

Digital Humanities in Japanese Studies

*Presentation for Council on East Asian Libraries
AAS, 2015*

Hoyt Long
University of Chicago

What are digital humanists doing?

1) Finding ways to archive and better curate digital collections

- Image Collections, Text Collections, Databases w/ richly linked metadata, New presentation platforms (Omeka, Neatline)

2) Developing methods for small and large scale data analysis

- Text Mining, Network Analysis, GIS, Image Analysis

3) Visualizing humanities data to facilitate data exploration

- Tools for search and pattern detection (Voyant, Palladio)

4) In Japanese Studies?

- Much of the focus still on #1, but slowly shifting to #2 and #3

How can librarians aid digital scholarship?

1) Learn about and help disseminate available tools & resources

- DH Centers and Labs, DH Conferences (esp. JADH), Lib Guides, Online Tutorials, Digital Collections

2) Work to enhance access to existing collections and make them available for data analysis

- Aozora, NINJAL corpus, HathiTrust, E-books, Electronic databases

3) Create workflows for building local digital collections and for transforming static print resources into dynamic archives

- Scanning, OCR, data hosting, online access, copyright issues → think of digital collections as resources for viewing and content analysis

Transforming Print to Database

Transforming structured data via hand input, scan/OCR, or born-digital material



現代詩 1920–1944: モダニズム詩誌作品要覧

100,000+ Poems

4,000+ Poets

166 Journals

草野 心平

蛙・生殖	「銅鑼」3 '25年発行 p6
病氣	「銅鑼」3 '25年発行 p7
蛙	「銅鑼」1 '25.4 p11
マリー・ローランサンとぼく 或る夢の貌	「銅鑼」1 '25.4 p12
蛙になる	「銅鑼」1 '25.4 p13
失恋者と蛙	「銅鑼」1 '25.4 p14
春	「銅鑼」1 '25.4 p15
春の刺激	「銅鑼」1 '25.4 p16
青い場面	「銅鑼」1 '25.4 p17
福田正夫の没落(評論)	「銅鑼」2 '25.5 p3
無題	「銅鑼」2 '25.5 p11
再びマンモス時代	「銅鑼」2 '25.5 p12
木登りをする女	「銅鑼」2 '25.5 p13
青い水たんぼ	「銅鑼」2 '25.5 p14
無題	「銅鑼」2 '25.5 p15
驟雨直後	「銅鑼」2 '25.5 p15
寸言	「銅鑼」2 '25.5 p16
蛙・中性	「銅鑼」4 '25.9 p9
無題(未定稿)	「銅鑼」4 '25.9 p10
編輯後記	「銅鑼」4 '25.9 p11
愛犬ルリパロと私の散歩	「銅鑼」5 '25.10 p34

The Database Version

A SQL database made accessible through a simple web-interface

Explore a Poet

JP Choose a Japanese Poet ▾

Choose a Foreign Poet ▾

Prolific Poets

- 岡崎 清一郎 (606 poems)
- 岩佐 東一郎 (543 poems)
- 滝口 武士 (494 poems)
- 前田 鉄之助 (489 poems)
- 西條 八十 (486 poems)
- 北園 克衛 (473 poems)
- 横山 青娥 (466 poems)
- 村野 四郎 (441 poems)
- 春山 行夫 (440 poems)
- 佐藤 惣之助 (414 poems)
- 竹中 郁 (409 poems)
- 安西 冬衛 (404 poems)
- 近藤 東 (394 poems)
- 阪本 越郎 (381 poems)
- 長田 恒雄 (373 poems)
- 城左門 (361 poems)
- 八十島 稔 (349 poems)
- 竹中 久七 (343 poems)
- 安藤 一郎 (341 poems)
- 白鳥 省吾 (291 poems)

Prolific Foreign Poets

- Rainer Maria Rilke (89 poems)
- Ko Ei (88 poems)
- Paul Valéry (88 poems)
- Jean Cocteau (84 poems)
- Paul Éluard (70 poems)
- Charles Baudelaire (61 poems)
- Francis Jammes (54 poems)
- André Gide (47 poems)
- Herbert Read (45 poems)

Explore a Journal

Choose a Journal ▾

Translation-heavy Journals (raw)

- 文芸汎論 (482 translations)
- 詩壇 (358 translations)
- 蝶人形 (352 translations)
- 詩と詩論 (348 translations)
- 四季 (338 translations)
- 新領土 (299 translations)
- セルバン (243 translations)
- PANTHÉON (180 translations)
- オルフェオン (172 translations)
- 文学 (148 translations)
- 愛誦 (127 translations)
- 椎の木(第三次) (121 translations)
- ドノゴトンカ (98 translations)
- 詩・現実 (96 translations)
- 近代風景 (95 translations)
- 詩洋 (83 translations)
- L'ESPRIT NOUVEAU (紀伊国屋書店) (80 translations)
- 新詩論 (76 translations)
- 日本詩壇 (68 translations)
- 奢濶都 (63 translations)

Referenced Foreigners

- Charles Baudelaire (88 refs)
- Jean Cocteau (69 refs)
- Rainer Maria Rilke (54 refs)
- Paul Valéry (53 refs)
- André Gide (52 refs)
- Stéphane Mallarmé (45 refs)
- Arthur Rimbaud (41 refs)
- Edgar Allan Poe (36 refs)
- T.C. Eliot (24 refs)

Long-Running Journals

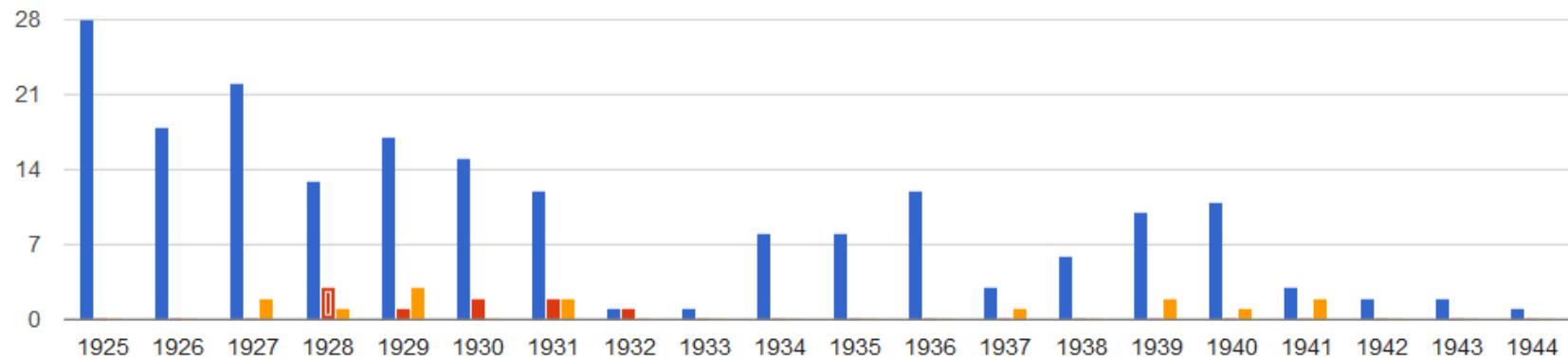
- 詩洋 (19 years)
- 森林 (18 years)
- 蝶人形 (14 years)
- 文芸汎論 (13 years)
- 地上樂園 (13 years)
- 四季 (11 years)
- 北方詩人 (11 years)
- 日本詩壇 (11 years)

Prolific Journals

- 日本詩壇 (10,680 poems)
- 愛誦 (9,354 poems)
- 蝶人形 (7,573 poems)
- 詩洋 (6,959 poems)
- 文芸汎論 (5,939 poems)
- 地上樂園 (4,292 poems)
- 詩神 (4,192 poems)
- 詩之家 (4,144 poems)

草野 心平 (クサノ・シンペイ)

193 poems 9 translations 14 references

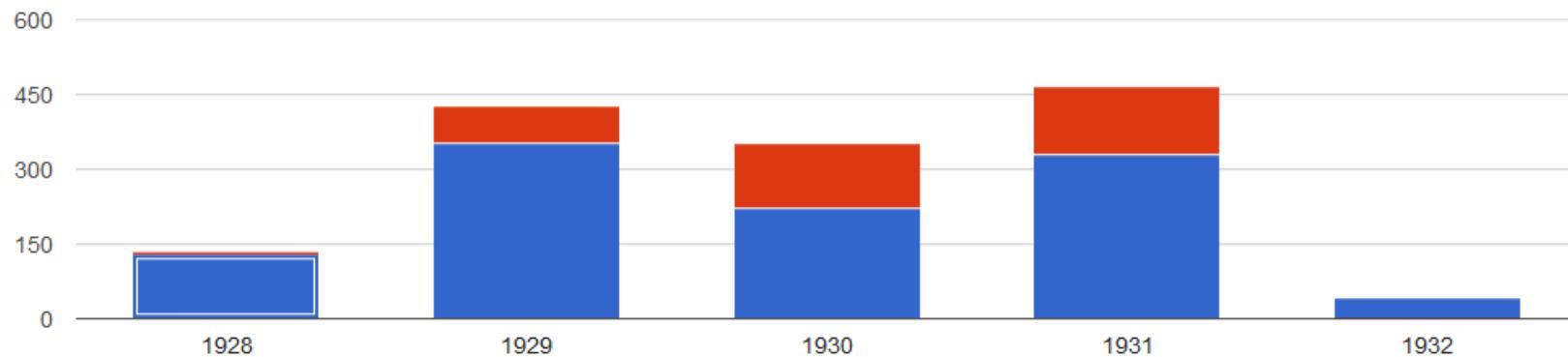


```
SELECT poemtitle, role, japname, originalauthor, volno, page, translator FROM japanmaster JOIN journals ON journals.id=japanmaster.journaltitle WHERE poempoet = 168020 AND datatype = 'work' AND role LIKE '%translate%' AND LEFT(pubdate, 4) = '1928' ORDER BY journaltitle;
```

- 「シカゴ」『銅鑼』Vol. 15 p. 4 (translation of カール・サンドバーグ)
- 「子供の天真性に就いて（「現代学校の起源と理想」第十五章）」『銅鑼』Vol. 15 p. 8 (translation of フランシスコ・フエラア)
- 「歩く者」『詩神』Vol. 4.11 p. 42 (translation of アルチュロ・ヂオヴァニッチ)

詩と詩論 (Shi to shiron): 1294 poems, 348 translations (21%)

View as: [Stacked](#) or [Grouped](#) | [Show/Hide SQL Queries](#)



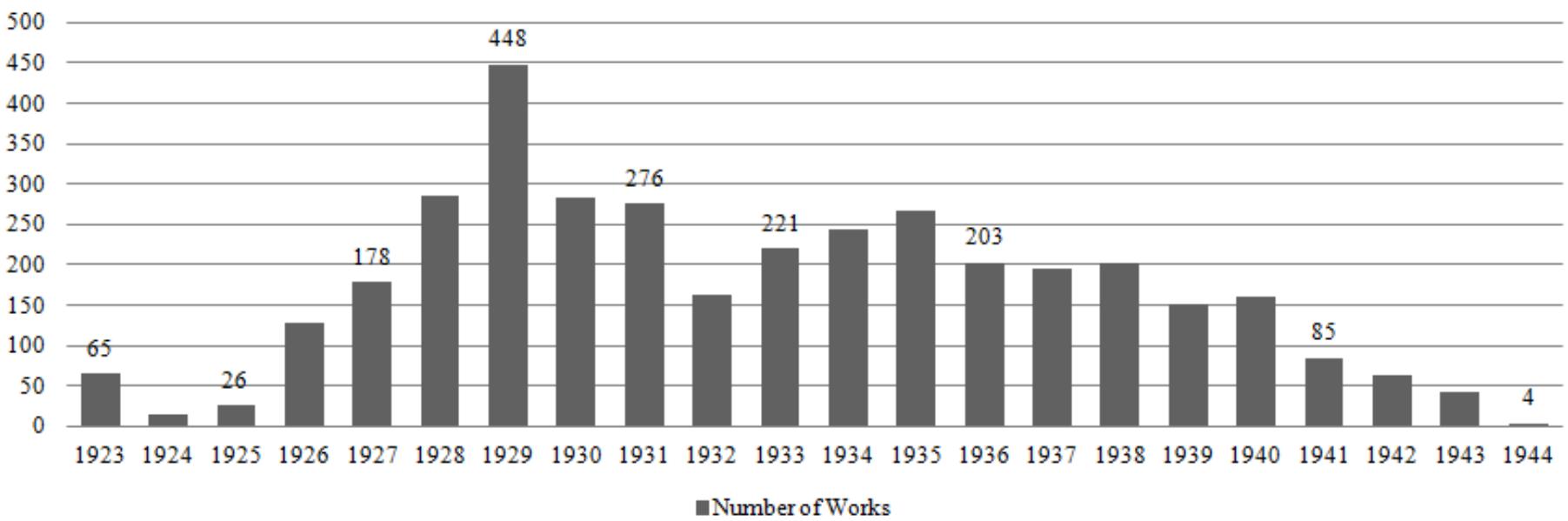
Sort by [print order](#) | [type](#)

- 神原 泰「未来派の自由語を論ず(一) ¥僚友Marinettiにおける」『詩と詩論>』1 (1928): 1
- 飯島 正「ジュウル・ロマンに関する覚え書 その一」『詩と詩論>』1 (1928): 21
- 三好 達治「ポール・ヴェルレーヌに就て」『詩と詩論>』1 (1928): 33
- 北川 冬彦「マツクス・ジャコブの散文詩論」『詩と詩論>』1 (1928): 36
- 上田 敏雄「TEXT SURREALISTE」『詩と詩論>』1 (1928): 42
- アラゴン, ルイ「TEXT SURREALISTE」『詩と詩論>』1 (1928): 42 (translated by 上田 敏雄(訳))
- 中村 喜久夫「純粹詩論」『詩と詩論>』1 (1928): 47
- 春山 行夫「日本近代象徴主義詩の終焉 (萩原／朔太郎・佐藤／一英両氏の象徴主義詩を検討す)」『詩と詩論>』1 (1928): 66
- 外山 卯三郎「詩学の基本的問題—言葉を中心とする詩学的思索—」『詩と詩論>』1 (1928): 85
- 北川 冬彦「空腹について」『詩と詩論>』1 (1928): 111
- 北川 冬彦「絶望の歌」『詩と詩論>』1 (1928): 112
- 北川 冬彦「萎びた筒」『詩と詩論>』1 (1928): 113
- 北川 冬彦「肉親の章」『詩と詩論>』1 (1928): 113
- 北川 冬彦「同じく」『詩と詩論>』1 (1928): 114
- 北川 冬彦「剃刀」『詩と詩論>』1 (1928): 115

This dataset helped me to initiate a research project focused on translated content within Japanese modernist journals. The results of this project will appear as “Fog and Steel: Mapping Communities of Literary Translation in an Information Age” (forthcoming in Journal of Japanese Studies, Summer 2015).

The data was useful for the following kinds of analysis:

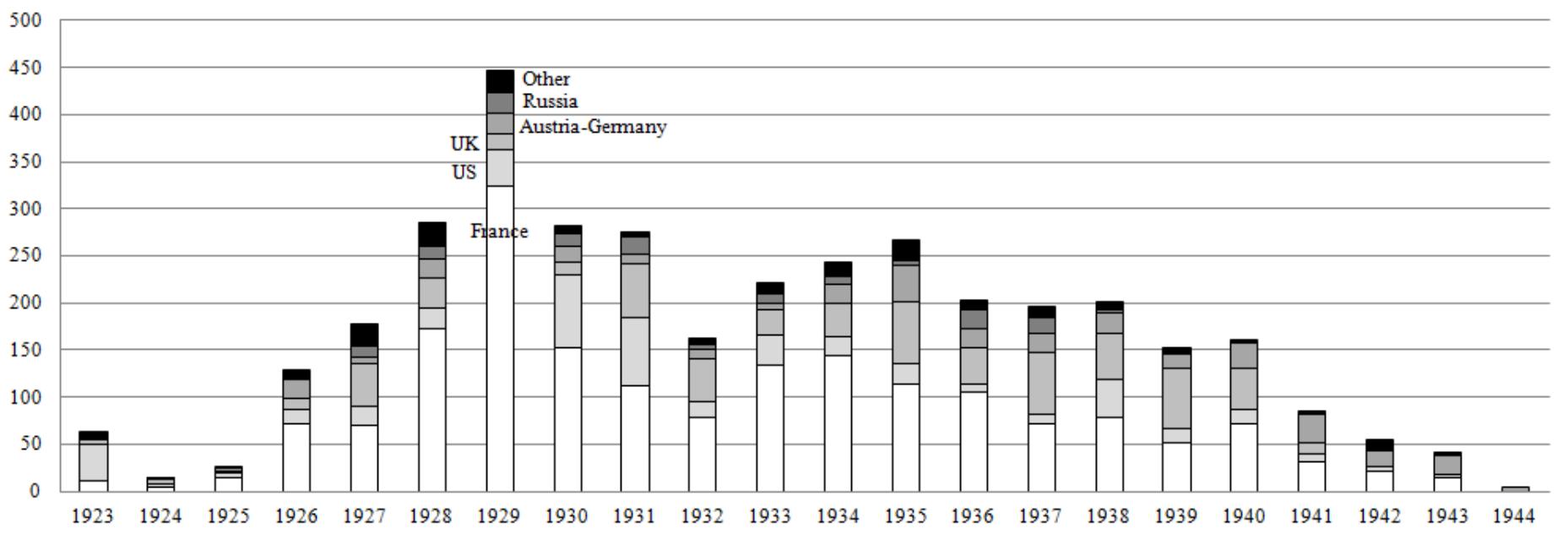
Translated Works by Year



1. Basic Statistical Analysis as a way to establish larger historical context and general trends

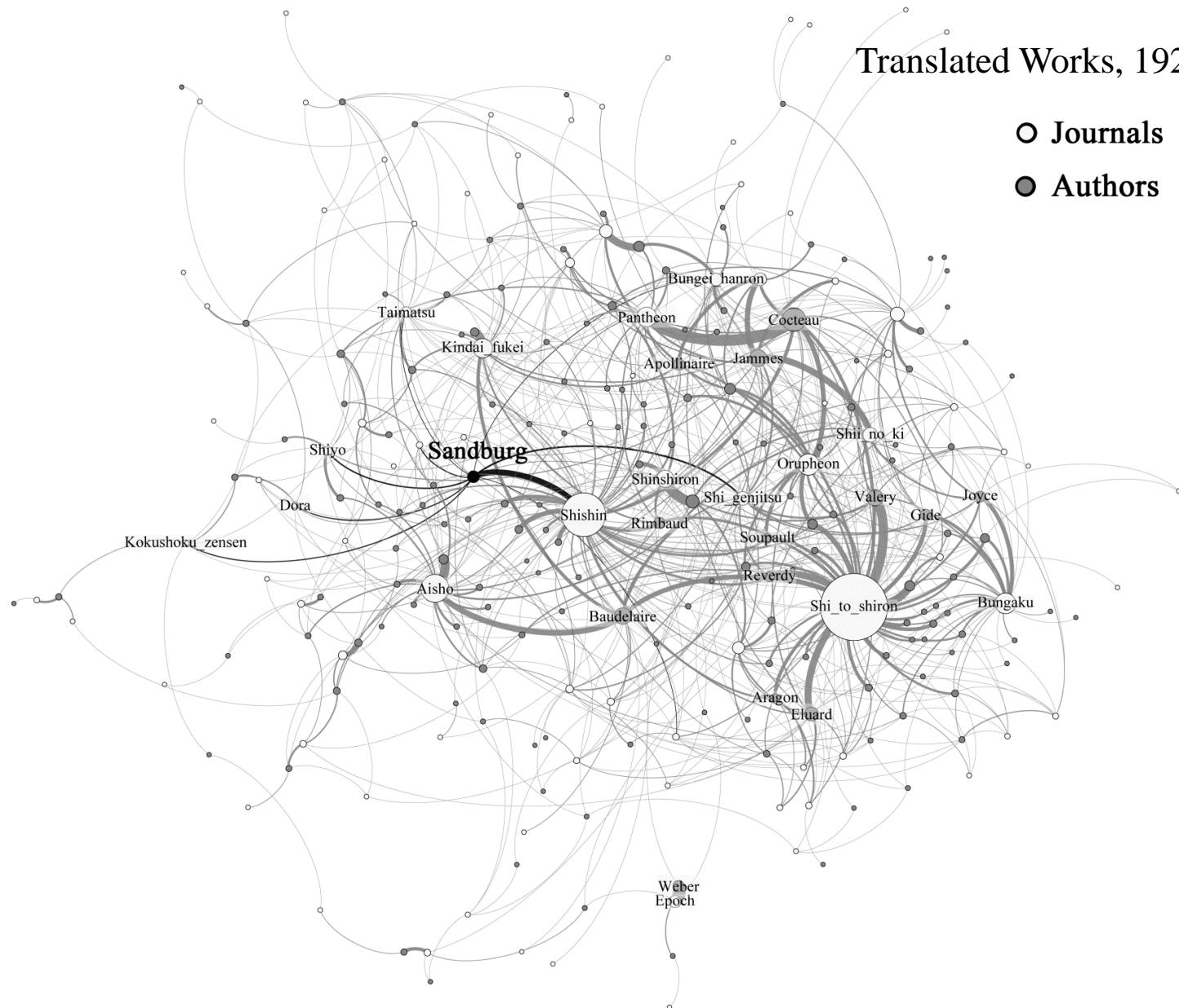
Translated Works by Region of Origin

□ France □ US □ UK □ Austria-Germany □ Russia ■ Other



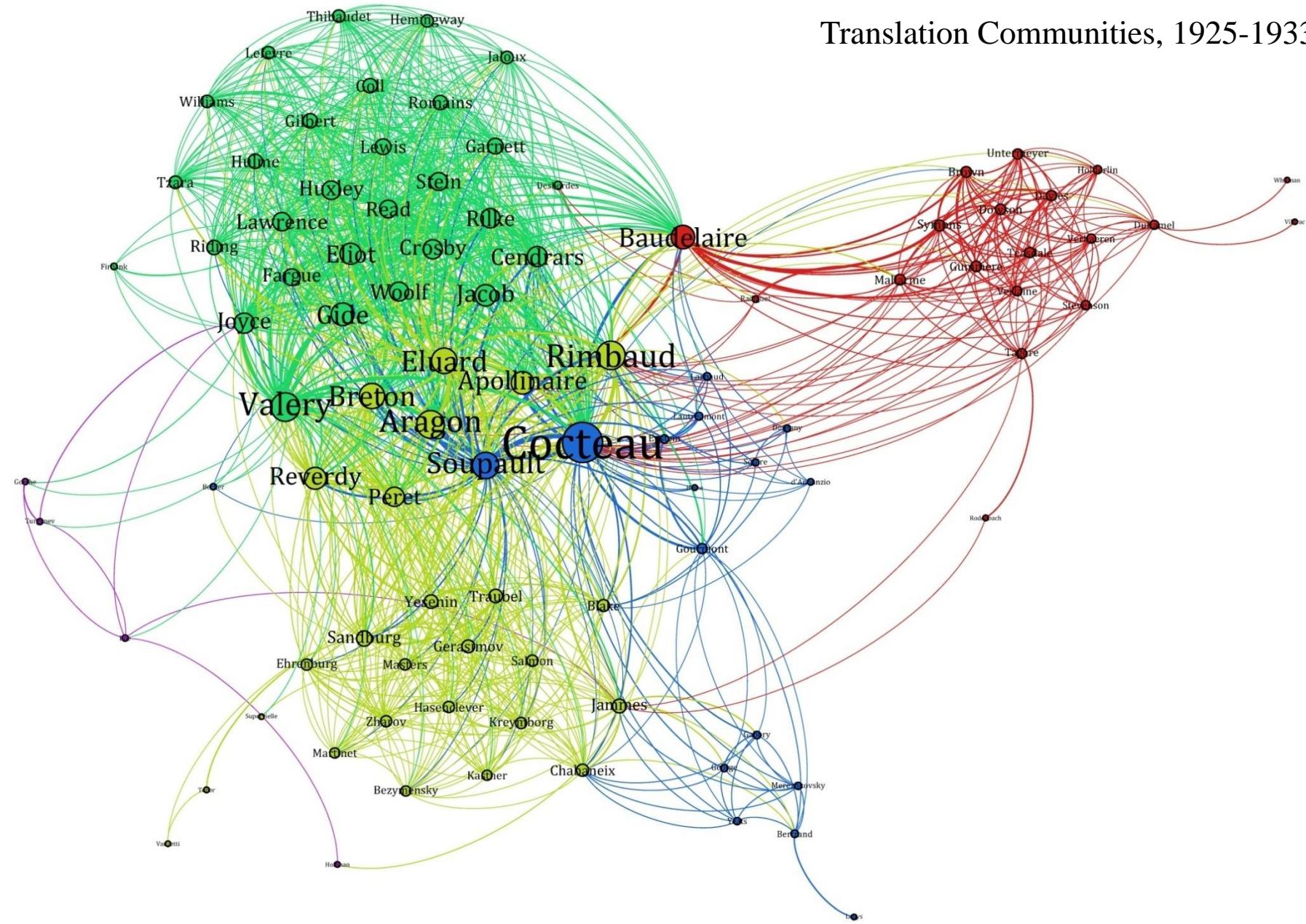
2. *Finer-grained statistical analysis using richer categories of metadata*

Translated Works, 1925-1933

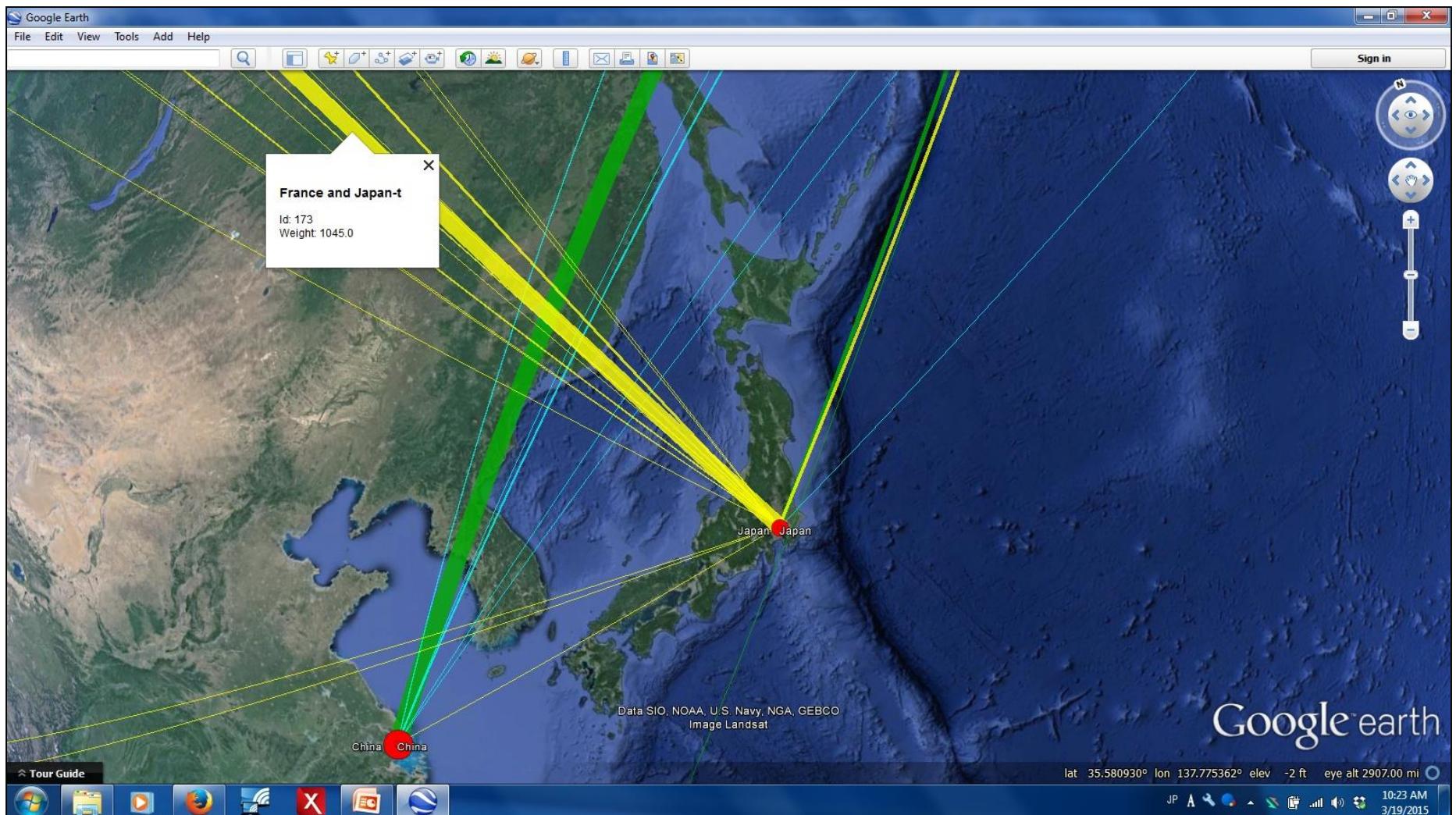


3. Complex network analysis to explore patterns of relation within the data (using Excel and Gephi)

Translation Communities, 1925-1933

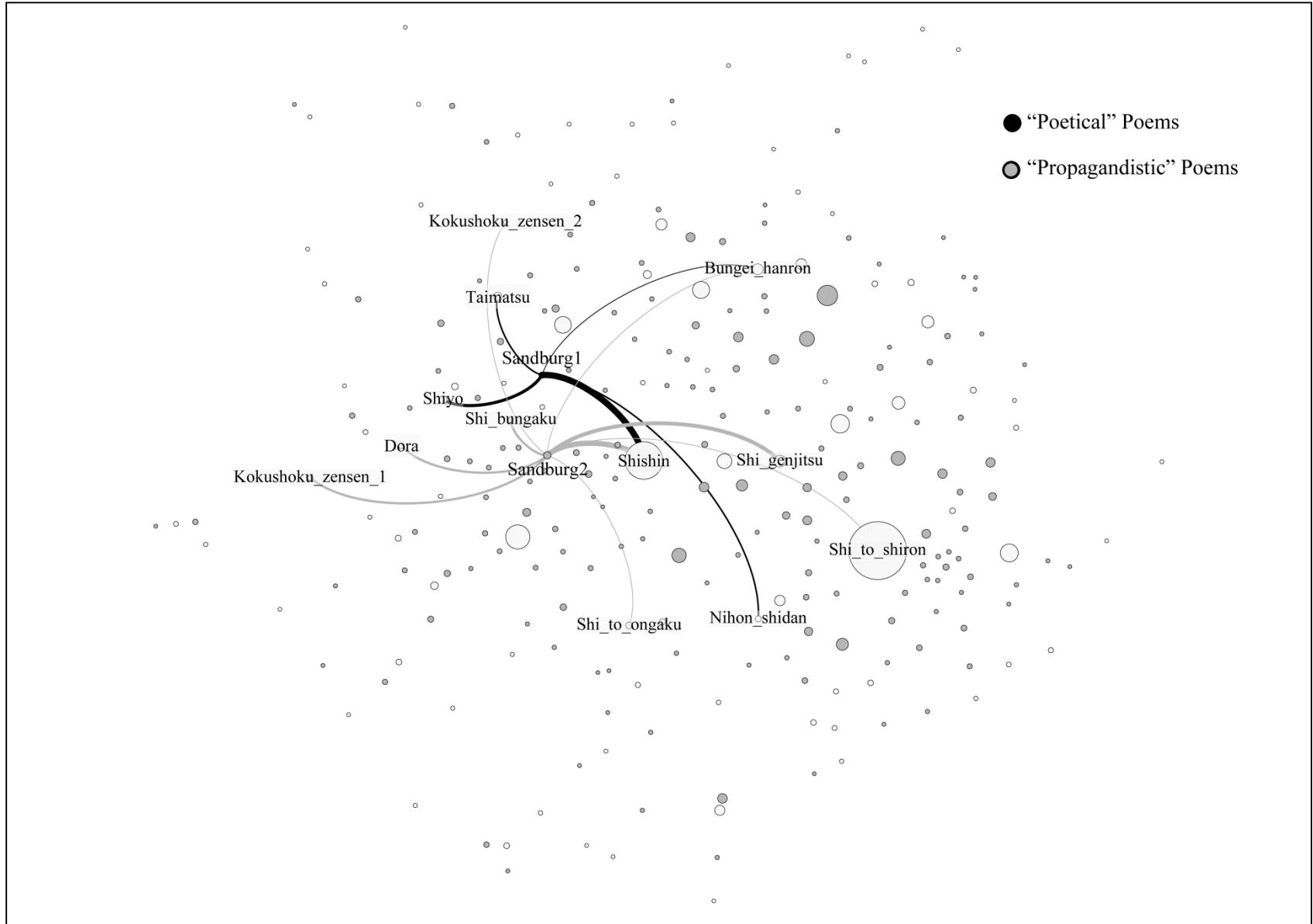


4. Transforming relational data to identify clusters or groupings of translated poets (Gephi)



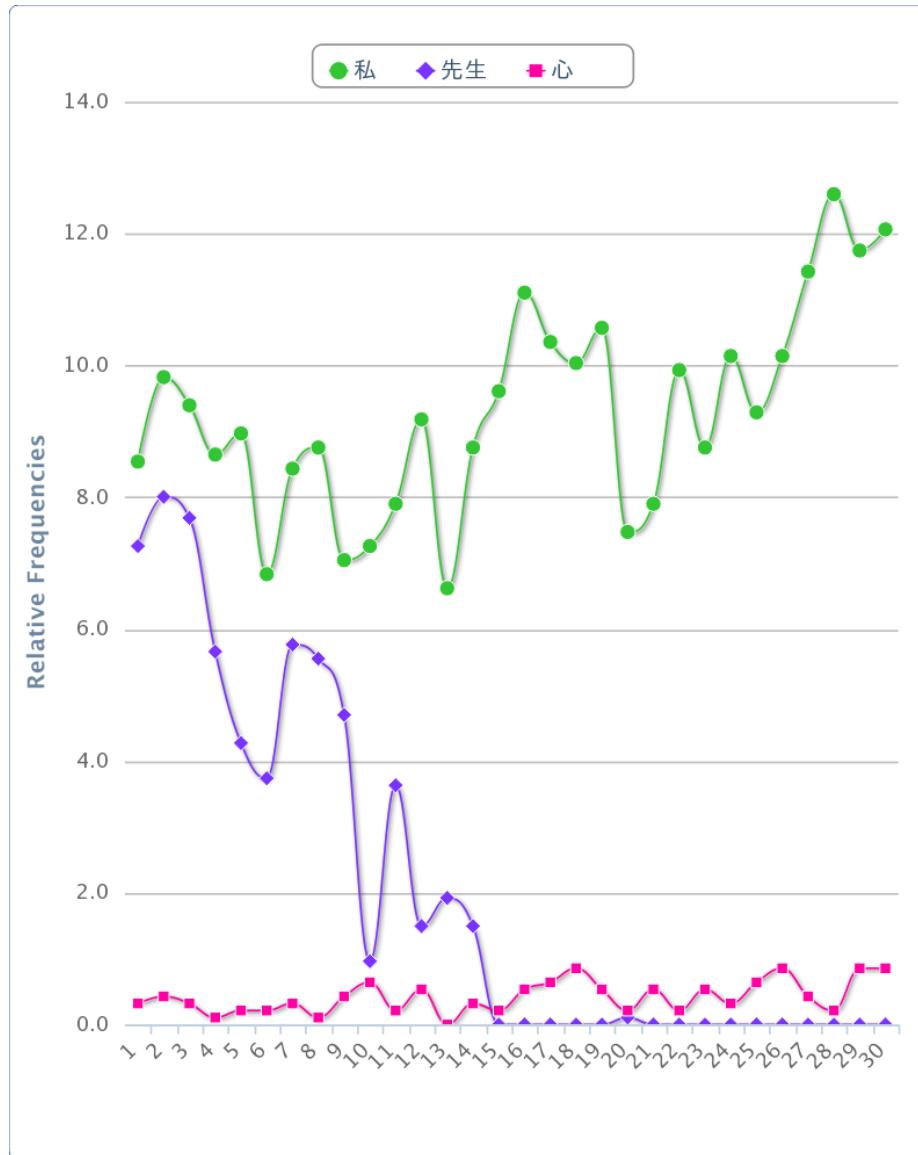
5. Comparative exploratory analysis of global translation patterns using other national datasets

But what about the texts? From network analysis to text analysis

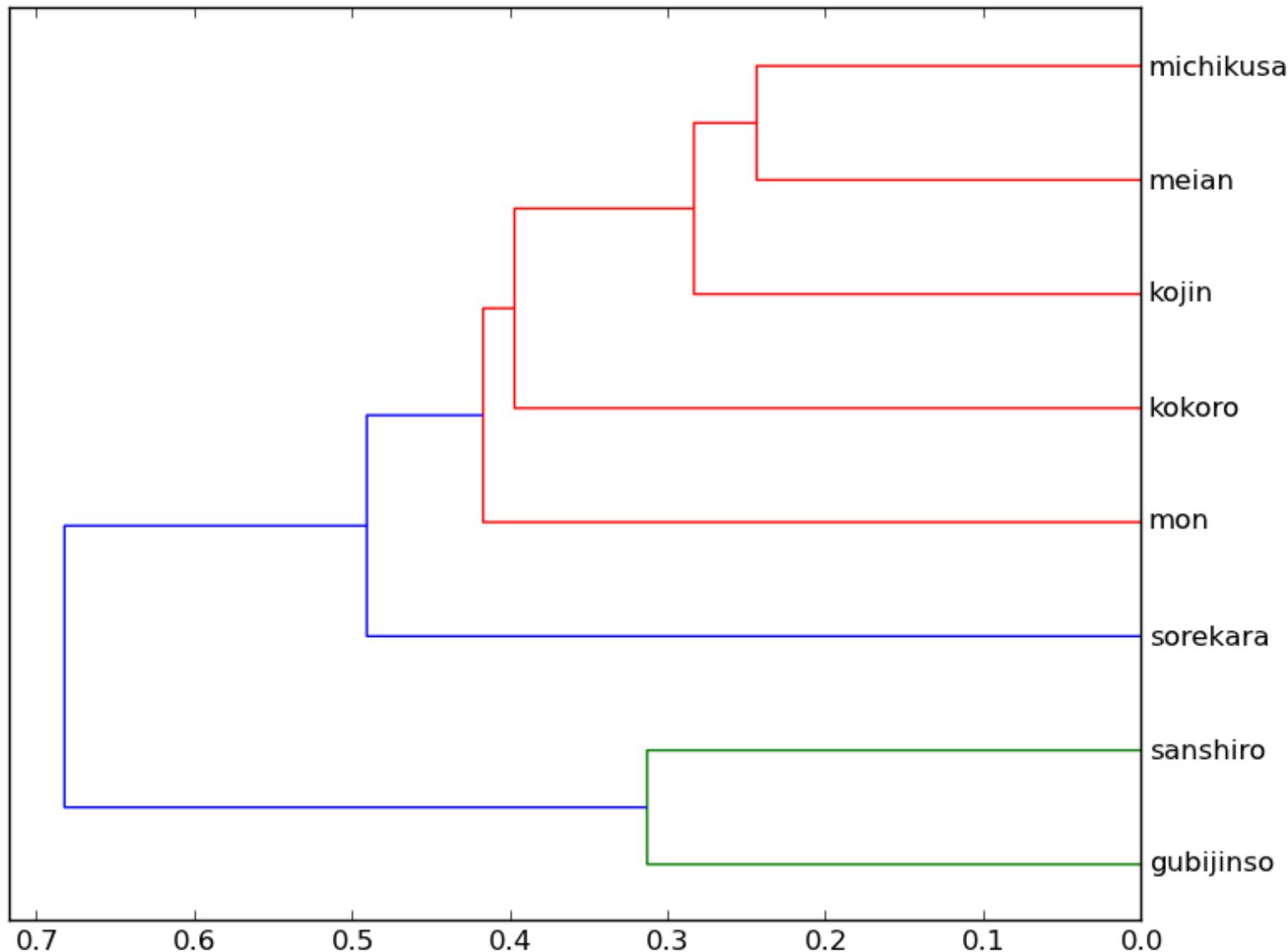


How do patterns of publication and collaboration intersect with aesthetic/stylistic patterns?

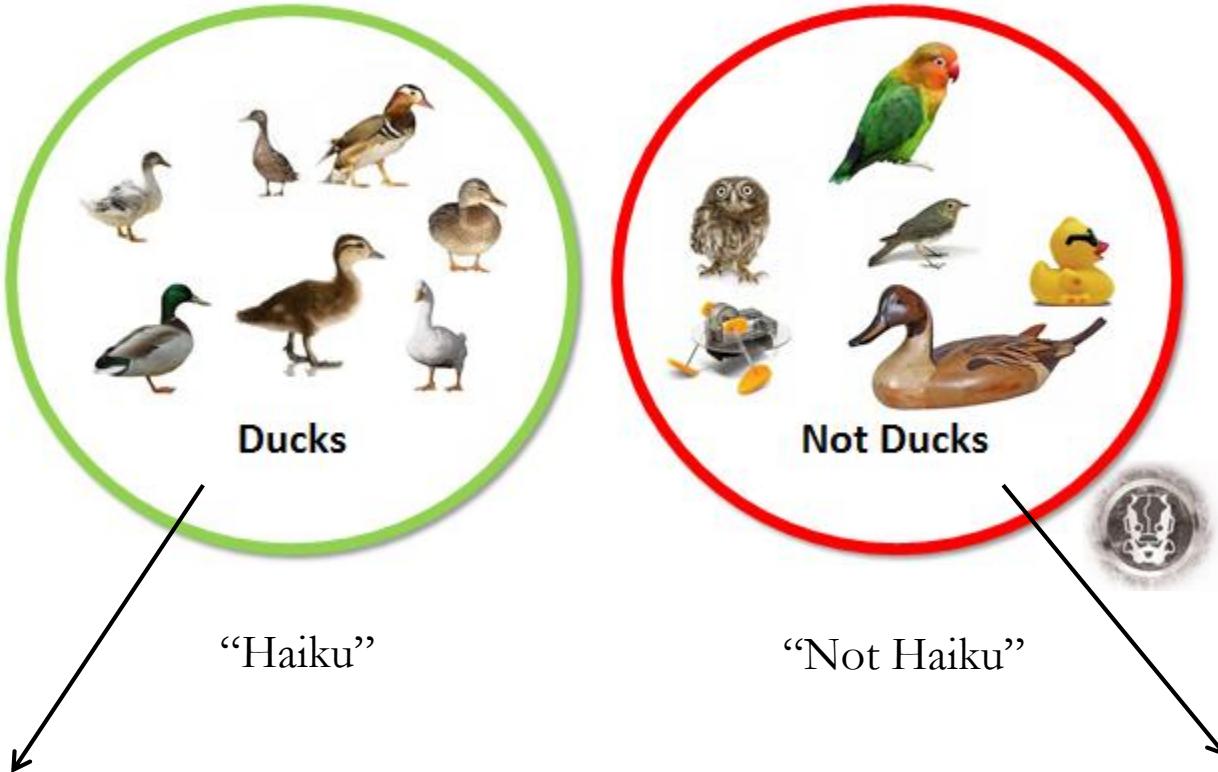
What kinds of analysis can we do with literary textual data?



1. Obtain basic corpus linguistics measures (e.g., n-grams) using digitized texts and online tools like (e.g., Voyant). Useful for individual texts and small corpora.



2. Simple comparisons of lexical difference using digitized texts and online python tutorials



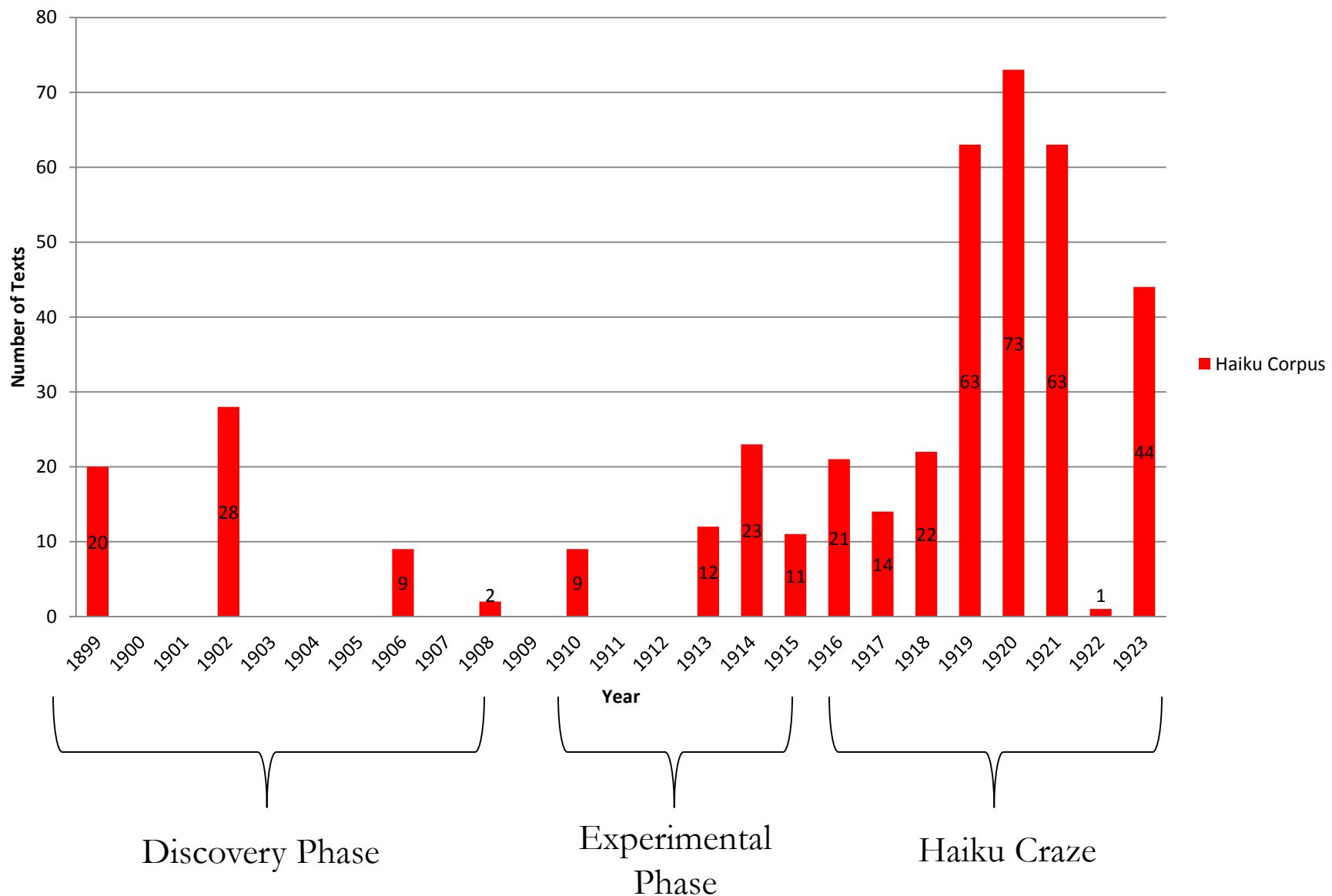
*Old pond
frogs jumped in
sound of water*

These words (features) are associated with each class

*Lo, in the wind what legions pass,
Phantom, innumerable, fleet!
Watching, I dream their pauseless feet
Are lights that run upon the grass.*

3. Text classification and exploratory analysis using algorithmic approaches like machine learning (large corpora and advanced programming). From “Literary Pattern Recognition: Modernism between Close Reading and Machine Learning,” forthcoming in *Critical Inquiry*.

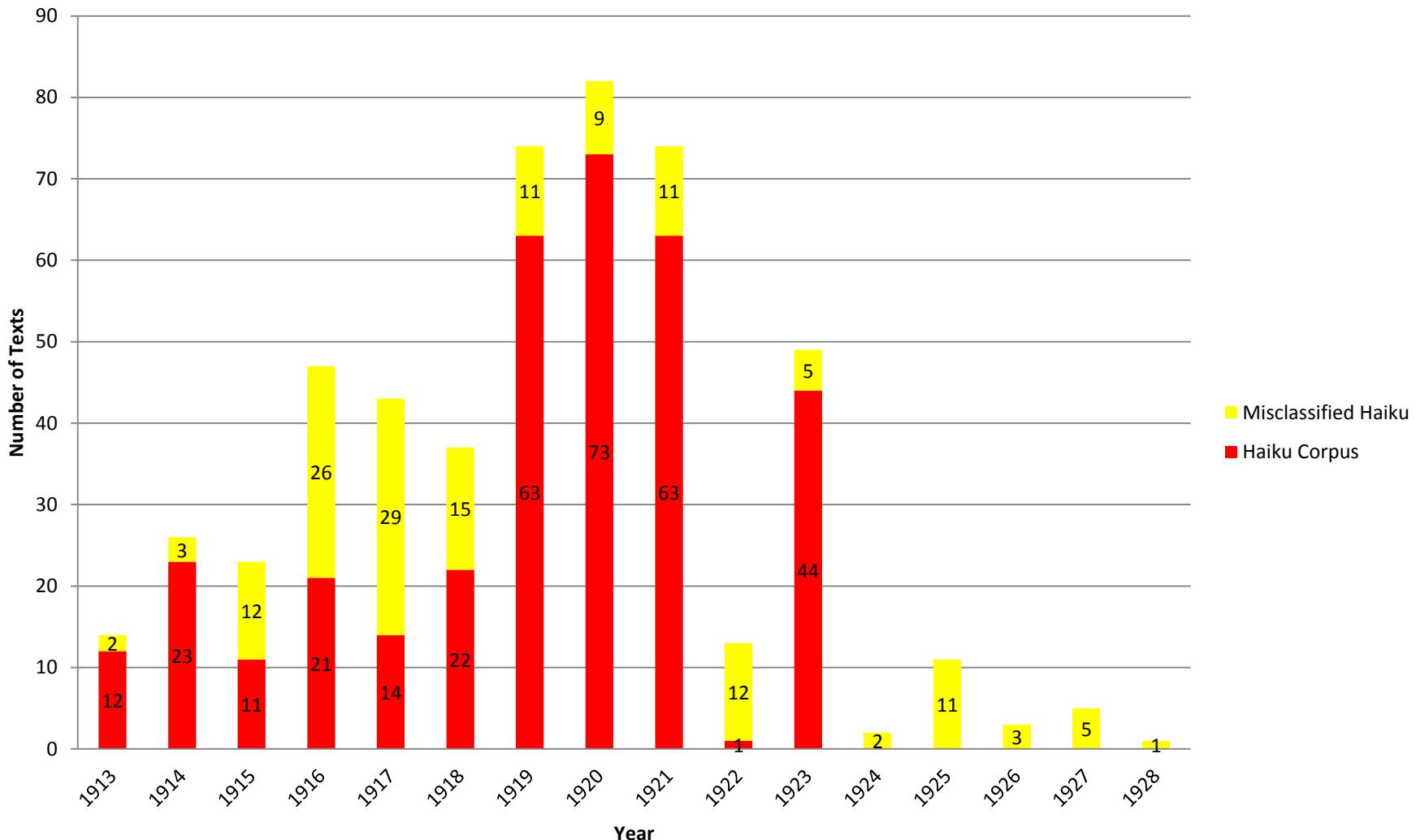
Assembling a Corpus of English-language Haiku

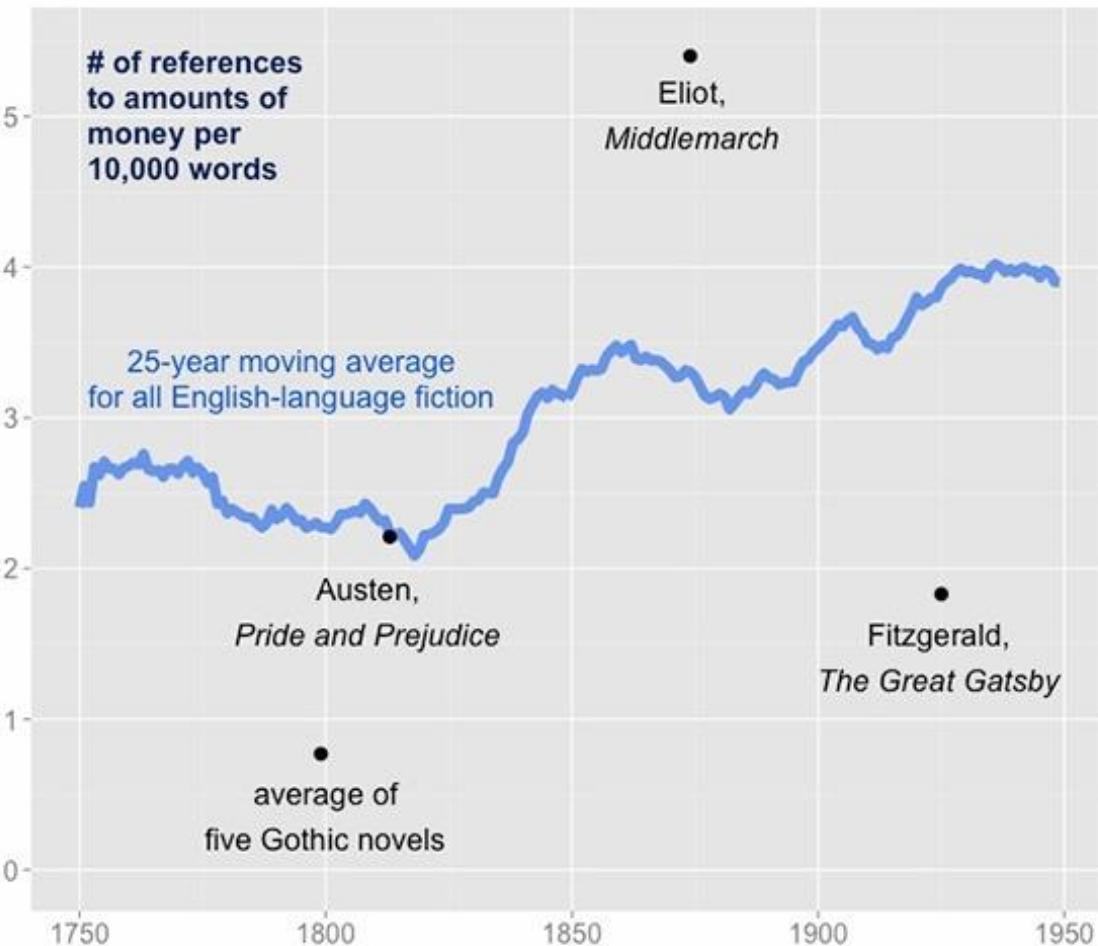


Finding Evidence of those Haiku in Larger Corpus of Poems

Magazine Corpora	Dates	Number of Short Poems
<i>Poetry Magazine</i> (Early)	1912-1917	222
<i>Poetry Magazine</i> (Late)	1918-1922	317
<i>Masses</i>	1911-1917	113
<i>Little Review</i>	1914-1922	119
<i>Others</i>	1915-1919	138
<i>Smart Set</i>	1915-1922	92
<i>Midland</i>	1915-1922	58
<i>Contemporary Verse</i>	1916-1922	256
<i>Lyric West</i>	1921-1922	106
Harlem Renaissance Magazines (<i>Crisis</i> , <i>Opportunity</i>)	1911-1929	268
Generalist Magazines (<i>Bookman</i> , <i>Century Magazine</i> , <i>Harper's Magazine</i> , <i>The Nation</i> , <i>New Republic</i> , and <i>Scribner's Magazine</i>)	1915-1922	230

Creating an Expanded Corpus of Machine Discovered Haiku





“Specific references to wealth and income were omnipresent in the literature of all countries before 1914,” [Piketty] writes, because money was a stable social reference point. After 1914, the inflationary shocks associated with two world wars briefly reduced inequality by devaluing capital. But inflation also rendered “the meaning of money ambiguous,” because prices now dated too rapidly to be meaningful for long. In response to this new fluidity, novelists simply stopped counting money; “money—at least in the form of specific amounts—virtually disappeared from literature.”

Thomas Piketty, *Capital in the Twenty-First Century* (2014)

3. Macroscale analysis of linguistic trends using much larger corpora (8,000 novels) and advanced statistical analysis. From “Cents and Sensibility,” published in *Slate* magazine in December, 2014. (w/ Richard So and Ted Underwood)

What are the prospects for the study of Japanese literature?

Short Term: Use existing collections like Aozora so as to begin exploratory work. There are over 12,000 works on Aozora, and about 1,000 of these are novel length. Need to build smaller sub-corpora of pre-processed texts.

青空文庫

著作権の保護期間
延長に反対します

70

インターネットの電子図書館、青空文庫へようこそ。

初めての方はまず「青空文庫早わかり」をご覧ください。

ファイル利用をお考えの方は、こちらをご一読ください。

プラウザでは読みにくいと思った方は、「青空文庫のXHTML, TEXTの読み方」をどうぞ。

メイン エリア

青空文庫早わかり

青空文庫の使い方と約束事を紹介しています。初めての方、ファイルやキャプチャーぞ。

総合インデックス

作家名、作品名の50音別に、公開作品と入力・校正作業中の作品を一覧できるインデックスです。

公開中 作家別: あ行 か行 さ行 た行 な行 は行
ま行 や行 わ行 わ行 他

公開中 作品別: あ か さ し た な は
い き し す つ に ひ
う きく す せ こと ぬ み
え けく せ そ と の ね
お けく せ そ と の ね
は ひ く ほ

作業中: 作家別・作品別

青空文庫 分野別リスト

分野別に公開作品を一覧できる、インデックスです。

www.aozora.gr.jp

メイン お知らせ 別館 資料 運営
www.aozora.gr.jp 内を検索
Google bing Naver Baidu goo

概要

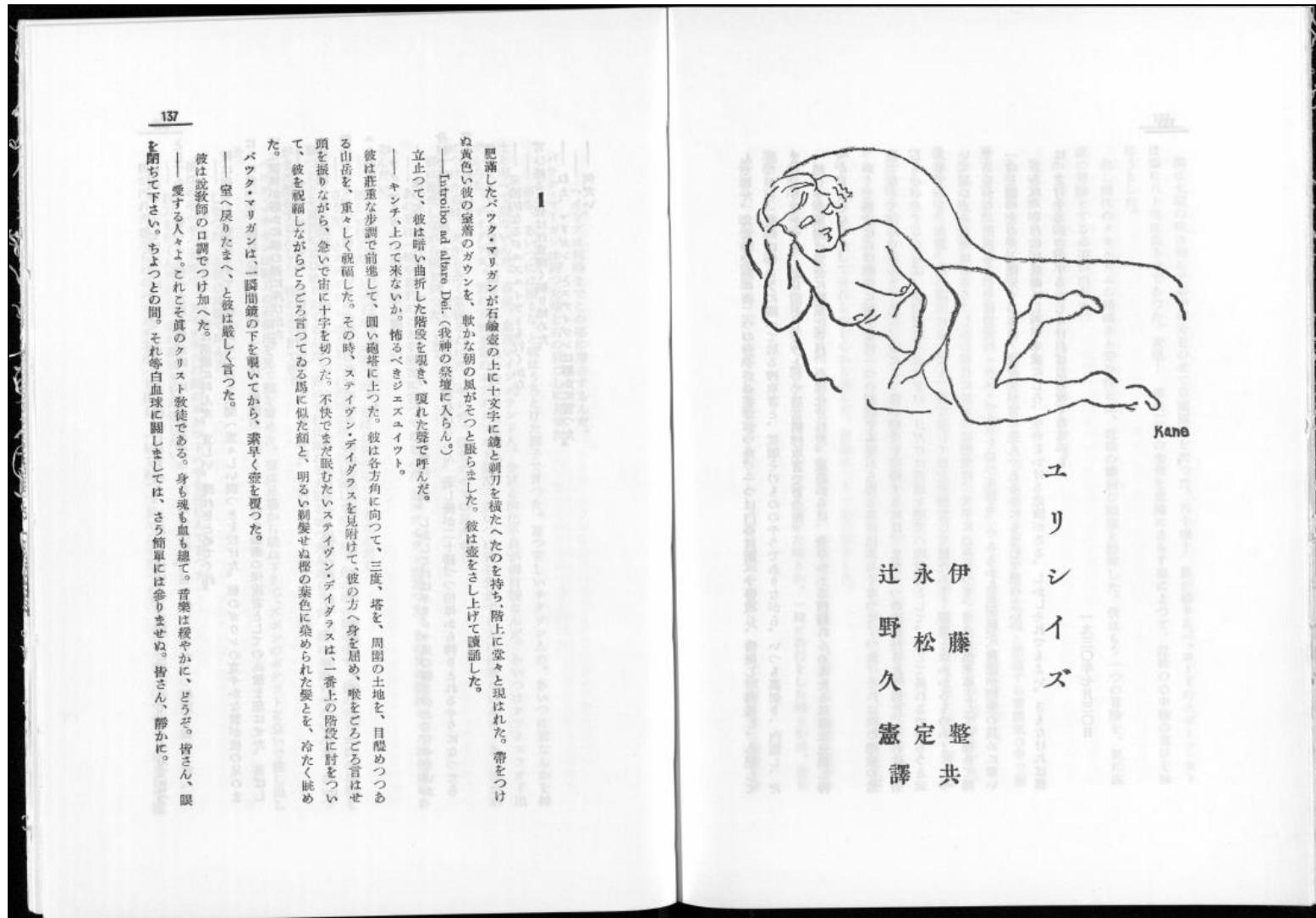
青空文庫の作品に対し、textsearch_jaで形態素解析を行ったCSVデータを無料で公開しています。CCライセンスの範囲内であれば、登録・報告作業の必要なしに、営利目的であってもご自由に本データをダウンロードしてご利用いただけます。

データ形式はCSV形式で下記のようなイメージになります。詳しくは [解析データ説明](#)をご覧ください。

【太宰 治 走れメロス 形態素解析データ ダウンロード】

row	num	word	type	subtype1	subtype2	subtype3	conttype	conjugation	basic	ruby	pronounce
1	1	メロス	名詞	一般					メロス	メロス	メロス
1	2	は	助詞	係助詞					は	ハ	ワ
1	3	激怒	名詞	サ変接続					激怒	ゲキド	ゲキド
1	4	し	動詞	自立			サ変・スル	連用形	する	シ	シ
1	5	た	助動詞				特殊・タ	基本形	た	タ	タ
1	6	。	記号	句点					。	。	。
2	1	必ず	副詞	一般					必ず	カナラズ	カナラズ
2	2	、	記号	読点					、	、	、
2	3	かの	連体詞						かの	カノ	カノ
2	4	邪智	名詞	一般					邪智	ジャチ	ジャチ
2	5	暴虐	名詞	一般					暴虐	ボウギャク	ボーギャク
2	6	の	助詞	連体化					の	ノ	ノ
2	7	王	名詞	一般					王	オウ	オー
2	8	を	助詞	格助詞	一般				を	ヲ	ヲ
2	9	除か	動詞	自立			五段・力行イ音便	未然形	除く	ノゾカ	ノゾカ
2	10	なけれ	助動詞				特殊・ナイ	仮定形	ない	ナケレ	ナケレ
2	11	ば	助詞	接続助詞					ば	バ	バ
2	12	なら	動詞	非自立			五段・ラ行	未然形	なる	ナラ	ナラ
2	13	ぬ	助動詞				特殊・ヌ	基本形	ぬ	ヌ	ヌ
2	14	と	助詞	格助詞	引用				と	ト	ト
2	15	決意	名詞	サ変接続					決意	ケツイ	ケツイ
2	16	し	動詞	自立			サ変・スル	連用形	する	シ	シ
2	17	た	助動詞				特殊・タ	基本形	た	タ	タ
2	18	。	記号	句点					。	。	。

Long Term: Build our own mini-collections geared to individual research interests or larger scale projects. This will involve creating a workflow to scan, OCR, and pre-process texts to make them available for the methods described previously. For a new project, I'm assembling a corpus of stream-of-consciousness inspired fiction from the early Showa period.



Translation of James Joyce's *Ulysses* (serialized in "Shi genjitsu" beginning in 1930)

肥満したバツク・マリガンが石鹼壺の上に十文字に鏡と剃刀を横たへたのを持ち、階上に堂々と現はれた。帯をつけぬ黃色い彼の室着のガウンを、軟かな朝の風がそつと脹らました。彼は壺をさし上げて讀誦した。

——Introibo ad altare Dei. (我神の祭壇に入らん。)

立止つて、彼は暗い曲折した階段を覗き、嗄れた聲で呼んだ。

——キンチ、上つて來ないか。怖るべきジエズユイツト。

彼は莊重な歩調で前進して、圓い砲塔に上つた。彼は各方角に向つて、三度、塔を、周圍の土地を、目醒めつつある山岳を、重々しく祝福した。その時、ステイヴン・ディダラスを見附けて、彼の方へ身を屈め、喉をごろごろ言はせ頭を振りながら、急いで宙に十字を切つた。不快でまだ眠むたいステイヴン・ディダラスは、一番上の階段に肘をついて、彼を祝福しながらごろごろ言つてゐる馬に似た顔と、明るい剃髪せぬ櫻の葉色に染められた髪とを、冷たく眺めた。

バツク・マリガンは、一瞬間鏡の下を覗いてから、素早く壺を覆つた。

——室へ戻りたまへ、と彼は嚴しく言つた。

彼は說教師の口調でつけ加へた。

——愛する人々よ。これこそ眞のクリスト教徒である。身も魂も血も總て。音樂は緩やかに、どうぞ。皆さん、眼を閉ぢて下さい。ちよつとの間。それ等白血球に關しましては、さう簡単には參りませぬ。皆さん、靜かに。

彼は斜め上方を眺めて呼び立てるやうな長い低い口笛を吹き、それから恍惚とした關心に囚はれて暫くやめてじつとしてゐた。彼のよく揃つた白い歯が處々に金の尖を光らした。黃金說教師。強い銳い口笛が二つ靜寂を破つて答へた

——おうい、ありがとう、と彼は活潑に叫んだ。宜しい。流れをとめてくれ。

彼は砲塔から飛び下りて、彼を眺めてゐる人間へ重々しい眼つきを投げた。彼のガウンの寛やかな襞が彼の足のまはりに集つた。脹れた蔭のある顔と不機嫌な橢圓形の頬とは、中世の藝術の保護者としての高僧を想はせた。喜ばしげな微笑が靜かに彼の唇に浮び出した。

——實に可笑しいね、と彼は快活に言つた。君の名は。古代ギリシヤ人ぢやないか。

ユリシーズ 1 肥満したバツク・マリガンが石鹼壺の上に十文字に鏡と剃刀を横たへたのを持ち、階上に堂々と現はれた。帶をつけぬ黃色い彼の室着のガウンを、軟かな朝の風がそつと脹らました。彼は壺をさし上げて讀誦した。 — Introibo ad altare Dei .(我神の祭壇に入らん。) 立止つて、彼は暗い曲折した階段を覗き、嗄れた聲で呼んだ。 — キンチ、上つて來ないか。怖るべきジエズユイツト。 彼は莊重な歩調で前進して、圓い砲塔に上つた。彼は各方角に向つて、三度、塔を、周圍の土地を、目醒めつつある山岳を、重々しく祝福した。その時、ステイヴン・ディダラスを見附けて、彼の方へ身を屈め、喉をごろごろ言はせ頭を振りながら、急いで宙に十字を切つた。不快でまだ眠むたいステイヴン・ディダラスは、一番上の階段に肘について、彼を祝福しながらごろごろ言つてゐる馬に似た顔と、明るい剃髪せぬ櫻の葉色に染められた髪とを、冷たく眺めた。 バツク・マリガンは、一瞬間鏡の下を覗いてから、素早く壺を覆つた。 — 室へ戻りたまへ、と彼は厳しく言つた。 彼は說教師の口調でつけ加へた。 — 愛する人々よ。これこそ眞のクリスト教徒である。身も魂も血も總て。音樂は緩やかに、どうぞ。皆さん、眼を閉ぢて下さい。ちよつとの間。それ等白血球に關しましては、さう簡単には參りませぬ。皆さん、靜かに。 彼は斜め上方を眺めて呼び立てるやうな長い低い口笛を吹き、それから恍惚とした關心に囚はれて暫くやめてじつとしてゐた。彼のよく揃つた白い歯が處々に金の尖を光らした。 黃金說教師。強い銳い口笛が二つ靜寂を破つて答へた — おうい、ありがとう、と彼は活潑に叫んだ。宜しい。流れをとめてくれ。 彼は砲塔から飛び下りて、彼を眺めてゐる人間へ重々しい眼つきを投げた。彼のガウンの寛やかな襞が彼の足のまはりに集つた。脹れた蔭のある顔と不機嫌な橢圓形の頬とは、中世の藝術の保護者としての高僧を想はせた。喜ばしげな微笑が靜かに彼の唇に浮び出た。 — 實に可笑しいね、と彼は快活に言つた。君の名は。古代ギリシヤ人ぢやないか。

Raw .txt of Ulysses processed with Python and MeCab to create tokenized text