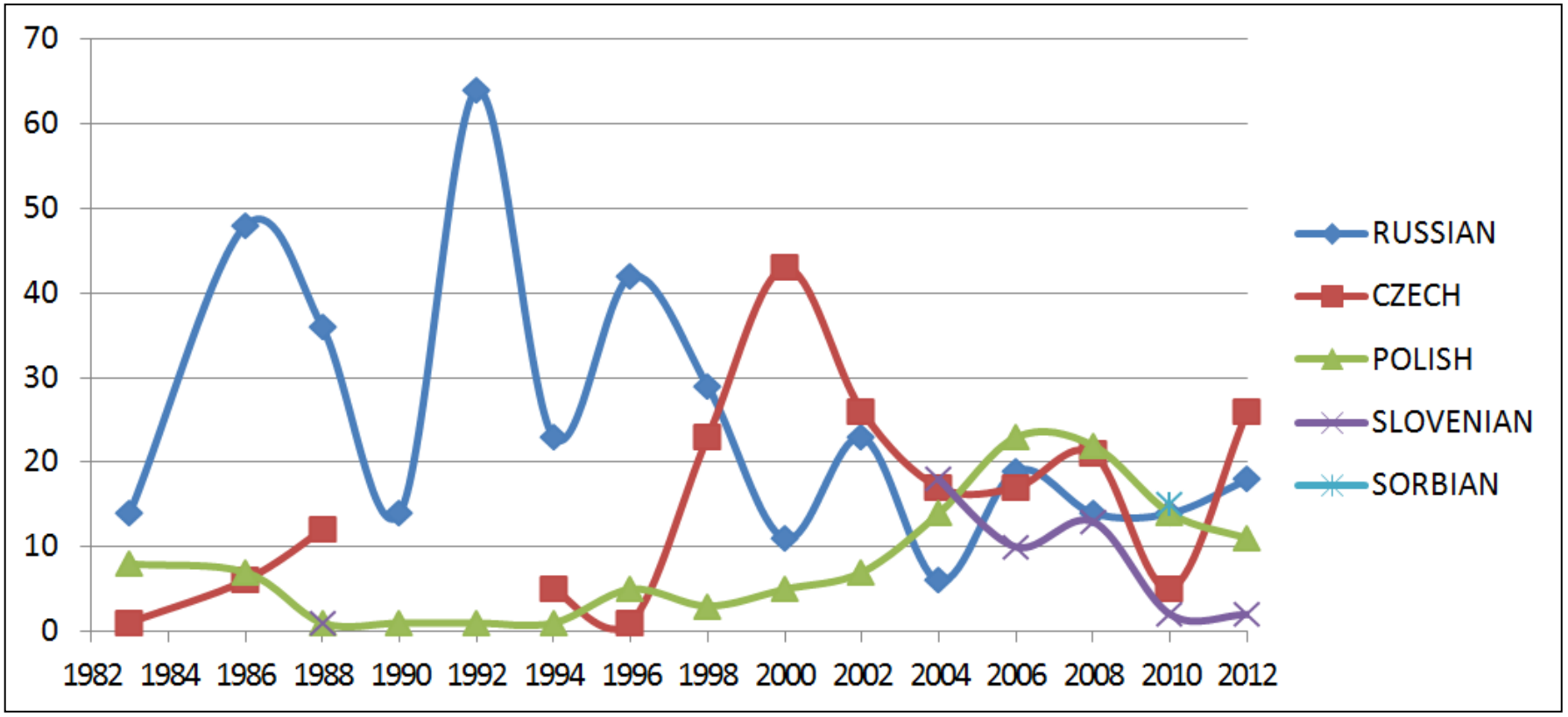
Trend for Germanic languages



Slavic languages

http://en.wikipedia.org/wiki/File:Slavic_languages.png

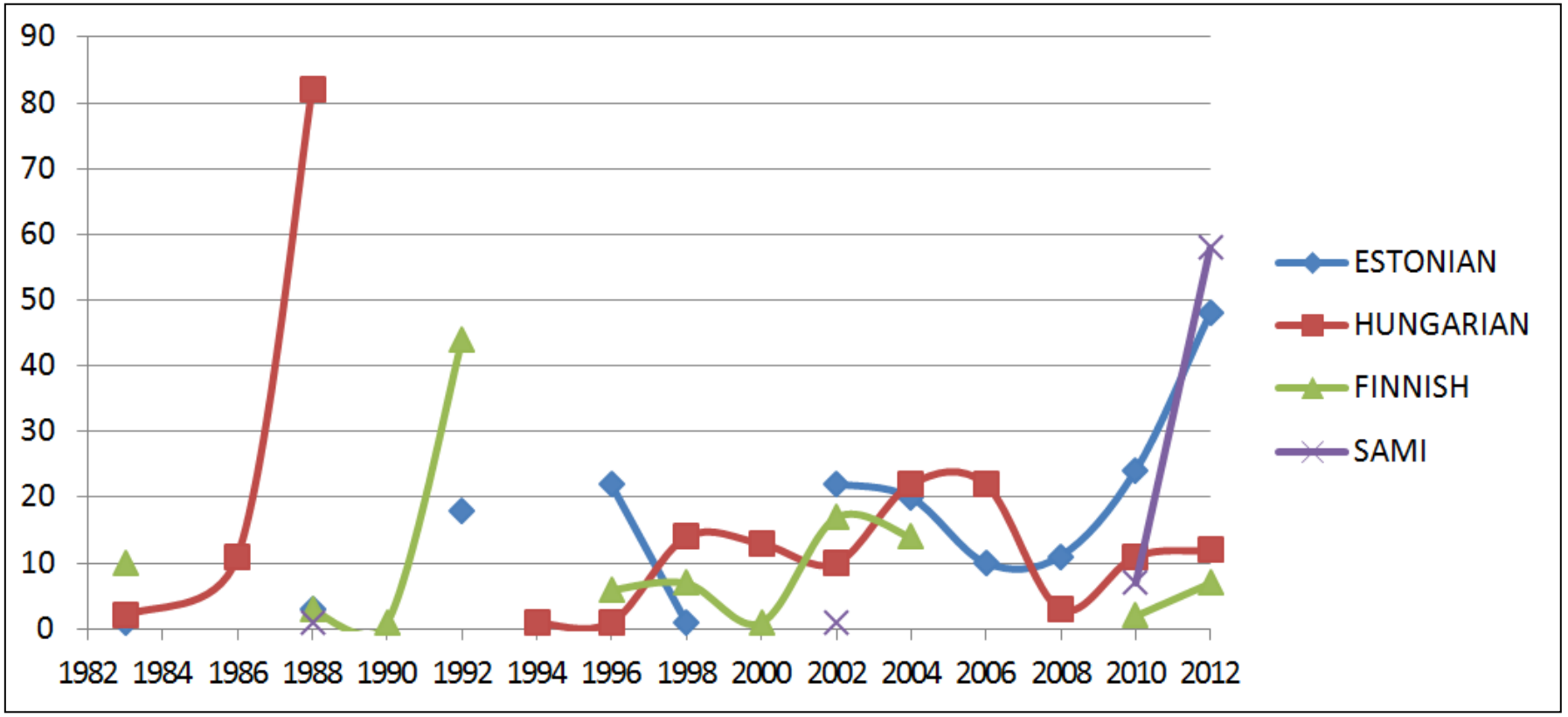
Trend for Slavic languages



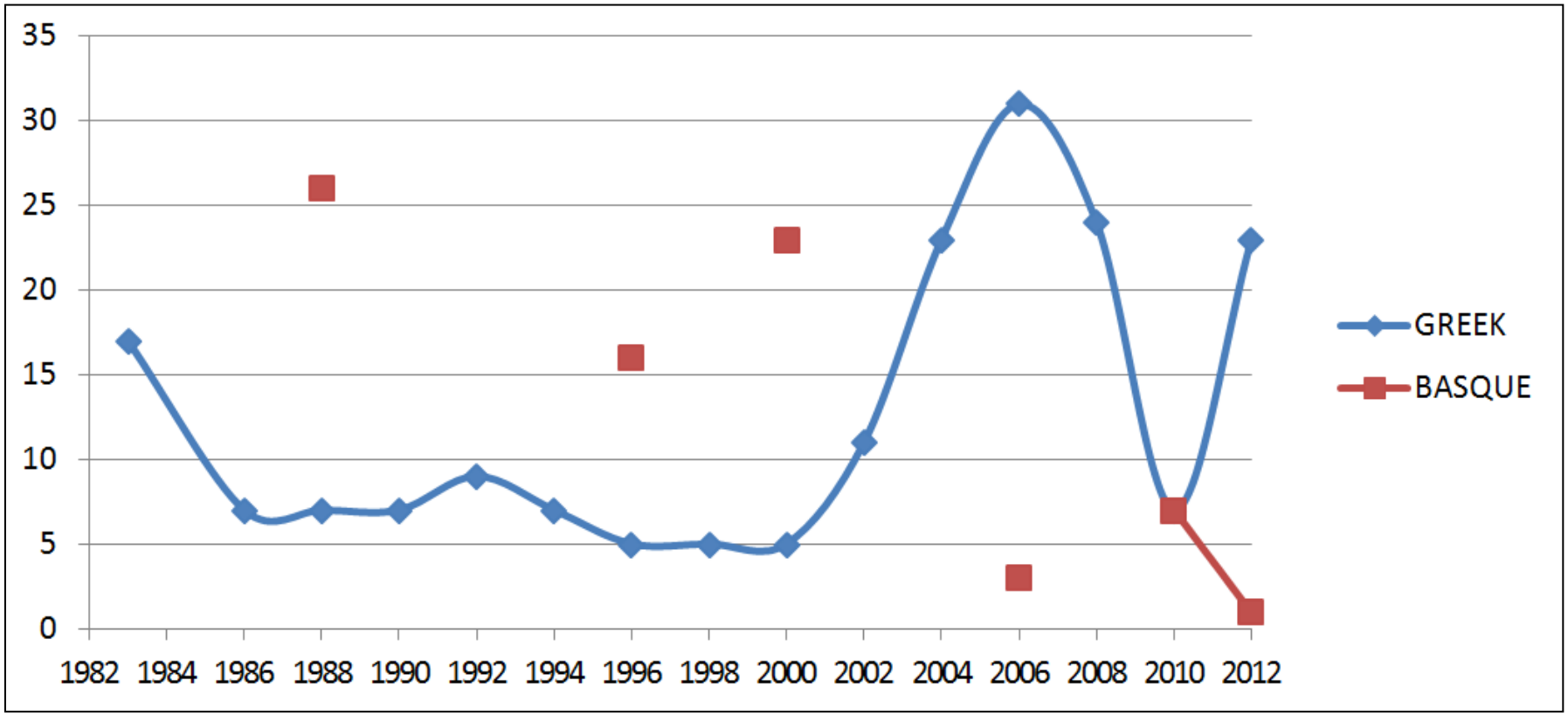
Uralic languages

http://en.wikipedia.org/wiki/File:Fenno-Ugrian_people.png

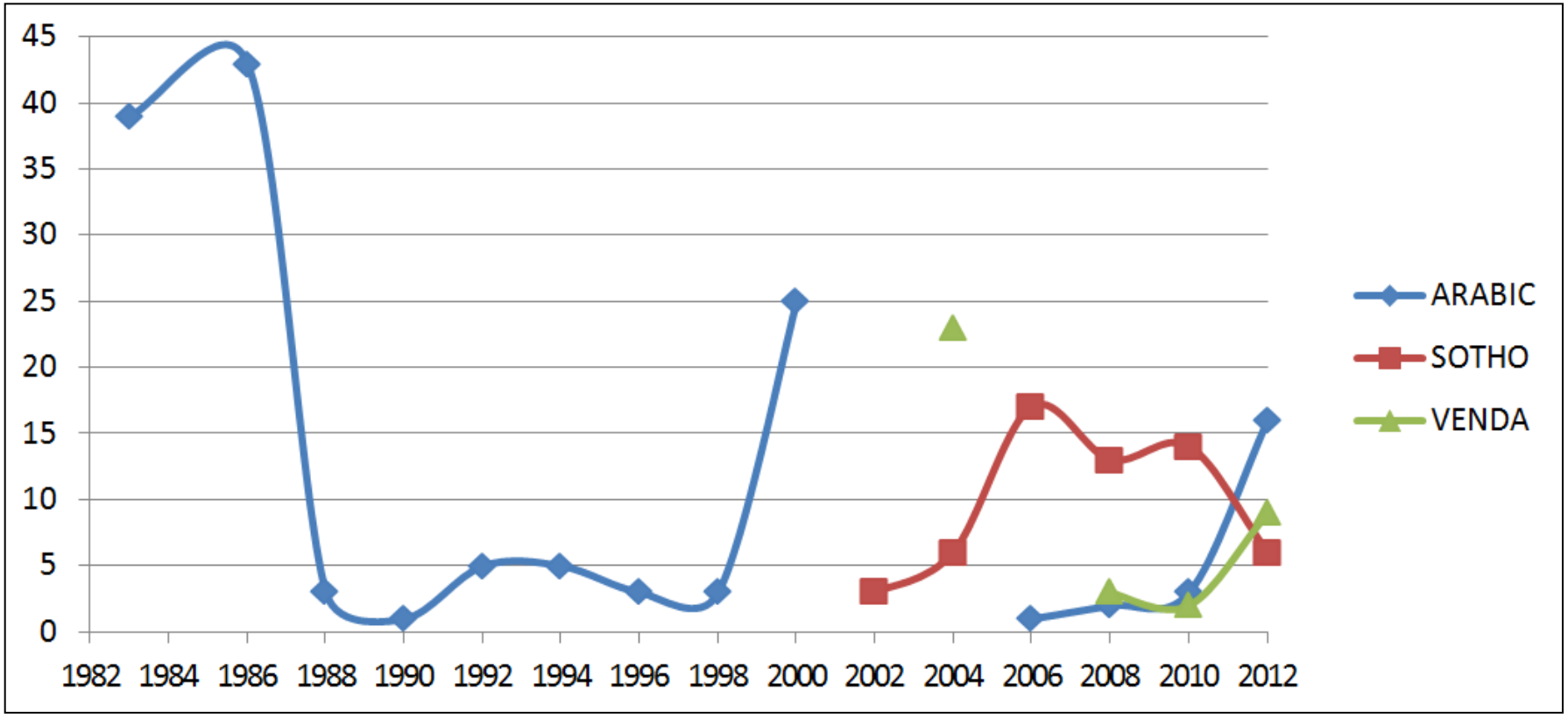
Trend for Uralic languages



Trend for Greek and Basque

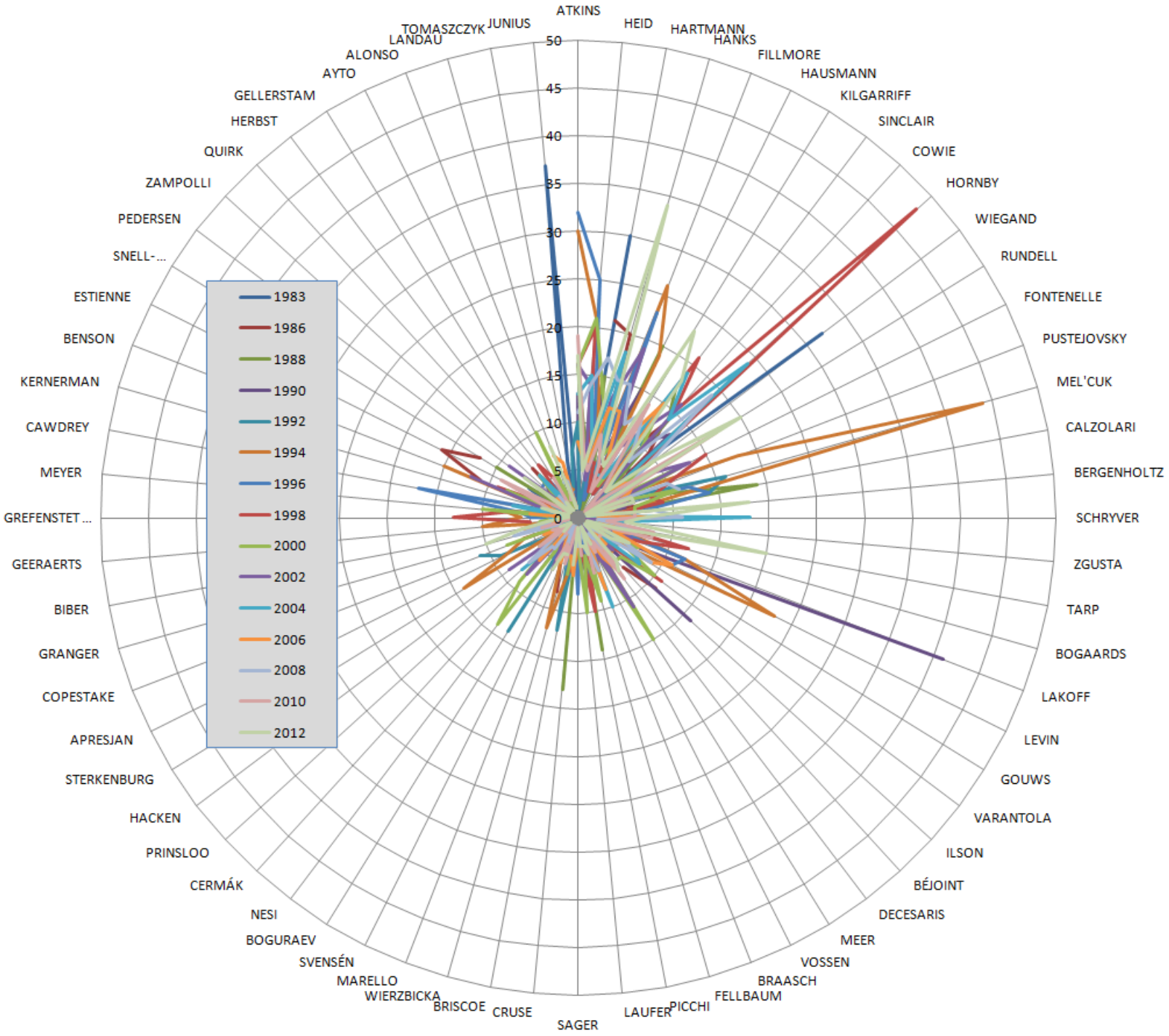


Trend for Arabic and two South African Bantu languages



Key scholars and their most significant congress impact(s)

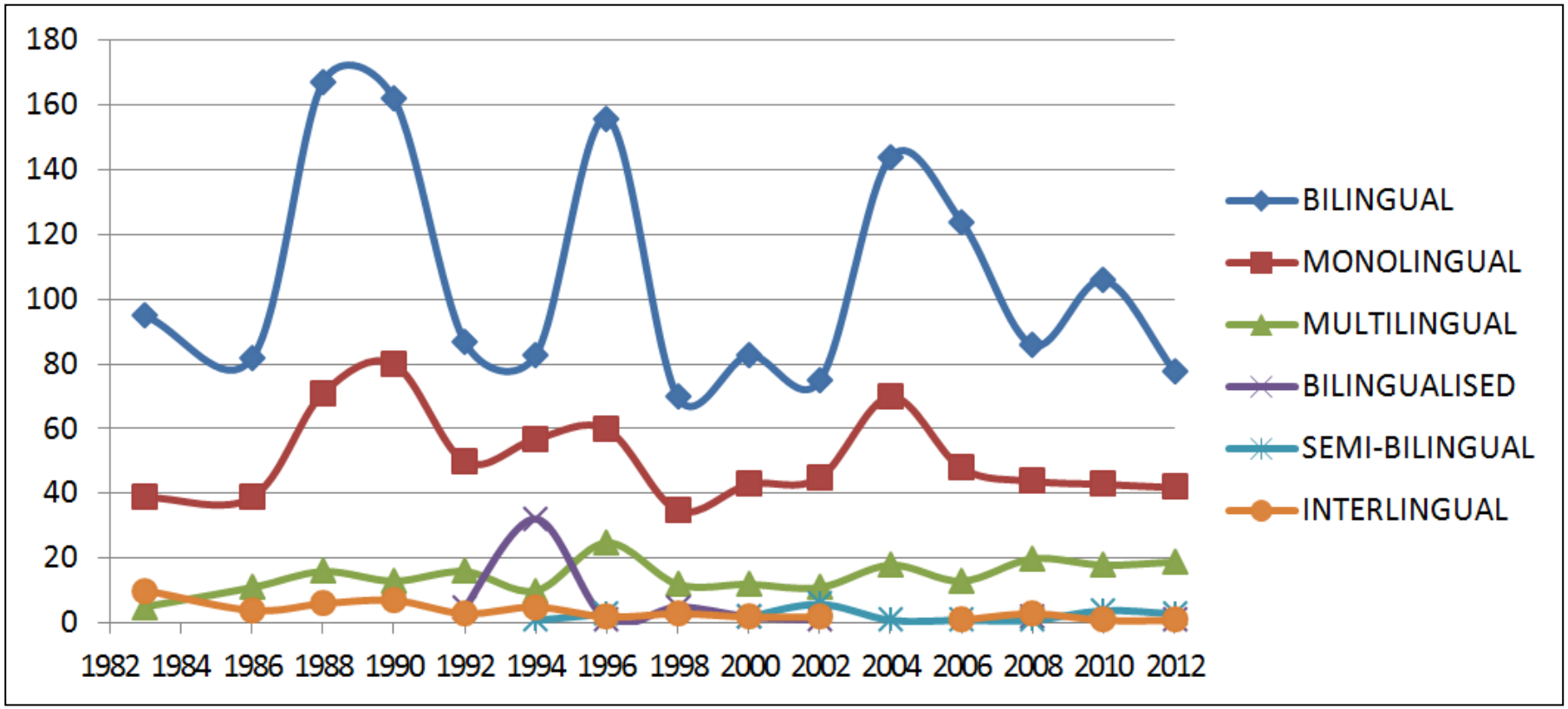
- A total of 68 family names are found in the top 1 000 keywords.
- May be visualised in a radar chart (see next).
  - in descending order, clockwise starting at 12 o'clock;
  - the impact of the various scholars on each congress is also shown.


ATKINS
HEID
HARTMANN
HANKS
FILLMORE
HAUSMANN
KILGARRIFF
SINCLAIR
COWIE
HORNBY
WIEGAND
RUNDELL
FONTENELLE
PUSTEJOVSKY
MEL'CUK
CALZOLARI
BERGENHOLTZ
SCHRYVER
ZGUSTA
TARP
BOGAARDS
LAKOFF
LEVIN
GOUWS
VARANTOLA
ILSON
BÉJOINT
DECESARIS
MEER
VOSSEN
BRAASCH
FELLBAUM
PICCHI
LAUFER
SAGER
CRUSE
BRISCOE
WIERZBICKA
MARELLO
SVENSÉN
BOGURAEV
NESI
CERMÁK
PRINSLOO
HACKEN
STERKENBURG
APRESJAN
COPESTAKE
GRANGER
BIBER
GEERAERTS
GREFENSTET…
MEYER
CAWDREY
KERNERMAN
BENSON
ESTIENNE
SNELL-…
PEDERSEN
ZAMPOLLI
QUIRK
HERBST
GELLERSTAM
AYTO
ALONSO
LANDAU
TOMASZCZYK
JUNIUS
50
45
40
35
30
25
20
15
10
5
0
1983
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012

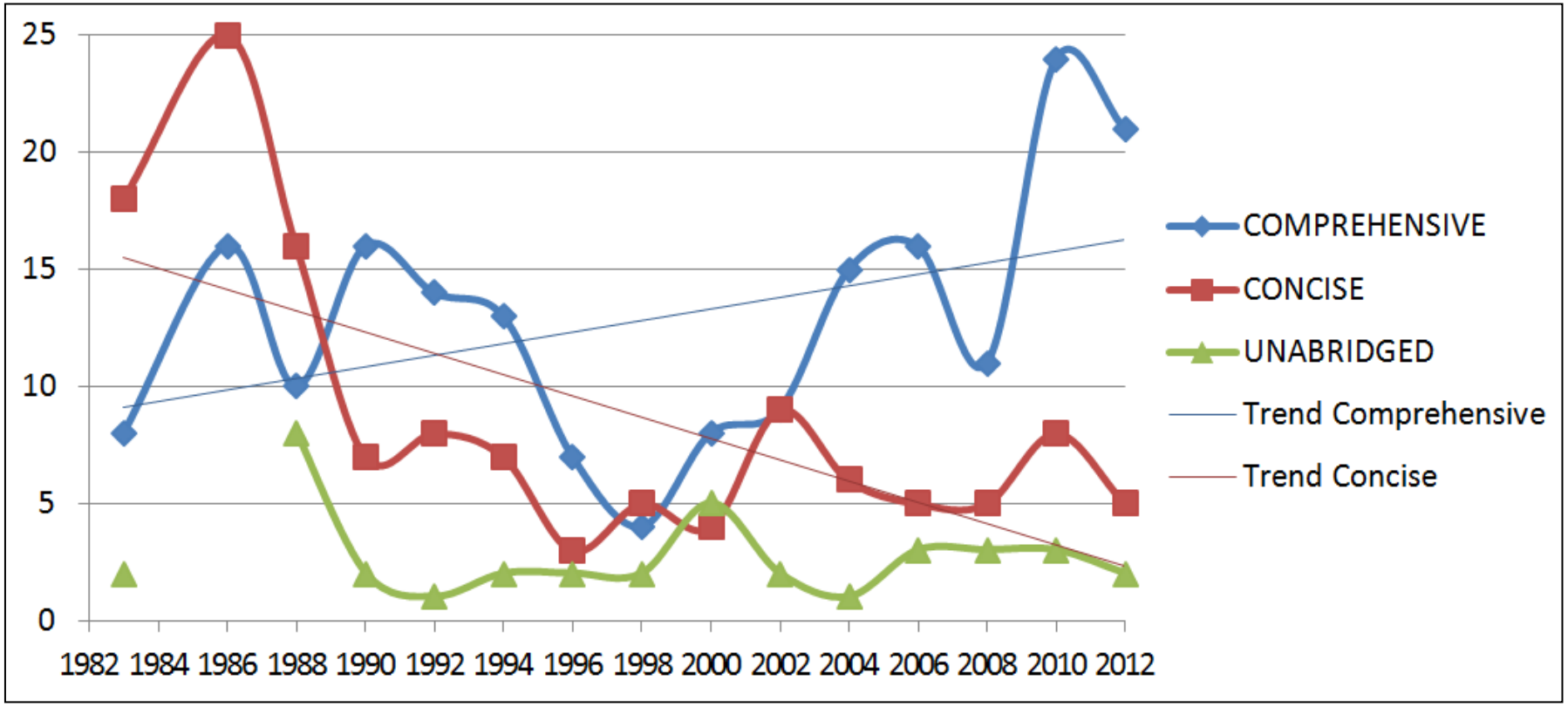
Discussion …

- Note that the great majority of the scholars that stand out in the corpus are:
  - (a) lexicographers (as compared to linguists at large);
  - (b) alive (as compared to the great lexicographers of the past); and
  - (c) very active at the EURALEX congresses themselves.

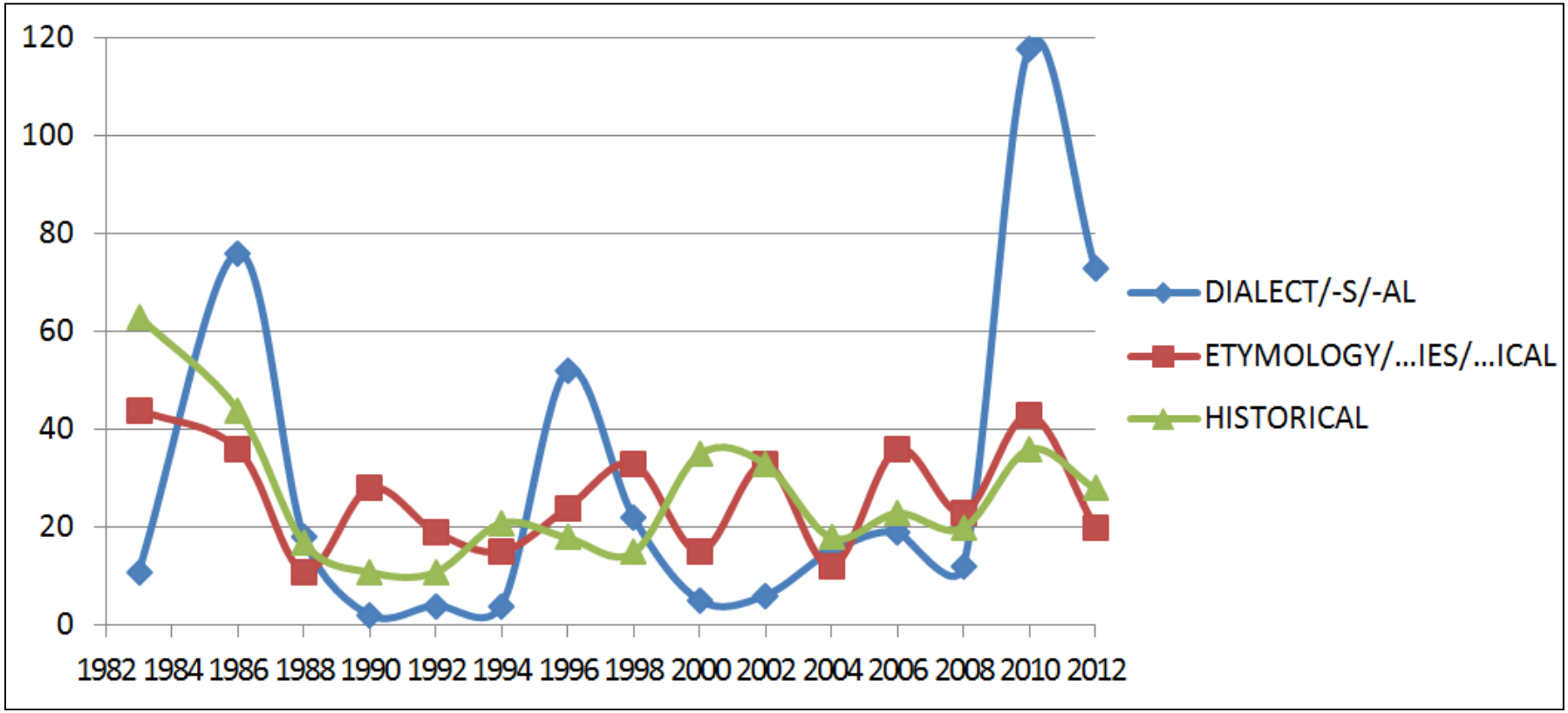
Trend for number of dictionary languages



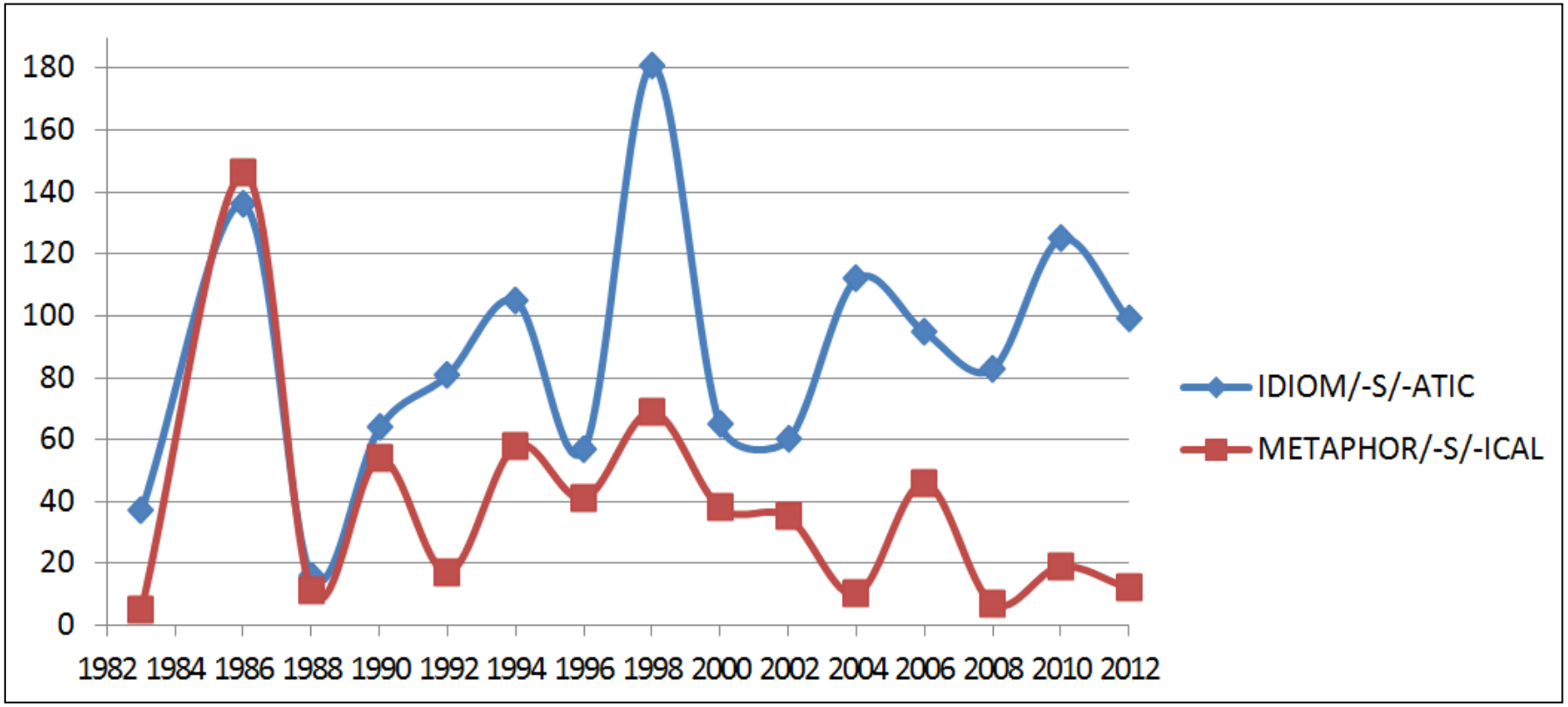
Trend for Comprehensive (and Unabridged) vs. Concise



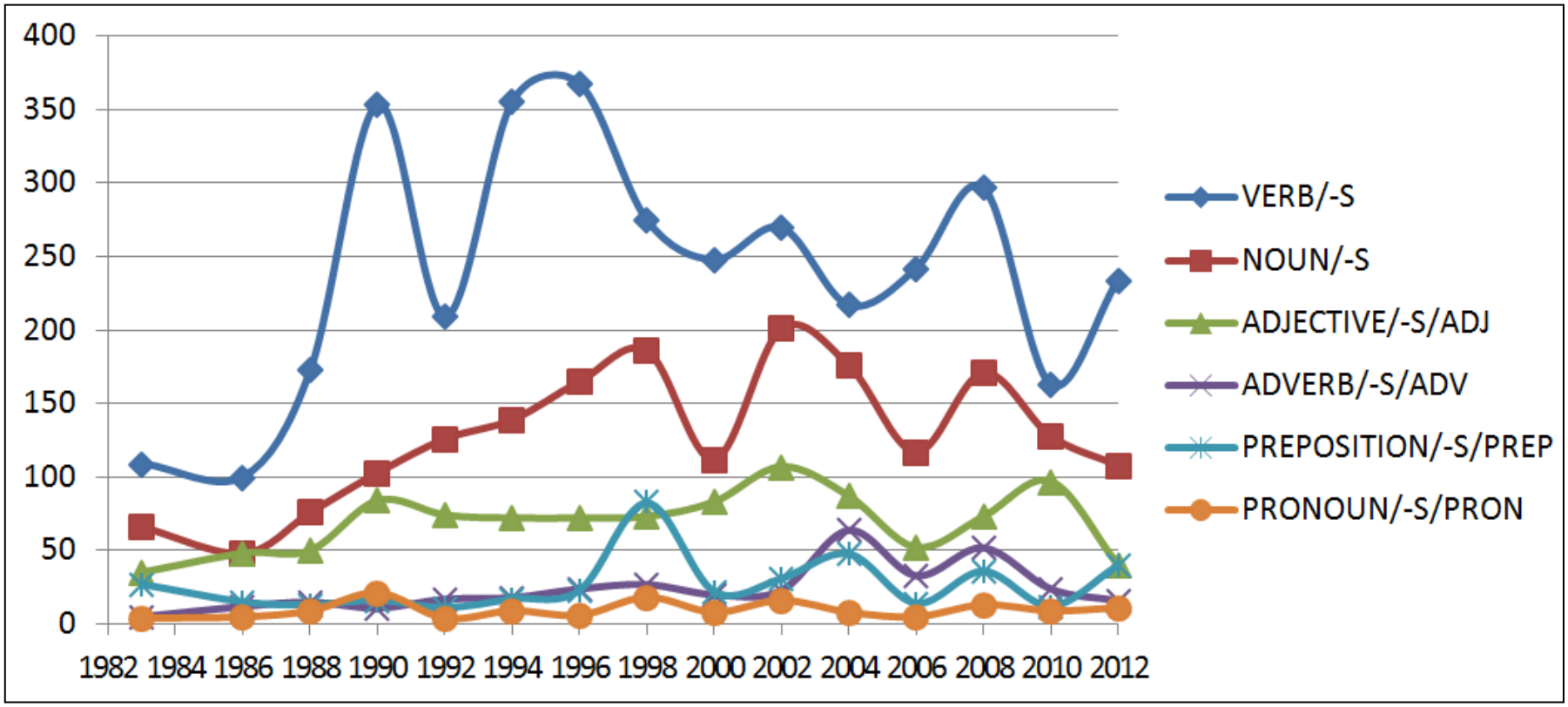
Trend for Dialect, Etymology, and Historical



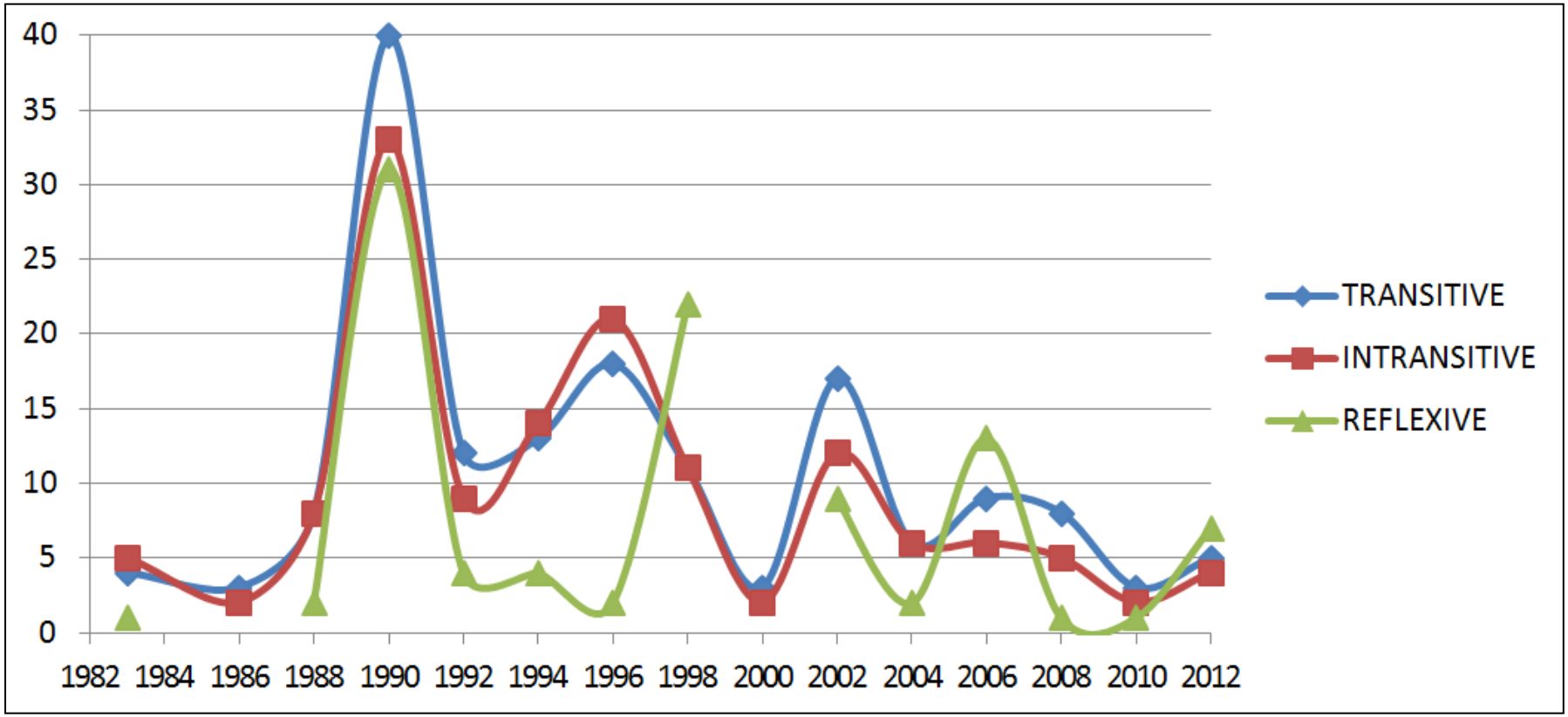
Trend for Idioms vs. Metaphors



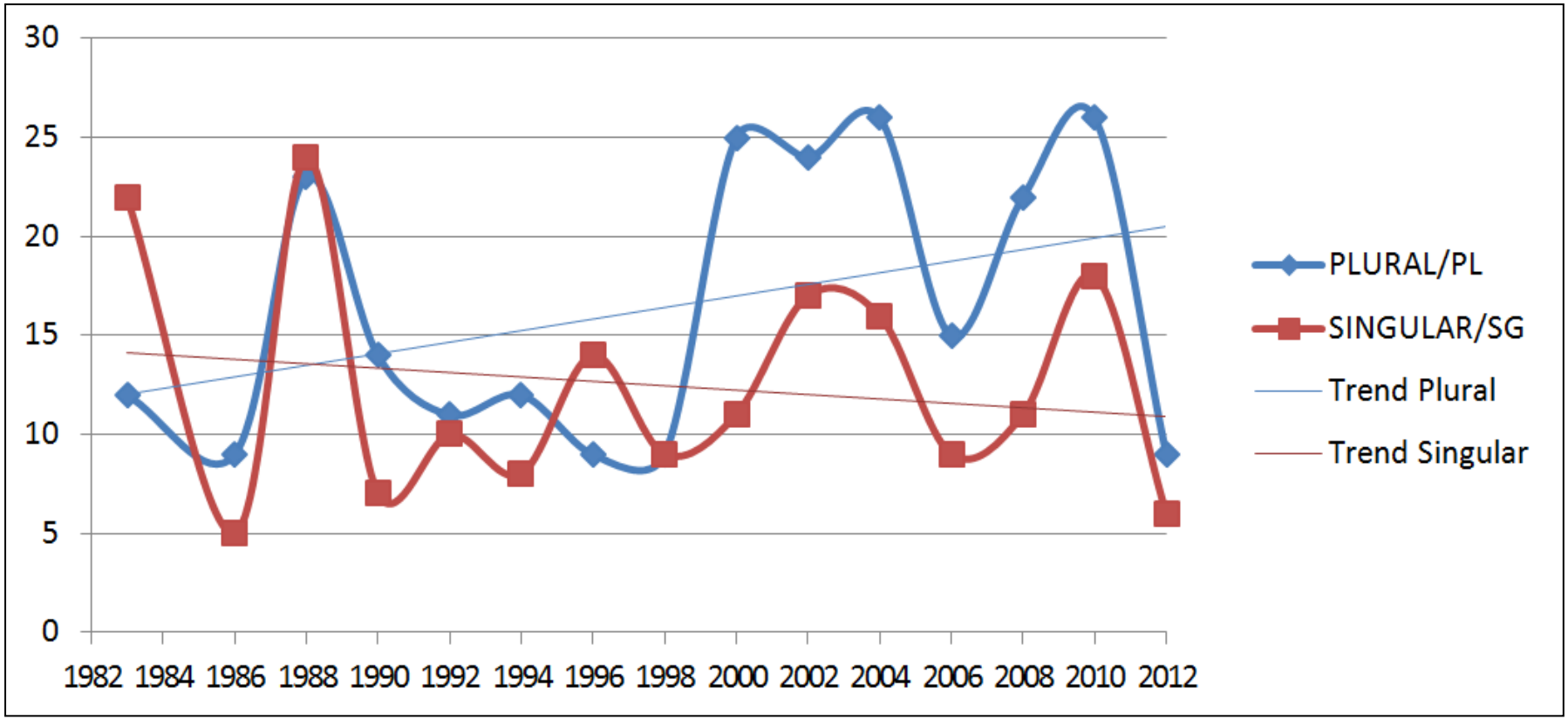
Trend for main parts of speech



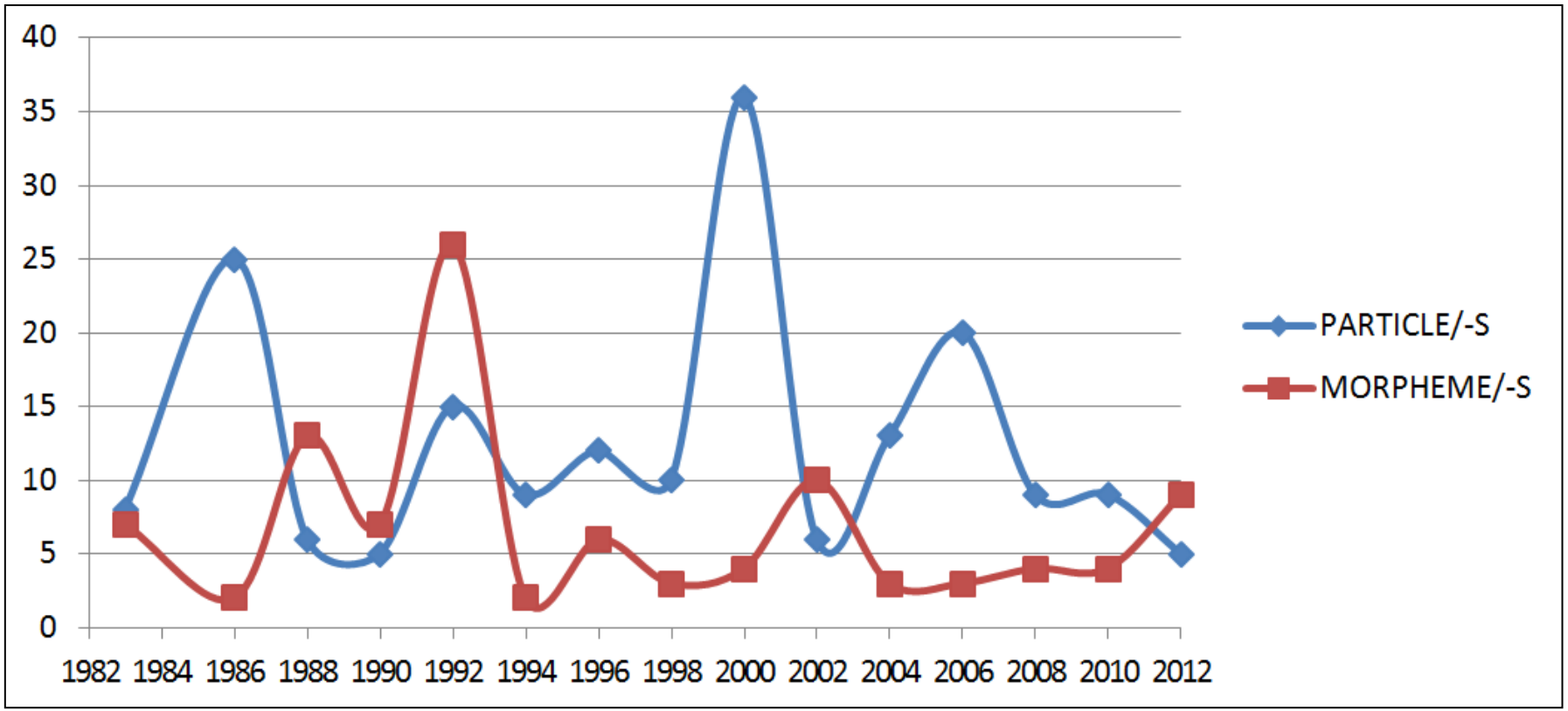
Trend for Transitive, Intransitive, and Reflexive



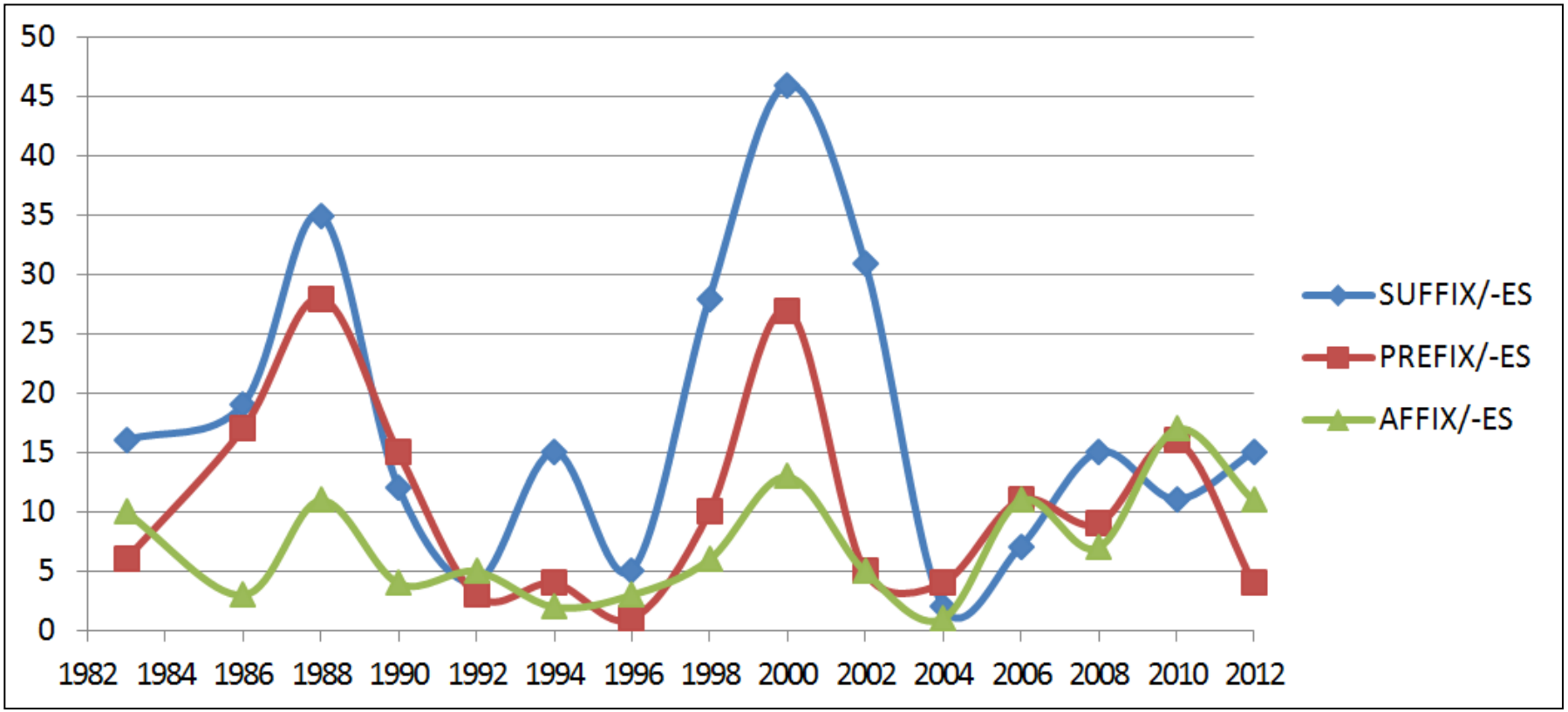
Trend for Plural vs. Singular



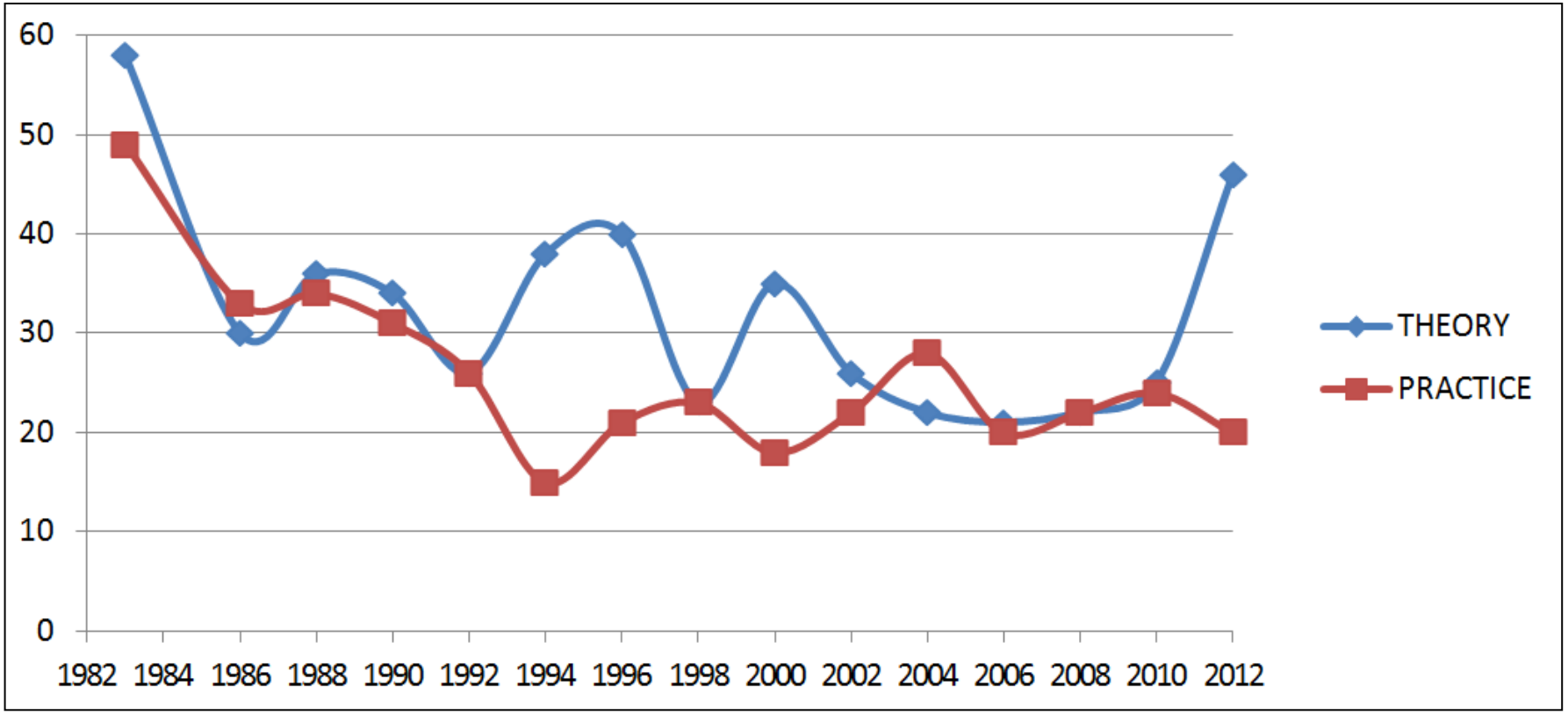
Trend for Particle and Morpheme



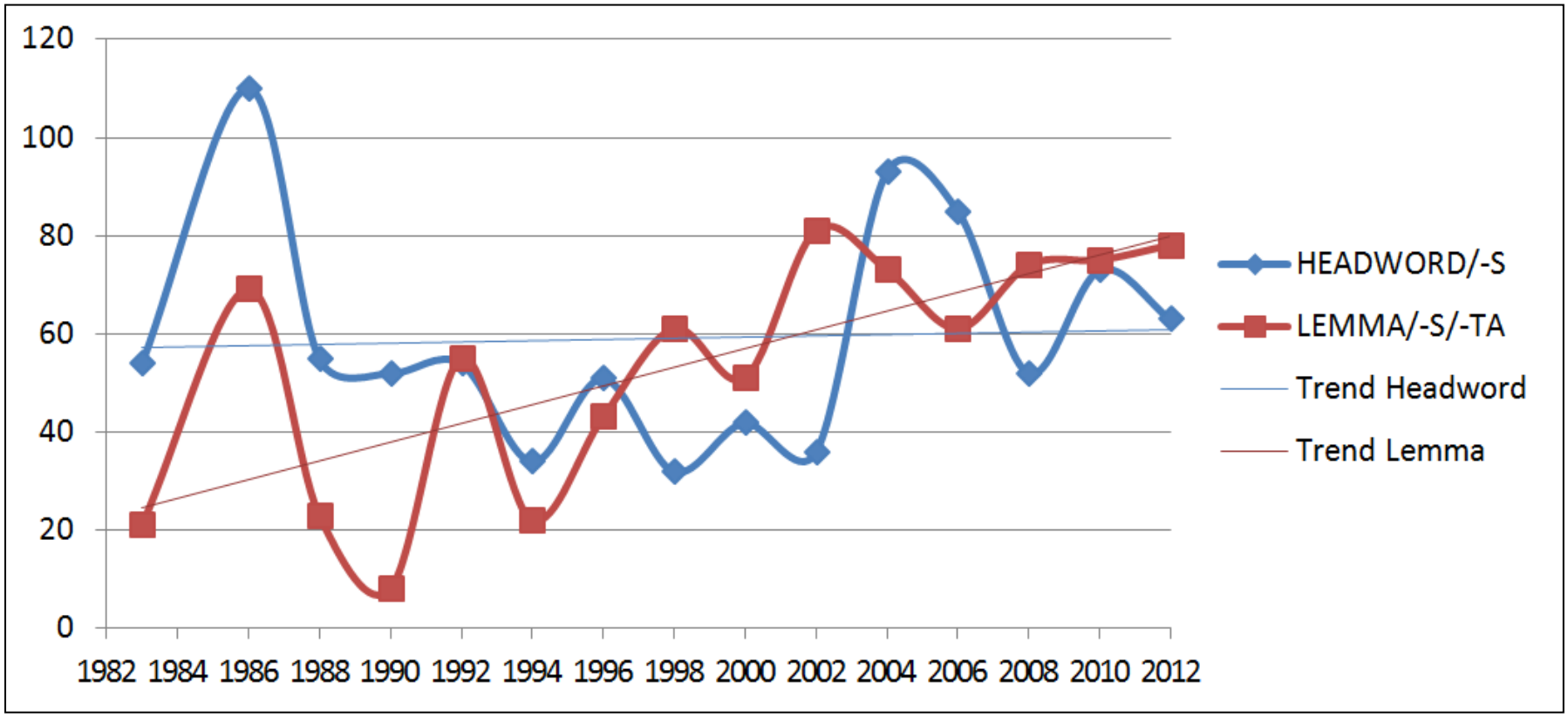
Trend for Suffix, Prefix and Affix



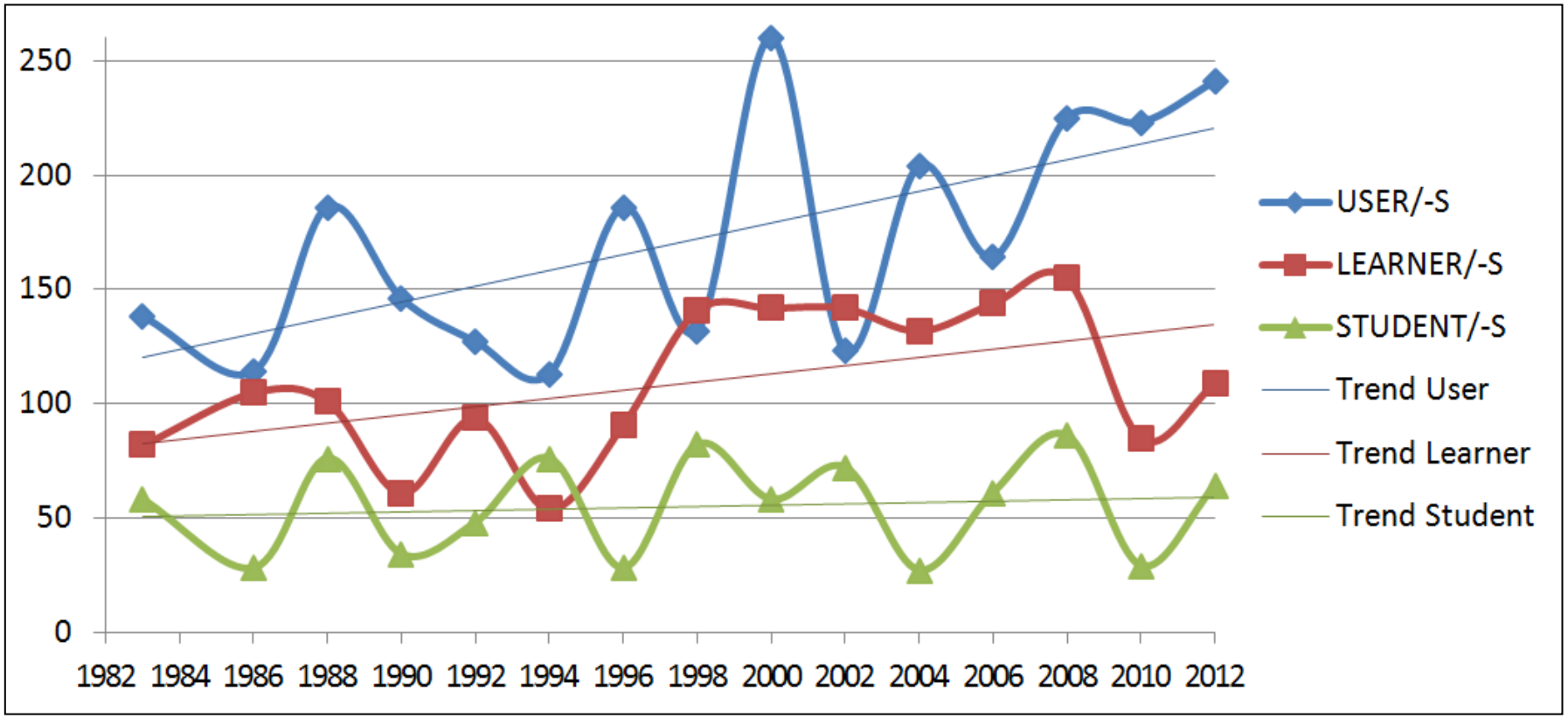
Trend for Theory and Practice



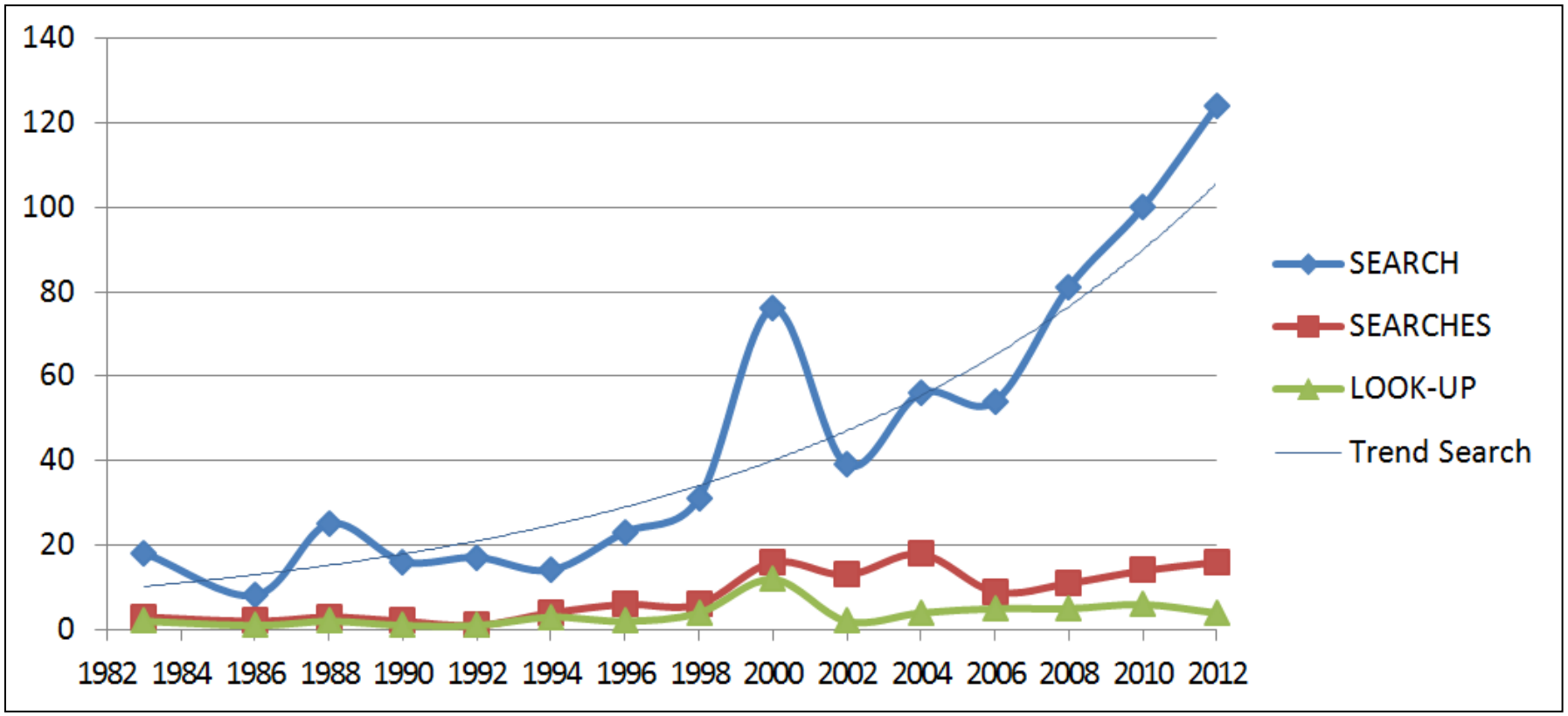
Trend for Headword vs. Lemma



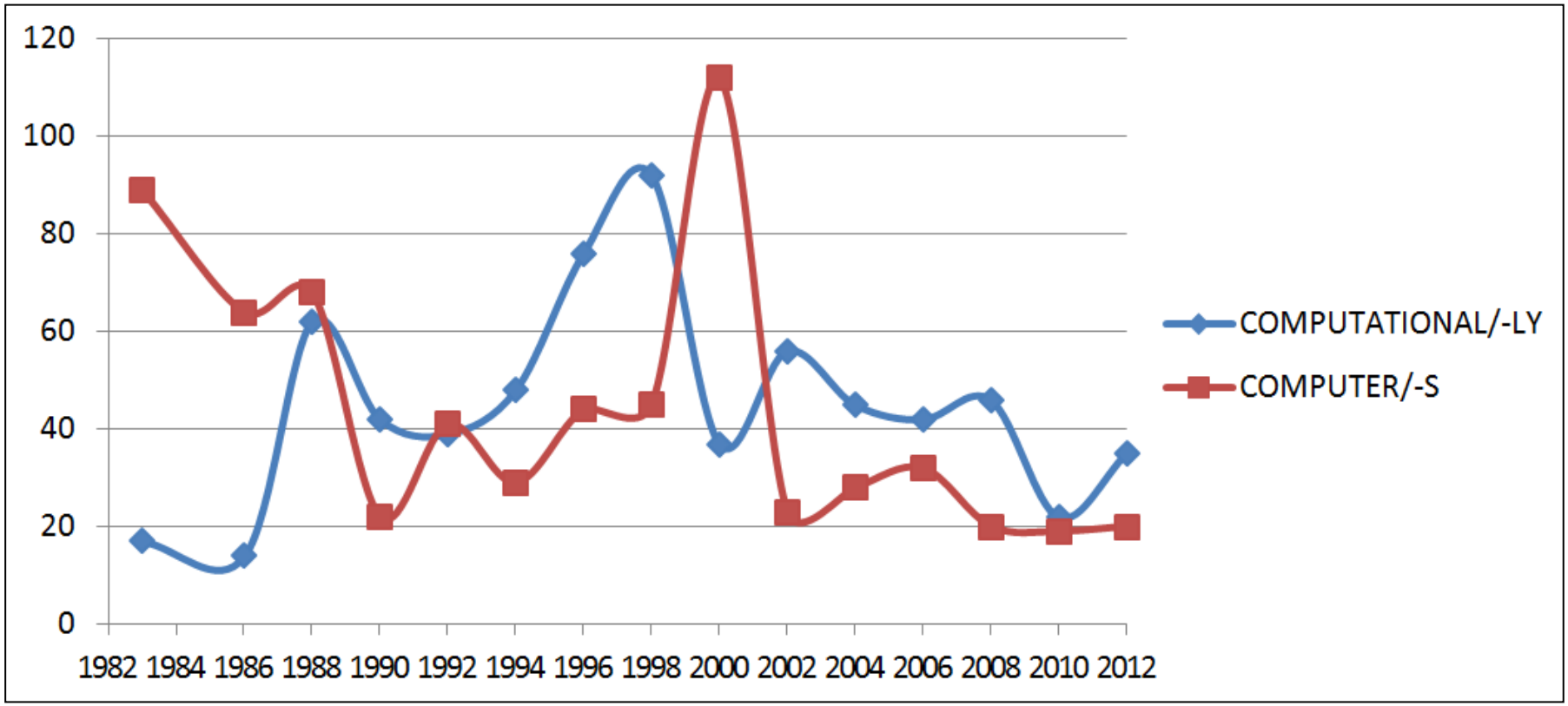
Trend for User, Learner and Student



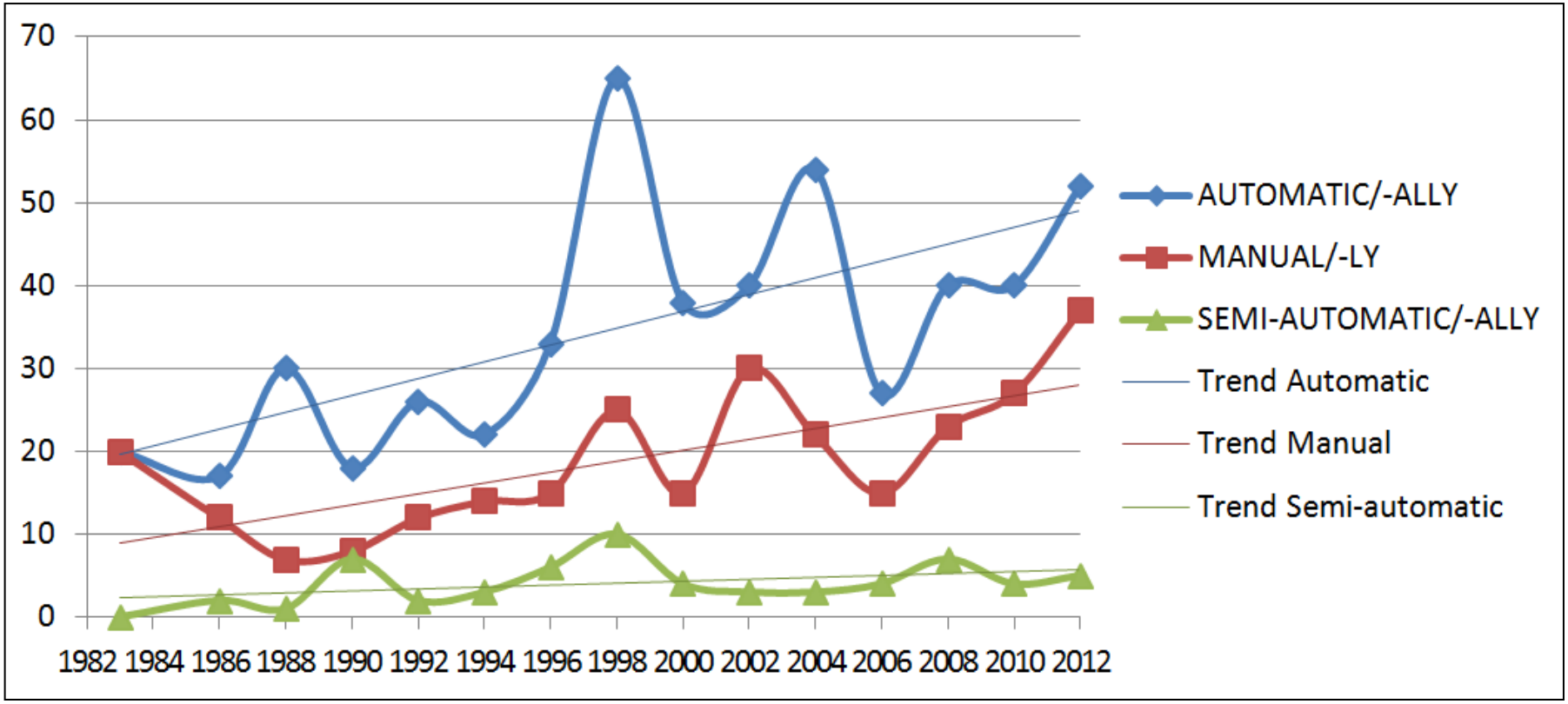
Trend for Search and Searches vs. Look-up



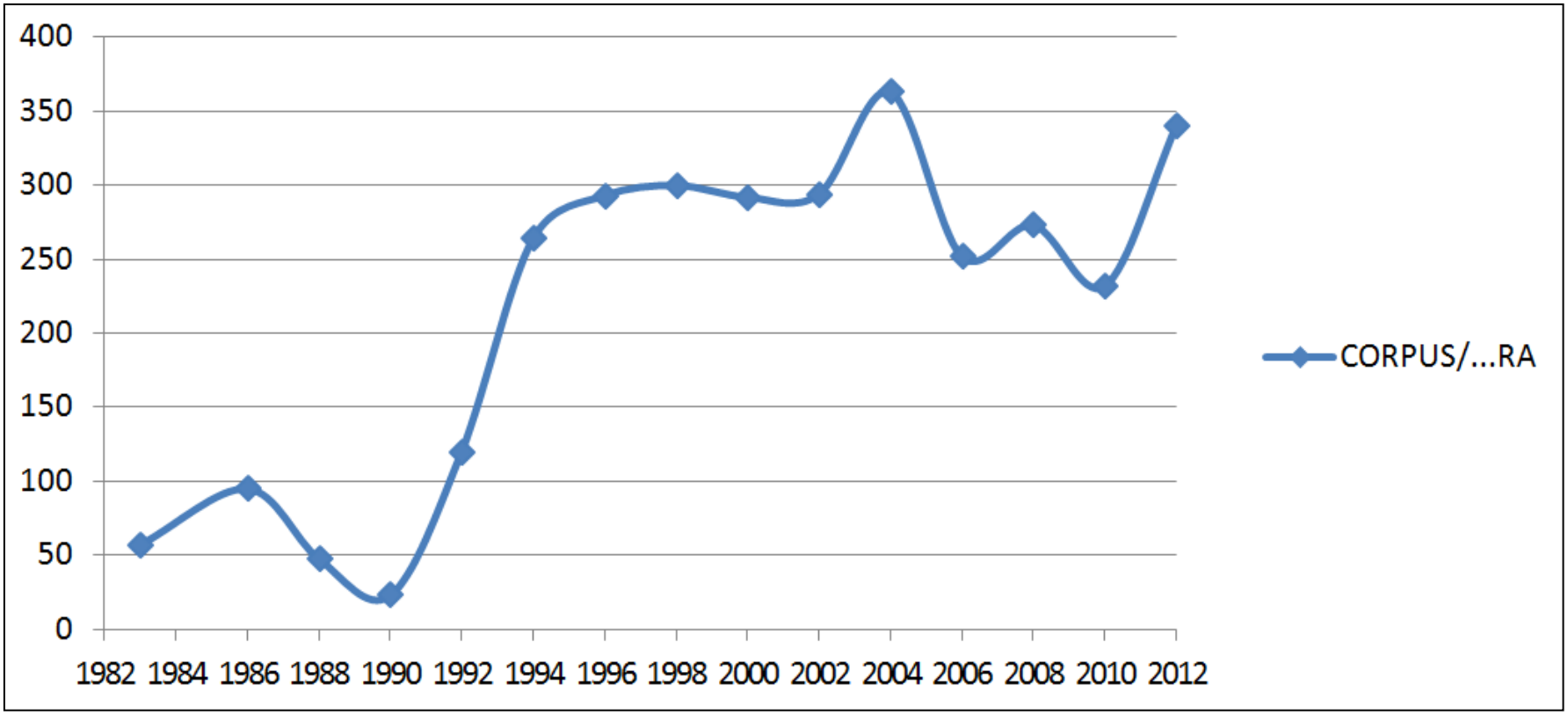
Trend for Computational and Computer



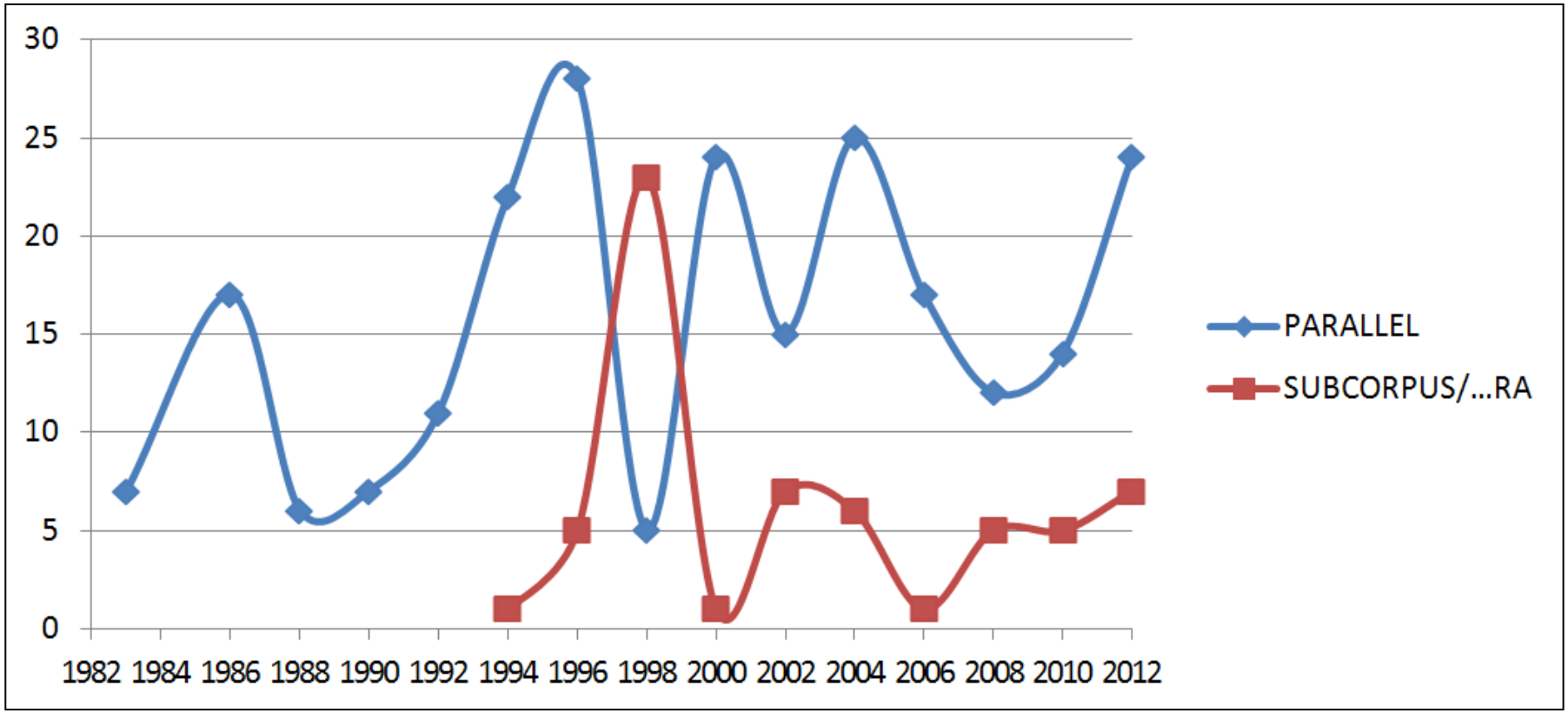
Trend for Automatic, Manual and Semi-automatic



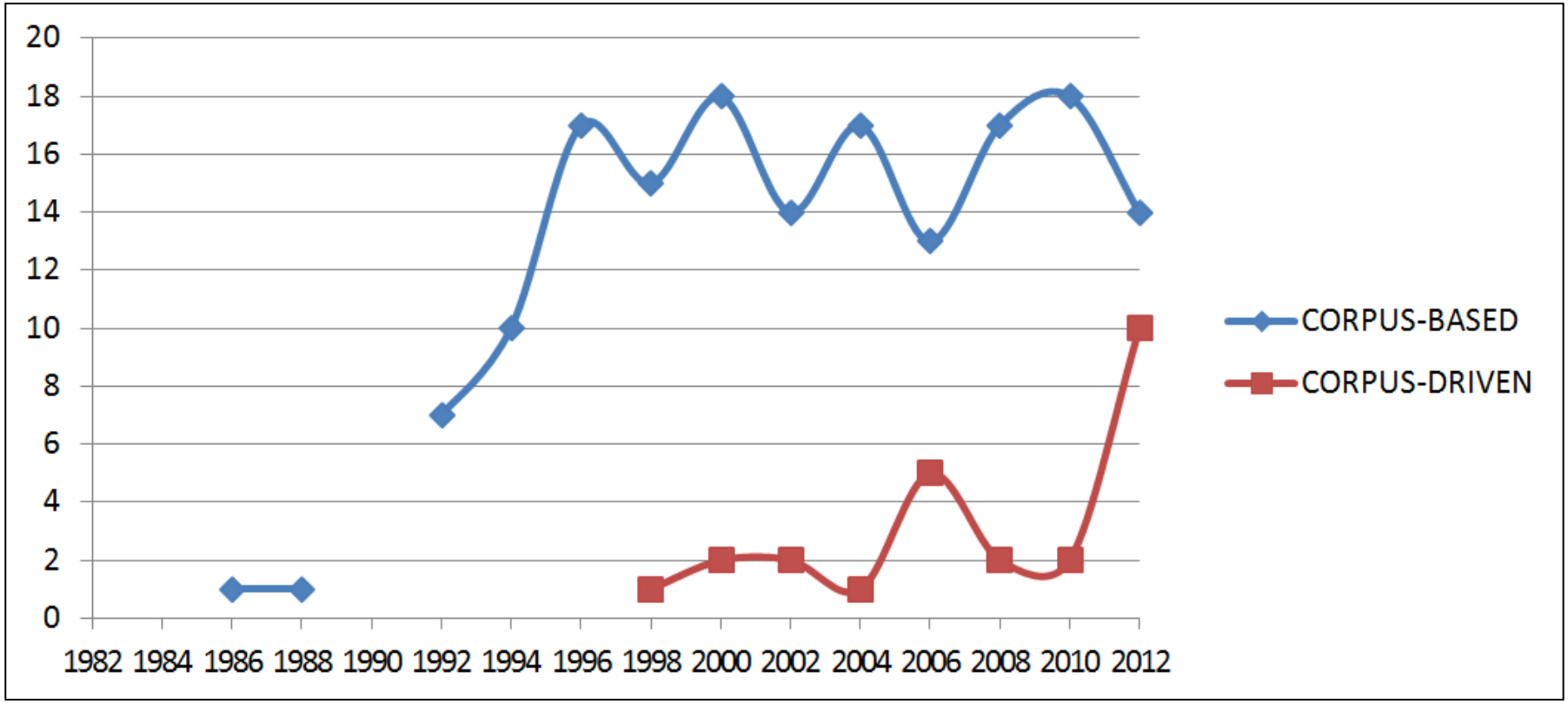
Trend for Corpus



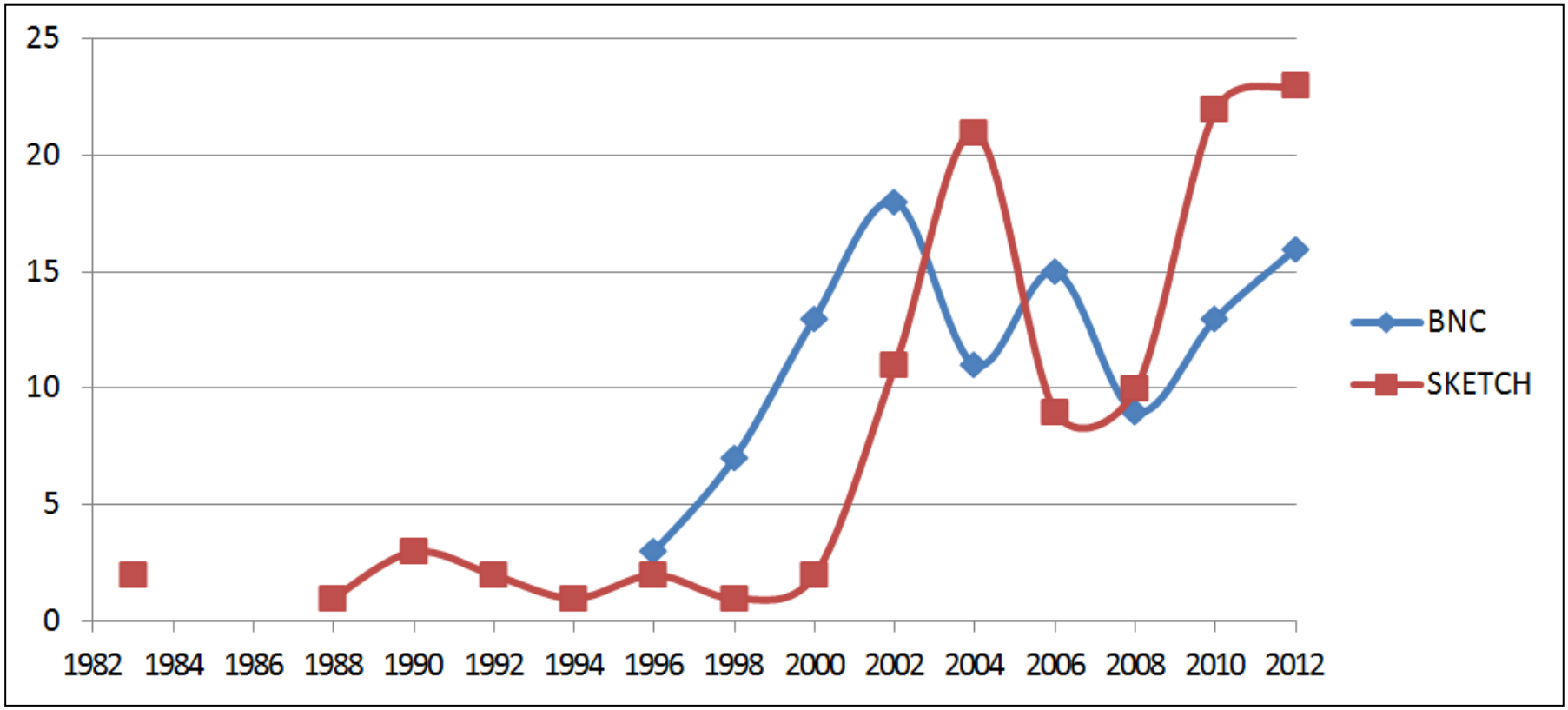
Trend for Parallel and Subcorpus



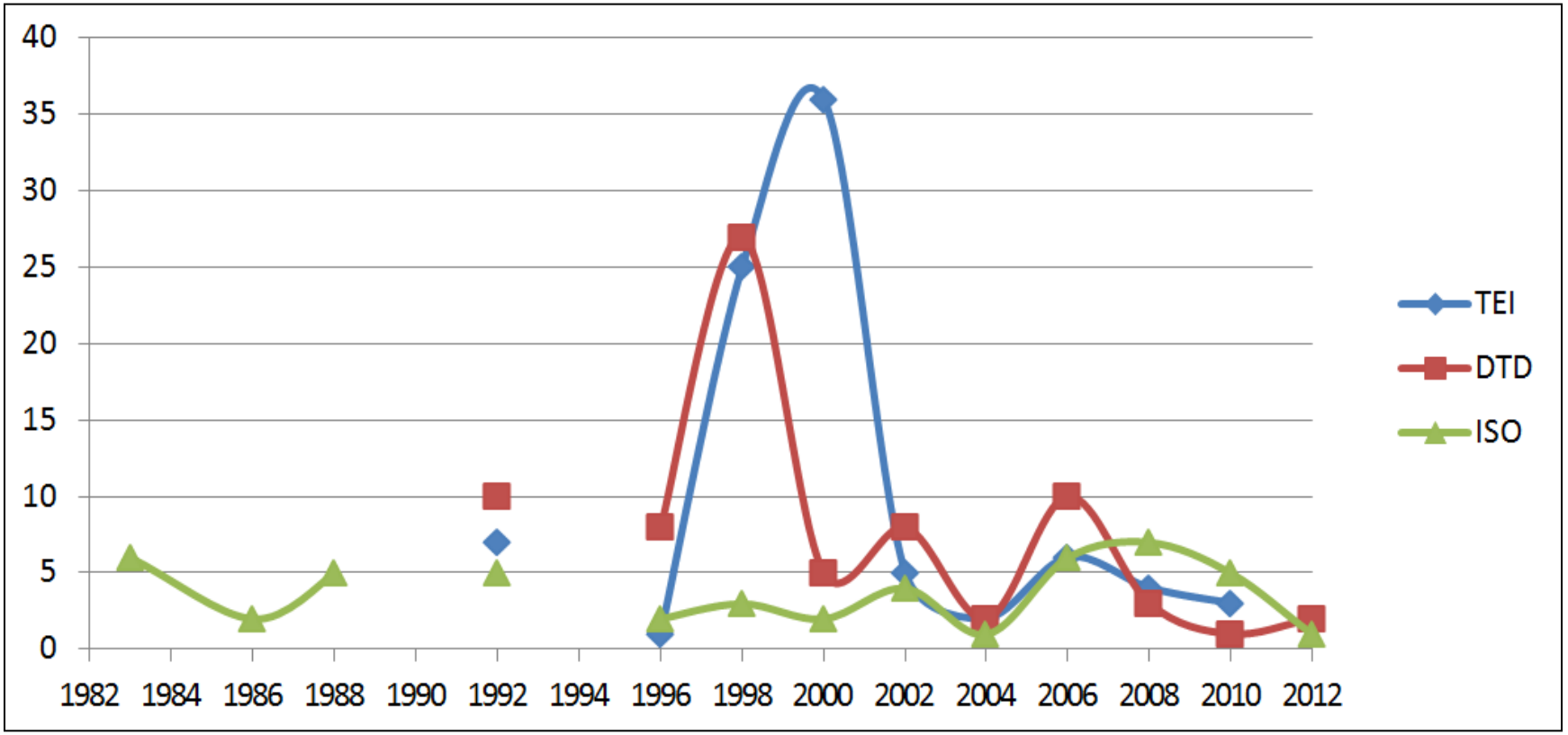
Trend for Corpus-based vs. Corpus-driven



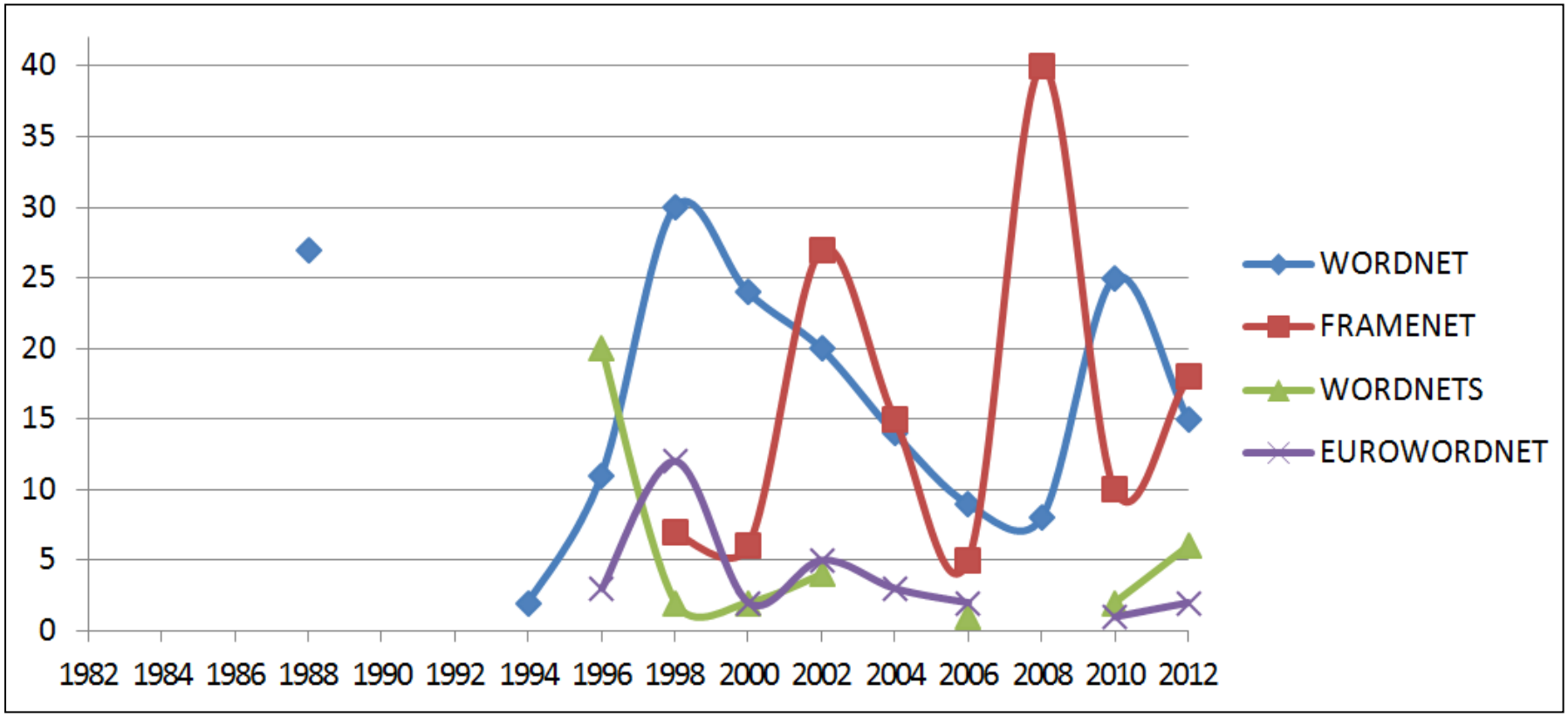
Trend for BNC and Sketch



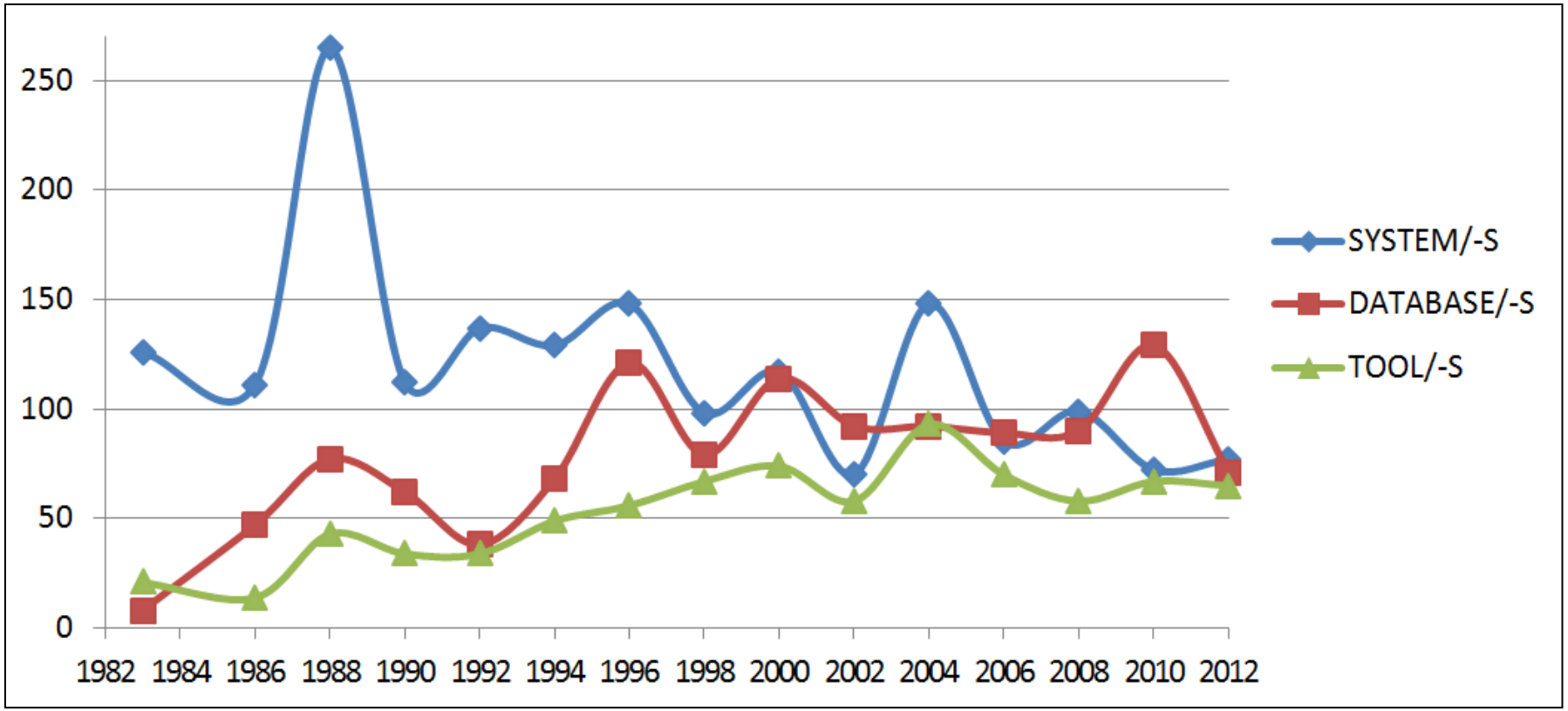
Trend for TEI, DTD and ISO



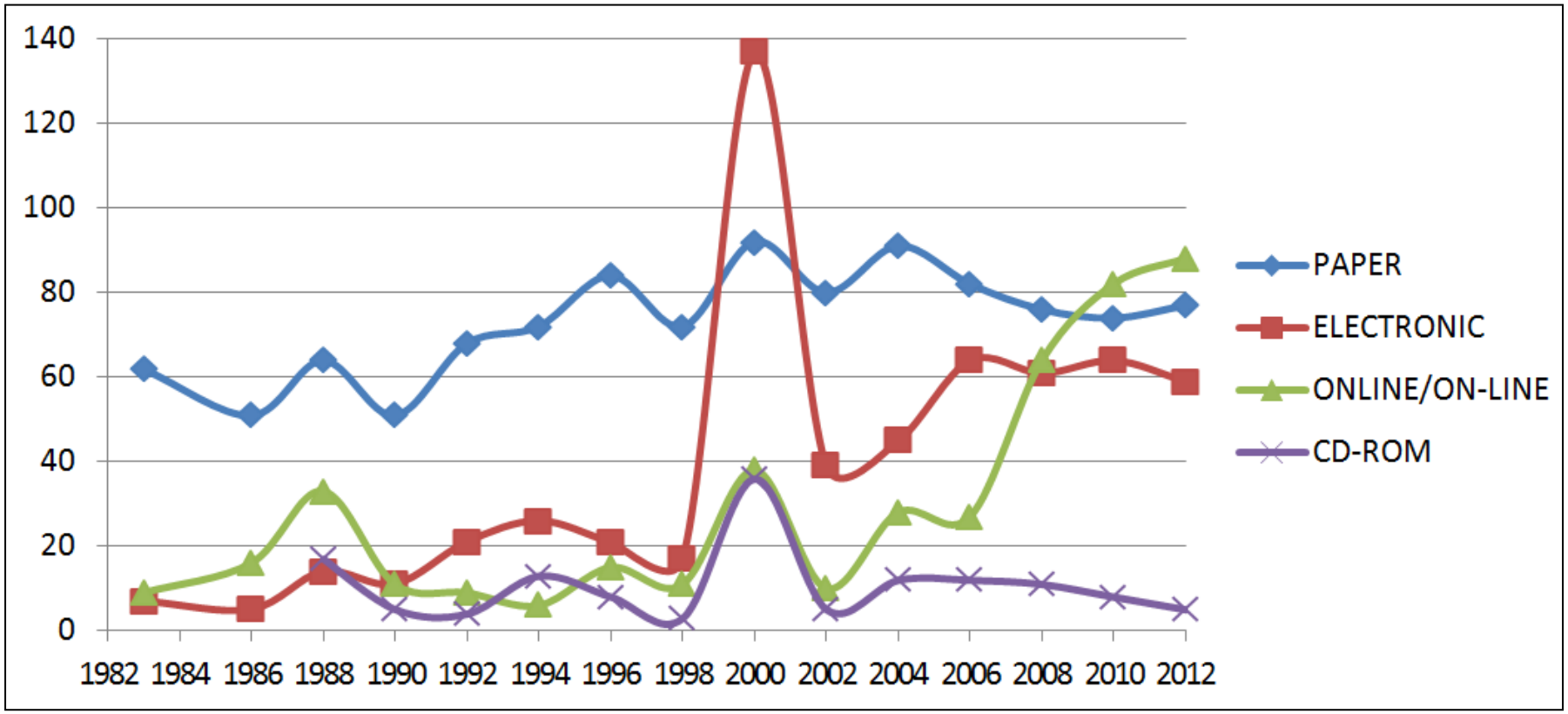
Trend for WordNet, FrameNet, wordnets, and EuroWordNet



Trend for System, Database and Tool



Trend for Paper, Electronic, Online, and CD-ROM



5. A CRYSTAL-CLEAR OUTLOOK

Fortune-telling the future

- Having reviewed the facts and trends in three decades of EURALEX congresses, one would assume that it is easy to now fortune-tell the future.
- Surprisingly, it is not.

Fortune-telling the future

- Yes lexicography is in transition from a highly traditional art and craft, typically funded by publishers seeking profit from publication in book form, to a new interdisciplinary science in which publishers, software houses, freelancers and university researchers from a variety of disciplines all participate in creating electronic resources for a wide variety of different applications, typically for online use.
- This is the present.

Fortune-telling the future

- The future will get rid of the book components altogether, and the form the online components will take will be driven by the ever-faster-evolving technological exploits.
- Some of this technology will have been conjured up by teams of lexicographers, but in most cases lexicographers will simply jump on the latest 'free' tool offered by a search engine, a social network, or a data-mining team.

Fortune-telling the future

- To the average user ‘the dictionary’ will simply disappear from view, and drown in a sea of advertisement — customised, of course, amongst others based on the searches in the lexicographic components of whatever tools or networks they use.
- In order to get rid of the pestering advertisements, users will be able to Go Premium, but Big Brother will still be watching them and continue to build their evolving profile in the cloud.

Fortune-telling the future

- That what we now call lexicographic databases will end up in a variety of social networks is a given.
- There lexicographic democracy will take its course.
- Machines, too, will automatically populate lexicographic databases.

Fortune-telling the future

- Stone Age lexicographers will try to compete with these mediocre sources, by painfully analysing unimaginably large amounts of real language data and crafting delicate summaries, mapping meaning onto use, focusing on the norms in order to better describe the exploitations, building frequency-supported patterns for the various word classes.
- But without a Publishing House, by then called a Marketing House, their efforts will be futile.

Fortune-telling the future

- The future will bring out both the best and the worst of today's lexicographic dreams.

Where will this leave today's academic lexicographers?

- They will be frantically looking for a theory of lexicography, to justify their research position.
- They will, of course, not succeed, unless they explain the plain obvious in some newly invented language of their own.
- Or else, if they keep on describing and categorizing what has already been lumped and split a million times before.
- Calling it a theory does not make it a theory.

Where will this leave today's academic lexicographers?

- The smarter colleagues will simply realise that lexicography is a synthetic science.
- This science will need ever more knowledge and (real!) theories from other disciplines in order to move forward scientifically.

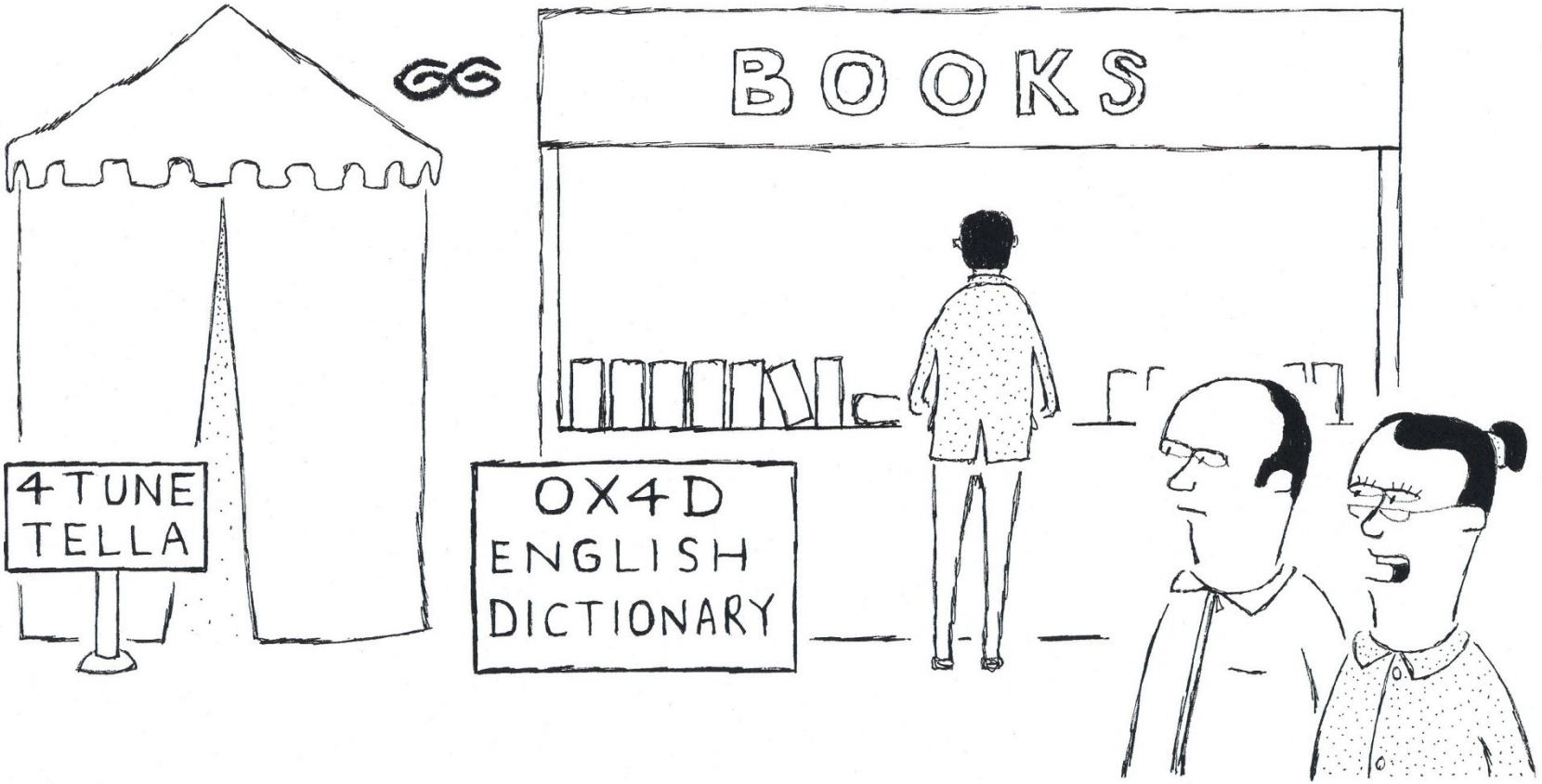
Where will this leave today's academic lexicographers?

- At congresses, ever more papers will be co-authored (the pressures to publish and be quoted will skyrocket).
- Each scholar will also be involved in as many papers as possible.
- The PEOPLELEX congresses — or whatever will succeed EURALEX and its sister associations, currently cut up along artificial borders — will be hosted by what is now the periphery.
- The current mainstream will come to realize that the lexicographies and solutions of the periphery have far more to contribute.

Where will this leave today's academic lexicographers?

- Lexicography as we know it today will cease to exist, lexicographers will be bringing together their data in entirely new ways, and dictionaries will change beyond all recognition.
- The times will still be as exciting as today, however, as we will be living in the future.

Takk!


BOOKS
4 TUNE TELLA
OX4D ENGLISH DICTIONARY
'WHAT MAKES YOU THINK THEY'RE MODERN FAKES?'