

What's in a Name? The Case of *Albanisch-Albanesisch* and Broader Implications

Alex Erdmann* Erhard Hinrichs** Brian Joseph*

* Department of Linguistics
The Ohio State University

** Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

On 8/7/15 5:25 AM, Joseph, Brian wrote:

Hi — I have a question about the German lexicon ...

I have noticed that whereas most early mentions of Albanian in American works (like those by William Dwight Whitney, Edgar Sturtevant, or Edward Sapir) use the designation “Albanian”, Leonard Bloomfield in his 1933 work *Language* uses “Albanese”. I wonder, of course, if Bloomfield was influenced by his German training and the German adjective “albanesisch”. So I am wondering if there is a way to gauge the relative frequency of German “albanisch” versus “albanesisch” between 1850 and 1950.

Thanks in advance for any help you can offer.

*In the same way, finding all these languages and groups (Sanskrit, Iranian, Armenian, Greek, **Albanese**, Latin, Celtic, Germanic, Baltic, Slavic) resemble each other beyond the possibility of mere chance, we call them the Indo-European family of languages.*

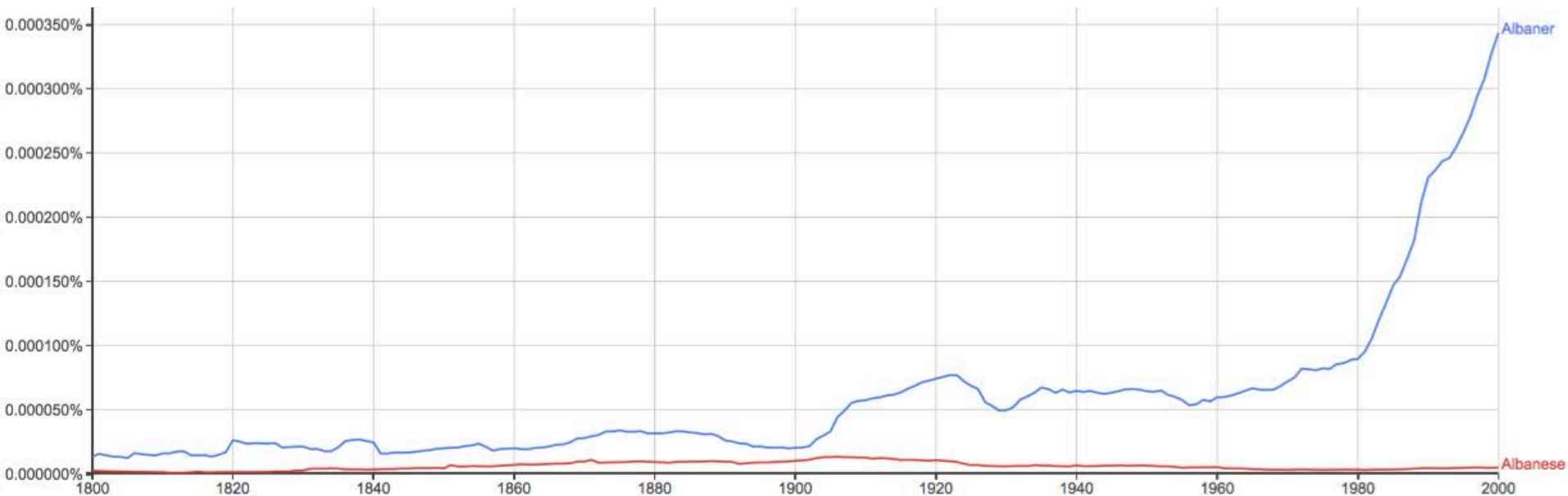
Leonard Bloomfield (1933). *Language*. The University of Chicago Press, p. 13f.

	Contemporary English	Bloomfield	Contemporary German
Country	<i>Albania</i>		<i>Albanien</i>
People	<i>Albanian(s)</i>		<i>Albaner</i>
Language	<i>Albanian</i>	<i>Albanese</i>	<i>Albanisch</i>
Adjective	<i>Albanian</i>		<i>albanisch</i>

- coverage: 1810 – 2000; size: 400 million words
- results for:
 - *Albanian*: 10 occurrences for 19th century
387 occurrences total
 - *Albanese*: 2 occurrences for 19th century
28 occurrences total

- Google books collection of digitized German books; coverage: 1800 – 2000; size: ???, but very large
<https://books.google.com/ngrams>
- Deutsches Textarchiv (DTA); coverage: 1610 -1900; size: 142 million words
<http://www.deutschestextarchiv.de/>
- Corpora of the Digitales Wörterbuch der deutschen Sprache (DWDS); coverage: since 1900
<http://www.dwds.de/d/korpora/>





Lemma	Frequency Count	Earliest Data Point	Latest Data Point
<i>Albanisch</i>	97	1650	1913
<i>Albaner</i>	61	1663	1913
<i>Albanesisch</i>	19	1789	1913
<i>Albanese</i>	36	1789	1913

Lemma	Frequency Count	Earliest Data Point	Latest Data Point
<i>Albanisch</i>	97	1650	1913
<i>Albaner</i>	61	1663	1913
<i>Albanesisch</i>	19	1789	1913
<i>Albanese</i>	36	1789	1913

Lemma	Frequency Count	Earliest Data Point	Latest Data Point
<i>Albanisch</i>	97	1650	1913
<i>Albaner</i>	61	1663	1913
<i>Albanesisch</i>	19	1789	1913
<i>Albanese</i>	36	1789	1913

Google Books German N-Gram Viewer: *albanische Sprache vs. albanesische Sprache*



19th century German

Contemporary German

Albanesisch



Albanisch

Albanesen



Albanier



Albaner

- (Relative) frequencies of corpus unigrams can be highly misleading
- Seamless keyword-in-context exploration is essential
- Data exploration tools with supporting visualizations such as DiaCollo are a real asset

Herodotos Project

<https://u.osu.edu/herodotos/>

aim: a comprehensive listing of group names mentioned in Classical sources and in secondary literature on those sources.

Thank you for your attention.