博士論文

中近世スペイン語古文書の統計的年代推定・場所推定

Statistical Methods for Dating and Geolocation of Medieval and Modern Spanish Documents

川崎　義史

要旨

中近世スペイン語古文書の統計的年代推定・場所推定

Statistical Methods for Dating and Geolocation of Medieval and Modern Spanish Documents

博士論文

東京大学大学院　総合文化研究科　言語情報科学専攻

川崎　義史

本研究の目的は，中近世スペイン語古文書の作成年代と作成場所を言語的特徴に基づき統計的に推定する方法を開発することである。現存する文献史料に基づく歴史学や通時言語学の研究において，作成年代と作成場所が不明の文献がいつ・どこで作成されたかを推定（特定）すること，及び文書の真贋を判定することは最重要課題である。正確な歴史の記述には，信頼できる文献資料が不可欠だからである。本研究では，文献史料を文書（document），文書の作成年代の推定を年代推定（dating），文書の作成場所の推定を場所推定（geolocation）と呼ぶことにする。データセットとして中近世スペイン語古文書コーパス CODEA（Corpus de Documentos Españoles Anteriores a 1700）を用いた。

本論文の貢献は，以下の四点である。

一つ目は，作成年代と作成場所を同時に推定する方法の提案である。管見の限り，同時推定を提案した研究は存在しない。先行研究では，年代推定・場所推定は個別に行われている。年代推定の研究では言語の空間的変異を無視している。同様に，場所推定の研究では言語の時間的変異を無視している。しかし，同年代における空間的変異や同地域における時間的変異が存在するので，言語の変異を扱う際には，時間軸と空間軸を同時に考慮する必要がある。実験により，作成年代と作成場所の個別推定に比べ，同時推定の方が常に予測精度が高くなることを示した。ただし，トレードオフとして，同時推定では計算量が増加する。

二つ目は，時空間カーネル平滑化の応用である。カーネル平滑化とは，カーネル関数を用いて，ある関数からより滑らかな関数を推定する方法である。本研究では，素性の出現頻度に対して，カーネル平滑化を適用した。カーネル平滑化では，関数に関して線形性やS字カーブなど特殊な性質を仮定する必要がない。カーネル平滑化により，データセットに点在する欠損値補間と頑健な推定が可能になる。時間軸と空間軸の各々においてカーネル平滑化を行う先行研究は存在するが，両者を組み合わせた時空間カーネル平滑化を文書分類のタスクに応用した研究は，管見の限り，存在しない。ただし，上述の利点がある一方で，本研究の実験では，カーネル平滑化の年代推定・場所推定への効果は限定的だった。

三つ目は，言語に依存しない年代推定法・場所推定法の提案である。素性として文字 *n*-gram を用いることで，単語毎に分かち書きされない言語（日本語や中国語など）の文書や，正書法が確立されていない時代の文書もそのまま扱うことができる。素性として単語 *n*-gram を用いる場合は，単語分割やステミングなどの技術開発が必要となる。また文字 *n*-gram は，単語 *n*-gram に比べ，素性数を大幅に削減できるという利点がある。

四つ目は，文献学的特徴に基づく計量的な年代推定法・場所推定法の提案である。スペイン語で書かれた文書の年代推定・場所推定は，スペイン語文献学の大きな目標の一つである。今日まで，記述的研究には大きな蓄積があるが，年代推定・場所推定を正面から扱った研究は存在しない。先行研究により，各年代や地点に特有の言語的特徴は，ある

程度，判明している。しかし，どれだけ細かな記述をしても，記述は記述に過ぎない。スペイン語史の記述から年代推定・場所推定という予測に移るには，計量的なアプローチが必要となる。重要性の異なる複数の証拠から総合的に判断するには，専門家の「勘」よりも，計量的手法の方が信頼性・実証性に勝るからである。「塵も積もれば山となる」というように，小さな証拠でも複数集まれば，大きな差異を生むことになる。本研究では，各々の証拠，つまり文献学的特徴の重みをデータから決定した。多くの文書に現れる特徴ほど，大きな重みが与えられる。この重みをプロットすることで，各特徴の年代推移・地理的変異を可視化することができる。これは，年代推定・場所推定の副産物として，スペイン語文献学への大きな貢献となる。

本論文の構成は以下の通りである。第 1 章では，研究背景の説明と問題定義を行った。第 2 章では，年代推定・場所推定に関連する先行研究を紹介した。第 3 章では，本研究で用いるコーパスの概要と記述的統計を示した。第 4 章では，素性として用いる文字 *n*-gram と文献学的特徴について説明した。第 5 章では，欠損値補間と頑健な推定を可能にするカーネル平滑化について説明した。第 6 章では *n*-gram 言語モデルによる年代推定法・場所推定法を，第 7 章では JS 情報量に基づく年代推定法・場所推定法を，第 8 章ではナイーブベイズ多変数ベルヌーイモデルによる年代推定法・場所推定法を説明した。第 9 章では，年代推定・場所推定の実験結果を示した。第 10 章で，結論と今後の課題を述べた。

キーワード :
中近世スペイン語古文書，CODEA，年代推定，場所推定，文字 *n*-gram，文献学的特徴，カーネル平滑化，*n*-gram 言語モデル，JS 情報量，ナイーブベイズ多変数ベルヌーイモデル

謝辞

本論文執筆にあたり，学部時代から温かいご指導を頂いた上田博人先生に心より感謝いたします。本論文のテーマである中近世スペイン語古文書の年代推定・場所推定という問題を知ったのは，ある国際学会の会場へ向かうバスの中での上田先生との会話でした。先行研究がほとんどなく，スペイン語文献学への大きな貢献となる研究テーマに大きな興味を持ちました。この研究テーマに取り組み始めてからは，上田先生の科研費の分担研究者として，スペインで開催された研究集会と学会に参加させていただきました。これらの機会を通じて，スペインでの研究の潮流を知り，海外の研究者と交流を持つことができました。

スペインのアルカラ大学の Pedro Sánchez-Prieto Borja 先生は，中近世スペイン語古文書コーパス CODEA（Corpus de Documentos Españoles Anteriores a 1700）のファイルを，本研究のために特別に提供してくださいました。この貴重なデータセットがなければ, 本論文を執筆することは不可能でした。このコーパスを作成したアルカラ大学の文献学研究グループ GITHE（Grupo de Investigación de Textos para la Historia del Español）の皆様にも感謝いたします。

言語情報科学専攻の田中伸一先生と寺澤盾先生，また地域文化研究専攻の和田毅先生には，博士論文の審査委員をお引き受けいただき，多くの示唆やコメントを頂きました。言語情報科学専攻事務室の国府田さつきさんと當間順子さんには，博士論文や科研費の事務手続きでお世話になりました。

上智大学のアントニオ・ルイス・ティノコ先生には，日本学術振興会特別研究員 PD として受けれていただき，快適な研究環境を提供していただきました。また，博士論文の審査委員をお引き受けいただき，多くの示唆やコメントを頂きました。上智大学研究支援センターの方々には，科研費関連の事務手続きでお世話になりました。

東京工業大学の高村大也先生と甲南大学の永田亮先生には，自然言語処理や機械学習に関する疑問点について丁寧に教えていただきました。

最後に，大学入学時から経済的な支援を惜しまず自由に勉強させてくれた家族と，温かい励ましをいつも送り続けてくれたパートナーの晴子さんに心から感謝いたします。

本研究は東京大学博士課程研究遂行協力制度（平成 23 年度，平成 24 年度）と JSPS 科研費 13J03408，15J04335 の助成を受けたものです。

目次

図目次

表目次

第1章 序論[1]

現存する文献史料に基づく歴史学や通時言語学の研究において，作成年代や作成場所が不明の文献がいつどこで作成されたかを推定（特定）すること，及び文書の真贋を判定することは最重要課題である。正確な歴史の記述や文献解釈には，信頼できる文献資料が不可欠だからである（Malkiel 1968；五神 2016）。本研究では，文献史料を文書（document），文書の作成年代の推定を年代推定（dating），文書の作成場所の推定を場所推定（geolocation）と呼ぶことにする。

年代推定法・場所推定法を開発するためには，当然のことながら，作成年代と作成場所が信頼できる文書が大量に必要となる。作成年代と作成場所が不正確な文書や少数の文書に基づく年代推定法・場所推定法は信頼性に欠けるからである。本研究では，作成年代と作成場所が信頼でき，ある程度の文書数を確保できるデータセットとして「中近世スペイン語古文書コーパス CODEA（Corpus de Documentos Españoles Anteriores a 1700）」を用いる。CODEA は，スペインのアルカラ大学の Sánchez-Prieto Borja 教授を中心とした文献学研究グループ GITHE（Grupo de Investigación de Textos para la Historia del Español）が作成しているもので，西暦 1100 年から 1700 年の間にスペイン各地で作成された 1500 の古文書（こもんじょ）からなる（Sánchez-Prieto Borja 2012；Sánchez-Prieto Borja *et al.* 2012；Díaz Moreno *et al.* 2015）。一部の古文書は中世ラテン語で書かれている。古文書とは，法令，権利，財産譲渡などの事項を明記した近代以前の手書きの行政・法律関係書類の総称である（Riesco Terrero 2004）。古文書には，一般的に作成日時，場所，差出人，宛先，作成主などが記載されている。そして現存する古文書の大多数は原本もしくはその控え（写し）であるので，古文書の作成年代や作成場所は信頼できる情報である[2]。

本研究の目的は，CODEA を用いて，中近世スペイン語古文書の作成年代と作成場所を言語的特徴に基づき統計的に推定する方法を開発することである。筆者の最終的目標は，古文書から抽出した言語的特徴に基づいて，古文書以外の文献（たとえば文学作品の写本）の年代推定や場所推定も可能にすることである[3]。そのために，まずは作成年代と作成場所が信頼できる古文書を用いて年代推定・場所推定の予測精度を調べる必要がある。本研究では，1100 年から 1700 年の間のスペイン語と中世ラテン語を，便宜的に「中近世スペイン語（español medieval y moderno）」と呼ぶことにする。ここで，スペイン語とは，主に今日のスペインを構成する領域で使用されている（いた）ロマンス語の総称である。また，言語的特徴とは，文字や単語などの出現頻度のことを指す。

言語学的に重要ではあるが，本研究では，以下の二つの問題は扱わない。

一つ目は，言語変化の原因や変異形の伝播の分析である（Lass 1980；Romaine 1982；Tuten 2003；Conde Silvestre 2007；Bybee 2015）。確かに，言語変化の問題（Coseriu 1978）は，言語学の本質的な問いの一つである。しかし，年代推定・場所推定のためには，言語変化の問題は考慮する必要がない。言語変化の結果が同じであれば，変化の原因やプロセスは

---

[1] 本論文は，川崎（2015a，2016）に基づいている。

[2] Sánchez-Prieto Borja 教授を中心とする研究グループ CHARTA（Corpus Hispánico y Americano en la Red：Textos Antiguos）は，作成年代や作成場所が信頼できる古文書コーパスを用いて実証的なスペイン語史を構築することを目指している。従来，スペイン語史の研究は，歴史的・文化的重要性の高いとされる文学作品の写本に基づき行われてきた。しかし，文学作品の写本には作成年代や作成場所の記載はなく，後世の写しという形でのみ現存している。したがって，作成年代や作成場所が曖昧な文学作品の写本に基づくスペイン語史は必ずしも信頼性に足るものではない。作成年代や作成場所が信頼できる古文書を用いることで，この問題に対処することができる。

[3] ここで，文学作品の年代推定・場所推定とは，現存する写本の作成年代・作成場所の推定であり，写本作成に先行する作品自体の成立年代・場所ではないことに注意されたい。作品が成立時のまま現存しているような場合を除き，作品の成立年代・場所の決定は，少なくとも言語学的見地からは，決着を見ない問題である。

本研究の目的とする年代推定・場所推定の結果には全く影響しないからである

二つ目は，言語の年代区分や方言区分を行うことである。通時言語学では，いつからいつまでを中世スペイン語とするかということが議論されることがある（Eberenz 1991，2009；Sánchez Lancis 2009；Granvik & Sánchez Lancis 2015）。同様に, 地理方言学では, どこに等語線を引くか, どのようにして方言区分を行うかというのが大きな問題となる（Pennny 2000；Ueda 2013c；Grieve 2016）。確かに，言語的連続体をいくつかの部分に分割する年代区分や方言区分というタスクは，言語変異を大まかに理解するための便宜的・教育的な意義がある。しかし，本研究の目的とする年代推定・場所推定の結果は，どのように年代区分や方言区分がなされたかには全く左右されない。したがって，言語の年代区分や方言区分は扱わない。

1.1　問題定義[4]

本研究は，第 2 章で紹介する先行研究に基づき，古文書の年代推定・場所推定を文書分類（Text Categorization，Text Classification）の問題として考える。以下，古文書を文書（ぶんしょ）と呼ぶ。文書分類とは，文書を一定の基準によって既存のクラス（カテゴリー）に仕分けることである（Sebastiani 2002）。クラスとしては，文書の作成者，内容（たとえば，スパムメール判定やニュース記事の分類），作成された時，作成された場所などが考えられる。本研究において，クラスは文書の作成年代と作成場所であり，どちらも離散変数とみなす。

文書分類は以下のように行われる。データセットをD，データセットDに含まれる文書数を$|D|$，データセットD内のj番目の文書を$d^{(j)} \in D$，文書$d^{(j)}$の属するクラスを$c^{(j)}$とする。このとき，データセットDは，文書$d^{(j)}$とその文書の属するクラス$c^{(j)}$のペア$\left(d^{(j)}, c^{(j)}\right)$の集合として表すことができる：

$$D = \left\{\left(d^{(1)}, c^{(1)}\right), \dots, \left(d^{(j)}, c^{(j)}\right), \dots, \left(d^{(|D|)}, c^{(|D|)}\right)\right\} \tag{1.1}$$

ここで，データセットD内の互いに異なるクラスの集合を$C = \{c_1, \dots, c_i, .., c_{|C|}\}$，総クラス数を$|C|(\leq |D|)$，クラス$c_i$に属するすべての文書を合体させ新たな一つの文書とみなした擬文書（pseudo-document）をd_iとする：

$$d_i = \bigcup_{d^{(j)} \in c_i} d^{(j)} \tag{1.2}$$

擬文書$d_i \in c_i$からなるデータセットをD'とする。D'は，訓練データと呼ばれる：

$$D' = \{d_1, \dots, d_i, \dots, d_{|C|}\} \tag{1.3}$$

文書分類を行う文書を$q \notin D$，文書qが属する真のクラスをc_q，文書qが属すると推定されるクラスを$\widehat{c_q}$，文書qとクラス$c \in C$の擬文書との類似度を表す関数を$\phi(q, c)$とする。関数$\phi(q, c)$を計算するには，すべての文書を数学的に表現する必要がある。本研究では，第 4 章で説明する文字 *n*-gram と文献学的特徴により文書を数学的に表現する。そして，文書qを，$\phi(q, c)$の値が最大となるクラスcに分類することにする：

$$\widehat{c_q} = \arg\max_{c \in C} \phi(q, c) \tag{1.4}$$

[4] 本研究で用いる数学記号や初歩的事項については，付録 A を参照。

式（1.4）は，$\phi(q,c)$の値が最大（max）となるクラスcに文書qを分類することを意味している。arg max は argument of the maximum の意味である。たとえば，$C=\{c_1,c_2,c_3\}$，$\phi(q,c_1)=1$，$\phi(q,c_2)=3$，$\phi(q,c_3)=5$のとき，$\max_{c\in C}\phi(q,c)=5$となり，文書qはクラスc_3に分類される。

本研究において，クラスは文書の作成年代と作成場所である。粒度を 1 年とした年代の集合を$T=\{1100,1101,\ldots,1700\}$とする（$|T|=601$）[5]。また，地点の集合を$L$とする。地点の粒度は，現代スペインの行政単位である県（provincia）と自治州（comunidad autónoma）の二つのレベルを設定した[6]。県レベルでは$|L|=52$，自治州レベルでは$|L|=19$となる（3.2 節を参照）。クラスの粒度を小さくするほど，データのスパースネスが大きくなる。これに対処するため，第 5 章で説明するカーネル平滑化を行う。

推定は，作成年代と作成場所を個別に推定する個別推定と，両者を同時に推定する同時推定の二つの方法で行う。後者の方法は，管見の限り，本研究が初めて提案する方法である。年代推定・場所推定を行う文書qの推定年代を$\widehat{t_q}\in T$，推定場所を$\widehat{l_q}\in L$とする。個別推定では，$|T|$個の年代から推定年代$\widehat{t_q}$を，$|L|$個の地点から推定場所$\widehat{l_q}$を独立に求める：

$$\begin{aligned}\widehat{t_q} &= \arg\max_{t\in T}\phi(q,t)\\ \widehat{l_q} &= \arg\max_{l\in L}\phi(q,l)\end{aligned} \tag{1.5}$$

ここで，$\phi(q,t)$は文書qと年代$t\in T$との類似度を，$\phi(q,l)$は文書qと地点$l\in L$との類似度を表す関数である。

一方，同時推定では，$|T|\times|L|$個のクラスから，作成年代と作成場所の組み合わせを同時に推定する。したがって，個別推定に比べ，計算量が増加する。地点の粒度が県のときクラス数は$|T|\times|L|=601\times52=31252$，粒度が自治州のときクラス数は$|T|\times|L|=601\times19=11419$となる。年代$t$と地点$l$の組み合わせを$(t,l)$，$(t,l)$の集合を$(T,L)$とする。このとき文書$q$の推定年代$\widehat{t_q}$と推定場所$l_q$の組み合わせ$(\hat{t}_q,\hat{l}_q)$は，$\phi(q,t,l)$が最大となるクラスになる：

$$(\hat{t}_q,\hat{l}_q)=\arg\max_{(t,l)\in(T,L)}\phi(q,t,l) \tag{1.6}$$

ここで，$\phi(q,t,l)$は，文書qと，年代tと地点lの組み合わせ(t,l)の類似度を表す関数である。同時推定のメリットは，同年代における空間的変異や同地域における時間的変異を考慮することができる点である。これにより，個別推定に比べ，推定精度が向上すると期待される。しかし，そのトレードオフで，計算量が大幅に増加する。個別推定ではクラス数が$|T|+|L|$であるのに対し，同時推定では$|T|\times|L|$個のクラスが存在するためである。

本研究では，関数$\phi(q,c)$として，*n*-gram 言語モデル（第 6 章），JS 情報量（第 7 章），ナイーブベイズ多変数ベルヌーイモデル（第 8 章）を用いる。関数$\phi(q,c)$は，*n*-gram 言語モデルではクラスcに文書qが属する事後確率，JS 情報量では文書qとクラスcの JS 情報量，ナイーブベイズ多変数ベルヌーイモデルではクラスcに文書qが属する事後確率を

[5] 1 年以下の粒度（月，週，日）での推定は，歴史的文書では非現実的である。逆に，粒度を 10 年，20 年，100 年のような期間とする場合，期間の分割開始時点の選択が恣意的になってしまう。たとえば，1 年を 3 分割するとき，1 月～4 月，5 月～8 月，9 月～12 月のように分割するべきか，4 月～7 月，8 月～11 月，12 月～3 月のように分割するべきかはタスク依存となる。また，期間を大きく取ると，期間の最初と最後で大きな年代差が生じてしまう。

[6] 県や自治州という行政単位は近代以降の制度であり，今日の県都や州都は，必ずしも昔の政治的・経済的中心地と一致しない。しかし本研究では，データ収集の容易さから，現代の行政単位を採用した。また，現在のコーパスの規模では，県より小さい市町村（municipio）の粒度での分析は非現実的である。

表している。いずれの方法も，真のクラスと分類されたクラスの誤差が最小になるように，関数$\phi(q, c)$のパラメータを調整する必要がある。誤差の評価指標については，第 9 章で説明する。

年代推定・場所推定を行うにあたり，以下の四つの仮定をしている。

一つ目は，作成年代や作成場所が同じ文書には類似した特徴が，また作成年代や作成場所が異なる文書には異なる特徴が存在するという仮定である。もし，このような特徴がないならば，文書の作成年代や作成場所を特定することは原理的に不可能である。

二つ目は，文書には後世の加筆・修正はなく，また，作成場所の言語的特徴を持つ写字生により作成されたと仮定している。この仮定により，文書の作成年代と作成場所を一意に決まる。

三つ目は，文書の内容，作成者，発行主などが異なっていても，作成年代や作成場所が同じ文書群には共通の言語的特徴があるという仮定である。この仮定は現実には正しくないが，現時点での文書数では，これらの属性をコントロールした上で実験を行うことは困難である。作成者が異なると言語的特徴が異なることもあるかもしれないが，だからといって同一と認められないほど異なるならば，年代推定・場所推定のように，ある程度の抽象化を伴う研究はそもそも不可能になる。また，文書の作成者が複数の場合には，共同作業に起因するシナジー効果（Kestemont *et al.* 2015）も考えられるが，問題が複雑になるのを防ぐだめに作成者は一人だと仮定する。

四つ目は，作成年代と作成場所は事前に設定したクラスのいずれかに属するという仮定である。文書分類においては，既定のクラス以外のクラスに分類することはできない。したがって，1100 年以前もしくは 1700 年以後に作成された文書や，地点集合Lに属さない地点で作成された文書の作成年代や作成場所は正しく推定することはできない。

図 1.1 に年代推定・場所推定の流れを示す。

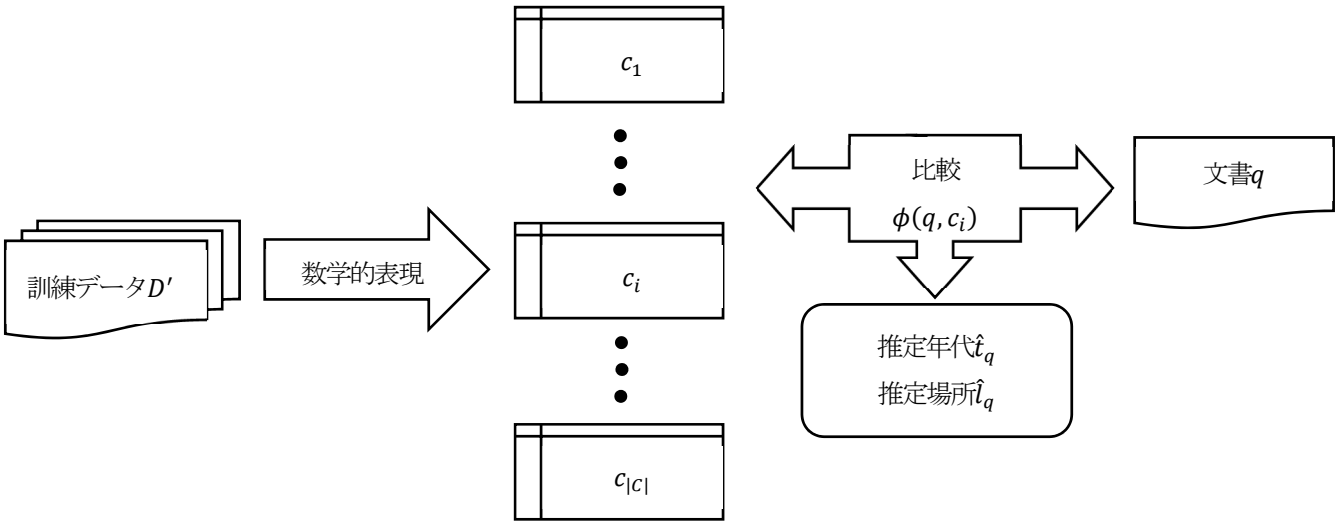


図 1.1　年代推定・場所推定の流れ

1.2　論文の貢献

本論文の貢献は，以下の四点である。

一つ目は，作成年代と作成場所を同時に推定する方法の提案である。管見の限り，同時推定を提案した研究は存在しない。先行研究では，年代推定・場所推定は個別に行われている。年代推定の研究では言語の空間的変異を無視している。同様に，場所推定の研究では言語の時間的変異を無視している。しかし，同年代における空間的変異や同地域における時間的変異が存在するので，言語の変異を扱う際には，時間軸と空間軸を同時に考慮する必要がある。実験によ

り，作成年代と作成場所の個別推定に比べ，同時推定の方が常に予測精度が高くなることを示した。ただし，トレードオフとして，同時推定では計算量が増加する。

二つ目は，時空間カーネル平滑化の応用である。カーネル平滑化とは，カーネル関数を用いて，ある関数からより滑らかな関数を推定する方法である。本研究では，素性の出現頻度に対して，カーネル平滑化を適用した。カーネル平滑化では，関数に関して線形性やS字カーブなど特殊な性質を仮定する必要がない。カーネル平滑化により，データセットに点在する欠損値補間と頑健な推定が可能になる。時間軸と空間軸の各々においてカーネル平滑化を行う先行研究は存在するが，両者を組み合わせた時空間カーネル平滑化を文書分類のタスクに応用した研究は，管見の限り，存在しない。ただし，上述の利点がある一方で，本研究の実験では，カーネル平滑化の年代推定・場所推定への効果は限定的だった。

三つ目は，言語に依存しない年代推定法・場所推定法の提案である。素性として文字 *n*-gram を用いることで，単語毎に分かち書きされない言語（日本語や中国語など）の文書や，正書法が確立されていない時代の文書もそのまま扱うことができる。素性として単語 *n*-gram を用いる場合は単語分割やステミングなどの技術開発が必要となる。また文字 *n*-gram は，単語 *n*-gram に比べ，素性数を大幅に削減できるという利点がある。

四つ目は，文献学的特徴に基づく計量的な年代推定法・場所推定法の提案である。スペイン語で書かれた文書の年代推定・場所推定は，スペイン語文献学の大きな目標の一つである。今日まで，記述的研究には大きな蓄積があるが，年代推定・場所推定を正面から扱った研究は存在しない。先行研究により，各年代や地点に特有の言語的特徴は，ある程度，判明している。たとえば，ラテン語の音連続-L-+[j]が，レオン地方では/ʎ/，アラゴン地方では/ʎ/，カスティーリャ地方では/ʒ/に対応することは，周知の事実である（Zamora Vicente 1967 : 146，244-245 ; Menéndez Pidal 1999 : § 53$_6$; Penny 2002 : 2.5.2.2）。同様に，16 世紀以降，譲歩を表す接続詞として maguer に代わり aunque が優勢になるということも広く知られている（Penny 2002 : 3.8.2）。確かに，現象の記述は重要である。しかし，どれだけ細かな記述をしても，記述は記述に過ぎない。スペイン語史の記述から年代推定・場所推定という予測に移るには，計量的なアプローチ（エイデン＆ミシェル 2016）が必要となる。重要性の異なる複数の証拠から総合的に判断するには，専門家の「勘」よりも，計量的手法の方が信頼性・実証性に勝るからである。「塵も積もれば山となる」というように，小さな証拠でも複数集まれば，大きな差異を生むことになる。本研究では，各々の証拠，つまり文献学的特徴の重みをデータから決定した。多くの文書に現れる特徴ほど，大きな重みが与えられる。この重みをプロットすることで，各特徴の年代推移・地理的変異を可視化することができる。これは，年代推定・場所推定の副産物として，スペイン語文献学への大きな貢献となる。

1.3　論文の構成

本論文の構成は以下の通りである。第 1 章では，研究背景の説明と問題定義を行った。第 2 章では，年代推定・場所推定に関連する先行研究を紹介した。第 3 章では，本研究で用いるコーパスの概要と記述的統計を示した。第 4 章では，素性として用いる文字 *n*-gram と文献学的特徴について説明した。第 5 章では，欠損値補間と頑健な推定を可能にするカーネル平滑化について説明した。第 6 章では *n*-gram 言語モデルによる年代推定法・場所推定法を，第 7 章では JS 情報量に基づく年代推定法・場所推定法を，第 8 章ではナイーブベイズ多変数ベルヌーイモデルによる年代推定法・場所推定法を説明した。第 9 章では，年代推定・場所推定の実験結果を示した。第 10 章で，結論と今後の課題を述べた。

第2章 関連研究

年代推定法・場所推定法は，①言語的特徴に基づく統計的手法，②言語的特徴に基づく非統計的手法，③非言語的特徴に基づく統計的手法，④非言語的特徴に基づく非統計的手法の 4 つに分類できる（表 2.1）。ここで，言語的特徴とは文字や単語などの出現頻度を，非言語的特徴はそれ以外の特徴を指す。また，統計的手法とは機械学習（machine learning）やパターン認識（pattern recognition）に基づく方法（Bishop 2006）を，非統計的手法とは専門家の設けた規則（論理式）に基づく方法（rule-based method）を指す。

	統計的手法	非統計的手法
言語的特徴	● 2.2 節，2.3 節，2.4 節で挙げた研究 ● 本研究	● 語彙（Chibnal 2000；Gazeau 2000；Tock 2000；Vincent 2000） ● 形態統語論（Azofra 2009：201-204）
非言語的特徴	● 放射性炭素年代測定（Oda *et al.* 2004，2007，2010a，2010b，2011；小田 2011） ● イメージ分析（He *et al.* 2014；Faigenbaum-Golovin *et al.* 2016）	● 人名（Chalmers 2000；Hillebrandt 2000） ● 歴史的事実（Cortina Gómez 1977） ● パレオグラフィー（Torrens Álvarez 1995；Veszprémy 2000） ● 印章（Harvey 2000）

表 2.1　年代推定法・場所推定法の分類

本研究は，①言語的特徴に基づく統計的手法に該当する。関連研究は，2.2 節，2.3 節，2.4 節で紹介する。

②言語的特徴に基づく非統計的手法としては，語彙に注目した研究（Chibnal 2000；Gazeau 2000；Tock 2000；Vincent 2000）や形態統語論に注目した研究（Azofra 2009：201-204）が挙げられる。

③非言語的特徴に基づく統計的手法としては，放射性炭素年代測定（Oda *et al.* 2004，2007，2010a，2010b，2011；小田 2011）や，イメージ分析（document image analysis）による方法（He *et al.* 2014；Faigenbaum-Golovin *et al.* 2016）が挙げられる。放射性炭素年代測定とは，文書の書かれている紙や皮紙に含まれる炭素 14 の存在比率に基づいて作成年代を特定する方法である。イメージ分析では，文書を OCR（Optical Character Recognition）で読み込み，そこで使用されている文字の形態や種類の比率から推定年代を計算する。

④非言語的特徴に基づく非統計的手法としては，人名に注目した研究（Hillebrandt 2000；Chalmers 2000），歴史的事実に注目した研究（Cortina Gómez 1977），パレオグラフィーに注目した研究（Torrens Álvarez 1995；Veszprémy 2000），印章に注目した研究（Harvey 2000）が挙げられる。どの手法にも一長一短があるため，様々な分野の研究者が協力して年代推定・場所推定を行うことが望ましい。

本研究が言語的特徴に注目する理由は，次の二点である。まず，スペイン語史の研究者として，言語的特徴に基づく年代推定法・場所推定法は興味深い課題である。第二に，大量の電子化された文書が利用可能になりつつある今日，定量化が容易な言語的特徴に注目するのは当然の流れといえる。人名，パレオグラフィー，印章などの非言語的特徴は定量化が難しく，一般的に，推定結果の範囲が広くなってしまう。適切な範囲に絞りこまれていない推定結果には，あ

まり意味がない。

本研究が統計的手法を用いる理由は，次の三点である。第一に，実証的である点である。非統計的手法では，最終的判断が研究者の「勘」に左右され，推定結果が主観的になりやすい。第二に，統計的手法により，言語に依存しない一般的な方法を構築できる点である。本研究は中近世スペイン語を対象としてるが，提案する年代推定法・場所推定法は言語に依存しない方法である。第三に，多数の特徴の組み合わせから有用な規則（論理式）を発見するのは，人間の頭にとって大変な作業である。

次節からは，言語的特徴に基づいて文書の属性を統計的に推定する関連研究を紹介する。まず，2.1 節では，著者推定の研究を紹介する。著者推定は本研究と直接関係はしないが，その方法論から学ぶところは多い。2.2 節では，統計的年代推定の研究を，2.3 節では統計的場所推定の研究を，2.4 節では本研究で使用する CODEA を用いた年代推定・場所推定の研究を紹介する。

2.1　著者推定

著者推定（Authorship attribution）は，著者不明の文書の著者を推定するための数理的・計量的方法を研究する分野である（Juola 2006；Koppel *et al.* 2009；Stamatatos 2009；村上 1994，2002，2004）。分析の対象は，文学作品・歴史的文書（Wake 1957；□llegård 1962；Brinegar 1963；Morton 1965；Mosteller & Wallace 1963；Hope 1994；Love 2002；Zhao & Zobel 2007；Jockers *et al.* 2008；Reynolds *et al.* 2012；Kestemont *et al.* 2015）や，犯罪関連の文書・電子メール（Li *et al.* 2006；Zheng *et al.* 2006；Abbasi & Hsinchun 2008；Iqbal *et al.* 2008；村上 2004：114-124；財津 & 金 2015）などである。著者自身のほか，著者の性別や年齢等の属性が推定の対象となることもある（Rao *et al.* 2010；Burger *et al.* 2011；Koppel *et al.* 2002；Peersman *et al.* 2011；Rybicki 2015）。

著者推定では，まず，出現頻度が高く各著者の文体的差異を反映すると考えられる特徴を抽出する。特徴としては，単語長（Mendenhall 1887），文長（Yule 1939），パラグラフ長（Li *et al.* 2006），語彙の豊富さ（Hoover 2003a），単語 *n*-gram・文字 *n*-gram・品詞 *n*-gram の頻度，機能語（前置詞，後置詞，接続詞，冠詞など）の頻度，句読点の頻度（Jin & Murakami 1993；Jin & Jiang 2012），文節パターン（金 2013）などが用いられる。これらの特徴は，文書の内容にはあまり依存せず，しかも無意識的に表出し，模倣するのは難しいものと考えられている（村上 1994：145）。次に，これらの特徴に基づき，著者不明の文書と各著者の文書集合との近さを測る。分類には，主成分分析（Binongo & Smith 1999；Binongo 2003；Kestemont *et al.* 2015）やクラスタリング（Hoover 2003）等の多変量解析，Delta（Burrows 2002；Evert *et al.* 2015），KL 情報量（Zhao & Zobel 2007），サポートベクターマシン（Diederich 2003），ナイーブベイズ（Clement & Sharp 2003），*k*-近隣法，ランダムフォレスト法（金・村上 2007）等の手法が用いられる。

2.2　統計的年代推定

統計的手法を用いて年代推定を行った研究を紹介する[7]。De Jong *et al.* (2005)，Tilahun（2011），Tilahun *et al.*（2012），Kumar *et al.*（2011），Kumar *et al.*（2012），Kumar（2013）は年代を離散変数とみなし，年代推定を文書分類の枠組みで扱った。一方，金（2009）と Bruster & Smith（2014）は年代を連続変数とみなし，年代推定を回帰の枠組みで扱った。

[7] 以下の数学的表記は，必ずしも原論文と同一ではない。

2.2.1　De Jong *et al.* (2005)

De Jong *et al.*（2005）は，年代推定の草分けであり，Ponte & Croft（1998）と Li & Croft（2003）に基づいた時間言語モデル（temporal language model）を提案した[8]。このモデルでは，

$$NLLR(X|Y) = \sum_{w \in X} P(w|X) \log \frac{P(w|Y)}{P(w|Z)} \tag{2.1}$$

で定義される正規化対数尤度比$NLLR$（Normalized Log-Likelihood Ratio ; Kraaij 2004）を用いて年代推定を行う。ここで，wは文書Xに含まれる単語，Zはデータセット全体の言語モデルである。正規化対数尤度比$NLLR(X|Y)$は，文書Xのユニグラム言語モデルとクラスYのユニグラム言語モデルの類似度を測る尺度である。ユニグラム言語モデルとは，単語の頻度で文書を表現したモデルである（付録 A の A10 を参照）。正規化対数尤度比が大きいほど，二つの言語モデルは類似しているとみなされる。少なくとも一つの$w \in X$に関して$P(w|Y) = 0$となる場合，$\log \frac{P(w|Y)}{P(w|Z)} = \log \frac{0}{P(w|Z)} = -\infty$となり正規化対数尤比が定義できなくなってしまう。この問題を回避するために，クラスYの言語モデルにスムージング（Zhai & Lafferty 2004）を行い，すべての$w \in X$に関して$P(w|Y) > 0$となるようにする（スムージングについては，6.3 節を参照）。$P(w|Z)$は，データセット全体の言語モデルなので常に$P(w|Z) > 0$となる。

年代推定には，事例ベースとクラスベースの二つの方法がある。事例ベースの推定では，年代推定を行う文書Xの言語モデルと訓練データ内の各文書dの言語モデルの比較を行う。年代の集合をT，正規化対数尤度比の値が最も大きいk個の文書の集合を$S_k \subseteq Z$，S_kに属する文書で年代$t \in T$に作成された文書を$d_t \in S_k$とすると，文書Xの推定年代$\hat{t_X}$は，文書Xと文書$d_t \in S_k$の正規化対数尤度比の合計が最大となる年代tとなる：

$$\hat{t_X} = \arg\max_{t \in T} \sum_{d_t \in S_k} NLLR(X|d_t) \tag{2.2}$$

一方，クラスベースの推定では，文書Xの言語モデルと年代tの言語モデルの比較を行う。年代tの言語モデルは，年代tに作成されたすべての文書dから構築される。このとき，文書Xの推定年代$\hat{t_X}$は，正規化対数尤度比が最大となる年代tとなる：

$$\hat{t_X} = \arg\max_{t \in T} NLLR(X|t) \tag{2.3}$$

実験は，オランダ語新聞の記事を用いて行われた。訓練データは 1999 年から 2005 年の間に書かれた二つのオランダ語新聞の記事（約 2GB），テストデータは別の 3 紙からランダムに選ばれた同一期間の 500 文書である。低頻度語の削除とステミングにより，語彙サイズは約 17 万語となった。クラスベースの場合，複数の粒度（3 か月，1 か月，1 週，

[8] この他の時間言語モデルの研究としては，以下の研究が挙げられる。Kanhabua & Nørvåg（2008, 2009, 2010）は，正規化対数尤度比モデルの拡張を行った。Chambers（2012）は，正規化対数尤度比モデルよりも対数線形モデルの方が高い精度を得ることを示した。Kotsakos *et al.*（2014）は，単語のバースト性に着目し対数線形モデルよりも実行時間が短く精度が高いモデルを提案した。Dalli & Wilks（2006）は，単語の出現頻度の周期性に着目し年代推定を行った。

2 日）によりクラス分割を行う。

実験の結果，事例ベースでは$k=1$の時に推定精度が最も高く，約 55%の文書が正しく年代推定された。クラスベースでは，粒度が小さいほど推定精度が高くなり，粒度が 2 日の場合，約 30%の文書が正しく年代推定された。事例ベースの推定の優位性として二つの要因が挙げられている。一つ目は，新聞記事という性質上，類似した内容の記事が複数存在することである。二つ目は，単語分布が異なる複数のトピック（政治，経済，スポーツ等）の記事が同一年代に存在しているため，クラスベースの推定ではパラメータの推定が不適切であることである。

推定年代の信頼度の尺度として，正規化対数尤度比が最大となる年代の値$NLLR_1$と二番目に大きい年代の値$NLLR_2$の対数比$\log(NLLR_1/NLLR_2)$を用いた。このとき，信頼度の高さと推定精度には正の相関が見られた。

2.2.2　Tilahun（2011）; Tilahun *et al.*（2012）

Tilahun らは，中世イングランドで作成されたラテン語の証書（charter）の年代推定法を提案した[9]。証書とは，土地や財産の所有権や譲渡を証明する法律文書で，主に教会や行政機関また王室などにより発行された文書である。これらの文書は DEEDS（Documents of Early England Data Set）と呼ばれるデータセットに含まれているものである[10]。Tilahun らの用いた 3353 文書の平均作成年代は 1246 年，最小値は 1089 年，最大値は 1438 年である。データセット全体の語彙数（タイプ）は約 5 万語，総単語数（トークン）は約 80 万語，平均文書長（単語数）は 237 語である。

各年代の文書群には特徴的な単語連続が存在するという仮定のもと，*k*-shingle と呼ばれる *k* 個（$k=1,2,\ldots$）の単語連続（単語 *n*-gram と等価）を素性として使用した。年代推定を年代による文書分類の問題とみなし，*k*-近隣法と，最尤法に基づいた Maximum Prevalence 法（以下，MP 法）より年代推定を行った。いずれの方法もカーネル関数と組み合わせて使用されている。

一つ目の *k*-近隣法は，文書同士の距離に基づき，分類を行う手法である。年代が既知の文書からなる訓練データをT，訓練データ内の文書を$d \in T$，テストデータをA，年代不詳の文書を$q \in A$とする。長さ *k* の *k*-shingle の出現頻度を素性としたときの文書qと文書dの類似度を$Sim_k(q,d)$とする。類似度は$0 \leq Sim_k(q,d) \leq 1$となるものを用いる必要がある。このとき，文書qと文書dの距離$d_k(q,d)$は$d_k(q,d) = 1 - Sim_k(q,d)$で与えられる[11]。$0 \leq Sim_k(q,d) \leq 1$より，距離$d_k(q,d)$も$0 \leq d_k(q,d) \leq 1$となる。訓練データ$T$内の文書のうち，文書$q$との距離$d_k(q,d)$が最小（つまり類似度$Sim_k(q,d)$が最大）となる m 個の文書の集合をS_mとする。これらの文書$d \in S_m$の持つ重みを，距離$d_k(q,d)$とカーネル関数K_{h_k}を用いて，$K_{h_k}\left(d_k(q,d)\right)$と定義する（カーネル関数については，5.1 節を参照）。ここで，hはバンド幅と呼ばれるパラメータである。異なる長さの *k*-shingle を *r* 個組み合わせたときの文書$d \in S_m$の *r* 次元カーネル重み$a(q,d)$は，

$$a(q,d) \equiv a(q,d|h_1,\ldots,h_r) = \prod_{k=1}^{r} K_{h_k}\left(d_k(q,d)\right) \tag{2.4}$$

で与えられる。このとき，文書qの推定年代$\hat{t_q}$は，文書$d \in S_m$の実年代t_dと文書qの持つ重み$a(q,d)$の平均として計算

[9] Tilahun ら以前の DEEDS を用いた研究については，Feuerverger *et al.* 2005，2008；Fiallos 1997，2000；Gervers 1997，2000a，2000b を参照。このほか，Thinniyam（2014）は，文書間の距離に基づいて文書の時系列化を行う方法を提案した。

[10] http://deeds.library.utoronto.ca/

[11] 用いる類似度$Sim_k(q,d)$によっては距離$d_k(q,d)$が三角不等式を満たさないので，$d_k(q,d)$は必ずしも数学的な意味での距離となるわけではない。

される：

$$\widehat{t_q} \equiv \arg \min_{t \in \{1089,\ldots,1438\}} \sum_{d \in S_m} (t_d - t)^2 a(q,d) = \frac{\sum_{d \in S_m} a(q,d) * t_d}{\sum_{d \in S_m} a(q,d)} \tag{2.5}$$

実験は，訓練データ$|T| = 2608$とテストデータ$|A| = 745$で行った。類似度$Sim_k(q,d)$としては，Jaccard 係数を，カーネルK_{h_k}にはガウスカーネルを用いた。S_mの文書数とバンド幅hは，グリッドサーチによる交差検定で決定する。m = 100 で 1-shingle と 2-shingle を組み合わせた時に，最良の結果となり，推定年代と実年代の絶対値誤差平均は 12.1 年，絶対値誤差中央値は 6.3 年，二乗平均平方根誤差は 20.2 年となった（これらの値の計算方法については，9.1 節を参照）。仮にテストデータAのすべての文書の推定年代を，訓練データ T の文書の平均作成年代 1246 年と推定した場合，絶対値誤差平均は 37 年，絶対値誤差中央値は 25 年，二乗平均平方根誤差は 47 年となる。

Gervers & Tilahun（2013）は，同様の方法で，文書の作成場所推定を行った。文書の作成場所は緯度・経度で表されている。作成場所不明の文書qの推定作成場所（$\widehat{Lat_q}, \widehat{Long_q}$）は，文書$d \in S_m$の緯度・経度（$Lat_d, Long_d$）と文書$d$の持つ重み$a(q,d)$の平均として計算される：

$$\widehat{Lat_q} = \frac{\sum_{d \in S_m} a(q,d) * Lat_d}{\sum_{d \in S_m} a(q,d)}$$

$$\widehat{Long_q} = \frac{\sum_{d \in S_m} a(q,d) * Long_d}{\sum_{d \in S_m} a(q,d)} \tag{2.6}$$

実験は，1066 年以降に作成された 4370 文書を用いて行われた（$|T| = 3670$，$|A| = 700$）。2-shingle を素性として用いた場合，実際の作成場所と推定地点との絶対値誤差平均は 87 ㎞，絶対値誤差中央値は 75 ㎞となった。文書の作成年代を考慮した場合，推定結果は悪化し，絶対値誤差平均は 92 ㎞，，絶対値誤差中央値は 80 ㎞となった。また，テストデータA内の文書の推定地点を訓練データT内の文書の平均緯度・経度と推定した場合，絶対値誤差平均は 134 ㎞，絶対値誤差中央値は 115 ㎞となった。

一方，MP 法は，文書の優勢度（prevalance）が最大となる年代を推定年代とする最尤法の一種である。ここで，優勢度とは尤度に相当するものである。年代が既知の文書からなる訓練データをT，訓練データ内の文書を$d \in T$，テストデータをA，年代不詳の文書を$q \in A$，文書qの推定年代を$\widehat{t_q}$，文書qに含まれる長さ k の k-shingle のトークンをs，その集合を$s_k(q)$，検証データをVとする。各年代tにおける文書qの優勢度$\pi_q(t)$は，sの出現確率$\pi_s(t)$の積で与えられる：

$$\pi_q(t) = \prod_{s \in s_k(q)} \pi_s(t) \tag{2.7}$$

各sは，各年代に固有の多項分布から独立に引かれたものとする。ただし，年代tにおけるsの真の出現確率$\pi_s(t)$は不明なので，訓練データTから推定する必要がある。$\pi_s(t)$の推定値を$\hat{\pi}_s(t)$とすると，優勢度$\pi_q(t)$の推定値$\hat{\pi}_q(t)$は，

$$\hat{\pi}_q(t) = \prod_{s \in s_k(q)} \hat{\pi}_s(t) \tag{2.8}$$

で与えられる。このとき，文書qの推定年代$\hat{t_q}$は，$\hat{\pi}_q(t)$が最大となる年代tとする：

$$\begin{aligned}\hat{t_q} &= \arg\max_{t\in\{1089,\ldots,1438\}} \hat{\pi}_q(t) \\ &= \arg\max_{t\in\{1089,\ldots,1438\}} \prod_{s\in s_k(q)} \hat{\pi}_s(t)\end{aligned} \tag{2.9}$$

$\hat{\pi}_s(t)$は，局所多項式ロジスティック回帰モデル（local polynomial logistic regression model）により推定される。ここで，各sは二項分布$\big(|s_k(q)|, \pi_s(t)\big)$に従うと想定している。$|s_k(q)|$は，文書$q$に含まれる長さ k の k-shingle のトークン数である。ただし，計算量の増加を抑えるために，0 次の多項式（つまり定数）により推定を行った。これは，第 5 章のカーネル平滑化と等価である。文書$d \in T$ におけるsの出現頻度を$n_s(d)$，k-shingle の総数を$N_s(d)$，文書dの推定年代を$\hat{t_d}$とする。このとき，年代tにおけるsの出現確率の推定値$\hat{\pi}_s(t)$はカーネル関数K_{h_k}を用いて，以下のように与えられる：

$$\hat{\pi}_s(t) = \frac{\sum_{d\in T} K_{h_k}(t_q - t_d) * n_s(d)}{\sum_{d\in T} K_{h_k}(t_q - t_d) * N_s(d)} \tag{2.10}$$

実験は，$|T| = 2608$，$|V| = 419$，$|A| = 326$で行った。カーネル関数K_{h_k}には t 分布カーネル関数を用い，パラメータ（バンド幅と自由度）は検証データVにより決定した。文書長は，年代tとは独立だと仮定した。2-shingle を用いたとき，テストデータAの結果が最良となり，推定年代と実年代の絶対値誤差平均は 9.0 年，絶対値誤差中央値は 6.0 年，二乗平均平方根誤差は 14.7 年となった。

2.2.3　Kumar *et al.*（2011）; Kumar *et al.*（2012）; Kumar（2013）

Kumar らは，文書の内容から文書にタイムスタンプを付与する方法を提案した（Kumar *et al.* 2011 ; Kumar *et al.* 2012 ; Kumar 2013）。タイムスタンプとは，文書が作成された年代もしくは文書が言及している年代のことである。タイムスタンプの付与には，ユニグラム言語モデル（Ponte & Croft 1998）と KL 情報量（Lafferty y Zhai 2001）に基づくモデルを用いた（KL 情報量については，付録 A の A10 を参照）。前者では尤度が最大，後者では KL 情報量が最小となる年代が文書のタイムスタンプとなる。いずれのモデルにおいても，スムージングを行いスパースネスに対処した。訓練データとして，生存期間が前 3800 年から後 2010 年の期間に含まれる人物についての Wikipedia の記事（約 22 万文書）を用いた。時間に関係する表現，数字，ストップワードを削除した結果，語彙サイズは約 37 万語となった。この訓練データを用いて，以下の三つの実験を行った。

一つ目の実験では，生存期間が前 3800 年から後 2010 年の間に含まれる人物（たとえば，プラトン）についての Wikipedia の記事から，その人物の生存期間（前 5〜4 世紀）の推定を行った。最良のモデルを用いたとき，テストデータ（約 8000 文書）での絶対値誤差平均は 37 年，絶対値誤差中央値は 22 年となった。

二つ目の実験では，グーテンベルクプロジェクトに含まれる英語の短編小説（1798 年〜2008 年）の出版年代の推定を行った。文書の平均単語数は約 14000 語である。最良のモデルを用いたとき，テストデータ（約 300 文書）での絶対値誤差平均は 23 年，絶対値誤差中央値は 19 年となった。

三つ目の実験では，前 500 年から後 2010 年の間のいずれかの年に起きた歴史的事象（戦争，王位継承，発明など）についての Wikipedia の記事から，それらが起きた年を推定した。最良のモデルを用いたとき，テストデータ（約 1300 文書）での絶対値誤差平均は 38 年，絶対値誤差中央値は 20 年となった。

2.2.4　金（2009：203-224）

金（2009 : 203-224）は，芥川龍之介（1892～1927）の作品の年代推定の実験を行った。データセットは，1912 年から 1927 年の間に執筆された 250 編の作品である。素性は計 39 個の助詞と読点の相対頻度とし，重回帰，サポートベクターマシン，決定木，ランダムフォレストにより年代推定を行った。実験は，ランダムに選んだ 25 作品をテストデータ，残りを訓練データとして使用し，これを 100 回行った。実験の結果，ランダムフォレストの絶対値誤差平均が最小となり，約 1.35 年であった。

2.2.5　Bruster & Smith（2014）

Bruster & Smith（2014）は，「制約付き対応分析（van de Velden *et al.* 2009）」を用いてシェークスピアの戯曲作品の年代推定を行った。制約付き対応分析では，まず，戯曲の文の長さに対して対応分析を行い，これにより得られた第一軸上のスコアに基づいて作品を相対的に時系列化する。次に，複数の作成年代既知の作品の作成年代を目的変数，上記のスコアを説明変数として単回帰を行う。ここで，作成年代既知の作品の作成年代が制約として働き，絶対的年代推定が可能になる。作成年代不詳の作品の推定年代は，対応分析から得られた第一軸上のスコアと回帰直線の係数に基づいて算出される。この手法は，説明変数（対応分析により得られた第一軸上のスコア）が，時間とともに単調増加するという非常に強い仮定を置いている。この研究では，シェークスピアの戯曲の文の長さが，作家のキャリアとともに単調増加しているという事実から内挿を行っている。

2.3　統計的場所推定

文書の場所推定とは，文書の内容からその文書の作成された場所を推定するタスクである[12]。データセットとしては，文書の作成場所（発信地）についての情報が得られる Twitter や Facebook などが用いられることが多い。場所推定は，文書分類の枠組みで考えることができる。素性としては，単語ユニグラムが一般的である。クラスの設定方法はタスク依存であるが，一般的には，一定の領域を格子状に分割した格子の一つ一つ（たとえば，50 ㎞×50 ㎞の格子）や政治的単位（市・県・州・国）がクラスとなる。クラス数は，範囲（アメリカ全土や全世界）や格子（geodesic grid）の大きさ（50 ㎞×50 ㎞，100 ㎞×100 ㎞）にもよるが，数万から数十万のオーダーになる。次に，各クラスに属する文書群から，単語の出現確率を推定する。ただし，クラス分割が微細になるほどデータのスパースネスが顕著になるので，隣接クラスに対して平滑化を行い，この問題に対処することがある（Serdyukov *et al.* 2009；Cheng *et al.* 2010；Lichman & Smyth 2014；Hulden *et al.* 2015）。

分類には，ナイーブベイズ（Wing & Baldridge 2011；Han *et al.* 2014；Hulden *et al.* 2015），KL 情報量による *k*-近隣法（Wing & Baldridge 2011；Roller *et al.* 2012；Han *et al.* 2014；Hulden *et al.* 2015），対数線形モデル（Han *et al.* 2014；Wing & Baldridge 2014），混合ガウスモデル（Priedhorsky *et al.* 2014），混合カーネル密度推定（Lichman & Smyth 2014），トピックモデル（Eisenstein *et al.* 2010）等のモデルが用いられる。文書の推定場所は，モデルから計算された事後確率や類似度が最大となるクラス（場所）となる（1.5）。モデルの性能は，精度（正しく分類された文書の割合）や，推定された場所と真の作成場所との距離の誤差（単位は㎞）の平均値・中央値・二乗平均平方根などで測られる。文書間距離から作成場

[12] ユーザーのリンクやフォロワーなどのメタデータに基づいて場所推定を行う方法については，Backstrom *et al.*（2010）や Jurgens *et al.*（2015）を参照。

所の緯度・経度を推定する方法については，2.2.2 項の Gervers & Tilahun（2013）を参照。

2.4　CODEA を用いた年代推定・場所推定

本研究で用いる CODEA をデータセットとして用いた年代推定・場所推定の研究としては，Ueda（2013b），Kawasaki（2013b，2014a，2014b，2015b，2015c）が挙げられる[13]。ここでは，Kawasaki（2014b）のみを紹介する。

Kawasaki（2014b）は，*k*-近傍法による年代推定・場所推定を行った。*k*-近傍法とは，文書間の類似度に基づいて分類を行う方法である。*k*-近傍法の利点は，単純な方法であるにもかかわらず，ある程度の分類性能を示すことである（Hastie *et al.* 2009：Chapter 13）。*k*-近傍法の欠点として，近傍点探索のための計算量の負荷と訓練データ全体の記憶の負荷が挙げられるが，本研究で用いたデータセットの大きさでは問題にならない。

文書qの属性Aが未知のときに，文書qとの類似度が高い上位 k 個の文書の属性のうち最も優勢な属性A'を文書qの属性として推定する[14]。$k=1$ の時，最も類似度が高い文書の属性A'を，文書qの属性とみなす。$k \geqq 2$ の場合は，多数決や重み付けにより分類を行う。*k*-近傍法では，類似度が高い文書同士は，類似した属性を持つと仮定している。

素性には，スペイン語文献学の先行研究（Zamora 1967；Sánchez-Prieto Borja 1998；Menéndez Pidal 1999；Penny 2002 など）に基づき，時間的変異・空間的変異を示す約 300 個の文献学的特徴（文字，音韻，形態論，統語論，語彙など）を使用した。文献学的特徴を素性として用いる利点は，素性の文献学的な解釈が可能である点である。

文書は，素性の有無（1 か 0）を要素とする二値ベクトル（binary vector）として表現した。文書間の類似度にはコサイン類似度を用いた。素性の集合を$F = \{f_1, \ldots, f_{|F|}\}$，年代推定・場所推定を行う文書$q$における素性$f \in F$の頻度を$n(f, q)$，文書$d$における素性$f \in F$の頻度を$n(f, d)$する。このとき，文書$q$と文書$d$のコサイン類似度$\mathrm{Cos}(q, d)$は，次式で与えられる：

$$\mathrm{Cos}(q, d) = \frac{\sum_{f \in F} n(f, q) * n(f, d)}{\sqrt{\sum_{f \in F} n(f, q)^2} \sqrt{\sum_{f \in F} n(f, d)^2}} \tag{2.11}$$

コサイン類似度には，非負性，同一性，対称性という性質がある：

$$0 \leq \mathrm{Cos}(q, d) \leq 1 \tag{2.12}$$

$$\mathrm{Cos}(q, d) = 1 \Leftrightarrow q = d \tag{2.13}$$

$$\mathrm{Cos}(q, d) = \mathrm{Cos}(d, q) \tag{2.14}$$

コサイン類似度の長所は，文書長の影響を除去することができる点である。これは，各文書ベクトルの長さが 1 になるようにユークリッド距離で正規化しているからである[15]。

文書qとのコサイン類似度が最大になる上位 k 個の文書の集合をS_kとする。文書qの推定年代$\hat{t_q}$は，S_kに属する文書$d' \in S_k$の実年代$t_{d'}$の加重平均

[13] ただし，CODEA の改訂が行われたので，これらの研究と本研究で用いるコーパスは同一のものではない。

[14] 「類は友を呼ぶ（A man is known by the company he keeps）」に基づいた分類方法だといえる。

[15] 一方，コサイン類似度の短所は，素性間の相関を無視している点，素性への重みづけが恣意的である点である。これらの欠点の修正法については，Mochihashi *et al.*（2004），Mikawa *et al.*（2011），Sidorov *et al.*（2014）を参照。

$$\widehat{t_q} = \sum_{d' \in S_k} \left(t_{d'} \times \frac{\mathrm{Cos}(q, d')}{\sum_{d'} \mathrm{Cos}(q, d')} \right) \tag{2.15}$$

から求めた。年代の範囲は 1109 年～1697 年である。コサイン類似度$\mathrm{Cos}(q, d')$が大きい（小さい）ほど文書d'の重みは大きく（小さく）なる。

文書qの推定場所$\widehat{l_q}$は，S_kに属する文書$d' \in S_k$の実作成場所$l_{d'}$のうち，コサイン類似度の和が最大となる場所$l \in L$とした：

$$\widehat{l_q} = \arg\max_{l \in L} \sum_{d'_l \in S_k} \mathrm{Cos}(q, d'_l) \tag{2.16}$$

ここで，$L = \{l_1, \ldots, l_{|L|}\}$は地点の集合，$d'_l$は$S_k$に属し且つ実作成場所が$l \in L$である文書である。地点の粒度は，県（41 地点）と地方（15 地点）の二つを設けた。

実験は，作成年代と作成場所がともに既知の 1026 文書を用いて行った。一個抜き交差検定の結果，$k = 6$ の時に最良の結果となり，実年代と推定年代との絶対値誤差平均は 21 年，絶対値誤差中央値は 14 年，二乗平均平方根誤差は 31 年となった。一方，場所推定の分類正解率は，県レベルで 48%，地方レベルで 63%となった。

2.5　考察

本研究が先行研究を踏襲している点と本研究の新規性を以下で述べる。

本研究が先行研究を踏襲しているのは，以下の三点である。

一点目は，年代と場所をともに離散変数とみなし，年代推定・場所推定を分類のタスクと設定している点である。これにより，特定の年代や場所のみに見られる局所的な現象をモデル化することが可能になる。確かに，年代と場所（緯度・経度）を連続変数とみなし，回帰を行うということも考えれる。しかし，線形回帰では，特定の年代や場所のみに見られる局所的な現象をモデル化することが難しい。なぜならば，線形回帰では大域的に線形な特徴を説明変数とする必要があるからである。非線形回帰による年代推定・場所推定は今後の課題とし，本研究では扱わない。

二点目は，分類器として *n*-gram 言語モデル，JS 情報量（KL 情報量の変種），ナイーブベイズという比較的単純なものを用いる点である。分類器自体には特別な工夫は施していない。対数線形モデルや SVM などの性能がより高いとされる分類器は，計算量が増えてしまうので，本研究では使用しない。また，分析のモデルが存在しない *k*-近傍法も用いない。

三点目は，素性として文字 *n*-gram を使用する点である。文字 *n*-gram の長所は，出現頻度が高く，言語に依存せず機械的に抽出できる点である。先行研究で用いられている単語 *n*-gram は文書の内容を捉えてしまうので，文書の内容に依存しない年代推定・場所推定を目指す本研究には適していない。ただし，文字 *n*-gram も，ある程度，文書の内容を捉えてしまっている。また，品詞 *n*-gram は文書の内容に依存しない素性であるが，本研究で用いるコーパスには品詞タグが付与されていないので，使用することはできない。

一方，本研究の新規性は，以下の三点である。

一点目は，文書の作成年代と作成場所の両者を推定するという点である。先行研究では，文書の作成年代もしくは作成場所のどちらかのみが推定対象となっていた。統計的年代推定の先行研究では，作者が同一の文書や地理的変異の小さい言語（ラテン語や現代英語など）で書かれた文書を用いているので，空間軸は考慮されていない。一方，統計的場所推定の先行研究では，ほぼ同時期に作成された文書を用いているので，時間軸は考慮されていない。しかし，本研

究で用いるコーパスは，時間軸と空間軸のどちらにおいても変異が存在するので，両者を考慮する必要がある。

二点目は，作成年代と作成場所の同時推定において，時間カーネル平滑化と空間カーネル平滑化を組み合わせた時空間カーネル平滑化を導入し，言語の時空間変異を考慮している点である。先行研究では，時間カーネル平滑化と空間カーネル平滑化は個別に用いられているだけある。時空間を同時に考慮するモデルは空間統計学（瀬谷 堤 2014：5.6；古谷 2011：9 章）では広く使用されているが，計算言語学への応用はないようである。

三点目は，素性として形態音韻論や形態統語論などの文献学的特徴を使用することである。これらの特徴は，筆者がスペイン語史の知見に基づき設定したものである。文献学的特徴を用いる利点は，言語学的に解釈がしやすい点である。文字 *n*-gram は，必ずしも言語学的に意味のある解釈ができるわけではない。また，文献学的特徴の年代推移を可視化することは，スペイン語文献学への大きな貢献となる。

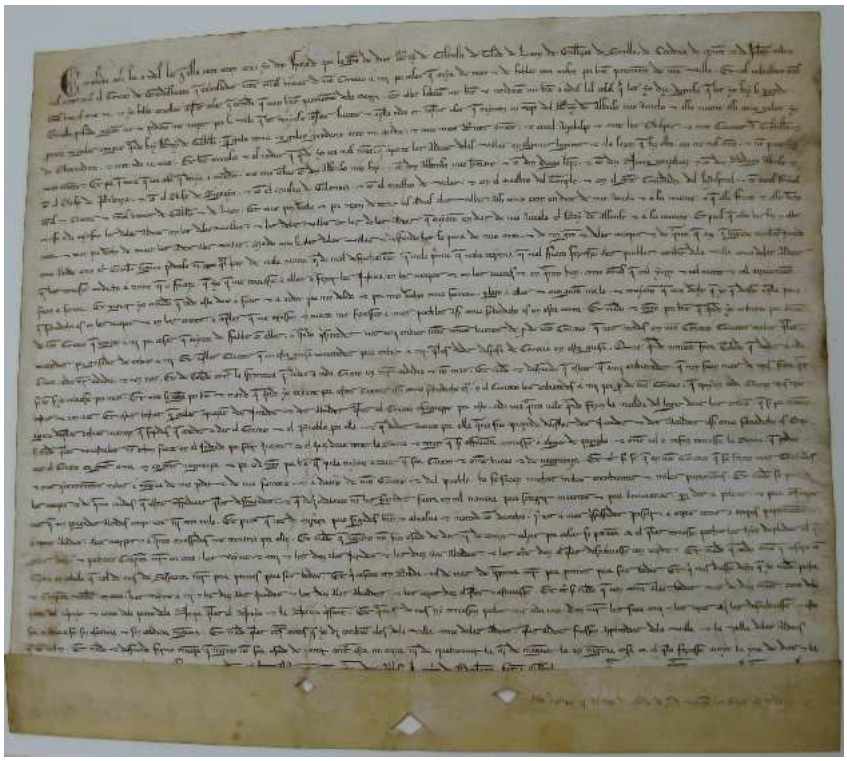
第3章 コーパス

3.1　概要

本研究では，中近世スペイン語古文書コーパス CODEA（Corpus de Documentos Españoles Anteriores a 1700）を使用した。このコーパスは，スペインのアルカラ大学の Sánchez-Prieto Borja 教授を中心とした文献学研究グループ GITHE（Grupo de Investigación de Textos para la Historia del Español）が作成しているもので，1100 年から 1700 年の間に，スペインで作成された 1500 の古文書からなる。ただし，今日のイタリアとポルトガルで作成されたスペイン語の文書もいくつか含まれている。これらの文書のうち，作成年代もしくは作成場所が不明な文書，オリジナルより 10 年以上後に作成された写し（traslado，copia），10 年以上前の文書を引用している文書（confirmación）などは除外し，作成年代と作成場所が一義的に決まる 1237 文書のみを使用する。これらの文書の作成年代と作成場所は，文書内に明示されている，もしくは文書の内容（人名や地名など）から確定することができるものである。

本研究では，パレオグラフィカ（versión paleográfica）とクリティカ（presentación crítica）と呼ばれる二つのバージョンを，適宜，使い分けた。前者は，略記法や文字の変異や単語の連結などもそのまま転写されており，文書をより忠実に表現している。クリティカは，略記法の展開，文字の変異の統一，単語の分割を行い，文書として読みやすくしたものである。両者とも，CHARTA（Corpus Hispánico y Americano en la Red：Textos Antiguos）による統一された方針のもとで転写されている。現時点では，品詞や語幹等の情報は付加されていない。以下に，文書 ID1（作成年代 : 1251 年，県 : Sevilla，自治州 : AN）の写真，パレオグラフィカ，クリティカを示す[16]。

写真



[16] http://corpuscodea.es/corpus/consultas.php（2015 年 10 月 11 日アクセス）。

パレオグラフィカ（versión paleográfica）

{h 1r} {1} Connoscida cosa sea a todos los q<ue> esta carta uieren como yo don Fferrando por la gr<aci>a de dios Rey de Castiella de Toledo de Leon de Gallizia de Seuilla de Cordoua de Murc<ia> & de Jahen enbie {2} mis cartas auos el Conceio de Guadalfaiara q<ue> enbiassedes u<uest>ros om<n>es buenos de u<uest>ro Conceio a mj. por cosas q<ue> auya de ueer & de fablar con uusco por bue<n> paramie<n>to de u<uest>ra villa. Et uos enbiastes u<uest>ros {3} om<n>es buenos ante mi. & yo fable conellos aq<ue>llas cosas q<ue> ente<n>di q<ue> eran bue<n> paramie<n>to dela tierra. Et ellos saliero<n> me bie<n> & recudiero<n> me bie<n> a todas las cosas q<ue> les yo dix. deguisa q<ue> les yo fuy so pagado. {4} Et esto passado rogaro<n> me & pidiero<n> me mercet por su villa q<ue> les touyesse aq<ue>llos fueros & aq<ue>lla uida et aq<ue>llos usos q<ue> ouyeran [*[it:mano 2 : :it]*ouyeron] en tie<n>po del Rey do<n> Alfonso mio Auuelo & assu muerte assi como gelos yo {5} p<ro>meti & gelos otorgue q<ua>ndo fuy Rey de Casti<e>lla. q<ue> gelo ternia & gelos guardaria ante mi Madre. & ante mios Ricos om<n>es. & antel Arcobispo & ante los Obispos. & ante Caueros d<e> Casti<e>lla & {6} de Estremadura. & ante toda mi corte. Et bie<n> connosco & es uerdat q<ue> q<ua>ndo yo era mas nin<n>o q<ue> aparte las Aldeas delas villas en algunos logares. & ala sazon q<ue> fiz esto. era me mas nin<n>o. & no<n> pare hy {7} tanto mie<n>tes. （中略） Et ma<n>do q<ue>las ot<ra>s cartas q<ue> yo dj tanbie<n> alos dela villa como delas Aldeas. q<ue>las aldeas fuessen Appartadas dela villa & la villa delas Aldeas. {33} q<ue> no<n> ualan. Et ma<n>do & deffiendo firme mie<n>tre q<ue> ni<n>guno no<n> sea osado de uenir cont<ra> esta mi carta nj<n> de quebrantar la nj<n> de me<n>guar la en ni<n>guna cosa ca el q<ue>lo fiziesse aurye la yra de dios & la {34} mia. & pechar mie en coto. mill. m<o>r<abedis>. [*[it:lat.:it]* : [!ffacta carta ap<u>d Sibilla Reg<e> exp<rimente>. xiij. die Ap<ri>lis.!] J<ohannes>. pet<ri> de Berla<n>ga fe<ci>t. [!ERA_M_CC_Lxxx_Nona!]]

クリティカ（presentación crítica）

{h 1r} {1} Coñocida cosa sea a todos los que esta carta vieren cómo yo don Ferrando, por la gracia de Dios rey de Castiella, de Toledo, de León, de Gallizia, de Sevilla, de Córdova, de Murcia e de Jaén, embié {2} mis cartas a vós el concejo de Guadalfajara que embiássedes vuestros omnes buenos de vuestro concejo a mí, por cosas que avía de veer e de fablar convusco por buen paramiento de vuestra villa. E vós embiastes vuestros {3} omnes buenos ante mí, e yo fablé con ellos aquellas cosas que entendí que eran buen paramiento de la tierra. E ellos saliéronme bien e recudiéronme bien a todas las cosas que les yo dix, de guisa que les yo fui so pagado. {4} E esto passado rogáronme e pidiéronme mercet por su villa que les toviesse aquellos fueros e aquella vida e aquellos usos que ovieran en tiempo del rey don Alfonso mio avuelo e a su muerte, assí como gelos yo {5} prometí e gelos otorgué cuando fui rey de Castiella que gelo ternía e gelos guardaría ante mi madre, e ante mios ricos omnes, e ant'el arçobispo, e ante los obispos, e ante caveros de Castiella e {6} de Estremadura e ante toda mi corte. E bien coñosco e es verdat que cuando yo era más niño que aparté las aldeas de las villas en algunos logares. E a la sazón que fiz esto érame más niño e non paré y {7} tanto mientes. （中略） E mando que las otras cartas que yo di tan bien a los de la villa como de las aldeas que las aldeas fuessen apartadas de la villa e la villa de las aldeas {33} que non valan. E mando e defiendo firmemientre que ninguno non sea osado de venir contra esta mi carta nin de quebrantarla nin de menguarla en ninguna cosa, ca el que lo fiziesse avrié la ira de Dios e la {34} mía, e pechar m'ié en coto mill morabedís. [!*Facta carta apud Sibilla, rege exprimente, XIII die aprilis.!] Johannes Petri de Berlanga fecit. [!Era MCCLXXX nona.!]*

{}内の文字や数字はフォリオや行番号である。[!...!]でマークされた箇所は文書の明示的な作成年代と作成場所である。*[it…:it]*のイタリック体の部分は，文書のメタ情報である。これらの文字列は，

本研究では，文書の属性のうち作成年代と作成場所のみに注目する。文書の作成者，文書の種類（国璽尚書，地方政府の文書，教会関係の文書，裁判関係の文書，私的文書），文字の種類（carolina，gótica，humanística，cortesana など）

といった属性は無視する。これらの属性も考慮した作成年代・場所推定法の開発は，今後の課題とする。

コーパスは，Excel 上で作成した（図 3.1）。スペインの研究チームから提供された Word 形式の文書データを整理し，Excel へインポートした。コーパス作成から後述の数理的分析まですべての作業は，Excel VBA（Visual Basic for Application）の自作プログラムで行った（付録 C を参照）。

	A	B	C	D	E	F
1	ID	年代	県	自治州	パレオグラフィカ（TRANSCRIPCIÓN PALEOGRÁFICA）	クリティカ（PRESENTACIÓN CRÍTICA）
2	1	1251	Sevilla	AN	{h 1r}	{h 1r}
3	1	1251	Sevilla	AN	{1} Connoscida cosa sea a todos los q<ue> esta carta uieren como yo don Fferrando por la gr<aci>a de dios Rey de Castiella de Toledo de Leon de Gallizia de Seuilla de Cordoua de Murc<ia> & de Jahen enbie	{1} Coñocida cosa sea a todos los que esta carta vieren cómo yo don Ferrando, por la gracia de Dios rey de Castiella, de Toledo, de León, de Gallizia, de Sevilla, de Córdova, de Murcia e de Jaén, embié
4	1	1251	Sevilla	AN	{2} mis cartas auos el Conceio de Guadalfaiara q<ue> enbiassedes u<uest>ros om<n>es buenos de u<uest>ro Conceio a mj. por cosas q<ue> auya de ueer & de fablar con uusco por bue<n> paramie<n>to de u<uest>ra villa. Et uos enbiastes u<uest>ros	{2} mis cartas a vós el concejo de Guadalfajara que embiássedes vuestros omnes buenos de vuestro concejo a mí, por cosas que avía de veer e de fablar convusco por buen paramiento de vuestra villa. E vós embiastes vuestros
5	1	1251	Sevilla	AN	{3} om<n>es buenos ante mi. & yo fable conellos aq<ue>llas cosas q<ue> ente<n>di q<ue> eran bue<n> paramie<n>to dela tierra. Et ellos saliero<n> me bie<n> & recudiero<n> me bie<n> a todas las cosas q<ue> les yo dix. deguisa q<ue> les yo fuy so pagado.	{3} omnes buenos ante mí, e yo fablé con ellos aquellas cosas que entendí que eran buen paramiento de la tierra. E ellos saliéronme bien e recudié ronme bien a todas las cosas que les yo dix, de guisa que les yo fui so pagado.
6	1	1251	Sevilla	AN	{4} Et esto passado rogaro<n> me & pidiero<n> me merçet por su villa q<ue> les touyesse aq<ue>llos fueros & aq<ue>lla uida et aq<ue>llos usos q<ue> ouyeran [[it:mano 2: :it]ouyeron] en tie<n>po del Rey do<n> Alfonso mio Auuelo & assu muerte assi como gelos yo	{4} E esto passado rogáronme e pidiéronme mercet por su villa que les toviesse aquellos fueros e aquella vida e aquellos usos que ovieran en tiempo del rey don Alfonso mio avuelo e a su muerte, assí como gelos yo

図 3.1　コーパス

3.2　記述的統計

上記の 1273 文書の平均文書長は 607 語，中央値は 444 語，最小値は 45 語，最大値は 3657 語である。文書長は各文書内の単語数とした。ここで単語とは，文頭記号とスペース，スペースとスペース，スペースと文末記号で区切られた文字連続である。したがって，必ずしも現代スペイン語の単語に相当するわけではない。たとえば，前置詞 de と定冠詞 la が一つになった dela は，一つの単語としてカウントした。文書長は裾が重い分布となっている（図 3.2）。総単語数（トークン）は約 77 万語，総語彙数（タイプ）は約 7 万語である。文書長の大小にかかわらず，一つの文書は一文書としてカウントする。

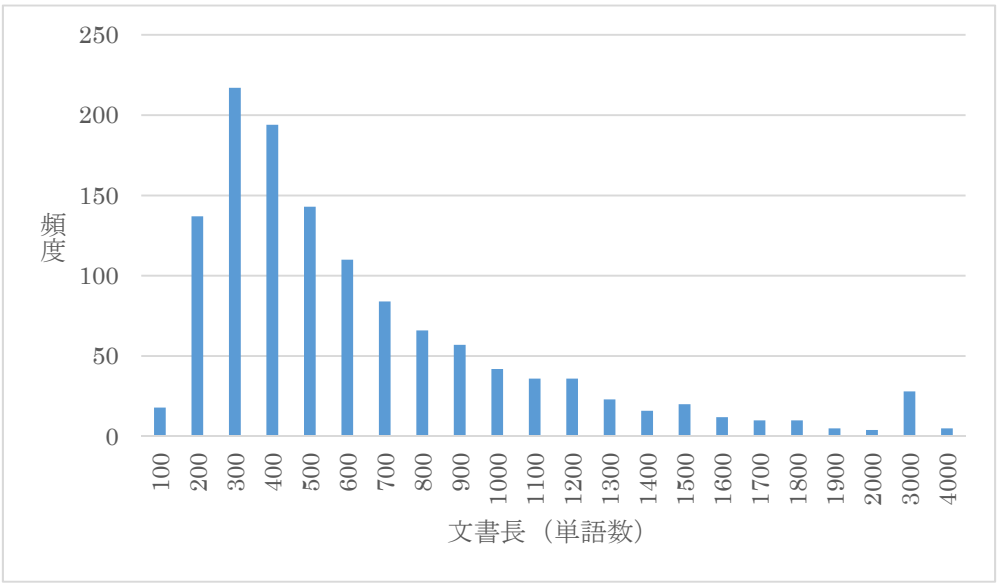


図 3.2　文書長別の文書数

文書の作成平均年代は 1393 年，中央値は 1369 年，最小値は 1097 年，最大値は 1697 年である。文書数の分布は均等ではなく，1300 年前後は多いが，12 世紀や 17 世紀は少ない（図 3.3）。

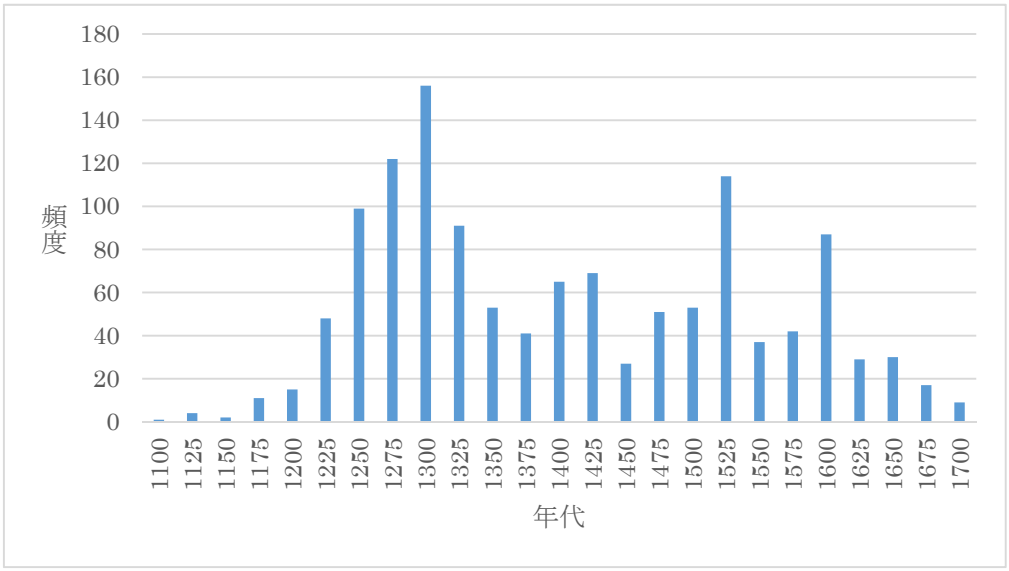


図 3.3　年代別の文書数

現代スペインの各自治州と所属県を，表 3.1 に示す。県名の後ろのカッコ内の数字は，図 3.4 の番号に対応している。州都の位置する県は太字にしてある。自治都市の Ceuta と Melilla の記載は省略した。

自治州	所属県（太字は州都の位置する県）
AN (Andalucía)	Almería (1), Cádiz (2), Córdoba (3), Granada (4), Huelva (5), Jaén (6), Málaga (7),**Sevilla** (8)
AR(Aragón)	Huesca (9), Teruel (10), **Zaragoza** (11)
AS (Asturias)	**Asturias** (49)
IB (Islas Baleares)	**Islas Baleares** (42)
CN (Islas Canarias)	**Las Palmas de Gran Canaria** (43), **Santa Cruz de Tenerife** (44)
CT (Cataluña)	**Barcelona** (27), Girona (28), Lérida (29), Tarragona (30)
CB (Cantabira)	**Cantabria** (12)
CL (Castilla y León)	Ávila (18), Burgos (19), León (20), Palencia (21), Salamanca (22), Segovia (23), Soria (24), **Valladolid** (25), Zamora (26)
CM (Castilla-La Mancha)	Albacete (13), Ciudad Real (14), Cuenca (15), Guadalajara (16), **Toledo** (17)
EX (Extremadura)	**Badajoz** (36), Cáceres (37)
GA (Galicia)	**A Coruña** (38), Lugo (39), Ourense (40), Pontevedra (41)
LR (La Rioja)	**La Rioja** (45)
MD (Madrid)	**Madrid** (31)
MU (Murcia)	**Murcia** (50)
NA (Navarra)	**Navarra** (32)
PV (País Vasco)	**Álava** (46), Guipúzcoa (47), Vizcaya (48)
VC (Valencia)	Alicante (33), Castellón (34), **Valencia** (35)

表 3.1　自治州の州都と所属県

図 3.4 に現代スペインの自治州と県の区分を示す[17]。自治州名は赤字で示している。番号は，表 3.1 の県に対応している。

[17] 地図作製のためのデータ（シェープファイル）は，Global Administrative Areas の GADM データベース（ver. 2.8）からダウンロードした。作図は，R（ver. 3.2.3）で行った。

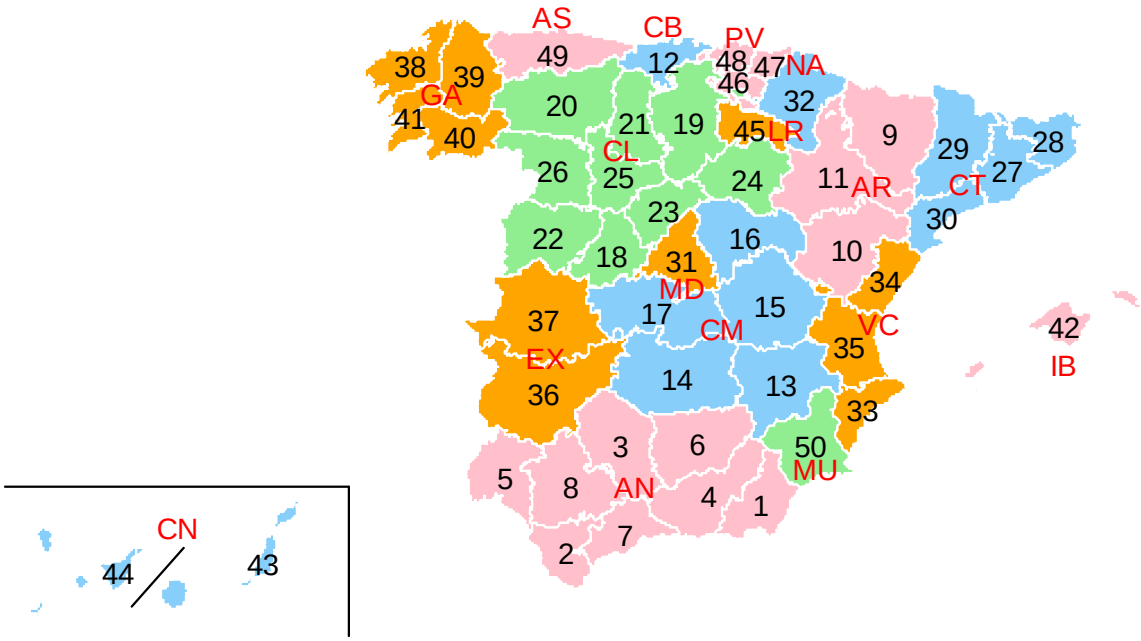


図 3.4　現代スペインの県と自治州

文書の作成場所は，県（provincia）の粒度で 52 地点（スペインの 50 県とイタリア，ポルトガル）である：A Coruña, Álava, Albacete, Alicante, Almería, Asturias, Ávila, Badajoz, Baleares, Barcelona, Burgos, Cáceres, Cádiz, Cantabria, Castellón, Ciudad Real, Córdoba, Cuenca, Girona, Granada, Guadalajara, Guipúzcoa, Huelva, Huesca, Italia, Jaén, La Rioja, Las Palmas, León, Lleida, Lugo, Madrid, Málaga, Murcia, Navarra, Ourense, Palencia, Pontevedra, Portugal, Salamanca, Santa Cruz de Tenerife, Segovia, Sevilla, Soria, Tarragona, Teruel, Toledo, Valencia, Valladolid, Vizcaya, Zamora, Zaragoza。イタリアとポルトガルに関しては，国全体を一つの地点とした。文書の作成場所は作成者の出身地（方言）と必ずしも一致するわけではないが，作成場所の方言が反映されているとみなした。表 3.2 は年代と県の文書数の分布を表している。年代は 25 年刻みとした。文書の分布は偏っており，Valladolid，León，Madrid，Navarra，Zaragoza は比較的文書数が多い一方，Alicante，Almería，Pontevedra，Valencia などはわずかである。文書が一つも存在しない年代と県も多数存在する。

	1100	1125	1150	1175	1200	1225	1250	1275	1300	1325	1350	1375	1400	1425	1450	1475	1500	1525	1550	1575	1600	1625	1650	1675	1700	合計
A Coruña																										0
Álava									3				1					2								6
Albacete																					3					3
Alicante																	1									1
Almería																				1						1
Asturias							4	25	1	5	3	2	1		3		1									45
Ávila							1	3	30	22								2								58
Badajoz											1								2		1					4
Baleares																										0
Barcelona																										0
Burgos					2		19	6	12	6	1	1			1	2	4	13			1					68
Cáceres									1		3	8	2	2					1		1					18
Cádiz								1		1				2		1	1	2			9					17
Cantabria				1	1	8	1	7	4	2	1	1	10	3		2	5	6								52
Castellón																										0
Ciudad Real																										0
Córdoba							1	1	2								1	2		1						8
Cuenca							1	1													1	3	6	1		13
Girona																										0
Granada																	3		3	5		2		3		16
Guadalajara								1	4	2			1		1	4	1	5	1	6	6	8	9	3	5	57
Guipúzcoa									1					1				2			3					7
Huelva																		1			2					3
Huesca				2	1			6	14	3	1	1	8		1	1	2									40
Italia																					11	1				12
Jaén											1					1	3	1		1		1	3			11
La Rioja				1		1	4	10	12	4	1	1		5	3		1	4								47
Las Palmas																										0
León		2	1	1	2	3	25	19	3	1			8	1	2	4	2	10				2	2	2	1	91
Lleida																										0
Lugo							2																			2
Madrid							1	1			4		3	2		1	4	7	8	8	25	5	2	4	2	77
Málaga																			1	1	2		1			5
Murcia																		1			1					2
Navarra						3	6	15	11	17	10	5	1	2	1		1	5								77
Ourense																										0
Palencia		1		1	6	24	13	6	2					6		3	1	2				1				66
Pontevedra					1																					1
Portugual																										0
Salamanca						3	5	4		13	5	4	2	5	4	7	2	4			2		2	4		66
Santa Cruz de Tenerife																										0
Segovia						1		1	13	1				1		6		5	1							29
Sevilla							7	6	3	1	3	5				1	2	5	4	7	12		1			57
Soria												1					1				1					3
Tarragona																										0
Teruel							1		1	1	2	1	10	14	2	5	1	2								40
Toledo					1	1		2	8		1	3	1	1	4	5	4	7	8	5		1				52
Valencia																									1	1
Valladolid	1	1	1	4		2	5	2	25	8	1	1	5	3		3	8	19	5	6	3	5	3			111
Vizcaya											1							4								5
Zamora					1		2	2	4	1	3		3	3	4	2			1		1					27
Zaragoza				1		2	1	3	2	3	11	7	8	18	1	3	4	3	2	1			1			71
合計	1	4	2	11	15	48	99	122	156	91	53	41	64	69	27	51	53	114	37	42	85	29	30	17	9	1270

表 3.2　年代と県の文書数の分割表

文書の作成場所は，自治州（comunidad autónoma）の粒度で 19 地点（スペインの 17 州とイタリア，ポルトガル）である：AN＝Andalucía，AR＝Aragón，AS＝Asturias，IB＝Islas Baleares，CN＝Islas Canarias，CT＝Cataluña，CB＝Cantabira，CL＝Castilla y León，CM＝Castilla-La Mancha，EX＝Extremadura，GA＝Galicia，IT＝Italia，LR＝La Rioja，MD＝Madrid，MU＝Murcia，NA＝Navarra，PT＝Portugal，PV＝País Vasco，VC＝Valencia。イタリアとポルトガルに関しては，国全体を一つの地点とした。文書の作成場所は作成者の出身地（方言）と必ずしも一致するわけではないが，作成場所の方言が反映されているとみなした。表 3.3 は年代と自治州の文書数の分布を表している。年代は 25 年刻みとした。文書の分布は偏っており，Andalucía，Aragón，Castilla y León は比較的文書数が多い一方，Galicia，Portugal，Valencia などはわずかである。文書が一つも存在しない年代と自治州も多数存在する。

	1100	1125	1150	1175	1200	1225	1250	1275	1300	1325	1350	1375	1400	1425	1450	1475	1500	1525	1550	1575	1600	1625	1650	1675	1700	合計
AN							8	8	5	2	4	5		2		3	10	11	8	16	25	3	5	3		118
AR				3	1	2	2	9	17	7	14	9	26	32	4	9	7	5	2	1			1			151
AS							4	25	1	5	3	2	1		3		1									45
CB				1	1	8	1	7	4	2	1	1	10	3		2	5	6								52
CL	1	4	2	6	11	33	70	43	89	52	10	7	18	19	11	27	18	55	7	6	8	8	7	6	1	519
CM					1	1	1	4	12	2	1	3	2	1	5	9	5	12	9	11	10	12	15	4	5	125
CN																										0
CT																										0
EX									1		4	8	2	2					3		2					22
GA					1		2																			3
IB																										0
IT																					11	1				12
LR				1		1	4	10	12	4	1	1		5	3		1	4								47
MD							1	1			4		3	2		1	4	7	8	8	25	5	2	4	2	77
MU																		1			1					2
NA						3	6	15	11	17	10	5	1	2	1		1	5								77
PT													1								2					3
PV									4		1		1	1				8			3					18
VC																	1								1	2
合計	1	4	2	11	15	48	99	122	156	91	53	41	65	69	27	51	53	114	37	42	87	29	30	17	9	1273

表 3.3　年代と自治州の文書数の分割表

第4章 素性

本研究では，二つの素性（feature）セットを用いて文書を数学的に表現する。一つ目は文字 *n*-gram の素性セット，二つ目は文献学的特徴の素性セットである。文字 *n*-gram の素性セットは，*n*-gram 言語モデル（第 6 章）と JS 情報量（第 7 章）による推定で用いる。文献学的特徴の素性セットは，ナイーブベイズ多変数ベルヌーイモデル（第 8 章）で用いる。

4.1　文字 *n*-gram

n-gram とは，ある文字列における *n* 個の文字，単語，もしくは品詞などの連続のことである。ここで単語は文頭記号とスペース，スペースとスペース，もしくはスペースと文末記号で区切られた一つ以上の文字連続とする。最小単位が文字のものを文字 *n*-gram，最小単位が単語のものを単語 *n*-gram，最小単位が品詞のものを品詞 *n*-gram という。特に，1-gram をユニグラム，2-gram をバイグラム，3-gram をトライグラムと呼ぶ。

本研究では，大文字小文字は区別せず，スペースも一つの文字としてカウントする。また，見やすさのために，スペースをアンダーバー（_）で表す。*n* が大きくなるほど，文字順や語順の情報を捉えることができる。スペースを 1 文字とした場合，文字 1-gram では，文字順も語順も完全に無視している。文字 2-gram では，文字順は部分的に考慮しているが語順は無視している。文字 *n*-gram（n≥3）では，文字順も語順もともに部分的に考慮している。

たとえば，「I live in Tokyo」という一文からなる文書の文字 1-gram，2-gram，3-gram と単語 1-gram，2-gram，3-gram の頻度は表 4.1 のようになる。このように，*n*-gram の頻度を要素として文書をベクトル表現する方法を bag-of-*n*-grams 表現と呼ぶ（付録 A の A11 を参照）。

文字 1-gram	i	_	l	v	e	n	t	o	k	y
頻度	3	3	l	1	1	1	1	2	1	1

文字 2-gram	i_	_l	li	iv	ve	e_	_i	in	n_	_t	to	ok	ky	yo
頻度	1	1	1	1	1	1	1	1	1	1	1	1	1	1

文字 3-gram	i_l	_li	liv	ive	ve_	e_i	_in	in_	n_T	_To	Tok	oky	kyo
頻度	1	1	1	1	1	1	1	1	1	1	1	1	1

単語 1-gram	I	live	in	Tokyo
頻度	1	1	1	1

単語 2-gram	I live	live in	in Tokyo
頻度	1	1	1

単語 3-gram	I live in	live in Tokyo
頻度	1	1

表 4.1　文書の bag-of-*n*-grams 表現

文字 *n*-gram の有効性は，多くの著者推定の研究において認められている（Kešelj *et al.* 2003；Peng *et al.* 2003；Koppel *et al.* 2009；Koppel *et al.* 2011；Luyckx & Daelemans 2011；Escalante *et al.* 2011；Stamatatos 2013；Sapkota *et al.* 2015）。

文字 *n*-gram の利点として，以下の四つが挙げられる。

一つ目は，対象言語についての知識を必要とせず，どの言語でも機械的に抽出することができる点である（Rosenfeld 2000）。これは，文書を単なる文字連続であるとみなしているからである。

二つ目は，単語 *n*-gram に比べスケールしやすく，素性数を大幅に削減できるからである。欧米言語の場合，特殊文字を含め文字数（アルファベットの数）は 30 個前後なので，ユニグラムは最大で約 30，バイグラムは約 900（$= 30^2$），トライグラムでも約 27000（$= 30^3$）程度に抑えることができる。一方，単語 *n*-gram では素性空間が数万～数十万のオーダーとなり，素性選択を行う必要が生じる。現時点では中近世スペイン語をレンマ化もしくはステミングする技術はないので，性数の異なる名詞・形容詞・冠詞の形や，人称や時制の異なる動詞の形は別の単語だと認識されてしまい，素性数の削減ができない。

三つ目は，単語 *n*-gram に比べ，スパースネスの度合いが小さい点である。文字 *n*-gram は，単語 *n*-gram に比べ出現頻度が高いため，安定的で頑健な特徴となる。特に，古文書のように文書長が短い文書を用いる場合，文字以外に出現頻度の高い素性を抽出することは困難である。また，出現頻度が高いので，誤字脱字や誤植などの散発的な誤り（ノイズ）に大きく影響されない。

四つ目は，文字 *n*-gram は，言語の違い，内容（語根），形態統語論的特徴（接辞や語幹），無意識的な表出行動といった様々なレベルの特徴を捉えることができる点である（Koppel *et al.* 2009；Koppel *et al.* 2011；Daelemans 2013；Kestemont 2014；Sapkota *et al.* 2015）。たとえば，um_や_sp という 3-gram の頻度が高いことは，文書がラテン語であることを意味している[18]。また，表記の揺れ（ni に対して nj や ny），略記法（que に対して q<ue>）（Ueda 2013b），単語の分割方法（de la に対して dela）などの無意識的な表出行動（村上 1994：145）は，クラスを区別するのに有用な特徴だと考えられている。ただし，すべての文字 *n*-gram が言語学的に上手く解釈できるわけではない。

文字 *n*-gram を抽出するために，以下の八つの前処理を各文書のパレオグラフィカに対して行う。クリティカは，使用しない。

1. 文頭記号（#）・文末記号（$）を挿入
2. 大文字を小文字へ変換（Sepan → sepan）
3. アクセント記号付の母音からアクセント記号を削除（áéíóú → aeiou）
4. アルファベット以外の文字（アラビア数字や.:;,!/-^?=*|(){}[]&）を削除
5. 文書中の明示的な作成年代や作成場所を削除
6. フォリオの番号（{h 1r}），行番号（{1}），本文への注釈箇所（[interlineado…], [margen:…]）を削除
7. 略記法の部分<…>を@へ置換（q<ue> → q@）
8. 二個以上のスペースの連続を一つに変換

一つ目の処理は，後述の *n*-gram 言語モデルで必要となる処理である。二つ目の処理は大文字と小文字の区別をしないこと，三つ目の処理は母音のアクセント記号を無視すること，四つ目の処理はアルファベット以外は無視することを意味している。五番目の処理は，作成年代や作成場所に関する明示的な情報は用いずに年代推定・場所推定を行うためのも

[18] スペイン語には um で終わる単語は基本的には存在しないが，ラテン語では第二変化名詞の対格の語尾が um となる（DOMINUM「支配者」など）。また，スペイン語には sp で始まる単語は基本的には存在しないが，ラテン語には存在する（SPATIUM「空間」など）。したがって，um_や_sp の存在は，文書がラテン語であることを示唆している。

のである。該当部分は，筆者があらかじめマニュアルでマーキング（[!...!]）しておいた。六番目の処理は，本文についてのメタな情報を無視することを意味している。七番目の処理は，省略部分の内容にかかわらず，省略の存在のみを情報として用いるということを意味している。八番目の処理は，二個以上のスペースの連続は一つとみなすことを意味している。以下に，前処理前と前処理後の文書 ID1（作成年代 : 1251 年，県 : Sevilla，自治州 : AN）の一部を示す。前処理前の太字部分は削除箇所を表している。

前処理前

{h 1r} {1} Connoscida cosa sea a todos los q<ue> esta carta uieren como yo don Fferrando por la gr<aci>a de dios Rey de Castiella de Toledo de Leon de Gallizia de Seuilla de Cordoua de Murc<ia> & de Jahen enbie **{2}** mis cartas auos el Conceio de Guadalfaiara q<ue> enbiassedes u<uest>ros om<n>es buenos （中略） Et ma<n>do & deffiendo firme mie<n>tre q<ue> ni<n>guno no<n> sea osado de uenir cont<ra> esta mi carta nj<n> de quebrantar la nj<n> de me<n>guar la en ni<n>guna cosa ca el q<ue>lo fiziesse aurye la yra de dios **&** la **{34}** mia. & pechar mie en coto. mill. m<o>r<abedis>. **[[it:lat.:it] : [!ffacta carta ap<u>d Sibilla Reg<e> exp<rimente>. xiij. die Ap<ri>lis.!]** J<ohannes>. pet<ri> de Berla<n>ga fe<ci>t. **[!ERA_M_CC_Lxxx_Nona!]]**

前処理後

#connoscida cosa sea a todos los q@ esta carta uieren como yo don fferrando por la gr@a de dios rey de castiella de toledo de leon de gallizia de seuilla de cordoua de murc@ de jahen enbie mis cartas auos el conceio de guadalfaiara q@ enbiassedes u@ros om@es buenos （中略） et ma@do deffiendo firme mie@tre q@ ni@guno no@ sea osado de uenir cont@ esta mi carta nj@ de quebrantar la nj@ de me@guar la en ni@guna cosa ca el q@lo fiziesse aurye la yra de dios la mia pechar mie en coto mill m@r@ j@ pet@ de berla@ga fe@t$

上記の前処理により，文字の 1-gram（種類数）は 32（アルファベット 26 文字，ñ，ç，@，_（スペース），#（文頭記号），$（文末記号）），2-gram の総数は最大 1024（= 32^2），3-gram の総数は最大 32768（= 32^3）となる。表 4.2 に，文字 2-gram の分布の一部を示す。

ID	1	2	3	4	5	6	7
年代	1251	1260	1262	1277	1278	1285	1295
県	Sevilla	Córdoba	Sevilla	Burgos	Segovia	Burgos	Valladolid
自治州	AN	AN	AN	CL	CL	CL	CL
a_	180	107	169	92	31	151	112
a#	0	0	0	0	0	0	0
a@	17	8	2	1	7	10	6
aa	0	0	0	0	0	0	0
ab	4	7	16	2	0	8	0
ac	0	2	7	0	1	0	0
aç	0	0	0	0	0	0	1
ad	35	26	60	10	12	24	33
ae	5	5	7	5	1	7	5
af	0	3	0	0	0	0	0
ag	6	13	9	15	1	24	6
ah	0	4	4	4	1	6	2
ai	0	2	4	1	0	5	2
aj	0	0	0	0	0	0	0
ak	0	0	0	0	0	0	0
al	73	45	82	43	16	52	36
am	12	12	30	8	0	12	14
an	31	45	94	40	8	50	48
añ	0	0	0	0	0	0	0
ao	1	0	0	0	0	0	0
ap	5	0	7	0	0	0	0
aq	11	0	11	0	0	3	3
ar	57	37	71	25	15	43	29
as	105	34	109	9	7	40	62
at	9	2	8	3	3	3	8
au	22	7	27	7	0	18	8
av	0	0	0	0	0	0	0
ax	0	0	0	0	0	0	0
ay	2	7	35	10	3	8	8
az	8	7	9	5	4	15	4

表 4.2　文字 2-gram の分布

4.2　文献学的特徴

本節では，スペイン語文献学の知見に基づき設定した文献学的素性セットについて説明する。この素性セットは，ナイーブベイズ多変数ベルヌーイモデル（第 8 章）で用いる。

まず，交替変数（alternating variables）と変異形（variant）を定義する。交替変数とは，意味や機能や効果の違いを生まない互いに異なる二つ以上の言語的表現の集合のことである（Kawasaki 2013b，2014a，2014b，2015b，2015c；Grieve 2016： 3.1）。変異形とは，ある交替変数について定義され，その交替変数に属する各々の言語的表現のことである。たとえば，動詞 tener「持つ」の点過去 1 人称単数という交替変数には，語幹母音が/u/となる **tuv**e と，語幹母音が/o/となる **tov**e の二つの変異形が存在する。同様に，硬口蓋鼻音/ɲ/を表す文字という交替変数には，少なくとも<ny>，<yn>，<nn>の三つの変異形が存在する。それぞれの交替変数に属する変異形の出現頻度は，年代，地点，その他の条件により異なるであろうが，基本的な意味や機能や効果の違いは存在しないとみなすことができる。

本研究では，交替変数の変異形の各々を一つの素性とし，すべての交替変数の変異形の集合を文献学的素性セットと呼ぶことにする。交替変数には，文書の内容や種類に依存しないと考えられるものを選んだ。本研究の目的は，文書の内容ではなく，文書の言語学的特徴によって，年代推定・場所推定を行うことだからである。この点は，文書の内容に大きく依存する単語 *n*-gram ではなく，文字 *n*-gram や品詞 *n*-gram を素性として用いる著者推定のタスクと共通している。したがって，交替変数には，文書の内容に大きく依存する内容語ではなく，動詞の活用語尾や機能語などの形式的なものが多く含まれる。ただし，文書の内容に依存するかしないかは程度問題であり，文書の内容に全く依存しないというわけではない。また，その認定は，研究者により揺れる可能性がある。

文献学的素性セットを用いる利点は，以下の二点である。

一つ目は，各特徴が文献学的に解釈できる点である。文字 *n*-gram は，必ずしも文献学的に意味のある解釈ができるわけではない。素性の理解のしやすさは，現象のモデル化における重要な要因の一つである。

二つ目は，文書の内容に依存しない文献学的素性セットを用いることで，古文書から抽出した素性で，古文書以外の文献史料の年代推定・場所推定もできる可能性があるということである。ただし，文献学的特徴が文書の内容や種類に依存しないというのは一つの仮定にすぎず，異なる立場の考えも存在する（Kabatek 2008）。

一方，文献学的素性セットを用いる短所は，以下の四点である。

一つ目は，出現頻度が小さいため，安定した統計値を得ることが難しい点である。出現頻度の大きい文字 *n*-gram に比べると，統計値の信頼性は低いと言わざるを得ない。したがって，出現頻度が高い素性を用いた場合に比べ，推定精度が低下する可能性がある，本研究では，カーネル平滑化により，出現頻度が小さいという問題に対処している。

二つ目は，言語ごとに素性セットを決定する必要がある点である。当然ながら，対象言語についての知識が必要とされる。これに対し，文字 *n*-gram は，言語に依存せず機械的に抽出が可能で，しかも網羅性もある。

三つ目は，特定の年代や地点において言語変異が小さいもしくは存在しない場合，年代推定・場所推定が困難になる点である。差異が存在しなければ，区別することはできない。また，一般に，言語変化の速度は緩慢なため，細かい粒度での推定が難しくなる。

四つ目は，文献学的素性セットは，年代推定・場所推定に役立つかもしれない文書の内容の変異という情報は捨ててしまっている点である。これに対し，文字 *n*-gram は，文書の内容の変異も，ある程度，捉えることができる。

本研究では，スペイン語史の先行研究（主に，Zamora Vicente 1967；Menéndez Pidal 1999；Penny 2002；Azofra 2009）に基づき，115 個の交替変数，273 個の変異形を設定した。ただし，重要な変異をすべて網羅できているわけではないと思われるので，今後，漏れを埋めていくことが必要である。表 4.3 は，文献学的素性セットの一覧を示している。

カテゴリー	交替変数		素性番号	変異形
形態統語論的特徴（名詞・形容詞・代名詞・冠詞）	f1	名詞「回」	f1a	vegada
			f1b	vez
	f2	形容詞「別の」	f2a	otro
			f2b	altro
	f3	形容詞「同じ」	f3a	mesmo
			f3b	meísmo
			f3c	mismo
	f4	形容詞・副詞「多くの」,「多く」	f4a	muncho
			f4b	mucho
	f5	不定代名詞「別のひと」	f5a	otri
			f5b	otre
			f5c	otrie
	f6	不定形容詞「ある～」，不定代名詞「誰か」	f6a	alguno
			f6b	dalguno
	f7	複合不定形容詞「どんな～」	f7a	qual manera quier
			f7b	qualquier manera
	f8	否定の不定代名詞	f8a	nada
			f8b	ren
	f9	否定の不定形容詞	f9a	nenguno
			f9b	ninguno
	f10	1 人称単数の主格代名詞	f10a	yo
			f10b	jo
			f10c	you
			f10d	eu
	f11	1 人称複数の主格・前置詞格代名詞	f11a	nós
			f11b	nosotros
	f12	2 人称複数の直接・間接目的格・再帰代名詞	f12a	vos
			f12b	os
	f13	1 人称複数と 2 人称複数の所有形容詞	f13a	nuestro
			f13b	nosso
	f14	1 人称複数・2 人称複数の間接目的格代名詞と 3 人称直接目的語の連辞	f14a	vos lo
			f14b	volo
	f15	3 人称間接目的格代名詞と 3 人称直接目的格代名詞の連辞	f15a	gelo
			f15b	selo
			f15c	lelo
	f16	不定詞に 3 人称直接・間接目的語が続く連辞	f16a	comprárlo
			f16b	comprállo
	f17		f17a	connusco

カテゴリー	交替変数		素性番号	変異形
		前置詞 con「～と」と 1 人称複数もしくは 2 人称複数の代名詞の連辞	f17b	con nós(otros)
	f18	前置詞句「私と」	f18a	comigo
			f18b	conmigo
	f19	近称の指示形容詞・代名詞	f19a	este
			f19b	aqueste
	f20	近称の指示形容詞・代名詞の語頭母音	f20a	iste
			f20b	este
	f21	3 人称単数・複数間接目的格代名詞の語末母音	f21a	le
			f21b	li
	f22	定冠詞の語頭母音の保持・消失	f22a	ela carta
			f22b	la carta
	f23	男性単数の定冠詞	f23a	el rey
			f23b	lo rey
	f24	/-n/や/-r/で終わる前置詞と/l-/で始まる定冠詞の連辞	f24a	en la
			f24b	ena
	f25	所有詞と名詞の連辞	f25a	la mía casa
			f25b	mi casa
			f25c	la casa mía
	f26	所有者が 1 人称単数の所有詞前置形の男性形	f26a	meu
			f26b	miou
			f26c	meo
			f26d	mio
			f26e	mi
	f27	所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の男性形	f27a	teu
			f27b	tou
			f27c	tuo
			f27d	to（男性形）
			f27e	tu
	f28	所有者が 1 人称単数の所有詞前置形の女性形	f28a	ma
			f28b	mia
			f28c	mie
			f28d	mi
	f29	所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の女性形	f29a	ta
			f29b	tua
			f29c	tue
			f29d	to（女性形）
			f29e	tu

カテゴリー	交替変数		素性番号	変異形
	f30	3 人称複数の所有形容詞	f30a	su
			f30b	lur
	f31	従属節における目的格代名詞と否定の副詞・主格代名詞の語順	f31a	que lo non mandó
			f31b	que non lo mandó
	f32	前置詞句における目的格代名詞と不定詞の語順	f32a	de lo hacer
			f32b	de hacerlo
(動詞)	f33	動詞 dar「与える」などの直説法現在 1 人称単数	f33a	dó
			f33b	doy
	f34	ラテン語の起動相に由来する動詞の直説法現在 1 人称単数と接続法現在	f34a	conosco
			f34b	conozco
	f35	半子音[j]の存在する音節の前の開母音[ɛ]が二重母音化	f35a	tengo
			f35b	tiengo
	f36	-er 動詞と-ir 動詞の線過去と過去未来の語尾	f36a	avía
			f36b	avié
	f37	andar, estar, tener, haber, placer, saber などの PYTA	f37a	ove
			f37b	uve
	f38	poder, poner の PYTA の PYTA	f38a	podiere
			f38b	pudiere
	f39	estar と andar の PYTA	f39a	estide
			f39b	estude
	f40	ser の PYTA	f40a	fui
			f40b	sove
	f41	traer の PYTA	f41a	traxe
			f41b	truxe
			f41c	troxe
	f42	hacer と venir の PYTA	f42a	fezo
			f42b	fizo
	f43	decir や traer の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形	f43a	dixieron
			f43b	dixeron
	f44	ser と ir の点過去	f44a	fuimos
			f44b	fuemos
	f45	ser と ir の PYTA	f45a	fuere
			f45b	fore
			f45c	fure
	f46	直説法点過去において弱変化する-er 動詞，-ir 動詞の 1 人称複数と 2 人称複数の語尾	f46a	comiemos
			f46b	comimos
	f47	直説法点過去の 2 人称複数の語尾	f47a	-stes
			f47b	-steis

カテゴリー	交替変数		素性番号	変異形
	f48	-ar 動詞の点過去 1 人称単数，3 人称単数	f48a	mandéi
			f48b	mandé
	f49	-ar 動詞の点過去 2 人称単数	f49a	comprastes
			f49b	comprestes
	f50	直説法点過去 3 人称複数の語尾	f50a	-aron
			f50b	-oron
	f51	-er 動詞と-ir 動詞の直説法点過去 3 人称複数,	f51a	vieron
		接続法過去，接続法未来の活用形	f51b	viron
	f52	tener などの直説法未来と過去未来	f52a	terné
			f52b	tenré
			f52c	tendré
	f53	ser の接続法現在	f53a	sía
			f53b	sea
			f53c	seya
	f54	valer などの直説法現在 1 人称単数と接続法現	f54a	vala
		在	f54b	valga
	f55	saber の接続法現在	f55a	sepa
			f55b	saba
	f56	接続法未来の 1 人称単数	f56a	oviero
			f56b	oviere
	f57	接続法未来の 1 人称複数と 2 人称複数	f57a	fuéremos
			f57b	fuermos
	f58	ver の語幹	f58a	veer
			f58b	ver
	f59	ser の語幹	f59a	seer
			f59b	ser
	f60	decir，morir などの不定形	f60a	dezir
			f60b	dizer
	f61	haber の 1 人称単数	f61a	é
			f61b	éy
	f62	haber の 1 人称複数，2 人称複数	f62a	avemos
			f62b	hemos
	f63	動詞「する」	f63a	fazer
			f63b	fer
	f64	動詞「置く」	f64a	dexar
			f64b	lexar
	f65	-er 動詞と-ir 動詞の現在分詞	f65a	aviendo
			f65b	oviendo

カテゴリー	交替変数		素性番号	変異形
	f66	-er 動詞と-ir 動詞の過去分詞	f66a	avido
			f66b	ovido
	f67	-er 動詞の過去分詞	f67a	-ido
			f67b	-udo
	f68	decir の過去分詞	f68a	decho
			f68b	dicho
	f69	強勢が後ろから二番目の音節にある動詞の直説法現在，直説法未来，接続法現在の 2 人称複数の語尾	f69a	amades
			f69b	amaes
			f69c	amáis
	f70	強勢が後ろから三番目の音節にある動詞の直説法現在，直説法未来，接続法現在の 2 人称複数の語尾	f70a	amábades
			f70b	amábais
	f71	関係節における未来の法	f71a	接続法未来
			f71b	直説法未来
（前置詞・副詞・接続詞）	f72	否定の接続詞	f72a	ne
			f72b	ni
			f72c	nen
			f72d	nin
	f73	期間を表す接続詞	f73a	mientre
			f73b	mientra(s)
	f74	譲歩の接続詞	f74a	maguer
			f74b	aunque
	f75	理由を表す接続詞	f75a	ca
			f75b	porque
	f76	条件を表す接続詞	f76a	si
			f76b	se
	f77	副詞を作る接尾辞	f77a	-mente
			f77b	-miente
			f77c	-mientre
	f78	前置詞・副詞「～まで」	f78a	fata
			f78b	fasta
	f79	前置詞「～のために」	f79a	para
			f79b	pora
	f80	前置詞「～なしに」	f80a	sin
			f80b	sen
			f80c	sien
			f80d	sienes
			f80e	sines

カテゴリー	交替変数		素性番号	変異形
	f81	前置詞・副詞「〜によると」	f81a	segundo
			f81b	segúnd
			f81c	segúnt
			f81d	según
	f82	否定の副詞	f82a	non
			f82b	no
	f83	副詞「とても」	f83a	muit
			f83b	muy
	f84	副詞「より〜」	f84a	más
			f84b	mais
	f85	副詞「そのように」	f85a	así
			f85b	ansí
			f85c	asín
	f86	副詞・接続詞・前置詞「〜のように」	f86a	como
			f86b	cuemo
			f86c	cumo
	f87	副詞「今」	f87a	agora
			f87b	ahora
	f88	副詞「後に」	f88a	empués
			f88b	aprés
			f88c	depués
			f88d	después
	f89	副詞「そのとき」	f89a	entonces
			f89b	estonces
	f90	場所を表す関係副詞	f90a	ó
			f90b	do
			f90c	ond(e)
			f90d	dond(e)
	f91	副詞「一緒に，同時に」	f91a	juntamente
			f91b	ensemble
形態音韻論的特徴	f92	単数形が二重母音/-ei/で終わる名詞の複数形	f92a	leys
			f92b	leyes
	f93	強勢短母音 ĕ の二重母音化	f93a	es
			f93b	yes
	f94	ラテン語の指小辞-ĕllŭ	f94a	capiella
			f94b	capilla
	f95	語末の母音/e/の保存・脱落	f95a	part
			f95b	parte

カテゴリー	交替変数		素性番号	変異形
	f96	最終音節の後舌母音	f96a	conventu
			f96b	convento
	f97	/kt/や/(u)lt/の子音連続	f97a	feito
			f97b	fecho
	f98	破裂音/p, b, k, g/もしくは摩擦音/f/と流音/l/の子音群	f98a	obrigar
			f98b	obligar
	f99	ラテン語の音連続-l-+[j]	f99a	fijo
			f99b	fillo
			f99c	fiyo
	f100	-d'c-，-t'c-の子音連続	f100a	judgar
			f100b	juzgar
			f100c	julgar
	f101	-m'n-の子音連続	f101a	nomne
			f101b	nomre
			f101c	nombre
	f102	-b't-，-p't-，-v't-などの子音連続	f102a	cibdad
			f102b	ciudad
	f103	母音間の-d-の保持・消失	f103a	odir
			f103b	oír
	f104	/d/の後ろの語末の/-e/の保持・消失	f104a	verdade
			f104b	verdad～verdat
	f105	複数形の語末	f105a	todos
			f105b	toz
	f106	/-a/で終わる女性形の名詞などの複数形の語末母音	f106a	coses
			f106b	cosas
	f107	20 と 30 を表す数詞	f107a	treinta
			f107b	trinta
			f107c	trenta
	f108	40～90 の 10 の倍数を表す数詞	f108a	quaraenta
			f108b	quarenta
			f108c	quaranta
表記的特徴	f109	語末の歯音の表記	f109a	mercet
			f109b	merced
	f110	不定形容詞 alguno「何らかの」の男性単数形の語尾の/-n/の表記	f110a	algúnd～algúnt
			f110b	algún
	f111	硬口蓋鼻音/ɲ/の表記	f111a	anyo
			f111b	ayno
			f111c	anno

カテゴリー	交替変数	素性番号	変異形
	f112　硬口蓋側面接近音/ʎ/の表記	f112a	ellyo
		f112b	eyllo
		f112c	ello
	f113　/kwa/，/gwa/の表記	f113a	qual
		f113b	quoal
	f114　鼻音＋両唇破裂音の表記	f114a	ambos
		f114b	anbos
	f115　非語源的<h->の表記	f115a	ordenar
		f115b	ordenar

表 4.3　文献学的素性セットの一覧

交替変数は，便宜的に，形態統語論的特徴，形態音韻論的特徴，表記的特徴の三つのカテゴリーに分類した。当然ながら，本研究の分類方法が最適なものとは限らない。しかし，本研究は，文献学的特徴の分類方法を提唱することを目指しているわけではない。また，分類方法の如何は，実験結果に全く影響しない。したがって，暫定的な分類で満足することにする。

検索には，コーパスの二つのバージョンを，適宜，使い分けた。文字の変異を見たい場合はパレオグラフィカを用いた。一方，/i/を表す<i>～<j>～<y>，/u/を表す<v>～<u>，/b/を表す<b>～<v>～<u>などのように，同じ音価を表す文字の変異は考慮せず変異を見たい場合はクリティカを用いた。検索は，自作のプログラムで，正規表現によるパターンマッチングにより実行した。現時点では，テキストに品詞タグやレンマなどの情報は付与されていないので，語順などの複雑なパターンを伴う変異の分析は行わなかった。

以下，使用する記号や表記上の注意を説明する：

- ラテン語は，ĕgō のように，小型英大文字で表記する。ĂĔĬŎŬ は短母音を，ĀĒĪŌŪ は長母音を表している。ラテン語の語源や母音の長短は，主に，水谷（2009）と Real Academia Española（2016）を参照した。
- ラテン語に由来する名詞類の語源は，基本的には，対格形で表すことにする。
- 形容詞は，特に断りのない場合は，男性単数形で代表させた。
- 動詞の活用形は，特に断りのない場合は，1 人称単数で代表させた。
- x~y は，x と y が互いに変異形であることを表している。
- x < y や y > x は，x が y を継承した形であることを表している。
- *x は，x が再建形であることを表している。
- /xxx/は，音韻的表記である。
- [xxx]は，音声学的表記である
- <x>は，文字（書記素）である。
- áéíóú は強勢母音を表している。
- [ǰ]は，ma**y**o の/maǰo/に見られるような，中硬口蓋摩擦音（mid-palatal fricative）を表している。
- 歯茎摩擦音[s]に対し，[s̪]は歯摩擦音を表している。
- -M(Ĩ)N-のような表記は，カッコ内の要素が消失することを表している。
- 「’」は母音消失を表している：-M’N- < -M(Ĩ)N-。

- 例文の後ろのカッコは，（文書 ID，作成年，作成県，作成自治州，文書の種類）を表している。

4.2.1　形態統語論的特徴

形態統語論的特徴とは，形態や統語に関して変異のあるものである。形態統語論的特徴は，便宜的に，名詞・形容詞・代名詞・冠詞，動詞，前置詞・副詞・接続詞の三種類に分類した。

4.2.1.1　名詞・形容詞・代名詞・冠詞

f1　名詞「回」

f1a　**vegada**

「回」を表す名詞には，変異形 vegada < VĪCĀTĂ が存在する：

… pechar nos ía en pena mill maravedís de la moneda nueva a cadaúno por cada *vegada* que contra ello les fuere o pasase en cualquier manera …（ID127，1340 年，Sevilla，AN，Cancilleresco）

f1b　**vez**

「回」を表す名詞には，変異形 vez < VĪCĔ が存在する：

… fasta el día que bos la faga sana e tantas *bezes* sea tenuda de vos pagar la dicha pena cuantas *bezes* en ella cayere …（ID137，1347 年，Cáceres，EX，Particular）

f2　形容詞「別の」

f2a　**otro**

「別の」を表す形容詞には，ĂL-が/o-/になった変異形 **o**tro < ĂLTRŬ が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.）：

…e ante los señores su presidente e oidores de su real consejo questán e residen en su corte e ante *otros* cualesquier justicias e juezes de cualesquier partes de los reinos e señoríos de su magestad ….（ID1489，1557 年，Sevilla，AN，Municipal）

f2b　**altro**

「別の」を表す形容詞には，ĂL-が保持された変異形 **al**tro < ĂLTRŬ が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.）：

… que *altre* ningún omen nin fema daquest segle ý donas de trebut nin de cens nin per ninguna *altra* manera nin razón daquest segle …（ID936，1357 年，Navarra，NA，Eclesiástico）

f3　形容詞「同じ」

f3a　mesmo

「同じ」を表す形容詞には，強勢音節の母音が/e/の変異形 m**e**smo < MĔD-ĬPSĬ(SSĬ)MŬ が存在する（Menéndez Pidal 1999：§ 98$_2$；Penny 2002：3.5.3.1）：

… contra cada vno de nos contra n<uest>ros bjenes por todo rrigor de derecho q<ue>les cayamos por ese *mesmo* fecho enla pena delos dichos veynte mjll m<a>r<auedi>s …（ID386，1412 年，Cáceres，EX，Eclesiástico）

f3b　meísmo

「同じ」を表す形容詞には，母音連続/ei/のある変異形 m**ei**smo < MĔD-ĬPSĬ(SSĬ)MŬ が存在する（Menéndez Pidal 1999：§ 98$_2$；Penny 2002：3.5.3.1）：

… pella gr<aci>a d<e> dios abbadessa del monesterio de Carrizo he el <con>uento desse *meismo* logar damos e otorgamos auos do<n> Monjo garcia …（ID459，1259 年，León，CL，Particular）

f3c　mismo

「同じ」を表す形容詞には，強勢音節の母音が/i/の変異形 m**i**smo < MĔD-ĬPSĬ(SSĬ)MŬ 存在する（Menéndez Pidal 1999：§ 98$_2$；Penny 2002：3.5.3.1）：

… maestro de la caualleria de sa<n>t Jague co<n> todo el cabildo de los freyres dessa *misma* orden el cambio es atal q<ue> do yo auos el Castiello la villa de veas ….（ID1200，1239 年，Burgos，CL，Cancilleresco）

f4　形容詞・副詞「多くの」，「多く」

f4a　muncho

「多くの」，「多く」を表す形容詞・副詞には，/n/が挿入された変異形 mu**n**cho が存在する（Zamora Vicente 1967：341, 361；Menéndez Pidal 1999：§ 69$_2$）：

… y esta çibdad Reç<i>bio *muncho* benefiçio como av<uest>ra alteza constara por las ynformaçiones que çerca dello sean fecho …（ID1387，1552 年，Madrid，MA，Cancilleresco）

f4b　mucho

「多くの」，「多く」を表す形容詞・副詞には，変異形 mucho < MŬLTŬ が存在する（Zamora Vicente 1967：341, 361；Menéndez Pidal 1999：§ 69$_2$）：

… despues q<ue> ouiemos uistas las cartas de la una part de la otra despues de *muchos* razonamie<n>tos todo el pleyto fue librado desta guisa q<ue> los freyres se partiero<n> de villa ….（ID1205，1243 年，Valladolid，CL，Cancilleresco）

f5　不定代名詞「別のひと」

f5a　otri

不定代名詞「別のひと」には，語末母音が/i/となる変異形 otr**i** < ĂLTRŬ が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.；Paredes García 2015）。otr**i** の語末母音/i/は，疑問詞・関係詞 qu**i**「誰」に影響された形であると考えられる：

… Et nos nj *otri* por nos nj<n> por Raço<n> d<e> nos nj<n> successores n<uest>ros non uos podamos toyller las d<i>tas …（ID895，1305 年，Navarra，NA，Eclesiástico）

f5b　**otre**

不定代名詞「別のひと」には，語末母音が/e/となる変異形 otr**e** < ĂLTRŭ が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.；Paredes García 2015）。otr**e** の語末母音/e/は，男性単数の指示形容詞・代名詞 est**e** などに影響された形であると考えられる：

… saluo si ffuere escripto entre los rrenglones desta Et rrazo<n> q<ue> digamos nos nj<n> *otre* por nos ant<e> qual q<ui>er Juez q<ue> nos no<n> uala …（ID93，1301 年，Ávila，CL，Particular）

f5c　**otrie**

不定代名詞「別のひと」には，語末母音が/ie/となる変異形 otr**ie** < ĂLTRŭ が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.；Paredes García 2015）。otr**ie** の語末母音/ie/は，疑問詞・関係詞の qu**ie**n「誰」に影響された形であると考えられる：

… E de todas las costas E dapn<n>os E menos cabos q<ue> vos o *otrie* por vos sobre esta dicha Razo<n> fizierdes …（ID1045，1490 年，Cádiz，AN，Particular）

f6　**不定形容詞「ある～」，不定代名詞「誰か」**

f6a　**alguno**

不定形容詞「ある～」，不定代名詞「誰か」には，それぞれ，変異形 alguno < ĂLĬQUĔ ŪNŭ，alguien（alguno が quien に影響された形）が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.；Zamora Vicente 1967：175）：

…Si *algún* omne aquesto que nós fiziemos quisier crebantar sea maledicto …（ID238，1224 年，Palencia，CL，Eclesiástico）

f6b　**dalguno**

不定形容詞「ある～」，不定代名詞「誰か」には，それぞれ，語頭に/d-/の付加された変異形 **d**alguno < /d-/ + alguno，**d**alguien < /d-/ + alguien が存在する（Menéndez Pidal 1999：§ 102$_3$；Penny 2002：3.5.5；Azofra 2009：3.7.1.1.；Zamora Vicente 1967：175）：

…E si *dalguno* omne también del so liñage como de outro estraño esti feito quisier quebrantar o corromper ….（ID435，1244 年，Asturias，AS，Eclesiástico）

f7　**複合不定形容詞「どんな～」**

f7a　**qual manera quier**

複合形容詞「どんな～」には，**qual** manera **quier** のように，qual と quier で名詞を挟んだ変異形が存在する（Pato 2012）：

… E cuanto derecho en esta dita viña ey o podría aver en *cual manera quier* todo lo renuncio e lo parto de mí e dólo a vós el dicho Joán Martínez …（ID787，1325 年，Zamora，CL，Particular）

f7b　**qualquier manera**

複合形容詞「どんな～」には，**qualquier** manera のように，名詞が qualquier の後ろに来る変異形が存在する（Pato 2012）：

… E si alguno o algunos ý oviere que contra ello les quissiere passar o venir en *cualquier manera* pechar nos ía en pena mil maravedís de la moneda nueva（ID134，1346 年，Zamora，CL，Cancilleresco）

f8　否定の不定代名詞

f8a　**nada**

否定の不定代名詞として，変異形 nada < (RĔM) NĀTĂ が存在する（Penny 2002：3.5.5, 5.1.1）：

… somos ende pagados e *nada* non remanece de dar escontra nós E de iste día adelantre sea tierra de nuestro poder …（ID517，1245 年，Salamanca，CL，Eclesiástico）

f8b　**ren**

否定の不定代名詞として，変異形 ren < RĔM (NĀTĂ)が存在する（Penny 2002：3.5.5, 5.1.1）：

…pero en tal manera que vós no ajades poder de vender e ni dempeñar ni dailenar en *ren* …（ID852，1238 年，Navarra，NA，Eclesiástico）

f9　否定の不定形容詞

f9a　**nenguno**

否定の不定形容詞には，語頭音節の母音が/e/の変異形 **ne**nguno < NĔC ŪNŬ が存在する（Menéndez Pidal 1999 : § 102$_2$; Penny 2002：3.5.5；Azofra 2009：3.7.1.2.）。ne**n**guno の/n/は ne**n** からの影響である：

… Et nenguno q<ue> esta mi carta crebantare nin menguare en *nenguna* cosa aya la yra de dios et la mia et peche ami …（ID1209，1241 年，Sevilla，AN，Cancilleresco）

f9b　**ninguno**

否定の不定形容詞には，語頭音節の母音が/i/の変異形 **ni**nguno < NĔC ŪNŬ が存在する（Menéndez Pidal 1999 : § 102$_2$; Penny 2002：3.5.5；Azofra 2009：3.7.1.2.）。ni**n**guno の/n/は ni**n** からの影響である：

… no<n> sea osado de hir contra esta mi carta deste mio donadio nin de q<ue>brantar la nin de minguar la en *ninguna* cosa …（ID1208，1253 年，Sevilla，AN，Cancilleresco）

f10　**1 人称単数の主格代名詞**

f10a　**yo**

1 人称単数の主格代名詞には，語頭子音が/ʝ/となる変異形 **y**o < ĔGŌ が存在する（Zamora Vicente 1967 : 101, 168-170 ; Menéndez Pidal 1999：§ 93$_1$）：

Sepan cuantos esta carta vieren cómo *yo* don Alfonso por la gracia de Dios rey de Castilla de León Toledo de Galizia de Sevilla …（ID1366，1281 年，León，CL，Cancilleresco）

f10b　**jo**

1 人称単数の主格代名詞には，語頭子音が/ʒ/となる変異形 **jo** < ĔGŌ が存在する（Zamora Vicente 1967 : 101, 168-170 ; Menéndez Pidal 1999：§ 93$_1$）：

In Dei nomine Sabuda cosa sia a totz homes que *jo* dona Cizilia dAlvernia priora del monasteri de Sant Peire de Ribes …（ID1495, 1282 年，Navarra，NA，Eclesiástico）

f10c　**you**

1 人称単数の主格代名詞には，変異形 you < ĔGŌ が存在する（Zamora Vicente 1967：101, 168-170；Menéndez Pidal 1999：§ 93$_1$）：

In Dei nomine amen Conozuda cousa sea a cuantos esta carta viren e oíren que *you* García Martínez taullero de Colombranos dou e outorgo cuanto derecho ey …（ID428，1274 年，León，CL，Eclesiástico）

f10d　**eu**

1 人称単数の主格代名詞には，変異形 eu < ĔGŌ が存在する（Zamora Vicente 1967：101, 168-170；Menéndez Pidal 1999：§ 93$_1$）：

Saban cuantos esta carta viren e oíren que *eu* don Arias pella gracia de Deus abat de Sant Andrés con el convento desse mismo lugar damos e outorgamos …（ID458，1270 年，León，CL，Eclesiástico）

f11　**1 人称複数の主格・前置詞格代名詞**

f11a　**nós**

1 人称複数の主格・前置詞格代名詞には，変異形 nós < NŌS が存在する（Menéndez Pidal 1999：§ 93$_1$；Penny 2002：3.5.1.1；Azofra 2009：3.6.1.4., 3.6.4）。2 人称複数の主格・前置詞格代名詞にも vós～vosotros の変異が存在したが，vós には 2 人称複数に加え，敬称の 2 人称単数としての用法もあり検索が複雑なため除外する：

… fago pleito con Martín Loçano de Goçón e con so mujer doña Ilana de la nuestra heredad que *nós* avemos ý de tierras de viñas e de casas …（ID419，1243 年，Palencia，CL，Eclesiástico）

f11b　**nosotros**

1 人称複数の主格・前置詞格代名詞には，変異形 nosotros < nós + otros が存在する（Menéndez Pidal 1999：§ 93$_1$；Penny 2002：3.5.1.1；Azofra 2009：3.6.1.4., 3.6.4）：

… toda la heredat e casas e huertos e todas las otras cosas que *nosotros* por nombre del dicho monesterio avemos en término de Val de Rey ….（ID432，1414 年，Valladolid，CL，Eclesiástico）

f12　**2 人称複数の直接・間接目的格・再帰代名詞**

f12a　**vos**

2 人称複数の直接・間接目的格・再帰代名詞には，変異形 vos < VŌS が存在する（Menéndez Pidal 1999：§ 94$_1$；Penny 2002：3.5.1）：

… e nós tovímoslo por bien Por que *vos* mandamos que del día questa nuestra carta os fuere notificada …（ID1448，1526 年，Sevilla，AN，Cancilleresco）

f12b　os

2 人称複数の直接・間接目的格・再帰代名詞には，vos の語頭の/b/が脱落した変異形 os < vos が存在する（Menéndez Pidal 1999：§ 94_1 ; Penny 2002：3.5.1）：

… e nós tovímoslo por bien Por que vos mandamos que del día questa nuestra carta *os* fuere notificada …（ID1448，1526 年，Sevilla，AN，Cancilleresco）

f13　1 人称複数と 2 人称複数の所有形容詞

f13a　nuestro

1 人称複数と 2 人称複数の所有形容詞には，変異形 nuestro < NŌSTRŬ，vuestro < VŌSTRŬ が存在する（Menéndez Pidal 1999：§ 51_1）：

… Abbad de Sant Fagunt hy el co<n>ujento deste mismo logar Fazemos las seellar co<n> *nuestros* sellos …（ID417，1236 年，León，CL，Eclesiástico）

f13b　nosso

1 人称複数と 2 人称複数の所有形容詞には，子音群/-str-/が/-s/-となった変異形 nosso < NŌSTRŬ，vosso < VŌSTRŬ が存在する（Menéndez Pidal 1999：§ 51_1）：

… ta<n> bien arras de *nossa* madre como h<er>edat de *nossos* ermanos como de *nossos* h<er>ederos fillos ye nietos ye nietas ye fillas …（ID761，1254 年，Asturias，AS，Eclesiástico）

f14　1 人称複数・2 人称複数の間接目的格代名詞と 3 人称直接目的語の連辞

f14a　vos lo

1 人称複数と 2 人称複数の間接目的格代名詞 nos と vos に 3 人称直接目的語（lo，la，los，las）が続く連辞には，前者の語尾の/s/が，後者の語頭の/l/に同化しない変異形が存在する：

… E si vos fueren demandadas o embargadas e yo non *vos las* fizier sanas que vos peche en pena cada día diez moravedís de la moneda sobredicha …（ID413，1323 年，Salamanca，CL，Eclesiástico）

f14b　volo

1 人称複数と 2 人称複数の間接目的格代名詞 nos と vos に 3 人称直接目的語（lo，la，los，las）が続く連辞には，/vos lo/ > /vollo～volo/のように，前者の語尾の/s/が，後者の語頭の/l/に同化して/ʎ/もしくは消失する変異形が存在する：

… e averemos daquí en delante de vos defender e amparar con ello en todo tiempo de quienquier que *vollo* demandasse o *vollo* contrallasse dello …（ID115，1301 年，Salamanca，CL，Particular）

f15　3 人称間接目的格代名詞と 3 人称直接目的格代名詞の連辞

f15a　gelo

3 人称単数・複数の間接目的格代名詞 le，les に 3 人称の単数・複数直接目的格代名詞 lo，la，los，las が続く連辞には，gelo，gela，gelos，gelas となる変異形が存在する（Menéndez Pidal 1999: § 94_3 ; Penny 2002: 3.5.1 ; Azofra 2009 :

3.6.2.；Ueda 2011）。これは，以下のプロセスによる：まず，ラテン語の 3 人称単数の間接目的格代名詞 ĭLLĪ に 3 人称の単数・複数直接目的格代名詞（たとえば，ĭLLŭM）が続く場合，/ll/ > /l/により，ĭLLĪ ĭLLŭ > /(e)ljelo/となった（ĭLLĪ ĭLLŭ が一語だとみなされ ĭLLĪ の Ī が語末から語中となったため，Ī > /e/ではなく Ī > /i/となった。単独では，ĭLLĪ > /(i)le/ > le となる）。続いて，語頭の/e/の脱落と/l/+[j] > /ʒ/により，/(e)ljelo/ > gelo となった。間接目的格代名詞が複数の場合（たとえば ĭLLĪS ĭLLŭ），ĭLLĪS の語末の/s/が半子音[j]の形成を妨げるため，*les lo となるはずであるが，単数形からの影響で gelo となった。したがって，gelo では間接目的語の単数・複数は中和される：

… para le dar fe e testimonio de lo que ante mí pasase e que *ge lo* diese por testimonio para guarda e conservación de su derecho …（ID1043，1483 年，Jaén，AN，Particular）

f15b　**selo**

3 人称間接目的格代名詞と 3 人称直接目的格代名詞の連辞には，変異形 selo が存在する。これは gelo の/ʒ/が無声化し/ʃ/となり，さらに/ʃ/が/s/と混同された形である（Menéndez Pidal 1999: § 94$_3$；Penny 2002: 3.5.1；Azofra 2009：3.6.2.；Ueda 2011）：

… e mando que *se la* paguen En las mandas que fago por Dios e por mi ánima …（ID173，1481 年，Guadalajara，CM，Judicial）

f15c　**lelo**

3 人称間接目的格代名詞と 3 人称直接目的格代名詞の連辞には，lelo のように，音変化が生じない変異形が存在する（Menéndez Pidal 1999: § 94$_3$；Penny 2002: 3.5.1；Azofra 2009： 3.6.2.；Ueda 2011）：

… yo mandéi*lela* dar en escrito e mandéi poner en ella mio seyello pendiente （ID586，1273 年，León，CL，Eclesiástico）

f16　**不定詞に 3 人称直接・間接目的語が続く連辞**

f16a　**comprárlo**

不定詞に 3 人称直接・間接目的語（le，lo，la，les，los，las）が続く連辞には，comprar+ lo > comprá**rl**o のように，変化が生じない変異形が存在する（Menéndez Pidal 1999：§ 94$_5$；Penny 2002：3.7.9.1；Azofra 2009：3.6.3.）：

… y allí estuvieron parados gran rato atalayando sin *poderlo* descubrir y después que se fueron salió el dicho don Hernando con hombres de a pie（ID1295，1569 年，Granada，AN，Particular）

f16b　**comprállo**

不定詞に 3 人称直接・間接目的語（le，lo，la，les，los，las）が続く連辞には，comprar+ lo > comprá**ll**o のように，不定詞の語尾の/-ɾ/が目的語の語頭の/l-/に同化して/ʎ/となる変異形が存在する（Menéndez Pidal 1999：§ 94$_5$；Penny 2002：3.7.9.1；Azofra 2009：3.6.3.）：

… Y por quel tesorero de la dicha iglesia a cuyo cargo están las campanas della y el *mandallas* tañer hizo tañer la queda en la dicha campana por avérselo mandado su cabildo …（ID1387，1552 年，Madrid，MA，Cancilleresco）

f17 前置詞 con「～と」と 1 人称複数もしくは 2 人称複数の代名詞の連辞

f17a connusco

前置詞 con「～と」の補語が 1 人称複数もしくは 2 人称複数の代名詞のとき，変異形 connusco < CŬM NŌSCŬM, convusco < CŬM VŌSCŬM が存在する（Menéndez Pidal 1999：§ 93_1;Penny 2002：3.5.1；Azofra 2009：3.6.1.5）：

… yo Loreinte Domingo fijo de don Laín de Serranos de Avianos otorgo e coñosco que fago tal postura *convusco* don Yagüe fijo de don Adeva que vos vendo todo cuanto heredamiento yo e mi muger avemos …（ID67，1285 年，Ávila，CL，Particular）

f17b con nós(otros)

前置詞 con「～と」の補語が 1 人称複数もしくは 2 人称複数の代名詞のとき変異形 con nos(otros)，con vos(otros)が存在する（Menéndez Pidal 1999：§ 93_1;Penny 2002：3.5.1；Azofra 2009：3.6.1.5）。それぞれ，前置詞 con と主格代名詞 nós(otros)，vós(otros)からなる形である：

… nós la priora e el conviento de Cañas fazemos cambio *con vós* don Pero Xeménez clérigo de Çarratón e dámosvos el nuestro solar que sale del nuestro palacio （688，1282 年，La Rioja，LR，Eclesiástico）

f18 前置詞句「私と」

f18a comigo

「私と」を表す前置詞句には，変異形 comigo（<CŬM MĒCŬM）が存在する（Menéndez Pidal 1999：§ 93_1;Penny 2002：3.5.1；Azofra 2009：3.6.1.5）：

… e prometo de enterrar ý mio cuerpo e mando ý *comigo* el quinto de cuanto mueble oviero al tiempo que yo finaro …（ID715，1284 年，Cantabria，CB，Eclesiástico）

f18b conmigo

「私と」を表す前置詞句には，変異形 co**n**migo が存在する（Menéndez Pidal 1999：§ 93_1;Penny 2002：3.5.1；Azofra 2009：3.6.1.5）。/n/のある co**n**migo は，co**n**tigo と co**n**sigo からの類推で作られたである：

… estando en la villa de Valladolit en las Cortes que agora ý fiz seyendo ý *conmigo* la reína doña María mi madre e el infante don Joán mio tío e mio adelantado mayor …（ID502，1307 年，Valladolid，CL，Cancilleresco）

f19 近称の指示形容詞・代名詞

f19a este

近称の指示形容詞・代名詞には，変異形 este < ĬSTĔ が存在する（Menéndez Pidal 1999：§ 98_3；Penny 2002：3.5.3.1；Azofra 2009：3.4.1.；Enrique-Arias 2012：101-103）：

… E porque lo suso dicho se pueda mejor guardar por *esta* dicha mi carta mando a vós las dichas justicias …（ID21，1471 年，Segovia，CL，Cancilleresco）

f19b aqueste

近称の指示形容詞・代名詞には，ĔCCŬM によって強化された変異形 aqueste < ĔCCŬ(M) ĬSTĔ が存在する（Menéndez Pidal 1999：§ 98_3；Penny 2002：3.5.3.1；Azofra 2009：3.4.1.；Enrique-Arias 2012：101-103）：

… E *aquestas* renunciaciones renunciamos en el caso pressent en tal manera que …（ID951，1458 年，Teruel，AR，Particular）

f20 近称の指示形容詞・代名詞の語頭母音

f20a iste

近称の指示形容詞・代名詞には，**i**ste < ĭSTĔ のように，語頭母音が/i-/となる変異形が存在する（Zamora Vicente 1967 : 176；Menéndez Pidal 1999：§ 99）：

… E de *iste* día adelante sea esta heredade de mio poder quita e in vuestro poder metida e confirmada …（ID508，1245 年，Salamanca，CL，Eclesiástico）

f20b este

近称の指示形容詞・代名詞として，**e**ste < ĭSTĔ のように，語頭母音が/e-/となる変異形が存在する（Zamora Vicente 1967：176；Menéndez Pidal 1999：§ 99）：

… E mandamos e defendemos que ninguno non sea osado de venir contra *este* nuestro privilegio pora crebantarlo nin pora minguarlo en ninguna cosa …（ID3，1262 年，Sevilla，AN，Cancilleresco）

f21 3 人称単数・複数間接目的格代名詞の語末母音

f21a le

3 人称単数・複数間接目的格代名詞には，語末母音が/e/となる変異形 l**e** < ĭLLĪ，l**e**s < ĭLLĪS が存在する。また，男性単数の指示代名詞・形容詞として，語末母音が/e/となる est**e**，aquest**e** などが存在する（Zamora Vicente 1967：111, 253, 25；Menéndez Pidal 1999：§ 93_{3n}, § 94_3）：

… E los omnes que allí poblaren o moraren que sean sus vassallos quitamientre e *le* sirvan e l*e* obedezcan e *le* fagan todas las cosas que vassallos solariegos deven fazer a señor …（ID48，1283 年，Ávila，CL，Municipal）

f21b li

3 人称単数・複数間接目的格代名詞として，語末母音が/i/となる l**i** < ĭLLĪ，l**i**s < ĭLLĪS が存在する。また，男性単数の指示代名詞・形容詞として，語末母音が/i/となる est**i**，aquest**i** などが存在する（Zamora Vicente 1967：111, 253, 25；Menéndez Pidal 1999：§ 93_{3n}, § 94_3）：

… e devemos *li* dar XXX soldos cada año pora su vestir en toda su vida（ID864，1279 年，Navarra，NA，Eclesiástico）

f22 定冠詞の語頭母音の保持・消失

f22a ela carta

ラテン語の対格に由来する定冠詞には，語頭の/e-/が保持された変異形 **e**lo < ĭLLŭ，**e**la < ĭLLĂ，**e**los < ĭLLŌS，**e**las < ĭLLĀS が存在する（Zamora Vicente 1967：166-167；Menéndez Pidal 1999：§ 100_3）：

… E sopiendo e entendiendo nós que la gracia e *ela* almosna que se da al <mon>asterio de Sant Çalvador de Leire que es servicio de Dios（ID893，1301 年，Navarra，NA，Eclesiástico）

f22b　**la carta**

ラテン語の対格に由来する定冠詞には，語頭の/e-/が消失した変異形 lo < elo，la < ela，los < elos，las < elas が存在する（Zamora Vicente 1967：166-167；Menéndez Pidal 1999：§ 100_3）：

… E del molino quel ganó ell abade don Martino sediendo en aquella casa Los qui tovieron *la* casa depués delle tovieron el molino por de la casa …（ID155，1229 年，Burgos，CL，Particular）

f23　**男性単数の定冠詞**

f23a　**el rey**

男性単数の定冠詞には，ラテン語の主格に由来する変異形 el < ĭLLĕ が存在する（Zamora Vicente 1967：166-167；Menéndez Pidal 1999：§ 100_3）：

… Otrosí *el* abat e *el* conviento prometieron al concello de Carcastiello ellos esto fiziendo de mantenerlos en lures fueros …（ID977，1281 年，Navarra，NA，Judicial）

f23b　**lo rey**

男性単数の定冠詞には，ラテン語の対格に由来する変異形 lo < ĭLLŭ が存在する（Zamora Vicente 1967：166-167；Menéndez Pidal 1999：§ 100_3）：

… E jo la dita priora e *lo* convent otorgam que si del I de vós dos devén en est coméi antz del sobredit terme …（ID1495，1282 年，Navarra，NA，Eclesiástico）

f24　***/-n/や/-r/で終わる前置詞と/l-/で始まる定冠詞の連辞***

f24a　**en la**

/-n/で終わる前置詞 en や con と/l-/で始まる定冠詞（たとえば la）の連辞には，en la のように，/l-/が/-n/に同化しない変異形が存在する。同様に，/-r/で終わる前置詞 per や por と/l-/で始まる定冠詞（たとえば la）の連辞には，por la のように，/-r/が/l-/に同化しない変異形が存在する（Zamora Vicente 1967：159-160, 166-167; Menéndez Pidal 1999：§ 100_4）：

… muchos servicios que fizieron a nós e a nuestro linage dámosles e otorgámosles que fagan dos ferias *en la* villa sobredicha de Guadalfajara por siempre jamás…（ID2，1260 年，Córdoba，AN，Cancilleresco）

f24b　**ena**

/-n/で終わる前置詞 en や con と/l-/で始まる定冠詞（たとえば la）の連辞には，en la > e**nn**a のように，/l-/が/-n/に同化する変異形が存在する。同様に，/-r/で終わる前置詞 per や por と/l-/で始まる定冠詞（たとえば la）の連辞には，por la > po**ll**a のように，/-r/が/l-/に同化する変異形が存在する（Zamora Vicente 1967 : 159-160, 166-167; Menéndez Pidal 1999：§ 100_4）：

… que lo ayades possedeades vendades donedes empenedes fagades delo toda vuestra volontad assí *ena* vida como *ena* morte …（ID780，1281 年，Zamora，CL，Eclesiástico）

f25 所有詞と名詞の連辞

f25a la mía casa

所有詞と名詞の連辞には，la mía casa のように，「定冠詞＋所有詞の前置形＋名詞」となる変異形が存在する（Menéndez Pidal 1999：§ 27, 95；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… E desto mandé dar a los de Cañas esta carta seellada con *el mio sello* de plomo …（ID673，1304 年，Burgos，CL，Cancilleresco）

f25b mi casa

所有詞と名詞の連辞には，mi casa のように，「定冠詞なし＋所有詞の前置形＋名詞」となる変異形が存在する（Menéndez Pidal 1999：§ 27, 95；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… yo el dicho Gil Pérez dóvos esta carta sellada con *mio sello* colgado …（ID672，1303 年，La Rioja，LR，Eclesiástico）

f25c la casa mía

所有詞と名詞の連辞には，la casa mía のように，「定冠詞＋名詞＋所有詞の後置形」となる変異形が存在する（Menéndez Pidal 1999：§ 27, 95；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… fiziemos fazer dos cartas semejables partidas por abecé e ambas seelladas con tres seellos con *el seello mío* de la chancellería（ID1200，1239 年，Burgos，CL，Cancilleresco）

f26 所有者が 1 人称単数の所有詞前置形の男性形

f26a meu

所有者が 1 人称単数の所有詞前置形の男性形には，語末母音が/-eu/となる変異形 m**eu** < MĔŬ が存在する（Zamora Vicente 1967：173-174）：

… de oye adelante e la dou e la entrego a Morerola como desuso é dicho e nen eu nen mia moller nin *meu* fillo nin omne de mia parte non seermos poderosos de la demandar ya maes …（ID1234，1255 年，Zamora，CL，Eclesiástico）

f26b miou

所有者が 1 人称単数の所有詞前置形の男性形には，語尾が/-ou/となる変異形 mi**ou** が存在する。miou は，tou < TŬŬ，sou < SŬŬ に影響された形である（Zamora Vicente 1967：173-174；Menéndez Pidal 1999：§ 27，§96$_1$）：

… dou e outorgo cuanto derecho ey enos mulinus que furon de *miou* padre don Martín Fernández e de mia madre doña Tereisa Sángez eno río de Sil cerca Ponferrada …（ID428，1274 年，León，CL，Eclesiástico）

f26c meo

所有者が 1 人称単数の所有詞前置形の男性形には，語尾が/-eo/となる変異形 m**eo** < MĔŬ が存在する（Menéndez Pidal 1999：§ 96$_2$）：

In Dei nomine et eius gratia Conocida cosa sea por este escrito que yo doña Pascuala e *meos* filios Domínico García e Fernando todos III de mancomún facemus karta de vendemiento …（ID517，1245 年，Salamanca，CL，Eclesiástico）

f26d　**mio**

所有者が 1 人称単数の所有詞前置形の男性形として，語尾が/-io/となる変異形 **mio** < MĔŬ が存在する（Menéndez Pidal 1999：§ 27，§ 96_1；Penny 2002：3.5.2；Azofra 2009：3.5.）：

…Ye esto fazemos yo Pedro Rodríguez ensembra con mia muyer doña María ye con *mios* fiyos ye con todos los otros herederos que desuso son nomrados …（ID1232，1243 年，Zamora，CL，Eclesiástico）

f26e　**mi（男性形）**

所有者が 1 人称単数の所有詞前置形の男性形として，語尾が/-i/となる変異形 **mi** が存在する（Menéndez Pidal 1999：§ 27，§ 96_1；Penny 2002：3.5.2；Azofra 2009：3.5.）：

Sepan cuantos esta carta vieren cómo yo Assensio Muñoz morador en Eça otorgo sobre mí e sobre todos *mis* bienes muebles e raízes los que agora é e abré cabadelante …（ID546，1279 年，Segovia，CL，Eclesiástico）

f27　**所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の男性形**

f27a　**teu**

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の男性形には，それぞれ，語尾が/-eu/となる変異形 **teu** < TŬŬ，**seu** < SŬŬ が存在する（Zamora Vicente 1967：173-174）：

…doule al prior de San Marcos e al convento desse menesmo lugar un corral con suas casas e con *seu* ixido e con todo *seu* pertenezemento en que mora Domingo Azedo …（ID1372，s.a.，s.l.，SL，Eclesiástico）

f27b　**tou**

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の男性形には，それぞれ，語尾が/-ou/となる変異形 **tou** < TŬŬ，**sou** < SŬŬ が存在する（Zamora Vicente 1967：173-174；Menéndez Pidal 1999：§ 27，§ 96_1）：

… dámusvos e outorgamus ela nuestra jugaría de Langre que ya de la nuestra cozina con todos *sous* dereitos cuantos le pertenezen en monte e en villa …（ID430，1256 年，Asturias，AS，Eclesiástico）

f27c　**tuo**

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の男性形には，それぞれ，語尾が/-uo/となる変異形 **tuo** < TŬŬ，**suo** < SŬŬ が存在する（Menéndez Pidal 1999：§ 96_2）：

… vendemos esto que avemos dicto al abat don Gonzalvo dAguilar e a *suo* convento e somos pagados de pretio …（ID221，1208 年，Palencia，CL，Eclesiástico）

f27d　**to（男性形）**

所有者が 2 人称単数と 3 人称単数の所有詞前置形の男性形には，それぞれ，語尾が/-o/となる変異形 **to** < TŬŬ，**so** < SŬŬ が存在する（Menéndez Pidal 1999：§ 27, 96_1；Penny 2002：3.5.2）：

…. E todo omne de *so* linage o de otra parte que esta carta crebantar sea maldito e dexcomungado amen …（ID334，1239 年，León，CL，Eclesiástico）

f27e　tu（男性形）

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の男性形として，それぞれ，語尾が/-u/となる変異形 t**u**, s**u** が存在する（Menéndez Pidal 1999：§ 27, 96$_1$；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… E dóvos la alcaría de Belnagech con su torre e con *sus* términos e con sus pertenencias …（ID649，1248 年，Sevilla，AN，Cancilleresco）

f28　所有者が 1 人称単数の所有詞前置形の女性形

f28a　ma

所有者が 1 人称単数の所有詞前置形の女性形には，語尾が/-a/となる変異形 m**a** < MĚĂ が存在する（Zamora Vicente 1967：50-51）：

… si per aventura fusa cosa que yo fallissei en toz les diz temps de *ma* vida o en algún dels que non laborasei la dita viña de les dites labos …（ID982，1358 年，Navarra，NA，Eclesiástico）

f28b　mia

所有者が 1 人称単数の所有詞前置形の女性形には，語尾が/-ia/となる変異形 m**ia** < MĚĂ が存在する（Zamora Vicente 1967：173-174；Menéndez Pidal 1999：§ 27，§ 96；Penny 2002：3.5.2；Azofra 2009：3.5.）：

…. dou e outorgo cuanto derecho ey enos mulinus que furon de miou padre don Martín Fernández e de *mia* madre doña Tereisa Sángez eno río de Sil cerca Ponferrada al monesterio de Santandrés dEspinareda …（ID428，1274 年，León，CL，Eclesiástico）

f28c　mie

所有者が 1 人称単数の所有詞前置形の女性形には，語尾が/-ie/となる変異形 m**ie** < m**ia** < MĚĂ が存在する（Zamora Vicente 1967：173-174；Menéndez Pidal 1999：§ 27，§ 96；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… fago almosna al hospital de los pobres de Sand Fagund de todas las *mies* viñas que yo compré en Sand Fagund …（ID421，1244 年，León，CL，Eclesiástico）

f28d　mi（女性形）

所有者が 1 人称単数の所有詞前置形の女性形には，語尾が/-i/となる変異形 m**i** < mie < mia が存在する（Menéndez Pidal 1999：§ 27, 96$_1$；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… E con *mi* mano propria róbrovos e confírmovos esta carta …（ID1209，1241 年，Sevilla，AN，Cancilleresco）

f29　所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の女性形

f29a　ta

所有者が 2 人称単数, 3 人称単数・複数の所有詞前置形の女性形には, それぞれ, 語尾が/-a/となる変異形 t**a** < TŬĂ, s**a** < SŬĂ が存在する（Zamora Vicente 1967：50-51）：

… E nós les diz Martín de Saldias, ortelán de Pampalona, e María d'Uani *sa* muller, estaranz en la pobla dels Horz de la Madalena, otorgam e venem de conoissutz e de manifest que recebem a trebut les dites viñes de vós …（ID936，1357 年，Navarra，NA，Eclesiástico）

f29b　**tua**

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の女性形には，それぞれ，語尾が/-ua/となる変異形 **tua** < TŬĂ，**sua** < SŬĂ が存在する（Zamora Vicente 1967：173-174；Menéndez Pidal 1999：§ 27，§ 96；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… E nós devandichos don Monio García e *sua* mulier doña Teresa Martínez prometemos e otorgamos de aprovechar esta pernomnada heredat …（ID459，1259 年，León，CL，Particular）

f29c　**tue**

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の女性形には，それぞれ，語尾が/-ue/となる変異形 **tue** < TŬĂ，**sue** < SŬĂ が存在する（Zamora Vicente 1967：173-174；Menéndez Pidal 1999：§ 27，§ 96；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… E por que esto sea firme damos a don Peidro e a *sue* fija Mari Pérez esta carta seellada con el nuestro seello …（ID685，1252 年，La Rioja，LR，Eclesiástico）

f29d　**to（女性形）**

所有者が 2 人称単数と 3 人称単数の所有詞前置形の女性形には，それぞれ，男性形の to，so が用いられる変異形が存在する（Menéndez Pidal 1999：§ 27, § 96_1；Penny 2002：3.5.2；Azofra 2009：3.5.）：

… E si Martín Loçano o *so* mujer se repintieren deste pleito que pechen XXX moravedís …（ID419，1243 年，Palencia，CL，Eclesiástico）

f29e　**tu（男性形）**

所有者が 2 人称単数，3 人称単数・複数の所有詞前置形の女性形には，それぞれ，語尾が/-u/となる変異形 **tu**，**su** が存在する（Menéndez Pidal 1999：§ 27，§ 96_1；Penny 2002：3.5.2；Azofra 2009：3.5.）：

…. E dóvos la presa de los molinos que dizen Silvar con *su* torre e con su cortijo e veinte arençadas de viñas e cuatro arençadas de huerto redor de la villa…（ID1202，1248 年，Sevilla，AN，Cancilleresco）

f30　**3 人称複数の所有形容詞**

f30a　**su**

3 人称複数の所有形容詞には，ラテン語の所有形容詞に由来する変異形 su，seu < SŬŬ などが存在する（Menéndez Pidal 1999：§ 97_2；Penny 2002：3.5.2）：

…e por non caer en olvido lo mandaron los reyes poner en escrito en *sus* privilegios por que los otros que regnassen depués dellos e toviessen el so logar fuessen tenudos de guardar aquello …（ID389，1289 年，Valladolid，CL，Cancilleresco）

f30b　**lur**

3 人称複数の所有形容詞には，ラテン語の指示代名詞・形容詞 ĬLLĔ の複数属格に由来する変異形 lur < ĬLLŌRŬ が存在する（Menéndez Pidal 1999：§ 97_2；Penny 2002：3.5.2）：

… dixieron que non les fazían fuerça nin tuerto mas que abrían las puertas de *lur* eglesia e que rancavan la serraya porque le pudiesen fer clauf por que fuesen salvas las cosas …（ID1363，1317 年，Zaragoza，AR，Eclesiástico）

f31 従属節における目的格代名詞と否定の副詞・主格代名詞の語順

f31a que lo non mandó

接続詞 que や si で導かれる従属節には，que **lo non** mandó や que **lo yo** mandó のように，目的格代名詞（lo など）が否定の副詞 non～no や主格代名詞（yo など）に先行する変異形が存在する（Nieuwenhuijsen 2006；川崎 2014）：

… e *si lo non* quesieren *que lo non* podades vender a iglesia nin a monesterio nin a dueña nin a cavallero …（ID468，1464 年，León，CL，Eclesiástico）

f31b que non lo mandó

接続詞 que や si で導かれる従属節には，que **non lo** mandó や que **yo lo** mandó のように，否定の副詞 non～no や主格代名詞（yo など）が目的格代名詞（lo など）に先行する変異形が存在する（Nieuwenhuijsen 2006；川崎 2014）：

… con tal pleito e condición que las tengáis en buen paramiento e *que no las* podáis vender ni trocar ni cambiar sin ser primero requerido el dicho nuestro monesterio …（ID1345，1505 年，León，CL，Eclesiástico）

f32 前置詞句における目的格代名詞と不定詞の語順

f32a de lo hacer

de や por などの前置詞に導かれる前置詞句には，de **lo hacer** のように，直接・間接目的格代名詞が不定詞に先行する変異形が存在する（Nieuwenhuijsen 2006；川崎 2014）：

… E mando que les dedes toda la sal que hobieren menester *para los poner* en pro e salvo …（ID101，1524 年，Huelva，AN，Particular）

f32b de hacerlo

de や por などの前置詞に導かれる前置詞句には，de **hacerlo** のように，不定詞が直接・間接目的格代名詞に先行する変異形が存在する（Nieuwenhuijsen 2006；川崎 2014）：

… E mando e defiendo firmemientre que ninguno non sea osado de venir contra esta mi carta nin *de quebrantarla* nin *de menguarla* en ninguna cosa （ID1，1251 年，Sevilla，AN，Cancilleresco）

4.2.1.2 動詞

強変化の完了幹が現われる直説法点過去と関連時制（接続法過去（-ra 形と-se 形），接続法未来）を総称して PYTA（perfectos y tiempos afines）とよぶ（Maiden 2011）。特に断りのない場合は，PYTA は点過去の 1 人称単数で代表させた。たとえば，haber「～がある」の PYTA では，直説法点過去（hube），接続法過去（hubiera/hubiese），接続法未来（hubiere）で，完了幹 hub-を共有している。

f33 動詞 dar「与える」などの直説法現在 1 人称単数

f33a dó

動詞 dar「与える」，estar「～である」，ser「～である」，ir「行く」の直説法現在 1 人称単数には，dó < DŌ，está < STŌ，só < SŬM，vó < VĀDŌ のように，語尾に-y が付かない変異形が存在する（Penny 2002：3.7.8.1.5）：

… Yo *dó* a ellos I solar que fu de don Bartolomé fi de Domingo Calvo …（ID455，1245 年，León，CL，Eclesiástico）

f33b **doy**

動詞 dar「与える」，estar「～である」，ser「～である」，ir「行く」の直説法現在 1 人称単数には，do**y** < dó + y，esto**y** < estó + y，so**y** < só + y，vo**y** < vó + y のように，語尾に/-i/が付く変異形が存在する（Penny 2002：3.7.8.1.5）：

… otorgo e coñosco por esta carta que *doy* todo mi poder complido a Miguel Ferraz pellitero e a Diego García escrivano vezinos de Benavente …（ID784，1420 年，Zamora，CL，Particular）

f34 ラテン語の起動相に由来する動詞の直説法現在 1 人称単数と接続法現在

f34a **conosco**

動詞 conocer < CŎGNŌSCĔRĔ「知っている」のように不定形が-SCĔRĔ で終わるラテン語の起動相に由来する動詞の直説法現在 1 人称単数と接続法現在には，cono**sc**o/konosko/，cono**sc**a/konoska/のように，語幹末子音が/sk/となる変異形が存在する（Menéndez Pidal 1999：§ 112_3；Penny 2002：3.5.8.1.3）：

… otorgo E *conosco* q<ue> ve<n>do A vos el honorable sen<n>or don iohan Ruyz de narbaez …（ID1043，1483 年，Jaén，AN，Particular）

f34b **conozco**

動詞 conocer < CŎGNŌSCĔRĔ「知っている」のように不定形が-SCĔRĔ で終わるラテン語の起動相に由来する動詞の直説法現在 1 人称単数と接続法現在には，cono**zc**o/konoʂko/～/konoθko/，cono**zc**a/konoʂka/～/konoθka/のように，語幹末子音が/ʂk/～/θk/となる変異形が存在する。これは，cono**sc**o，cono**sc**a などの音節末の/s/（<s>で表記）が，cono**ç**es，cono**cí** などその他の活用形に見られる/ʂ/～/θ/（<ç>や<c>で表記）の影響で，/ʂ/～/θ/（<z>で表記）となった形である（Menéndez Pidal 1999：§ 112_3；Penny 2002：3.7.8.1.3）：

… Et yo Mig<ue>ll p<er>ez de Lardero el sobredicho otorgo *conozco* q<ue> vendo estas casas estas dos pieças esta vin<n>a sobredicha …（ID890，1287 年，Navarra，NA，Particular）

f35 半子音[j]の存在する音節の前の開母音[ɛ]が二重母音化

f35a **tengo**

動詞 tener「持つ」，venir「来る」の直説法現在 1 人称単数と接続法現在には，半子音[j]の存在する音節の前の開母音[ɛ]が二重母音化しない変異形 **te**ngo < [tɛn(j)o] <TĔNĔŌ，**ve**ngo < [βɛn(j)o] < VĔNĬŌ が存在する（Zamora Vicente 1967：89，98，181，218，258）：

… fiziemos fer dos cartas partidas por abc de las cuales yo fray Sancho *tengo* la una e yo Miguel Gonçálvez *tengo* la otra（ID855，1270 年，Navarra，NA，Particular）

f35b **tiengo**

動詞 tener「持つ」，venir「来る」の直説法現在 1 人称単数と接続法現在には，半子音[j]の存在する音節の前の開母音[ɛ]が二重母音化する変異形 **tie**ngo < [tɛn(j)o] <TĔNĔŌ，**vie**ngo < [βɛn(j)o] < VĔNĬŌ が存在する（Zamora Vicente 1967：89，98，181，218，258）：

… jo don Miguel de Uncastiello atorgo e *viengo* de manifiesto que é recebido e me *tiengo* bien pagado de los D soldos …（ID961, 1274 年，Navarra，NA，Particular）

f36 -er 動詞と-ir 動詞の線過去と過去未来の語尾

f36a avía

-er 動詞と-ir 動詞の線過去と過去未来の語尾には，それぞれ，av**ía** < HĂBÉ(B)ĂM，**-ía** < (HĂB)É(B)ĂM のように，語尾が/-ía/となる変異形が存在する（Menéndez Pidal 1999：§ 117；Penny 2002：3.7.8.3；Azofra 2009：4.4.3.）：

… E el dicho Domingo Gil deiziendo que los dichos heredamientos que los dichos Ferrando e Gonçalo *avían* en Val de Palacios que eran e *pertenecían* seer de los dichos Ferrando e Gonçalo …（ID280，1353 年，Cáceres，EX，Judicial）

f36b avié

-er 動詞と-ir 動詞の線過去と過去未来の語尾には，それぞれ，av**ié** < av**ía**，**-ié** < **-ía** のように，語尾が/-íe/となる変異形が存在する（Menéndez Pidal 1999：§ 117；Penny 2002：3.7.8.3；Azofra 2009：4.4.3.）。強勢が/-íe/と/-ié/のどちらかは綴りからは不明であるが，便宜上，/-ié/とみなすことにする：

… e dixo que *fazié* saber que Mateos Sánchez fijo de Sancho Blásquez que *avié* vendido la heredad e derecho que *avié* en el aldea de Passaron …（ID199，1351 年，Cáceres，EX，Particular）

f37 andar，estar，tener，haber，placer，saber などの PYTA

f37a ove

andar「歩く」，estar「～である」，tener「持つ」，haber「～がある」，placer「気に入る」，saber「知っている」などの PYTA には，幹母音が/o/になる変異形が存在する：and**o**ve（ove をモデルにした形），est**o**ve（ove をモデルにした形），t**o**ve（ove をモデルにした形），**o**ve < HABUĪ，pl**o**gue < PLACUĪ，s**o**pe < SAPUĪ（Menéndez Pidal 1999：§ 120；Penny 2002：3.7.8.6.2；Azofra 2009：4.5.3.）：

… porque en la verdad en el dicho concejo nunca *ovo* ni se hizo número descrivanos ni tal se provará ni mostrará …（ID1170，1526 年，Granada，AN，Municipal）

f37b uve

andar「歩く」，estar「～である」，tener「持つ」，haber「～がある」，placer「気に入る」，saber「知っている」などの PYTA には，幹母音が/u/になる変異形が存在する：and**u**ve，est**u**ve，t**u**ve，**u**ve，pl**u**gue，s**u**pe（Menéndez Pidal 1999：§ 120；Penny 2002：3.7.8.6.2；Azofra 2009：4.5.3.）：

… ell arçobispo de Toledo á en ell almaxerifadgo de Toledo el cual préstamo me *uvo* dado don Sancho arçobispo（ID481，1287 年，Toledo，CM，Eclesiástico）

f38 poder，poner の PYTA の PYTA

f38a podiere

完了幹の語幹母音が ŏ（PŎTU-，PŎSU-）である poder「～できる」，poner「置く」の PYTA には，幹母音が/o/になる変異形 p**o**diere，p**o**siere が存在する（Penny 2002：3.7.8.6.2；Azofra 2009：4.5.3.）：

.. e assí mesmo con condición que ninguno non *podiese* plantar en las dichas faças nin en algunas dellas ningund árbol …（ID759, 1466 年，Cantabria，CB，Eclesiástico）

f38b　**pudiere**

完了幹の語幹母音が ŏ（PŎTU-，PŎSU-）である poder「～できる」，poner「置く」の PYTA には，幹母音が/u/になる変異形 p**u**diere，p**u**siere が存在する（Penny 2002：3.7.8.6.2；Azofra 2009：4.5.3.）：

… E si redrar e amparar e defender e vos lo fazer todo sano non quisiere o non *pudiere* …（ID1082，1472 年，Cádiz，AN，Particular）

f39　**estar と andar の PYTA**

f39a　**estide**

estar「～である」と andar「歩く」の PYTA には，それぞれ，変異形 estide，andude が存在する（Menéndez Pidal 1999：§ 120；Penny 2002：3.7.8.6.2）。est**ide** は，STĔTĪ の変異形 STĒTĪ に由来する。and**ide** は，est**ide** をモデルにした形である：

… fallamos que el suelo de don Fernando que estava bien cuemo *estido* de primero e mandamos que *estidiesse* assí …（ID256, 1242 年，Palencia，CL，Eclesiástico）

f39b　**estude**

estar「～である」と andar「歩く」の PYTA には，それぞれ，p**ude** < PŎTUĪ をモデルにした変異形 est**ude**，and**ude** が存在する（Menéndez Pidal 1999：§ 120；Penny 2002：3.7.8.6.2）：

… E por razón que Domingo Pérez e María Pérez e su marido Pero Blasco non son en la tierra nin *estudieron* presentes a esta véndida dicha（ID98，1301 年，Ávila，CL，Particular）

f40　**ser の PYTA**

f40a　**fui**

ser「～である」の PYTA には，ラテン語から継承した変異形 fui < FŬĪ が存在する（Menéndez Pidal 1999：§ 120；Penny 2002：3.7.8.6.2）：

... De la cual dita partición e fiadura son testigos que presentes *fueron* en el logar don Pedro Gil de la Raga fillo de don García dOllo …（ID853，1266 年，Navarra，NA，Judicial）

f40b　**sove**

ser「～である」の PYTA には，ove < HABUĪ をモデルにした形として，変異形 s**ove** が存在する（Menéndez Pidal 1999：§ 120；Penny 2002：3.7.8.6.2）：

… E desto son testigos rogados de amas las partes que *sovieron* y que lo vieron e que lo oyeron ell alcalde do Yagüe de Cereso Sancho Ferrández de Paganos …（ID632，1287 年，La Rioja，LR，Eclesiástico）

f41 traer の PYTA

f41a traxe

traer「持ってくる」の PYTA には，幹母音が/a/になる変異形 tr**a**xe < TRAXĪ が存在する（Menéndez Pidal 1999 : § 120 ; Penny 2002 : 3.7.8.6.2）:

… han dedar los d<ic>hos sen<n>ores pan harto p<ar>a los carreteros que *traxieren* las Rentas al dicho conuento（ID326, 1501 年，Palencia，CL，Eclesiástico）

f41b truxe

traer「持ってくる」の PYTA には，幹母音が/u/になる変異形 tr**u**xe が存在する（Menéndez Pidal 1999 : § 120）:

… E agora sabed que delas d<ic>has aueriguaçiones q<ue>se *truxeron* detodo el Reyno para hazer la d<ic>ha yguala y Rep<ar>timi<ento>（ID1042，1564 年，Madrid，MA，Municipal）

f41c troxe

traer「持ってくる」の PYTA には，幹母音が/o/になる変異形 tr**o**xe が存在する（Menéndez Pidal 1999 : § 120 ; Penny 2002 : 3.7.8.6.2）:

… q<ue> corten madera rrama p<ar>a faz<er> puentes por do pasen ellos los ganados q<ue> *troxieren* …（ID131，1350 年，Sevilla，AN，Cancilleresco）

f42 hacer と venir の PYTA

f42a fezo

hacer「する」と venir「来る」の PYTA には，点過去 1 人称単数（fize < FĒCĪ，vine < VĒNĪ）を除いて，幹母音が/e/となる f**e**zo < FĒCIT, v**e**no < VĒNIT のような変異形が存在する（Penny 2002 : 3.7.8.6.2 ; Azofra 2009 : 4.5.3.）:

… Reuere<n>do sen<n>or padre abbad suso d<ic>ho al q<ua>l Rogamos q<ue>las escreviese o *feziese* escreujr las sygnase co<n> su sygno …（ID1360，1507 年，León，CL，Eclesiástico）

f42b fizo

hacer「する」と venir「来る」の PYTA には，点過去 1 人称単数（fize < FĒCĪ，vine < VĒNĪ）を除いて，幹母音が/i/となる f**i**zo < FĒCIT, v**i**no < VĒNI のような変異形が存在する（Penny 2002 : 3.7.8.6.2 ; Azofra 2009 : 4.5.3.）:

… al qual Rogamos quela escriviese o *fiziese* escrevir la signase consu Signo …（ID576, 1511 年, Salamanca, CL, Eclesiástico）

f43 decir や traer の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形

f43a dixieron

decir「言う」や traer「持ってくる」の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形には，dix**i**eron < *DĪXÉRUNT や trax**i**eron < *TRĂXÉRUNT（Penny 2002 : Table 3.39）のように，硬口蓋音/ʃ/の直後に半子音[j]が存在する変異形が存在する：

… E luego anbas las dichas p<ar>tes *dixieron* al dicho bachiller vicario q<ue> sobre Razon de vn pl<e>ito q<ue> entre ellos yera …（ID1377，1457 年，León，CL，Judicial）

f43b　**dixeron**

decir「言う」や traer「持ってくる」の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形には，dixeron < dix**ie**ron < *DĪXĔRUNT や traxeron < trax**ie**ron < *TRĂXĔRUNT（Penny 2002：Table 3.39）のように，dix**ie**ron や trax**ie**ron の半子音[j]が直前の硬口蓋音/ʃ/（<x>で表される）に吸収され消失した変異形が存在する：

… y luego el d<ic>ho sen<n>or prior monges co<n>bento del d<ic>ho mon<esterio> *dixero*<*n*> q<ue> vista la petiçio<n> delos suso d<ic>hos de cada vno dellos …（ID727，1503 年，Cantabria，CB，Eclesiástico）

f44　**ser と ir の点過去**

f44a　**fuimos**

動詞 ser「～である」と ir「行く」の点過去には，1 人称単数 f**ui**，2 人称単数 f**ui**ste，1 人称複数 f**ui**mos，2 人称複数 f**ui**stes のように，強勢母音が二重母音/ui/となる変異形が存在する（Menéndez Pidal 1999：§ 120$_5$; Zamora Vicente 1967：190, 274；Penny 2002：3.7.8.6.2）。3 人称単数（fue）・複数（fueron）では，変異は存在しない：

… no<n> *ffuimos* bie<n> pagados de todos estos m<o>r<auedis> ssobredich<o>s …（ID329，1290 年，Valladolid，CL，Eclesiástico）

f44b　**fuemos**

動詞 ser「～である」と ir「行く」の点過去には，1 人称単数 fu**e**，2 人称単数 f**ue**ste，1 人称複数 f**ue**mos，2 人称複数 f**ue**stes のように，強勢母音が二重母音/ue/となる変異形が存在する（Menéndez Pidal 1999：§ 120$_5$; Zamora Vicente 1967：190, 274；Penny 2002：3.7.8.6.2）。3 人称単数（fue）・複数（fueron）では，変異は存在しない：

… los cuales dichos dineros vós a nós diestes e pagastes E nós de vós los recibiemos onde *fuemos* bien pagados e somos ende manifiesto e del precio e del aliala …（ID891，1296 年，Navarra，NA，Eclesiástico）

f45　**ser と ir の PYTA**

f45a　**fuere**

動詞 ser「～である」と ir「行く」の PYTA には，1 人称単数以外で，強勢母音が/ue/となる変異形 f**ue**ste < FŬ(Ī)STĪ，f**ue** < FŬ(Ī)T，f**ue**ron < FŬ(Ē)RŬNT などが存在する（Menéndez Pidal 1999：§ 120$_5$）：

… El qui non *fuere* vezinu que de otra part viniere morar entre vós dé una cuarta de almud de pan cada un año a nós …（ID163，1237 年，Burgos，CL，Eclesiástico）

f45b　**fore**

動詞 ser「～である」と ir「行く」の PYTA には，1 人称単数以外で，強勢母音が/o/となる変異形 f**o**ste < FŬ(Ī)STĪ，f**o** < FŬ(Ī)T，f**o**ron < FŬ(Ē)RŬNT などが存在する（Menéndez Pidal 1999：§ 120$_5$）。これらの形は，ŭ の後ろの ī もしくは ē が脱落した形（FŬ(Ī)STĪ，FŬ(Ī)T，FŬ(Ī)MŬS，FŬ(Ī)STĪS，FŬ(Ē)RŬNT など）に由来する：

… E por esta dicha carta pedimos e rogamos a los joizes o joez alcalle o alcalles que oy día son o *foren* de aquí adelantre de los concellos e logares onde estovieren los dichos bienes que los ponga …（ID587，1446 年，Asturias，AS，Eclesiástico）

f45c　**fure**

動詞 ser「～である」と ir「行く」の PYTA には，1 人称単数以外で，強勢母音が/u/となる変異形 **fu**ste < FŬ(Ī)STĪ, **fu** < FŬ(Ī)T，**fu**ron < FŬ(Ē)RŬNT などが存在する（Menéndez Pidal 1999：§ 120$_5$）。これらの形は，ŭ の後ろの ĭ もしくは ē が脱落した形（FŬ(Ī)STĪ，FŬ(Ī)T，FŬ(Ī)MŬS，FŬ(Ī)STĪS，FŬ(Ē)RŬNT など）に由来する：

… yo mando e demáis mando que todos juizos e todos arbridios e todas sentencias que entre ellos *furan* atá aquí sobre este pleito que todos sean per aquí rematadas …（ID586，1273 年，León，CL，Eclesiástico）

f46　**直説法点過去において弱変化する-er 動詞，-ir 動詞の 1 人称複数と 2 人称複数の語尾**

f46a　**comiemos**

直説法点過去において弱変化する-er 動詞，-ir 動詞には，1 人称複数と 2 人称複数の語尾が，それぞれ，**-ie**mos，**-ie**ste となる変異形が存在する。これは，3 人称複数の語尾**-ie**ron に影響された形である（Penny 2002：3.7.8.6.1；Azofra 2009：4.4.4.）：

…. E por mayor firmedumbre deste pleito *fiziemos* fazer dos cartas semejables partidas por abecé e ambas seelladas con tres seellos con el seello mío.（ID1200，1239 年，Burgos，CL，Cancilleresco）

f46b　**comimos**

直説法点過去において弱変化する-er 動詞，-ir 動詞には，1 人称複数と 2 人称複数の語尾が，それぞれ，**-i**mos < -ĬMŬS と**-i**ste < -ĬSTĪS となる変異形が存在する（Penny 2002：3.7.8.6.1；Azofra 2009：4.4.4.）：

… E dímosvos la dicha carta de véndida que de las dichas tierras *fezimos* en vuestro poder （ID1362，1413 年，Cádiz，AN，Particular）

f47　**直説法点過去の 2 人称複数の語尾**

f47a　**-stes**

直説法点過去の 2 人称複数の語尾には，変異形-st**e**s < -STĪS が存在する（Azofra 2009：4.4.4.）：

El Rey Venerable y deuoto padre Vi la carta que me *scriuistes* a tres del presente con el Licenciado Huergo mi Capellan y el Abbad fray Juan de Pedraza …（ID1329，1591 年，Madrid，MA，Cancilleresco）

f47b　**-steis**

直説法点過去の 2 人称複数の語尾には，変異形-st**ei**s が存在する（Azofra 2009：4.4.4.）。これは，他の活用の 2 人称複数（たとえば，am**éi**s < amedes，com**éi**s < comedes，amar**éi**s < amaredes，amás**ei**s < amades など）からの類推で，-st**e**s の/e/が二重母音/éi/になった形である（Azofra 2009：4.4.4.）。この変化は，少なくとも，他の活用の 2 人称複数において，/-edes/ > /-éis/のように母音間の/d/が消失した後に生じたものである：

… p<ar>a que ssepueda aueriguar ssi *llevasteys* algo demassiado sopena quelo que de otra manera lleuaredes Lopagareys conel quatro tanto …（ID206，1625 年，Toledo，CM，Judicial）

f48 -ar 動詞の点過去 1 人称単数，3 人称単数

f48a mandéi

-ar 動詞の点過去 1 人称単数と 3 人称単数には，それぞれ，mand**éi** < MĂNDĀ(V)Ī，mand**óu** < MĂNDĀV(Ī)T のように強勢母音が二重母音/-éi/，/-óu/となる変異形が存在する（Zamora Vicente 1967：101, 185）：

… <E> yo *mandéil*ela dar en escrito e *mandéi* poner en ella mio seyello pendiente …（ID586, 1273 年, León, CL, Eclesiástico）

f48b mandé

-ar 動詞の点過去 1 人称単数と 3 人称単数には，mand**é** < MĂNDĀ(V)Ī，mand**ó** < MĂNDĀV(Ī)T のように強勢母音が単母音/-é/，/-ó/となる変異形が存在する（Zamora Vicente 1967：184）：

… E yo tove por bien de lo mandar saber e *mandé* por mi carta al obispo de Astorga e al abat de Carrazedo …（ID423，1270 年，León，CL，Eclesiástico）

f49 -ar 動詞の点過去 2 人称単数

f49a comprastes

-ar 動詞の点過去 2 人称単数には, compr**a**stes のように強勢母音が/á/となる変異形が存在する（Zamora Vicente 1967：184）：

… por razón del heredamiento que vós señor obispo sobredicho *comprastes* en Váguilafuente e en Turuégano para complir los testamentos de doña Sancha Gómez mi madre …（ID1174，1295 年，Segovia，CL，Eclesiástico）

f49b comprestes

-ar 動詞の点過去 2 人称単数には，compr**e**stes のように強勢母音が/é/となる変異形が存在する。これは，1 人称単数 compré に，ほかの活用において 2 人称単数を表す形態素/-s/が付いた形である（Zamora Vicente 1967：184）：

… me quito e me parto de todo este heredamiento sobredicho e de todo los bienes que ende nos *levestes* o vuestro padre do Rodrigo Fernández o vuestro hermano Ramir Rodríguiz …（ID1238，1278 年，Zamora，CL，Particular）

f50 直説法点過去 3 人称複数の語尾

f50a -aron

直説法点過去 3 人称複数の語尾には，pas**aron**，met**ieron**，pid**ieron** のように，/-aron/ < -ĀRŬNT，/-ieron/ < -ĔRŬNT となる変異形が存在する（Zamora Vicente 1967：184-185，259，269；Penny 2002：3.7.8.6.1）：

…　porque los prelados nos *rogaron* que las tres ayudas que nos *prometieron* a dar de sus vassallos por razón de la guerra …（ID589，1276 年，Burgos，CL，Cancilleresco）

f50b -oron

直説法点過去 3 人称複数の語尾には，pas**oron**，met**ioron**，pid**ioron** のように，3 人称単数（pas**ó**，met**ió**，pid**ió**）＋/-ron/となる変異形が存在する（Zamora Vicente 1967：184-185，259，269；Penny 2002：3.7.8.6.1）：

… Ot<ro>ssi les q<ui>tamos las penas las amparas en q<ue> *cayoro*<*n*> por Razo<n> delas entregas delos Judios ffasta el dia q<ue> esta carta es ffecha …（ID71，1288 年，Álava，PV，Cancilleresco）

f51　-er 動詞と-ir 動詞の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形

f51a　vieron

-er 動詞と-ir 動詞の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形には，**vie**ron < VĪDĒRŬNT のように，強勢母音が二重母音/ie/となる変異形が存在する（Díaz Moreno *et al*. 2015）：

Coñuçuda cosa sea a cuantos esta carta *vieren* que yo do Rodrigo prior de la casa de Santo Toribio con el conviento des mismo logar damos a vós…（ID591，1253 年，Cantabria，CB，Eclesiástico）

f51b　viron

-er 動詞と-ir 動詞の直説法点過去 3 人称複数，接続法過去，接続法未来の活用形において，**vi**ron < VĪDĒRŬNT のように，強勢母音が単母音/i/となる変異形が存在する（Díaz Moreno *et al*. 2015）：

In Dei nomine amen Conoçuda cosa seja a todos cuantos esta carta *viren* e *oíren* que yo Gutier García fago carta de vendición a vós Ruy Guibón …（ID441，1253 年，Asturias，AS，Eclesiástico）

f52　tener などの直説法未来と過去未来

f52a　terné

動詞 poner「置く」，tener「持つ」，venir「来る」の直説法未来と過去未来には，それぞれ，po**rné** < pon(e)ré，te**rné** < ten(e)ré，ve**rné** < ven(i)ré のように，幹母音の脱落と音位転換の生じた変異形が存在する（Menéndez Pidal 1999：§ 123；Penny 2002：3.7.8.4.3；Azofra 2009：4.6.2.）：

… E encara que vós ý dó e lexo poder e señoría que si aquellos que daquí adelant lo *ternán* no lo tenían como devían o non querían o non porían pagar los dichos III soldos …（ID625，1345 年，Teruel，AR，Judicial）

f52b　tenré

動詞 poner「置く」，tener「持つ」，venir「来る」の直説法未来と過去未来には，それぞれ，po**nré** < pon(e)ré，te**nré** < ten(e)ré，ve**nré** < ven(i)ré のように，幹母音の脱落が生じた変異形が存在する（Menéndez Pidal 1999：§ 123；Penny 2002：3.7.8.4.3；Azofra 2009：4.6.2.）：

… en caso que vós o quiquiere que por tiempo la dita viña *tenrá* e posedirá cessávades de pagar el dito trehúdo por el dito mes de agosto …（ID843，1385 年，Huesca，AR，Particular）

f52c　tendré

動詞 poner「置く」，tener「持つ」，venir「来る」の直説法未来と過去未来には，それぞれ，po**ndré** < pon(e)ré，te**ndré** < ten(e)ré，ve**ndré** < ven(i)ré のように，幹母音の脱落と/-d-/の挿入が生じた変異形が存在する（Menéndez Pidal 1999：§ 123；Penny 2002：3.7.8.4.3；Azofra 2009：4.6.2.）：

… dó yo a vós la dita viña a trehúdo a generación que vós o quiquiere que por tiempo la dita viña *tendrá* e posederá dedes e paguedes e siades tenidos dar e pagar a mí （ID843，1385 年，Huesca，AR，Particular）

f53 ser の接続法現在

f53a sía

動詞 ser「～である」の接続法現在には，強勢母音が/í/となる変異形 s**í**a < SĔDĔĂM が存在する（Zamora Vicente 1967 : 190, 264）：

… Item mandamos y queremos que por cadaúno de nós nos *sían* pagados e satisfechos todos nuestros deudos tuertos ….（ID757, 1522 年，Teruel，AR，Particular）

f53b sea

動詞 ser「～である」の接続法現在には，強勢母音が/é/となる変異形 s**e**a < SĔDĔĂM が存在する（Menéndez Pidal 1999 : § 113_a）：

… E si alguno contra este fecho vinier el abad de San Fagund que *sea* tenido de sanar este solar de todo omne a Pedro Palentia o a sos heredadores（ID455，1245 年，León，CL，Eclesiástico）

f53c seya

動詞 ser「～である」の接続法現在には，強勢母音が/é/で，この母音と次の音節の母音/a/の間に子音[j]が入る変異形 se**y**a～se**j**a < SĔDĔĂM が存在する（Zamora Vicente 1967：190, 264）：

… Se alguno contra este nuestro fecho venier en contrario *seya* maldito e escomungado e con Judas en inferno dampnado e peche a vós L morabedís …（ID570，1267 年，Asturias，AS，Eclesiástico）

f54 valer などの直説法現在 1 人称単数と接続法現在

f54a vala

動詞 caer「落ちる」，oír「聞こえる」，salir「出る」，traer「持ってくる」，valer「価値がある」などの直説法現在 1 人称単数と接続法現在には，cayo < CĂDŌ，caya < CĂDĂM，oyo < ĂŬDĬŌ，oya < ĂŬDĬĂM，salo < SĂL(Ĭ)Ō，sala < SĂL(Ĭ)IĂM，trayo < TRĂHŌ，traya < TRĂHĂM，valo < VĂL(Ĕ)Ō，vala < VĂL(Ĕ)ĂM などのように，それぞれ，語幹＋/-o/，語幹＋/-a/ となる変異形が存在する（Menéndez Pidal 1999：§ 113_{2b}；Penny 2002：3.7.8.1.1，3.7.8.1.3；Azofra 2009：4.5.1.）：

… e los testigos de la carta deven ver fazer la paga en dinero o oro o plata o otra cosa que lo *vala*…（ID513, 1472 年, Salamanca, CL，Eclesiástico）

f54b valga

動詞 caer「落ちる」，oír「聞こえる」，salir「出る」，traer「持ってくる」，valer「価値がある」などの直説法現在 1 人称単数と接続法現在には，cai**g**o，cai**g**a，oi**g**o，oi**g**a，sal**g**o，sal**g**a，trai**g**o，trai**g**a，val**g**o，val**g**a などのように，それぞれ，語幹＋/-go/，語幹＋/-ga/となる変異形が存在する（Menéndez Pidal 1999：§ 113_{2b}；Penny 2002：3.7.8.1.1，3.7.8.1.3；Azofra 2009：4.5.1.）。これらは，ten**g**o，ten**g**a のように，直説法現在 1 人称単数と接続法現在が，それぞれ，語幹＋/-go/，語幹＋/-ga/となる形からの類推形である：

… E como los ditos dos huertos *valgan* muito más e puedan de aquellos fallar muito mayor sens …（ID768，1379 年，Teruel，AR，Particular）

f55 saber の接続法現在

f55a sepa

saber「知る」の接続法現在には，ラテン語の形を継承した変異形 sepa < SĂPĬAM が存在する（Penny 2002 : 3.7.8.1.1）：

… *Sepades* que yo recibo en mi guarda e en mio defendimiento e en mi comienda el monesterio de Santo Turibio …（ID702, 1271 年，Palencia，CL，Cancilleresco）

f55b saba

saber「知る」の接続法現在には，ほかの-er 動詞からの類推に基づく変異形 saba が存在する：

Saban cuantos esta carta viren e oíren que eu don Arias pella gracia de Deus, abat de Sant Andrés con el convento d'esse mismo lugar, damos e outorgamos a vós …（ID458，1270 年，León，CL，Eclesiástico）

f56 接続法未来の 1 人称単数

f56a oviero

接続法未来の 1 人称単数には，ovier**o** < HĂBŬĔRŌ のように，語末母音が/-o/になる変異形が存在する（Menéndez Pidal 1999：§ 118$_5$; Penny 2002：3.7.8.4.4）。これは，ラテン語の直説法未来完了に由来する形である：

… E si esto non *cumpliro* ante que yo *finaro* que lo cumplan mios herederos al día del mio enterramiento assí como sobredicho es …（ID715，1284 年，Cantabria，CB，Eclesiástico）

f56b oviere

接続法未来の 1 人称単数には，ovier**e** < HĂBŬĔRĬM のように，語末母音が/-e/になる変異形が存在する（Menéndez Pidal 1999：§ 118$_5$; Penny 2002：3.7.8.4.4）。これは，ラテン語の接続法完了に由来する形である：

… otorgo e prometo de dar a vós, el abad e el convento sobredichos, o a aquellos que ý fueren en el monesterio al tiempo que yo *finare* mil moravedís de los de la guerra（ID330，1284 年，Valladolid，CL，Eclesiástico）

f57 接続法未来の 1 人称複数と 2 人称複数

f57a fuéremos

接続法未来の 1 人称複数と 2 人称複数には，fué**re**mos < FŬĔRĬMŬS，fué**re**des < FŬĔRĬTĬS のように，強勢音節の後の音節の母音/e/が保持された変異形が存在する（Penny 2002：3.7.3.2, 3.7.8.4.4）：

… E si vós o vuestros sucessores *oviéredes* de vender el dicho solar que non seades poderosos de lo vender salvo que primero nos lo fagades saber（ID1185，1461 年，Palencia，CL，Particular）

f57b fuermos

接続法未来の 1 人称複数と 2 人称複数には，fuermos < fué**r**(e)mos，fuerdes < fué**r**(e)des のように，強勢音節の後の音節の母音/e/が脱落した変異形が存在する（Penny 2002：3.7.3.2, 3.7.8.4.4）：

… E si por aventura venier tiempo que vós e los que de vós venieren *quesierdes* vender el dicho majuelo con su tierra que seades tenudos de lo fazer saber primero a nós …（ID1233，1455 年，Zamora，CL，Eclesiástico）

f58 ver の語幹

f58a veer

動詞 ver < VĪDĒRĔ「見える」には，不定形 **ve**er，現在分詞 **ve**yendo のように，語幹が ve-となる変異形が存在する（Menéndez Pidal 1999：§ 31；Penny 2002：3.7.9.1，3.7.9.2，3.7.9.3）：

… Vuestra merced lo sabrá y considerará lo que digo y *veerá* lo que más combendrá hazer …（ID1093，1591 年，Madrid，MA，Particular）

f58b ver

動詞 ver < VĪDĒRĔ「見える」には，不定形 **v**er，現在分詞 **v**iendo のように，語幹が v-となる変異形が存在する（Menéndez Pidal 1999：§ 31；Penny 2002：3.7.9.1，3.7.9.2，3.7.9.3）：

… E renuncio la ley que dice que los testigos de la carta deven *ver* fazer la paga de dineros o de cosa que lo vala …（ID360，1314 年，Valladolid，CL，Eclesiástico）

f59 ser の語幹

f59a seer

動詞 ser < SĔDĒRĔ「～である」には，不定形 **se**er，現在分詞 **se**yendo，過去分詞 **se**ído のように，語幹が se-となる変異形が存在する（Menéndez Pidal 1999：§ 31；Penny 2002：3.7.9.1，3.7.9.2，3.7.9.3）：

… nos avéis fecho e considerando el provecho e utilidad del dicho monesterio e *seyendo* verdadera mente informado de personas dignos de fe …（ID610，1498 年，La Rioja，LR，Particular）

f59b ser

動詞 ser < SĔDĒRĔ「～である」には，不定形 **s**er，現在分詞 **s**iendo のように，過去分詞 **s**ido のように，語幹が s-となる変異形が存在する（Menéndez Pidal 1999：§ 31；Penny 2002：3.7.9.1，3.7.9.2，3.7.9.3）：

Este día en Logroño *siendo* ý presentes el onrado religioso don Guido prior del monasterio de Santa María de Nágera …（ID719，1368 年，La Rioja，LR，Eclesiástico）

f60 decir，morir などの不定形

f60a dezir

動詞 decir「言う」，morir「死ぬ」などの不定形には，-er 動詞の変異形 diz**er** < DĪCĔRĔ，morr**er** < MŎRĪRĔ に対し，-ir 動詞の変異形 dez**ir** < DĪCĔRĔ，mor**ir** < MŎRĪRĔ が存在する（Zamora Vicente 1967：177-178）：

… nós nin otre por nós non podamos *dezir* nin razonar en nengún tiempo del mundo que non fuimos bien pagados de todos estos moravedís（ID329，1290 年，Valladolid，CL，Eclesiástico）

f60b dizer

動詞 decir「言う」，morir「死ぬ」などの不定形には，-ir 動詞の dez**ir** < DĪCĔRĔ，mor**ir** < MŎRĪRĔ に対し，-er 動詞の変異形 diz**er** < DĪCĔRĔ，morr**er** < MŎRĪRĔ が存在する（Zamora Vicente 1967：177-178）：

… Et destas cosas sobredichas algunos dellos uiron los p<ri>uilegios dellos oyro<n> *dizer* q<ue>los auien todos ensenbla acordaro<n> dize<n> q<ue> assi lo uiro<n> usar en tie<n>po del Rey （ID423，1270 年，León，CL，Eclesiástico）

f61 haber の 1 人称単数

f61a é

動詞 haber「持っている」の 1 人称単数には，変異形 é < HĂBĔŌ が存在する（Zamora Vicente 1967：101，191-192）：

… dó e otorgo al prior e al convento de San Marcos de León cuanto que yo *é* en Eslava ela cual yaz cabe Roda e las casas que *é* en Roda …（ID1456，1247 年，León，CL，Cancilleresco）

f61b éy

動詞 haber「持っている」の 1 人称単数には，変異形 éy < HĂBĔŌ が存在する（Zamora Vicente 1967：101，191-192）：

… dó e vendo toda la meetad de las nove partes que *ey* e aver devo en aquellos logares que de suso son dichos…（ID415，1235，León，CL，Particular）

f62 haber の 1 人称複数，2 人称複数

f62a avemos

動詞 haber「持っている」の 1 人称複数，2 人称複数には，それぞれ，変異形 avemos < HĂBĒMŬS，avedes < HĂBĒTĪS が存在する（Penny 2002：3.7.8.1.5）：

… damos a Martín Pérez e a don Pedro so primo II terras que *avemos* enna vega de Aguilar en camio por otra tierra que az en el cuérnago de los molinos de Valcarrero …（ID254，1238 年，Palencia，CL，Eclesiástico）

f62b hemos

動詞 haber「持っている」の 1 人称複数，2 人称複数には，それぞれ，変異形 hemos < HĂBĒMŬS，hedes < HĂBĒTĪS が存在する（Penny 2002：3.7.8.1.5）：

… damos ye otorgamos a vós doña María López ye a vós free Gonzalo Sánchez cuanto nós *emos* enna Veliella e cuanto *emos* in Río de Vimne e cuanto *emos* in Santa Buenia …（ID1457，1253 年，León，CL，Eclesiástico）

f63 動詞「する」

f63a fazer

「する」を表す動詞には，変異形 fazer < FĂCĔRĔ が存在する（Zamora Vicente 1967：261；Menéndez Pidal 1999：§ 106_4；Penny 2002：3.7.7.1，3.7.8.4.3，3.7.9.1）：

… los dichos monesterios e eglesias eran benidas en grant pobredat e se non podían mantener nin *fazer* aquel servicio que a Dios devían por las almas de aquellos que las fundaron …（ID614，1380 年，Valladolid，CL，Cancilleresco）

f63b fer

「する」を表す動詞には，変異形 fer < FĂCĔRĔ，far < *FĀRĔ が存在する（Zamora Vicente 1967：261；Menéndez Pidal 1999：§ 106_4；Penny 2002：3.7.7.1，3.7.8.4.3，3.7.9.1）：

… e prometemos por nós e por los vicarios e clérigos que son e por tiempo serán *fer* el dito aniversario a siempre e ultra aquello pagar los ditos dos soldos X dineros del aniversario…（ID768，1379 年，Teruel，AR，Particular）

f64 動詞「置く」

f64a dexar

「置く」を表す動詞には，変異形 dexar < LĂXĀRĔ が存在する（Penny 2002：1.2.3）：

Sepan cuantos esta carta vieren cómo yo Diego López de Haro señor de Vizcaya tengo por bien de *dexar* a vós don Yuc de Mont Pie prior de Santa María de Nágera …（ID708，1298 年，La Rioja，LR，Particular）

f64b lexar

「置く」を表す動詞には，変異形 lexar < LĂXĀRĔ が存在する（Penny 2002：1.2.3）：

… Item *lexo* al espital de Ruvielos dos soldos jaqueses item *lexo* a Joán de Peña mi fijo mi gramaja cárdena …（ID766，1277 年，Teruel，AR，Particular）

f65 -er 動詞と-ir 動詞の現在分詞

f65a aviendo

-er 動詞と-ir 動詞の現在分詞には，**av**iendo のように現在幹（**av**-er）から形成される変異形が存在する（Menéndez Pidal 1999：§ 120_6；Zamora Vicente 1967：262；Penny 2002：3.7.9.2；Pato & O'Neill 2013）：

… yaciendo enferma de cuitada e grave enfermedat e temiendo morir e *aviendo* miedo de las penas de los infiernos e copdiciando ir a la gloria del Santo Paradiso si a Dios place …（ID799，1378 年，Teruel，AR，Particular）

f65b oviendo

-er 動詞と-ir 動詞の現在分詞には，**ov**iendo のように完了幹（**ov**-e）から現在分詞が形成される変異形が存在する（Menéndez Pidal 1999：§ 120_6；Zamora Vicente 1967：262；Penny 2002：3.7.9.2；Pato & O'Neill 2013）。完了幹から形成された現在分詞のうち，pudiendo のみが現代語でも用いられている：

… Onde nós los ditos pecheros de Maquirriaín por nós e por todos los ditos nuestros sucessores *oviendo* por firmes ratas estables e valederas a perpetuo las ditas avenencias tratos e composiciones …（ID937，1382 年，Navarra，NA，Eclesiástico）

f66 -er 動詞と-ir 動詞の過去分詞

f66a avido

-er 動詞と-ir 動詞の過去分詞には，**av**ido のように現在幹（**av**-er）から形成される変異形が存在する（Menéndez Pidal 1999：§ 122_3；Zamora Vicente 1967：263）：

… obligamos a vosotros y a los vuestros todos nuestros bienes assí mobles como sedientes *avidos* y por aver en todo lugar e especialmente vos obligamos una casa nuestra sitia en el dicho lugar …（ID835，1502 年，Zaragoza，AR，Particular）

f66b **ovido**

-er 動詞と-ir 動詞の過去分詞には，**ov**ido のように完了幹（**ov**-e）から形成される変異形が存在する（(Menéndez Pidal 1999：§ 122$_3$；Zamora Vicente 1967：263）：

… de nuestras ciertas ciencias atorgamos e conocemos de vós Martín Quílez comprador susodicho aver *ovido* e recebido e en contantes en poder nuestro pasado todos quellos tres florines de oro en oro fino …（ID834, 1500 年, Zaragoza, AR, Particular）

f67 **-er 動詞の過去分詞**

f67a **-ido**

-er 動詞の過去分詞には，ten**ido** のように語尾が/-ído/となる変異形が存在する（Menéndez Pidal 1999：§ 121；Penny 2002：3.7.9.3；Azofra 2009：4.7.2.）：

… si por aventura alguno de los herederos fuesse contra este mi donadío que sobredicho es que fuesse *tenido* de pechar dos mil moravedís al que regnasse en Castilla …（ID665，1262 年，La Rioja，LR，Eclesiástico）

f67b **-udo**

-er 動詞の過去分詞には，ten**udo** のように語尾が/-údo/となる変異形が存在する（Menéndez Pidal 1999：§ 121；Penny 2002：3.7.9.3；Azofra 2009：4.7.2.）。

… e somos *tenudos* de fazer paret o tejado e toda lavor que sea de fazer en las casas sobredichas…（ID538，1273 年，Toledo，CM，Eclesiástico）

f68 **decir の過去分詞**

f68a **decho**

動詞 decir「言う」の過去分詞には，強勢母音が/é/となる変異形 **de**cho < DĬCTŬ が存在する（Menéndez Pidal 1999：§ 122$_3$）：

…Ye yo doña María Suáriz con otorgamiento de mios fillos e de mias fillas ya *dechos* damos a vós abat don Froila ye al convento ya *decho* cuanto heredamiento avemos in Villandax …（ID440，1256 年，Asturias，AS，Eclesiástico）

f68b **dicho**

動詞 decir「言う」の過去分詞には，強勢母音が/í/となる変異形 **di**cho < DĬCTŬ が存在する（Menéndez Pidal 1999：§ 122$_3$）：

… E yo Macía Gutiérrez notario deván *dicho* escriví esta carta per mandado de Álvar García escriván del rey e notario maor del León …（ID260，1260 年，León，CL，Judicial）

f69 **強勢が後ろから二番目の音節にある動詞の直説法現在，直説法未来，接続法現在の 2 人称複数の語尾**

f69a **amades**

強勢が後ろから二番目の音節にある-ar 動詞，-er 動詞，-ir 動詞の直説法現在，直説法未来，接続法現在の 2 人称複数には，それぞれ，-á**d**es < -ĀTĬS，-é**d**es < -ḖTĬS，-í**d**es < -ĪTĬS のように，語尾に/-d-/がある変異形が存在する（Penny 2002：3.7.3.1.；Azofra 2009：4.4.1.）。

… Et si Algo les *auedes* tomado o pendrado q<ue> gelo *tornedes* luego todo Et uos njn ellos no<n> *fagades* end Al …（ID540, 1291 年, Toledo, CM, Cancilleresco）

f69b **amaes**

強勢が後ろから二番目の音節にある-ar 動詞, -er 動詞, -ir 動詞の直説法現在, 直説法未来, 接続法現在の 2 人称複数には, それぞれ, **-áe**s < -á(d)es, **-ée**s < -é(d)es, **-íe**s < -í(d)es のように, 語尾の/-d-/が脱落し母音連続/áe/, /ée/, /íe/が生じた変異形が存在する（Penny 2002：3.7.3.1.；Azofra 2009：4.4.1.）:

… rogamos e mandamos a bós el reverendo señor abat de Cruz que *fagaes* autorizar esta santa bulla por juez eclegiástico signado por escrivano …（ID325, 1482 年, Palencia, CL, Eclesiástico）

f69c **amáis**

強勢が後ろから二番目の音節にある-ar 動詞, -er 動詞, -ir 動詞の直説法現在, 直説法未来, 接続法現在の 2 人称複数には, それぞれ, **-ái**s < -á(d)es, **-éi**s < -é(d)es, **-í**s < -í(d)es のように, 語尾の/-d-/が脱落し二重母音/ái/, /éi/, 短母音/í/が生じた変異形が存在する（Penny 2002：3.7.3.1.；Azofra 2009：4.4.1.）:

…E desto otorgamos dos escritura<s> en un tenor la una para que vós el dicho señor arçobispo *levéis* al dicho señor rey …（ID374, 1458 年, Toledo, CM, Municipal）

f70 **強勢が後ろから三番目の音節にある動詞の直説法現在，直説法未来，接続法現在の 2 人称複数の語尾**

f70a **amábades**

強勢が後ろから三番目の音節にある-ar 動詞の線過去, 過去未来, 接続法過去, 接続法未来の 2 人称複数には, それぞれ, amába**d**es < ĂMĀBĀTĬS < ĂMĀBĀTĬS, amaría**d**es < amar + -íades, amára**d**es < ĂMĀRĀTĬS < ĂMĀ(VĔ)RĀTĬS / amáse**d**es < ĂMĀSSĒTĬS < ĂMĀ(VĬ)SSĒTĬS, amáre**d**es < ĂMĀRĬTĬS < ĂMĀ(VĔ)RĬTĬS のように, 語尾に/-d-/がある変異形が存在する（Penny 2002 : 3.7.3.2 ; Azofra 2009 : 4.4.1.）。-er 動詞, -ir 動詞でも同様の変異形が存在する : comía**d**es, comería**d**es, comiéra**d**es/comiése**d**es, comiére**d**es ; vivía**d**es, viviría**d**es, viviéra**d**es/viviése**d**es, viviére**d**es。本コーパスでは, amáb**ae**s のように母音連続/áe/がある変異形は見られなかった :

… si en alguna dellas *huuiesedes* incurrido contal que no hayais permanecido enellas en vn an<n>o …（ID1342, 1634 年, Madrid, MA, Eclesiástico）

f70b **amábais**

強勢が後ろから三番目の音節にある-ar 動詞の線過去, 過去未来, 接続法過去, 接続法未来の 2 人称複数には, それぞれ, amáb**ai**s < amába(d)es, amarí**ai**s < amaría(d)es, amár**ai**s < amára(d)es / amás**ei**s < amáse(d)es, amár**ei**s < amáre(d)es のように, 語尾の/-d-/が脱落し二重母音/ái/, /éi/がある変異形が存在する（Penny 2002 : 3.7.3.2 ; Azofra 2009 : 4.4.1.）。-er 動詞, -ir 動詞でも同様の変異形が存在する : comí**ai**s, comerí**ai**s, comiér**ai**s/comiés**ei**s, comiér**ei**s ; viví**ai**s, viviría**i**s, viviér**ai**s/viviés**ei**s, viviér**ei**s :

…Y a los que *hallareis* culpados prenderéisles los cuerpos y tomaldes sus confesiones haciéndoles cargo de las culpas …（ID992, 1596 年, Madrid, MA, Judicial）

f71 関係節における未来の法

f71a 接続法未来

関係節において，cuantos esta present carta **vieren** のように，接続法未来が用いられる変異形が存在する：

Sepan cuantos esta carta *vieren* cómo yo don Mateos fi de Juan García de Miguel Ivañes vendo a vós doña Isabel de Sant Román …（ID478，1280 年，Segovia，CL，Particular）

f71b 直説法未来

関係節において，cuantos esta present carta **verán** のように，直説法未来が用いられる変異形が存在する：

Sepan cuantos esta present carta *verán* e *odrán* que yo don Pascual dArive vezino de Sangüessa vengo de manifiesto que me tiengo apagado …（ID962，1276 年，Navarra，NA，Particular）

4.2.1.3 前置詞・副詞・接続詞

f72 否定の接続詞

f72a ne

否定の接続詞には，変異形 ne < NĔC が存在する（Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）：

… E se dalgunu ovier que non querá lavrar *ne* aprovechar assí cumo deve tomar el orden el sou quiñón de aquél …（ID1464，1253 年，León，CL，Eclesiástico）

f72b ni

否定の接続詞には，変異形 ni < NĔC が存在する（Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）：

… Si per ventura de esta enfermedat passaro de est sieglo al otro que filios mios *ni* filias *ni* parient ninguno *ni* omne del sieglo no los embargue en estas heredades …（ID884，1234 年，Navarra，NA，Eclesiástico）

f72c nen

否定の接続詞には，変異形 nen < NĔC が存在する（Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）。ne**n** の語末の /-n/は no**n** の影響である：

… e me quito dela de oye adelante e la dou e la entrego a Morerola como desuso é dicho e *nen* eu *nen* mia moller *nin* meu fillo *nin* omne de mia parte non seermos poderosos de la demandar ya maes …（ID1234，1255 年，Zamora，CL，Eclesiástico）

f72d nin

否定の接続詞には，変異形 nin < NĔC が存在する（Zamora Vicente 1967：277；Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）。ni**n** の語末の/-n/は no**n** の影響である：

… e mando que nenguno non sea osado de los fazer pesar nin tuerto *nin* demás ca aquél que lo fiziesse al cuerpo e a todo cuanto que oviesse me tornaría por ello …（ID828，1260 年，Guadalajara，CM，Cancilleresco）

f73 期間を表す接続詞

f73a mientre

期間を表す接続詞には，語末母音が/-e/となる変異形 mientr**e** < (do)mientre < DŬM ĬNTĔRĬM が存在する（Menéndez Pidal 1999：§ 128$_2$；Penny 2002：3.8.2）：

… e si pora ventura la dicha doña Especiosa finar en ante de los cuatro años complidos e *mientre* lo nós toviermos lo que sobredicho es …（ID381，1319 年，Valladolid，CL，Particular）

f73b mientra(s)

期間を表す接続詞には，語末母音が/-a/となる変異形 mientr**a**(s) < mientre が存在する（Menéndez Pidal 1999：§ 128$_2$；Penny 2002：3.8.2）。語末の/-a(s)/は， fuera(s)「そとで」，nunca(s)「決して～ない」など/-a(s)/で終わる副詞からの影響である：

Oy Domingo de Ramos *mientras* estaba todo el convento en la Passión se á ido dél el licenciado Antonio Cavallero que estaba aquí …（ID1055，1637 年，Cuenca，CM，Eclesiástico）

f74 譲歩の接続詞

f74a maguer

譲歩の接続詞には，変異形 maguer(a) < ビザンティン・ギリシア語 μακάρι「～ならばいいのになあ」が存在する（Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）：

… *maguer* que vós la dicha Inés digades que me lo non agradecedes o que seades desobidiente o desagradeciente…（ID553，1430 年，Salamanca，CL，Particular）

f74b aunque

譲歩の接続詞には，変異形 aunque < aun + que が存在する（Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）：

… Y por tanto no dexaré de suplicar a vuestra señoría *aunque* yo no meresca tanta merced …（ID1487，1557 年，Sevilla，AN，Cancilleresco）

f75 理由を表す接続詞

f75a ca

理由を表す接続詞には，変異形 ca < QŬĬĂ が存在する（Menéndez Pidal 1999 : § 130 ; Penny 2002 : 3.8.2 ; Azofra 2009 : 5.4.2.）：

… e otrosí mando que les den a los frailes e a sus ombres que allá fueren possada *ca* mi voluntad es que assí se cumpla …（ID101，1524 年，Huelva，AN，Particular）

f75b porque

理由を表す接続詞には，変異形 porque < por que が存在する（Menéndez Pidal 1999 : § 130 ; Penny 2002 : 3.8.2 ; Azofra 2009 : 5.4.2.）：

… Yo hize luego lo que el consejo manda y fui al monasterio de Santispiritus y *porque* era ya tarde dije a la socomendadora que haze oficio de comendadora durante la vacante que otro día …（ID1059，1597 年，Salamanca，CL，Eclesiástico）

f76 条件を表す接続詞

f76a si

条件を表す接続詞には，母音が/i/となる変異形 **si** < SĪ が存在する（Zamora Vicente 1967：169,277；Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）：

… E *si* algún omne viniere de nuestros o de estraños que este nuestro fecho quisier demandar peche al rey …（ID516，1251 年，Salamanca，CL，Eclesiástico）

f76b se

条件を表す接続詞には，母音が/e/となる変異形 **se** < SĪ が存在する（Zamora Vicente 1967：169,277；Menéndez Pidal 1999：§ 130；Penny 2002：3.8.2）：

… E *se* algún omne la contrariar a vós o al monesterio o yo meismo quinquer que for sea por ende maldito …（ID439，1255 年，Asturias，AS，Eclesiástico）

f77 副詞を作る接尾辞

f77a -mente

形容詞から副詞を作る接尾辞には，firme**mente** のように，強勢短母音 Ĕが二重母音化せず/e/となった変異形-m**e**nte < MĔNTĔ が存在する（Menéndez Pidal 1999：§ 128$_3$；Penny 2002：3.4；Azofra 2009：5.1.2.2.）：

… Mando defiendo *firme mente* q<ue> nj<n>guno no<n> sea osado delos embargar nj<n> delos contrallar …（ID5，1278 年，Segovia，CL，Cancilleresco）

f77b -miente

形容詞から副詞を作る接尾辞には，firme**miente** のように，強勢短母音 Ĕが二重母音化し/ié/となった変異形-m**ie**nte < MĔNTĔ が存在する（Menéndez Pidal 1999：§ 128$_3$；Penny 2002：3.4；Azofra 2009：5.1.2.2.）：

… Et por q<ue> sea mas ffirm<e> rrenu<n>çiamos *espresa mient<e>* a q<ue>l dereh<o> …（ID106，1301，Ávila，CL，Particular）

f77c -mientre

形容詞から副詞を作る接尾辞には，firme**mientre** のように，強勢短母音 Ĕが二重母音化し/ié/となり，/r/が挿入された変異形-m**ie**nt**r**e < MĔNTĔ が存在する（Menéndez Pidal 1999：§ 128$_3$；Penny 2002：3.4；Azofra 2009：5.1.2.2.）。/r/の挿入は，接続詞(do)mientre < DŬM ĬNTĔRĬM からの影響だと考えられている：

… otorgamos de complir todas aq<ue>stas cosas de suso esc<ri>ptas bien e *leal mientre* …（ID1371，1249 年，León，CL，Particular）

f78 前置詞・副詞「～まで」

f78a fata

「～まで」を表す前置詞・副詞には，語中に/s/がない変異形 fata < アラビア語 ḥatta が存在する（Menéndez Pidal 1999：§ 129；Penny 2002：3.8.1）：

… E mandámosvos que los que an comprado *fata* aquí que les constringades que los vendan al plazo que el arçobispo e el cabildo les pusiere …（ID533，1294 年，Valladolid，CL，Cancilleresco）

f78b fasta

「～まで」を表す前置詞・副詞には，語中に/s/がある変異形 fasta < アラビア語 ḥatta が存在する（Menéndez Pidal 1999：§ 129；Penny 2002：3.8.1）：

… E otrosí renuncio la exepción del derecho en que diz que fasta dos años complidos es omne tenudo de provar la paga que fiziere …（ID137，1347 年，Cáceres，EX，Particular）

f79 前置詞「～のために」

f79a para

「～のために」を表す前置詞には，語頭母音が/a/の変異形 p**ar**a < PRŌ ĂD が存在する（Menéndez Pidal 1999：§ 129；Penny 2002：3.8.1）：

… Ordenamos damos <con>ffirmamos otorgamos les estas cosas *para* siemp<re> jamas …（ID7，1295 年，Valladolid，CL，Cancilleresco）

f79b pora

「～のために」を表す前置詞には，語頭母音が/o/の変異形 p**o**ra < pora が存在する（Menéndez Pidal 1999：§ 129；Penny 2002：3.8.1）：

… uos do q<ue> lo ayades libre quito por juro de h<er>edat *pora* siemp<re> iamas *pora* uos *pora* u<uest>ros fijos *pora* u<uest>ros Nietos pora quantos de uos uiniere<n> …（ID1208，1253 年，Sevilla，AN，Cancilleresco）

f80 前置詞「～なしに」

f80a sin

「～なしに」を表す前置詞には，母音が/i/となる変異形 **si**n < SĬNĔ が存在する（Menéndez Pidal 1999：§ 128$_4$，§ 129）：

… esta present carta con mi propria mano escriví con una otra su parella de una tenor e de una forma *sin* más e *sin* menos e fiz en ella este mi signo acostumbrado en testimonio de verdat …（ID935，1369 年，Navarra，NA，Eclesiástico）

f80b sen

「～なしに」を表す前置詞には，母音が/e/となる変異形 **se**n < SĬNĔ が存在する（Menéndez Pidal 1999：§ 128$_4$，§ 129）：

… Se per aventura vendir o empeñar *sen* mandado del convento tólganja ie venga pora la casa …（ID321，1227 年，Valladolid，CL，Eclesiástico）

f80c **sien**

「～なしに」を表す前置詞には，母音が二重母音/ie/となる変異形 **sie**n < SĪNĔ が存在する（Zamora Vicente 1967：218）：

… Si per aventura per tal necessidad me venier que non poda escusar de vendela devo a cuando quesier *sien* toda contradicha ye depués de mio finamiento mando que recudades a mio fiyo …（ID1371，1249 年，León，CL，Particular）

f80d **sienes**

「～なしに」を表す前置詞には，第一音節の母音が二重母音/ie/，語末が/-es/となる変異形 **sienes** < SĪNĔ が存在する（Zamora Vicente 1967：218）。語末の/-es/は，menos，más などの語末が/-s/となる副詞からの影響である：

... E por mayor confirmación que a vós salve e faga salvo con nós e *sienes* de nós el dito campo damos a vós fiança de salvedat a segunt buen fuero de Aragón …（ID975，1283 年，Huesca，AR，Particular）

f80e **sines**

「～なしに」を表す前置詞には，第一音節の母音が/i/，語末が/-es/となる変異形 **sines** < SĪNĔ が存在する（Menéndez Pidal 1999：§ 128_4，§ 129）。語末の/-es/は，menos，más などの語末が/-s/となる副詞からの影響である：

… E yo doña Ferrera otorgo todo lo que se continúa desuso en esta carta de atener planament *sines* contradicho ninguno …（ID889，1273 年，Navarra，NA，Eclesiástico）

f81 **前置詞・副詞「～によると」**

f81a **segundo**

「～によると」を表す前置詞・副詞には，語末が/-do/となる変異形 segun**do** < SĔCŬNDŬ が存在する（Menéndez Pidal 1999：§ 63_2，§ 129；Penny 2002：3.8.1）：

… obligamos nos todas n<uest>ras bonas mobles no<n> mobles p<or>a co<n>prir todas estas cosas *segundo* q<ue>las de suso p<ro>metemos …（ID380，1276 年，León，CL，Eclesiástico）

f81b **segúnd**

「～によると」を表す前置詞・副詞には，語末が/-nd/となる変異形 segú**nd** < SĔCŬNDŬ が存在する（Menéndez Pidal 1999：§ 63_2，§ 129；Penny 2002：3.8.1）：

… estando Juntados e<n>las casas de n<uest>ro ayuntamjento Enel dia ordinario *segund* q<ue>lo thenemos de Vso y costunbre …（ID269，1551 年，Toledo，CM，Eclesiástico）

f81c **segúnt**

「～によると」を表す前置詞・副詞には，語末が/-nt/となる変異形 segú**nt** < SĔCŬNDŬ が存在する（Menéndez Pidal 1999：§ 63_2，§ 129；Penny 2002：3.8.1）：

… estando ayutadas en su cabilldo a canpana tanjda *segu<n>t* q<ue> lo han de vso de costu<n>bre …（ID485，1439 年，Toledo，CM，Eclesiástico）

f81d **según**

「～によると」を表す前置詞・副詞には，語末が/-n/となる変異形 segú**n** < segúnd～segúnt が存在する（Menéndez Pidal 1999：§ 63_2，§ 129；Penny 2002：3.8.1）。

… estando juntos y ayuntados por su mun<n>jdor *segun* q<ue>lo an de vso de se ayuntar …（ID194，1522 年，Guadalajara，CM，Judicial）

f82 否定の副詞

f82a **non**

否定の副詞には，語末の/n/が保持された変異形 no**n** < NŌN が存在する（Moreno Bernal & Horcajada 1997；Menéndez Pidal 1999：§ 62_3；Penny 2002：3.4）：

... Pero es mi voluntad que cuando los cavalleros se armaren *non* se estienda esta premática a ellos …（ID274，1515 年，Burgos，CL，Cancilleresco）

f82b **no**

否定の副詞には，語末の/n/が脱落した変異形 no < NŌN が存在する（Moreno Bernal & Horcajada 1997；Menéndez Pidal 1999：§ 62_3；Penny 2002：3.4）：

… e si por ventura *no* lo fiziesse que vós que seades poderosos de pendrar los bienes de los bivos e del muerto …（ID295，1275 年，Palencia，CL，Eclesiástico）

f83 副詞「とても」

f83a **muit**

「とても」を表す副詞には，/t/が保持された変異形 mui**t** < MŬLT(Ŭ)が存在する（Menéndez Pidal 1999：§ 47_{2c}；Penny 2002：2.5.2.4）：

... vj ley vna carta esc<ri>pta en pargamj<n>o seyllada co<n> el siello dela *muyt* noble seynora do<n>a M<aria> mug<er> q<ue> fue del jnfante don joh<a>n …（ID1081，1330 年，Navarra，NA，Particular）

f83b **muy**

「とても」を表す副詞には，/t/が脱落した変異形 mu**y** < MŬLT(Ŭ)が存在する（Menéndez Pidal 1999：§ 47_{2c}；Penny 2002：2.5.2.4）：

… E otrosí renuncio la ley del justo precio que el *muy* noble rey don Alfonso que Dios perdone fizo e ordenó en las cortes de Alcalá de Henares …（ID346，1407 年，Valladolid，CL，Eclesiástico）

f84 副詞「より～」

f84a **más**

「より～」を表す副詞には，強勢母音が/á/の変異形 m**á**s < MĂGĬS が存在する（Menéndez Pidal 1999：§ 128_1；Penny 2002：3.4）：

… E si *más* vale destos dichos tres mil maravedís fazemos dello donación e dámoslo a la dicha eglesia …（ID310，1363 年，Cáceres，EX，Eclesiástico）

f84b **mais**

「より～」を表す副詞には，f86 に対して，強勢母音が二重母音/ái/の変異形 m**ais** < MĂGĪS が存在する（Menéndez Pidal 1999：§ 128_1；Penny 2002：3.4）：

… Et si *mays* val q<ue>l dich<o> p<re>çio q<ui>tamos uos la mayoria damos uos lo en pura donasçio<n> …（ID598，1319 年，Asturias，AS，Eclesiástico）

f85 **副詞「そのように」**

f85a **así**

「そのように」を表す副詞には，変異形 así < ĂD SĪC が存在する（Menéndez Pidal 1999：§ 128_4；Penny 2002：3.4）：

… E por nuestra sentencia difinitiva juzgando *así* lo juzgamos pronunciamos y mandamos en estos escritos …（ID682，1492 年，Burgos，CL，Eclesiástico）

f85b **ansí**

「そのように」を表す副詞には，変異形 a**n**sí < así が存在する（Menéndez Pidal 1999：§ 128_4；Penny 2002：3.4）。語中の/-n-/は，前置詞 en からの影響である：

… Y yo el dicho Juan Domínguez por mí e mi moger e quien mi voz obere *ansí* recibo e la dicha metá de granja e cosas sobredichas en fuero en la manera que dicha es …（ID583，1501 年，León，CL，Eclesiástico）

f85c **asín**

「そのように」を表す副詞には，変異形 as**ín** < así + /-n/が存在する（Menéndez Pidal 1999 : § 128_4 ; Penny 2002 : 3.4）。語末の/-n/は，/-n/で終わる副詞 no**n**，bie**n** や前置詞 **en**，co**n**，si**n** などからの影響である：

… E yo dito Fortún Gonçálvez en el nombre mío proprio e *asín* como procurador de Gracia Ferrández muger mia legítima dó a vós ditos prior e capítol …（ID659，1382 年，Zaragoza，AR，Particular）

f86 **副詞・接続詞・前置詞「～のように」**

f86a **como**

「～のように」を表す副詞・接続詞・前置詞には，語頭母音が/o/となる変異形 **co**mo < QUŌMŎDŎ が存在する（Menéndez Pidal 1999：§ 39_4；Penny 2002：3.4，3.8.2）：

… obligo me de guardallas de co<n>plillas en toda mj vida assi *como* dichas son …（ID327，1274 年，Palencia，CL，Eclesiástico）

f86b **cuemo**

「～のように」を表す副詞・接続詞・前置詞には，語頭母音が二重母音/ue/となる変異形 c**ue**mo < QUŌMŎDŎ が存在する（Menéndez Pidal 1999：§ 39_4；Penny 2002：3.4，3.8.2）：

… Et dogelo pora dar uender camiar enagenar pora ffaz<er> dello todo lo q<ue> q<ui>siere<n> *cuemo* de lo suyo mismo en tal manera q<ue> no<n> lo vendan ni<n> lo den ni<n> lo enagene<n> …（ID562，1255 年，Palencia，CL，Cancilleresco）

f86c **cumo**

「～のように」を表す副詞・接続詞・前置詞には，語頭母音が/u/となる変異形 **cu**mo < QUŌMŎDŏ が存在する（Menéndez Pidal 1999：§ 39$_4$；Penny 2002：3.4，3.8.2）：

… ma<n>damos fazer dos cartas tal la una *cumo* la ot<ra> selladas co<n> n<uest>ros seellos …（ID301，1276 年，Palencia，CL，Eclesiástico）

f87 副詞「今」

f87a **agora**

「今」を表す副詞には，/-g-/が保持された変異形 a**g**ora < HĀC HŌRĀ が存在する（Menéndez Pidal 1999：§ 128$_3$；Penny 2002：3.4）：

… el dicho Diego de Cea escrivano público respondió e dixo que lo que hasta *agora* á dado del dicho proceso á sido por mandamiento de la justicia …（ID1488，1557 年，Sevilla，AN，Judicial）

f87b **ahora**

「今」を表す副詞には，/-g-/が脱落した変異形 aora～ahora < agora が存在する（Menéndez Pidal 1999：§ 128$_3$；Penny 2002：3.4）：

… No les escrivo *ahora* porque toman mis cartas en mal que a todo el reino an escrito …（ID1486，1520 年，Valladolid，CL，Eclesiástico）

f88 副詞「後に」

f88a **empués**

「後に」を表す副詞には，変異形 empués～empós < ĬN PŌST が存在する（Menéndez Pidal 1999：§ 128$_2$；Zamora Vicente 1967：276；Penny 2002：3.8.1）：

… E si por ventura vós e los vuestros e qui *empués* vós las ditas viñas tenrá e posedirá non lavrávades las ditas viñas …（ID986，1379 年，Huesca，AR，Particular）

f88b **aprés**

「後に」を表す副詞には，変異形 aprés < ĂD PRĔSSŬ が存在する（Zamora Vicente 1967：276）：

… e prometemos cantar e celebrar el dito aniversario por todos tiempos de la nuestra vida e *aprés* de nós los vicarios e clérigos que serán en la dita eglesia e lugar de Ruvihuelos …（ID772，1384 年，Teruel，AR，Eclesiástico）

f88c **depués**

「後に」を表す副詞には，変異形 depués～depós < DĒ PŌST が存在する（Menéndez Pidal 1999：§ 128$_2$；Zamora Vicente 1967：276；Penny 2002：3.8.1）：

… vos damos e vos otorgamos que lo ayades libre e quito por siempre jamás vós e los que vernán *depués* de vós en el monesterio de Aguilar por juro de heredad …（ID295，1275，Palencia，CL，Eclesiástico）

f88d **después**

「後に」を表す副詞には，変異形 después～depós < DĒ ĔX PŌST が存在する（Menéndez Pidal 1999：§ 128_2；Zamora Vicente 1967：276；Penny 2002：3.8.1）：

… Otrossí mandamos que si los fijos partieren con el padre *después* de muerte de su madre que el padre aya por sí sus escusados e los fijos por sí los suyos …（ID3，1262 年，Sevilla，AN，Cancilleresco）

f89 **副詞「そのとき」**

f89a **entonces**

「そのとき」を表す副詞には，変異形 entonces < ĬNTŬNCCĔ＋/-s/が存在する（Menéndez Pidal 1999：§ 128_2；Penny 2002：3.4）：

… ca así desde *entonces* para agora e de agora para *entonces* lo otorgamos e lo emos e avremos por firme e valedero bien ansí …（ID513，1472 年，Salamanca，CL，Eclesiástico）

f89b **estonces**

「そのとき」を表す副詞には，変異形 estonces < ĔXTŬNCCĔ＋/-s/が存在する（Menéndez Pidal 1999：§ 128_2；Penny 2002：3.4）。

… e coñocí e otorgué que non valía *estonces* más desto e en caso que más valiese fize dello donación al dicho monesterio…（ID386，1412 年，Cáceres，EX，Eclesiástico）

f90 **場所を表す関係副詞**

f90a **ó**

場所を表す関係副詞には，変異形 ó < ŭBI が存在する（Menéndez Pidal 1999：§ 128_1；Penny 2002：3.4；Azofra 2009：7.4.2.）。所在「～で」と起点「～から」の用法の区別は行わない：

… e todos los logares que Ferrant Martínez el sobredicho avié en el açogue mayor *ó* venden el pescado que son en linde de la calle e la plaça que es ante la carnicería …（ID1173，1277 年，Segovia，CL，Eclesiástico）

f90b **do**

場所を表す関係副詞には，変異形 do < de+ ó が存在する（Menéndez Pidal 1999：§ 128_1；Penny 2002：3.4；Azofra 2009：7.4.2.）。所在「～で」と起点「～から」の用法の区別は行わない：

… Esta carta fue fecha a la carnicería *do* venden el pescado Yo Pero García escrivano público la escreví e fiz este mi signo en esta carta …（ID1173，1277 年，Segovia，CL，Eclesiástico）

f90c **ond(e)**

場所を表す関係副詞には，変異形 ond(e) < ŭNDE が存在する（Menéndez Pidal 1999：§ 128_1；Penny 2002：3.4；Azofra 2009：7.4.2.）。所在「～で」と起点「～から」の用法の区別は行わない：

… quiero que ayan para sí las penas en lugar dellos salvo en aquellas cibdades e villas e lugares *onde* yo é fecho merced de las dichas penas a otros …（ID282，1448 年，Burgos，CL，Cancilleresco）

f90d **dond(e)**

場所を表す関係副詞には，変異形 dond(e) < de ond(e)が存在する（Menéndez Pidal 1999：§ 128$_1$；Penny 2002：3.4；Azofra 2009：7.4.2.）。所在「〜で」と起点「〜から」の用法の区別は行わない：

… E in continenti el dicho señor corregidor por ante mí el escrivano fue a las casas *donde* murió el doctor don Pedro Fernández Pando para efecto de inventariar los vienes que an quedado…（ID819，1688 年，Madrid，MA，Judicial）

f91 **副詞「一緒に，同時に」**

f91a **juntamente**

「一緒に，同時に」を表す副詞には，juntamente が存在する：

… fue acordado que de aquí adelante vós el dicho concejo *juntamente* con el nuestro corregidor o juez de residencia que agora es o fuere en el dicho condado …（ID1425，1506 年，Vizcaya，PV，Cancilleresco）

f91b **ensemble**

「一緒に，同時に」を表す副詞には，ensemble < フランス語 ensemble が存在する（Zamora Vicente 1967：276）：

… jo don García abat de Fitero *ensemble* con el convent damos a vós doña Taresa lo que vós nos diestes cun todo lo al que nós avemos …（ID852，1238 年，Navarra，NA，Eclesiástico）

4.2.2　形態音韻論的特徴

形態音韻論的特徴は，形態や音韻に関して変異のあるものである。

f92 **単数形が二重母音/-ei/で終わる名詞の複数形**

f92a **leys**

buei < BŎVĔ，lei < LĒGĔ，rei < RĒGĔ のように単数形が二重母音/-ei/で終わる名詞の複数形には，bue**is**，le**is**，re**is** のように，単数形＋/-s/となる変異形が存在する（Menéndez Pidal 1999：§ 75$_3$；Penny 2002：3.5.1）：

… E renunciamos a todas *leis* raçones e defensiones que contra esti fecho desta carta vengan en juicio nin fuera de juicio …（ID670，1292 年，La Rioja，LR，Eclesiástico）

f92b **leyes**

buei < BŎVĔ，lei < LĒGĔ，rei < RĒGĔ のように単数形が二重母音/-ei/で終わる名詞の複数形には，bue**ies**，le**ies**，re**ies** のように，単数形＋/-es/となる変異形が存在する（Menéndez Pidal 1999：§ 75$_3$；Penny 2002：3.5.1）：

… Et rrenu<n>çio todas las ot<ra>s *leyes* ffueros derech<o>s esc<ri>ptos no<n> esc<ri>ptos estableçimie<n>tos husos costumbres …（ID86，1296 年，Ávila，CL，Particular）

f93 強勢短母音 Ĕ の二重母音化

f93a es

動詞 ser「～である」の直説法現在 3 人称単数，線過去，接続詞「～と」には，それぞれ，e(s) < ĔST，era < ĔRAM，e < ĔT のように，強勢短母音 Ĕ が二重母音化しない変異形が存在する（Zamora Vicente 1967 : 95-97，190-191，264，267）：

En el nombre de Dios que *es* Padre e Fijo e Espíritu Santo que son tres Personas e un Dios …（ID6，1285 年，Burgos，CL，Cancilleresco）

f93b yes

動詞 ser「～である」の直説法現在 3 人称単数，線過去，接続詞「～と」には，それぞれ，ye(s) < ĔST，yera < ĔRAM，ye < ĔT のように，強勢短母音 Ĕ が二重母音化する変異形が存在する（Zamora Vicente 1967 : 95-97，190-191，264，267）：

… E nós abat e conviento ya dicho otorgamos lo que en esta carta *ye* escrito e nós partes ya dichas otorgamos tener e gardar e complir todo lo que en esta carta ye escrito …（ID581，1310 年，Asturias，AS，Eclesiástico）

f94 ラテン語の指小辞-ĔLLŬ

f94a capiella

ラテン語の指小辞-ĔLLŬ を持つ語には，cap**ie**lla < CAPĔLLĂ，cast**ie**llo < CASTĔLLŬ などのように，/-ieʎo/となる変異形が存在する（Menéndez Pidal 1999：§ 10_2；Penny 2002：2.4.2.5）：

… Et q<ue> ot<ro> ni<n>guno no<n> sea enterrado en esta *capiella* dicha ssino<n> los q<ue> fuere<n> de u<uest>ro linage …（ID355，1295 年，Valladolid，CL，Eclesiástico）

f94b capilla

ラテン語の指小辞-ĔLLŬ を持つ語には，cap**i**lla < CAPĔLLĂ，cast**i**llo < CASTĔLLŬ などのように，/-iʎo/となる変異形が存在する（Menéndez Pidal 1999：§ 10_2；Penny 2002：2.4.2.5）。これは，/-ieʎo/の/e/が後ろの/ʎ/に吸収された形である：

… segu<n>d q<ue>lo avemos de vso de costu<n>br<e> enla calostra del dicho monesterio çerca dela *capilla* de ssant françisco nos todos en vno ot<or>gamos conosçemos por esta c<art>a q<ue> damos afuero avos …（ID677，1413 年，Zamora，CL，Particular）

f95 語末の母音/e/の保存・脱落

f95a part

part < PĂRTĔ や present < PRĂĔSĔNTĔ などのように，子音や子音連続の後ろの語末の母音/e/が脱落する変異形が存在する（Penny 2002：2.4.3.2；Ueda 2015）。検索は，part，present に限定した：

… E vi la *present* cláusula de la nota original del último testament del dicho Nicolao Bernat notario público …（ID620，1494 年，Zaragoza，AR，Particular）

f95b **parte**

parte < PĂRTĔ，presente < PRĂĔSĔNTĔ などのように，子音や子音連続の後ろの語末の母音/e/が残る変異形が存在する（Penny 2002：2.4.3.2；Ueda 2015）。検索は，parte，presente に限定した：

… E desde oy día en adelante por esta *presente* carta vos dó e otorgo el juro la tenencia posesión propiedat e señorio destos dichos bienes …（ID584，1393 年，León，CL，Eclesiástico）

f96 **最終音節の後舌母音**

f96a **conventu**

最終音節の後舌母音には，convent**u** < CŎNVĔNTŬ や som**u**s < SŬMŬS のように，/u/になる変異形が存在する（Zamora Vicente 1967：111-113；Menéndez Pidal 1999：§ 29_1）：

… yo don Monio Paiz prior del monesterio de San Marcos ensembla con el *conventu* des mismo lugar damos ye otorgamos a vós doña María López ye a vós free Gonzalo Sánchez …（ID1457，1253 年，León，CL，Eclesiástico）

f96b **convento**

最終音節の後舌母音には，convent**o** < CŎNVĔNTŬ や som**o**s < SŬMŬS のように，/o/になる変異形が存在する（Zamora Vicente 1967：111-113；Menéndez Pidal 1999：§ 29_1）：

… yo don Moñio Paiz pela gracia de Deus prior de San Marcus ensembra cono *convento* des mismo lugar damus a vós Domingo Calvo e Domingo Tesso e Pedro Saco e don Migael de las Eras …（ID1464，1253 年，León，CL，Eclesiástico）

f97 **/kt/や/(u)lt/の子音連続**

f97a **feito**

/kt/や/(u)lt/の子音連続には，fe**it**o < FĂCTŬ や mu**it**o < MŬLTŬ のように，音節末の/k/や/l/が母音化した[i̯]が直後の/t/を口蓋化させずに残る変異形が存在する（Zamora Vicente 1967 : 150-152，241-242，276 ; Menéndez Pidal 1999 : § 47_{2c}，§ 50_1，§ 122_3；Penny 2002：2.5.2.4）：

… e coando las avrán *feitas* mandamos a vosotros capellanos que prengades uno o dos de vuestros parroquianos …（ID979，1265 年，Navarra，NA，Eclesiástico）

f97b **fecho**

/kt/や/(u)lt/の子音連続には，fe**ch**o < [fei̯to] < FĂCTŬ や mu**ch**o < [mui̯to] < MŬLTŬ のように，/kt/や/(u)lt/の子音連続において音節末の/k/や/l/が母音化した[i̯]が直後の/t/を/ʧ/に口蓋化させた変異形が存在する（Zamora Vicente 1967：150-152，241-242，276；Menéndez Pidal 1999：§ 47_{2c}，§ 50_1，§ 122_3；Penny 2002：2.5.2.4）：

… vi una carta del rey seellada con so seello de la poridat *fecha* en esta manera …（ID1369，1286 年，León，CL，Cancilleresco）

f98 **破裂音/p, b, k, g/もしくは摩擦音/f/と流音/l/の子音群**

f98a **obrigar**

破裂音/p，b，k，g/もしくは摩擦音/f/と流音/l/からなる子音群には，ob**r**igar < ŎBLĬGĀRĔ，comp**r**ir < CŎMPLĒRĔ のように，流音/l/が/r/となる変異形が存在する（Zamora Vicente 1967：137-138）：

… E yo Domingo Aparicio otorgo todas estas cosas desuso dichas e dó pora *comprir* esto por fiador de mancomún comigo e …（ID380，1276 年，León，CL，Eclesiástico）

f98b **obligar**

破裂音/p，b，k，g/もしくは摩擦音/f/と流音/l/からなる子音群には，ob**l**igar < ŏBLĭGĀRĔ，compl**l**ir < CŏMPLĒRĔ のように，流音/l/が保持される変異形が存在する（Zamora Vicente 1967：137-138）：

…. prometemos e otorgamos por nós e por los que vernán depués de nós de *complir* e de fazer *complir* aqueste capellán por las ánimas de los sobredichos …（ID295，1275 年，Palencia，CL，Eclesiástico）

f99 ラテン語の音連続-L-+[j]

f99a **fijo**

ラテン語の音連続-L-+[j]には，fi**j**o < FĪLĬŬ のように，-L-+[j] > /ʒ/（<i>～<j>～<g>で表記）となる変異形が存在する（Zamora Vicente 1967：146，244-245；Menéndez Pidal 1999：§ 53_6；Penny 2002：2.5.2.2）。検索は，簡単のため，conce**j**o < CŏNCĬLĬŬ，conse**j**o < CŏNSĬLĬŬ，fi**j**o < FĪLĬŬ，me**j**or < MĔLĬŌRE，mu**g**(i)er < MŬLĬĔRĔ に限定した：

… yo Peidro *fiio* uendo a uos don Joan de piliella a u<uest>ra *mugier* don<n>a helena el mio solar …（ID246，1228 年，Burgos，CL，Particular）

f99b **fillo**

ラテン語の音連続-L-+[j]には，fi**ll**o < FĪLĬŬ のように，-L-+[j] > /ʎ/（<ll>で表記）となる変異形が存在する（Zamora Vicente 1967：146，244-245；Menéndez Pidal 1999：§ 53_6；Penny 2002：2.5.2.2）。検索は，簡単のため，conce**ll**o < CŏNCĬLĬŬ，conse**ll**o < CŏNSĬLĬŬ，fi**ll**o < FĪLĬŬ，me**ll**or < MĔLĬŌRE，mu**ll**er < MŬLĬĔRĔ に限定した：

… nos Las ditas maria Ruiz dona toda Ruiz su *filla* façemos donacion damos ajtorgamos ponemos en tenje<n>ça en corporal possession …（ID893，1301 年，Navarra，NA，Eclesiástico）

f99c **fiyo**

ラテン語の音連続-L-+[j]には，fi**y**o < FĪLĬŬ のように，-L-+[j] > /ʃ/（<y>で表記）となる変異形が存在する（Zamora Vicente 1967：146，244-245；Menéndez Pidal 1999：§ 53_6；Penny 2002：2.5.2.2）。検索は，簡単のため，conce**y**o < CŏNCĬLĬŬ，conse**y**o < CŏNSĬLĬŬ，fi**y**o < FĪLĬŬ，me**y**or < MĔLĬŌRE，mu**y**er < MŬLĬĔRĔ に限定した：

… yo Maria ff<e>rr<and>ez *fiya* de M<aria> dom<i>ng<ue>z de Aruos do Auos frey M<arti>no p<ri>or delos frayres p<re>dicadores …（ID780，1281 年，Zamora，CL，Eclesiástico）

f100 -D'C-，-T'C-の子音連続

f100a **judgar**

母音消失により生じた-D'C-，-T'C-の子音連続には，ju**d**gar < JŪD(Ĭ)CĀRĔ，porta**d**go < PŎRTĀT(Ĭ)CŬ などのように，最初の子音が/-d-/となる変異形が存在する（Menéndez Pidal 1999：§ 60_3；Penny 2002：2.5.5）：

… p<ar>a q<ue> ellos anbos junta mente ayan de ber entend<e>r *judgar* en <e>stos dichos dapn<n>os E los aberiguar …（ID1274，1478 年，Granada，AN，Cancilleresco）

f100b **juzgar**

母音消失により生じた-D'C-，-T'C-の子音連続には，ju**z**gar < ju**d**gar，porta**z**go < portadgo などのように，最初の子音が/-d-/から/-s̱-/（<z>で表記）となった変異形が存在する（Menéndez Pidal 1999：§ 60$_3$；Penny 2002：2.5.5）：

… suplico a v<uestra>s<eñoria> se sirua de aduertirlo al consejo para que no *juzgue* que es falta de atençion mia en obedeçerle sino neçesidad del tiempo …（ID1243，1649，Cuenca，CM，Particular）

f100c **julgar**

母音消失により生じた-D'C-，-T'C-の子音連続には，ju**l**gar < JŪD(Ī)CĀRĔ，-a**l**go < PŎRTĀT(Ĭ)CŬ などのように，最初の子音が/-l-/となる変異形が存在する（Zamora Vicente 1967：152；Menéndez Pidal 1999：§ 60$_3$）：

… si el dich<o> p<ri>or o el co<n>ue<n>to o otri por ellos o algu<n>o por su ma<n>dado viniesse cont<ra> ne<n>gu<n>a cosa q<ue> los dichos arbitros & amigos ma<n>dassen o *julgassen* segund dich<o> es …（ID615，1286 年，La Rioja，LR，Judicial）

f101 **-M'N-の子音連続**

f101a **nomne**

母音消失により生じた-M'N-の子音連続には，ho**mn**e < HŎM(Ĭ)NĔ などのように，変化の生じない変異形/-mn-/が存在する（Menéndez Pidal 1999：§ 59$_2$；Penny 2002：2.5.5）：

… viendo auos Conuie<n>to de suso decho todas estas h<er>edades *nomnadas* q<ua>nto yo enellas ey o auer deuo por p<re>çio *numnado* L M<orauedis> de Moneda Real …（ID434，1233 年，Asturias，AS，Eclesiástico）

f101b **nomre**

母音消失により生じた-M'N-の子音連続には，ho**mr**e < HŎM(Ĭ)NĔ などのように，/n/が/r/になった変異形/-mr-/が存在する（Menéndez Pidal 1999：§ 59$_2$；Penny 2002：2.5.5）：

… fiz esc<ri>uir esta carta & q<ue> no<n> ue<n>ga en dulda esc<ri>uj enella mjo *nomre* …（ID1238，1278 年，Zamora，CL，Particular）

f101c **nombre**

母音消失により生じた-M'N-の子音連続には，ho**mbr**e < HŎM(Ĭ)NĔ などのように，/n/が/r/になり，さらに/-b-/が挿入された変異形/-mbr-/が存在する（Menéndez Pidal 1999：§ 59$_2$；Penny 2002：2.5.5）：

En el *nombre* de dios q<ue> es padre fijo sp<irit>u s<an>c<t>o q<ue> son tres p<er>sonas vn dios …（ID6，1285 年，Burgos，CL，Cancilleresco）

f102 **-b't-，-p't-，-v't-などの子音連続**

f102a **cibdad**

母音消失により生じた-B'T-，-P'T-，-V'T-などの子音連続には，子音連続の前の母音が後舌母音でない場合，de**bd**a < DĒB(Ĭ)TĂ，ca**bd**al < CĂP(Ĭ)TĀLĔ，ci**bd**ad < CĪV(Ĭ)TĀTĔ などのように，変異形/-bd-/が存在する（Menéndez Pidal 1999：§ 60$_1$；Penny 2002：2.5.5）：

… pedimos e requerimos al muy magnífico señor don Nuño de la Cueva corregidor en esta *cibdad* de Badajoz …（ID1380, 1549 年, Badajoz, EX, Municipal）

f102b **ciudad**

母音消失により生じた-B'T-, -P'T-, -V'T-などの子音連続には，子音連続の前の母音が後舌母音でない場合，de**ud**a < debda，ca**ud**al < ca**bd**al，ci**ud**ad < ci**bd**ad などのように，変異形/-ud-/が存在する（Menéndez Pidal 1999 : § 60$_1$; Penny 2002 : 2.5.5）。これは，/-bd-/の第一要素/b/が母音化し/u/になった形である：

… dixo que pedia asumerçed q<ue>le mandasse dar la sen<tenç>ja signada pues la parte dela dicha *çiudad* se avia desistido dela apelaçion …（ID383, 1528 年, Cáceres, EX, Judicial）

f103 **母音間の-D-の保持・消失**

f103a **odir**

ラテン語で母音間の-D-を持つ語には，o**d**ir < ĂŬDĪRĔ, pose**d**ir < PŎSSĬDĒRĔ, se**d**er < SĔDĒRĔ, ve**d**er < VĬDĒRĔ のように，母音間の-D-が保存される変異形が存在する（Zamora Vicente 1967 : 230-231）。検索は，odir, posedir, seder, veder に限定した：

… E cuanto a esto atorgo me *seder* bien pagado e conviengo e primero a vós e a los vuestros *seder* de manifiesto e catar de todo daño agora e todos tiempos dios obligación de todos mis bienes …（ID804, 1291 年, Huesca, AR, Eclesiástico）

f103b **oír**

ラテン語で母音間の-D-を持つ語には，oír < ĂŬDĪRĔ, poseer < PŎSSĬDĒRĔ, s(e)er < SĔDĒRĔ, v(e)er < VĬDĒRĔ のように，母音間の-D-が脱落する変異形が存在する（Zamora Vicente 1967 : 230-231）。検索は，oír, poseer, s(e)er, v(e)er に限定した：

…e nós atorgamos *seer* bien pagados de todo el derecho perteneciente al dito capítol …（ID762, 1320 年, Teruel, AR, Particular）

f104 **/d/の後ろの語末の/-e/の保持・消失**

f104a **verdade**

verdad**e** < VĒRĬTĀTĔ のように，/d/の後ろで語末の/-e/が保持される変異形が存在する（Zamora Vicente 1967 : 117）：

… pus en ella mio sinal en testimonio de *uerdade* Et mandou me el dea<n> sobred<i>c<t>o que desse A joha<n> garcia …（ID586, 1273 年, León, CL, Eclesiástico）

f104b **verdad～verdat**

verda**d**～verda**t** < verdad**e** < VĒRĬTĀTĔ のように，/d/の後ろで語末の/-e/が消失する変異形が存在する（Zamora Vicente 1967 : 117）：

… esc<ri>uj estas cartas p<or> mia mano fiz enellas mia sinal en testimonio de *verdat* …（ID380, 1276 年, León, CL, Eclesiástico）

f105 複数形の語末

f105a todos

複数形の語末には，母音＋/-s/の変異形が存在する（Zamora Vicente 1967：248-250；González Ollé 1996：285）：

… E otrosí renuncio, e me parto, e quito e dexo de toda ley, e de todo fuero … e de *todos usos*, e de *todas costumbres*, e de *todas razones* …（ID1045，1490 年，Cádiz，AN，Particular）

f105b toz

複数形の語末には，/-ls/（els），/-ns/（vezins），/-rs/（labors），/-ts/（totz～toz）のような子音連続が生じる変異形が存在する（Zamora Vicente 1967：248-250；González Ollé 1996：285）：

… yo lo dit Sancho Miguel de Sandaña obliguéi *toz* mes *bens* mobles e *heredaz presenz* e per venir en *toz locs* que yo les ayéi e trobar se puissen renunciant *toz fors toz us* e totes costumnes de for seglar …（ID982，1358 年，Navarra，NA，Eclesiástico）

f106 /-a/で終わる女性形の名詞などの複数形の語末母音

f106a coses

/-a/で終わる女性形の名詞などの複数形の語末母音には，**coses** < cosa + s のように，語末が/-es/となる変異形が存在する（Zamora Vicente 1967：113-116）：

… e de *totes les altres dones* del dit convent *les cuales dites viñes* devem tenir nós les diz Martín de Saldias e María dUvani sa muller ortelans de Pampalona …（ID936，1357 年，Navarra，NA，Eclesiástico）

f106b cosas

/-a/で終わる女性形の名詞などの複数形の語末母音には，cos**as** < cosa + s のように，語末が/-as/となる変異形が存在する（Zamora Vicente 1967：113-116）：

… E nós sobre esto queriendo saber verdat de *las cosas sobreditas* mandamos a *las ditas partidas* que presentasen ante nós *todas* e c*oalesquiere cartas* …（ID938，1360 年，Navarra，NA，Judicial）

f107 20 と 30 を表す数詞

f107a treinta

20 と 30 を表す数詞には，それぞれ，v**ei**nte < VĪGĬNTĪ，t**rei**nta < TRĪGĬNTĀ のように，強勢母音が二重母音/éi~eí/となる変異形が存在する（Zamora Vicente 1967：166，252，261；Menéndez Pidal 1999：§ 89_3；Penny 2002：3.6.1）：

… E véndovoslas con entradas e con salidas e con todas sus pertenencias por *treinta* maravedís de la moneda de la guerra …（ID 1225，1287 年，Segovia，CL，Particular）

f107b trinta

20 と 30 を表す数詞には，それぞれ，v**i**nte < VĪGĬNTĪ，t**ri**nta < TRĪGĬNTĀ のように，強勢母音が/í/となる変異形が存在する（Zamora Vicente 1967：166，252，261；Menéndez Pidal 1999：§ 89_3；Penny 2002：3.6.1）：

… vos viendo por precio que recebí de vós *trinta* e un moravedís de los dineros quel rey don Alfonso mandó fazer …（ID662，1343 年，Asturias，AS，Particular）

f107c **trenta**

20 と 30 を表す数詞には，それぞれ，**v**ente < VĪGĬNTĪ，t**r**enta < TRĪGĬNTĀ のように，強勢母音が/é/となる変異形が存在する（Zamora Vicente 1967：166，252，261；Menéndez Pidal 1999：§ 89$_3$；Penny 2002：3.6.1）：

… E condennamos a los ditos lavradores dOrorvía a pagar los ditos *trenta* nueve cafizes de trigo a los ditos abat e convento o al mostrador desta sentencia …（ID938，1360 年，Navarra，NA，Judicial）

f108 **40〜90 の 10 の倍数を表す数詞**

f108a **quaraenta**

40〜90 の 10 の倍数を表す数詞には，quar**ae**nta < QŬĂDRĀGĬNTĀ，cinqu**ae**nta < QŬĪNQŬĀGĬNTĀ，ses**ae**nta < SĔXĀGĬNTĀ，set**ae**nta < SĔPTŬĀGĬNTĀ，och**ae**nta < ŎCTŌGĬNTĀ，non**ae**nta〜nov**ae**nta < NŌNĀGĬNTĀ のように，語尾が/-aenta/となる変異形が存在する（Zamora Vicente 1967：166，252，261；Menéndez Pidal 1999：§ 89$_3$；Penny 2002：3.6.1）：

… quanto nos de uemos heredar uendemos robramos toda questa heredad por c *sesaenta* morauedis …（ID209，1186 年，Palencia，CL，Eclesiástico）

f108b **quarenta**

40〜90 の 10 の倍数を表す数詞には，quar**e**nta < QŬĂDRĀGĬNTĀ，cinqu**e**nta < QŬĪNQŬĀGĬNTĀ，ses**e**nta < SĔXĀGĬNTĀ，set**e**nta < SĔPTŬĀGĬNTĀ，och**e**nta < ŎCTŌGĬNTĀ，non**e**nta〜nov**e**nta < NŌNĀGĬNTĀ のように，語尾が/-enta/となる変異形が存在する（Zamora Vicente 1967：166，252，261；Menéndez Pidal 1999：§ 89$_3$；Penny 2002：3.6.1）：

… e esta dicha vin<n>a uos uendo por precio no<n>brado *setenta* çi<n>co m<aravedis> desta mon<eda> vssada q<ue> agora corre …（ID444，1345 年，Salamanca，CL，Eclesiástico）

f108c **quaranta**

40〜90 の 10 の倍数を表す数詞には，quar**a**nta < QŬĂDRĀGĬNTĀ，cinqu**a**nta < QŬĪNQŬĀGĬNTĀ，ses**a**nta < SĔXĀGĬNTĀ，set**a**nta < SĔPTŬĀGĬNTĀ，och**a**nta < ŎCTŌGĬNTĀ，non**a**nta〜nov**a**nta < NŌNĀGĬNTĀ のように，語尾が/-anta/となる変異形が存在する（Zamora Vicente 1967：166，252，261；Menéndez Pidal 1999：§ 89$_3$；Penny 2002：3.6.1）：

… si vos olos v<uest>ros falleceredes e fallecera<n> enla paga delos ditos *quaranta* sol<do>s del dito cens en<e>l dito t<er>mino ….（ID950，1454 年，Zaragoza，AR，Eclesiástico）

4.2.3　表記的特徴

表記的特徴とは，ある音素を表すのに用いられる文字（書記素）の変異に関するものである。

f109 **語末の歯音の表記**

f109a **mercet**

CĪVĬĀTĔ や MĔRCĒDĔ のような名詞や，ĬNVĬĀTĔ のような 2 人称複数への命令形など，母音＋-TĔ や母音＋-DĔ で終わる語には，ciuda**t**，merce**t**，envia**t** のように，語末の歯音が<t>で表記される変異形が存在する（Sánchez-Prieto Borja 1998；Torrens Álvarez 1998；Penny 2002：2.5.3.2.4；Kawasaki 2013a）：

…Et pedio me por merced q<ue> yo ma<n>dasse saber la *uerdat* de q<ua>les p<ri>uillegios de q<ua>les usos ouuieran q<ue>los ma<n>dasse tener en ellos …（ID423，1270 年，León，CL，Eclesiástico）

f109b **merced**

CĪVĪĀTĔ や MĔRCĒDĔ のような名詞や，ĬNVĪĀTĔ のような 2 人称複数への命令形など，母音＋-TĔ や母音＋-DĔ で終わる語には，ciuda**d**，merce**d**，envia**d** のように，語末の歯音が<d>で表記される変異形が存在する（Sánchez-Prieto Borja 1998；Torrens Álvarez 1998；Penny 2002：2.5.3.2.4；Kawasaki 2013a）：

…Et yo P<er>o ortiç notario del deua<n>tdito Abbat co<n> *uolu*<*n*>*tad* ma<n>damie<n>to de eyl del Conue<n>to de Mo<n>tArago<n> fiç esta carta por Abc partida …（ID816，1260 年，Huesca，AR，Eclesiástico）

f110 不定形容詞 alguno「何らかの」の男性単数形の語尾の/-n/の表記

f110a **algúnd～algúnt**

「何らかの」を表す不定形容詞の男性単数形には，変異形 algú**nd**～algú**nt** < algun(o) < ĂLĬQUĪS ŪNŬ が存在する。語末の<-nd～-nt>は，/-nd~-nt/ではなく/-n/を表している。これは，語源的に<-nd～-nt>が存在する segú**nd**～segú**nt**（ただし/según/と発音される）において，<-nd～-nt>が/-n/を表わす「表記上の資格 (habilitación gráfica)」を持つようになり，非語源的に algú**nd**～～algú**nt** でも用いられるようになったためである（Sánchez-Prieto 1998：142；Sánchez-Prieto 2006：252-253）。同様の現象は，否定の不定代名詞 ningún「誰も～ない」にも見られる。男性単数形以外では，非語源的表記は見られない。

… E oto<r>go E prometo de aue<r> por firme E estable E valedera esta d<ic>ha donaçio<n> E de no<n> yr nj<n> venjr yo nj<n> ot<r>e por mj cont<ra> ella nj<n> cont<ra> p<ar>te della en *nj*<*n*>*gu*<*n*>*d* nj<n> en *algu*<*n*>*d* t<ien>po por alg<una> man<era> …（ID1346，1414 年，Madrid，MA，Particular）

f110b **algún**

「何らかの」を表す不定形容詞の男性単数形には，語源的な表記を持つ変異形 algún < algun(o) < ĂLĬQUĪS ŪNŬ が存在する（Sánchez-Prieto 1998：142；Sánchez-Prieto 2006：252-253）：

… Et p<ro>metemos nos obligamos que si por bentura por *algu*<*n*> t<iem>po algu<n>a p<er>ssona bos pusiere embargo empacho o malla boç enla d<i>ta bjna …（ID994，1449 年，Navarra，NA，Eclesiástico）

f111 硬口蓋鼻音/ɲ/の表記

f111a **anyo**

硬口蓋鼻音/ɲ/の表記には，a**ny**o < ĂNNŬ のように，変異形<ny>が存在する：

… Et vos dandoy pagando el dicho trehudo como dicho es cadaun *anyo* enel dicho dia …（ID749，1519 年，Zaragoza，AR，Eclesiástico）

f111b **ayno**

硬口蓋鼻音/ɲ/の表記には，a**yn**o < ĂNNŬ のように，変異形<yn>が存在する（González Ollé 1996：313）：

… Et assi uos p<ro>metemos lealme<n>t abona fe q<ue> nos cada *ayno* p<er>petualme<n>t dizremos las dichas dos mjssas …（ID985，1337 年，Navarra，NA，Particular）

f111c **anno**

硬口蓋鼻音/ɲ/の表記には，a**nn**o < ĂNNŬ のように，変異形<nn>が存在する（Ueda 2013a）：

… El qui ouiere un bue do bestia con ke pueda labrar o heredat pora un bue denos kada un *anno* un almud e medio de pan lo medio de trigo elo medio de ordio e nuef dineros …（ID163，1237 年，Burgos，CL，Eclesiástico）

f112 **硬口蓋側面接近音/ʎ/の表記**

f112a **ellyo**

硬口蓋側面接近音/ʎ/の表記には，e**lly**o < ĬLLŬ のように，変異形<<ly>が存在する：

… Et prometo me obligo a vos o aqualquiere vicario quj por tiempo sera del dito Monasterio todas *aquellyas* tener complir obseruar enla manera forma que de suso ditas son …（ID844，1435 年，Huesca，AR，Particular）

f112b **eyllo**

硬口蓋側面接近音/ʎ/の表記には，e**yll**o < ĬLLŬ のように，変異形<yl>が存在する（González Ollé 1996：313）：

… io don P<er>o arceiz de arroniz estando en mj memoria bona mando dono *aqueylla* heredat de Ceruera de Andion con sos *coyllaços* e coanto uenja en Nauarra de Garcia Ceruera …（ID884，1234 年，Navarra，NA，Eclesiástico）

f112c **ello**

硬口蓋側面接近音/ʎ/の表記には，e**ll**o < ĬLLŬ のように，変異形<ll>が存在する：

… encomiendo mi anima an<uest>ro Senyor dios creador de *aquella* al qual suplico por su infinita clemencia que sienpre que la querra lebar deste mundo …（ID747，1526 年，Zaragoza，AR，Particular）

f113 **/kwa/，/gwa/の表記**

f113a **qual**

/kwa/，/gwa/の表記には，qual < QŬĀLĔ のように，<o>が現れない変異形<qua>，<gua>が存在する（González Ollé 1996：312-313）：

… Ca *qual* q<ui>er q<ue> lo fiziesse pechar nos ye en coto mill m<a>r<auedis> …（ID44, 1273 年, Toledo, CM, Cancilleresco）

f113b **quoal**

/kwa/，/gwa/の表記には，qu**o**al < QŬĀLĔ のように，<o>が現れる変異形<quoa>，<guoa>が存在する（González Ollé 1996：312-313）：

… la *quoal* cosa mucho pesa anos por que aqueilla cosa amamos specialme<n>t sobre todas las otras …（ID979，1265 年，Navarra，NA，Eclesiástico）

f114 **鼻音＋両唇破裂音の表記**

f114a **ambos**

鼻音＋両唇破裂音の表記には，a**mb**os，e**mp**erador のように，変異形<mb>，<mp>が存在する（Douvier 1995）：

… otorgamos dos cartas de la d<ic>ha hermandad *ambas* en vn thenor p<ar>a cada cabildo …（ID449，1499 年，Salamanca, CL，Eclesiástico）

f114b **anbos**

鼻音＋両唇破裂音の表記には，a**nb**os，e**np**erador のように，変異形<nb>，<np>が存在する（Douvier 1995）：

… por q<ue> esto synado no<n> venga en duda *Anbas* las p<a>r<te>s otorgamos desto dos c<art>as *Anbas* fechas en vn tenor por ant<e> Jua<n> ma<r>q<ue>z …（ID425，1459 年，León，CL，Eclesiástico）

f115 **非語源的<h->の表記**

f115a **ordenar**

母音で始まる語には，**o**rdenar < ORDĬNĀRĔ のように，母音の前で非語源的な<h->が表記されない変異形が存在する（Sánchez-Prieto Borja 1998：119-120）：

… fago & *ordeno aqu<e>sti* n<uest>ro *vltimo* Testame<n>t mjo & d<e>la d<i>ta donya Cathaljna mj mug<er> …（ID748, 1409 年，Teruel，AR，Particular）

f115b **hordenar**

母音で始まる語には，**h**ordear < ORDĬNĀRĔ のように，母音の前で非語源的な<h->が表記される変異形が存在する（Sánchez-Prieto Borja 1998：119-120）：

… Los dos juntam<en>t et hu<ius> anjme conforme fazemos e *hordenamos* este n<uest>ro vltimo testame<n>to et n<uest>ra *hultima* volu<n>tat （ID757，1522 年，Teruel，AR，Particular）

第5章 カーネル平滑化

本章では，カーネル平滑化（Kernel smoothing）について説明する。カーネル平滑化とは，カーネル関数（Kernel function）を用いて，関数$f(x)$から，より滑らかな関数$\hat{f}(x)$を推定する方法である（Bishop 2006 2.5.1；Hastie *et al.* 2009：Chapter 6；Tilahun 2011；Tilahun *et al.* 2012）。カーネル平滑化では，$f(x)$に関して線形性やS字カーブなどの特殊な性質を仮定する必要がない。本論文において，$f(x)$は，各クラスにおける素性の出現頻度に当たる。カーネル平滑化により，データセットに点在する欠損値の補間と，ノイズには大きく影響されない頑健な推定が可能になる。実際に観測されたデータのみを見るのではなく，観測される可能性のあったデータも考慮することで，推定精度が向上すると期待される。（Omi *et al.* 2013）。

以下, カーネル関数, 時間カーネル平滑化（Temporal kernel smoothing）, 空間カーネル平滑化（Spatial kernel smoothing）, 時空間カーネル平滑化（Spatio-temporal kernel smoothing）について説明する。

5.1　カーネル関数

カーネル関数$K(x, x')$は，2 つの引数を取りスカラーを返す関数である。$x' \in \mathbb{R}^1$を中心としたとき，x'からの距離に応じて，観測点xへ付与される重みを表している。したがって，着目している点x'に近い（遠い）点$x \in \mathbb{R}^1$ほど大きな（小さな）重みが与えられる。最大の重みは自分自身に与えられる。観測点は等間隔で並んでいることが望ましい。等間隔で並んでいない場合，平滑化の効果が減退する。カーネル関数には，全定義域で非負$K(x, x') \geq 0$となるような任意の関数を用いることができる。本論文では，カーネル関数$K(x, x')$としてガウスカーネルを用いる：

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{5.1}$$

ガウスカーネルは釣り鐘の形をしており，着目している点x'に関して左右対称である。カーネル関数は近傍の幅を決めるパラメータ（バンド幅ともよばれる）により特徴づけられるが，ガウスカーネルでは標準偏差σがこれにあたる。σが小さすぎるとノイズに過敏となり，大きすぎると過剰な平滑化を引き起こしてしまうので，適切なσの選択が重要となる[19]。σの決定の仕方については，9.1 節で述べる。本論文では，すべての着目点x'において同一のσを用いた[20]。

図 5.1 は，$x' = 0$に注目してσを変化させた時のカーネル関数$K(x, 0)$を表している。σが小さいほど尖った形になり，大きいほど平坦な形になる：

[19] すべての観測点に同一の重み（たとえば1）を与えた場合，各注目点における重み付き平均はデータ全体の算術平均と一致し，一様分布となる。これはガウスカーネルにおいて，$\sigma = \infty$の場合である。一方，$\sigma \to 0$での平滑化は，元の分布と一致する。

[20] 各着目点x'において，異なるσを用いることも可能である。たとえば，着目点x'における標準偏差$\sigma_{x'}$を，x'についての関数（$\sigma_{x'} = f(x')$）とすることが考えられる。

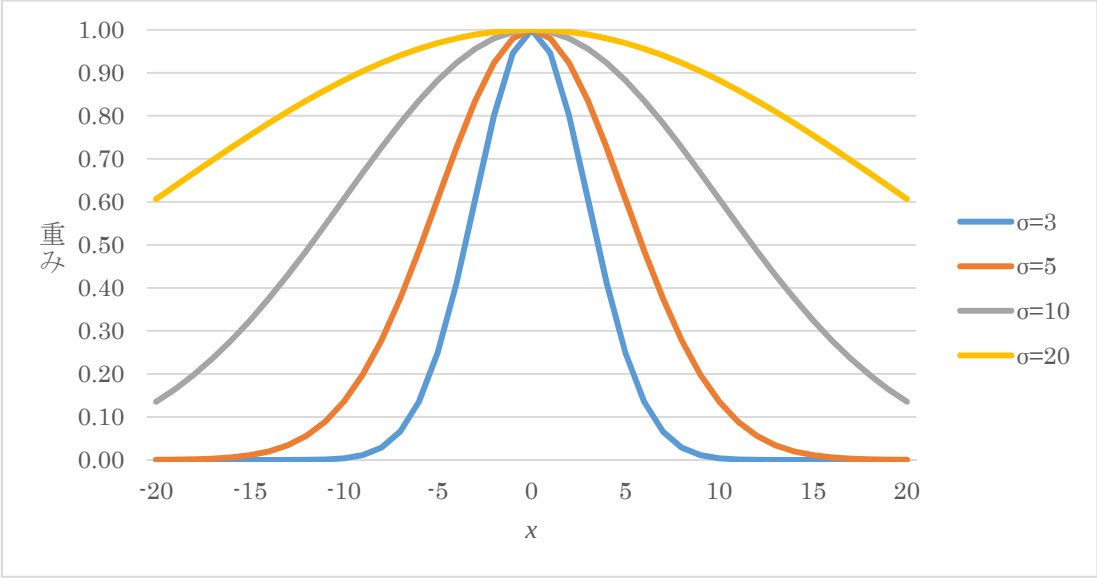


図 5.1　カーネル関数 $K(x, 0)$

5.2　時間カーネル平滑化

年代の集合を$T = \{1295, 1296, \ldots, 1305\}$，着目する年代を$t' \in T$，観測年代を$t \in T$とする。着目年代$t'$に対する年代$t$の重みを表す時間カーネル関数$K(t, t')$を，ガウスカーネルを用いて，

$$K(t, t') = \exp\left(-\frac{\|t - t'\|^2}{2{\sigma_t}^2}\right) \tag{5.2}$$

と定義する。時間カーネル平滑化パラメータσ_tは，t'やtには依存せず一定とする。したがって，$K(t, t')$は 2 つの年代間の差のみに依存し，t'からtへの差が小さい（大きい）ほど，重みは大きく（小さく）なる。

たとえば，$t' = 1300$，$\sigma_t = 10$とした場合，$K(1290, 1300) = \exp\left(-\frac{(1290-1300)^2}{2\times10^2}\right) \approx 0.61$，$K(1295, 1300) = \exp\left(-\frac{(1295-1300)^2}{2\times10^2}\right) \approx 0.88$となり，1295 年への重みのほうが 1290 年への重みより大きくなる。

ある素性の年代tにおける出現頻度をn_tとすると，年代t'におけるカーネル平滑化後の頻度$\hat{n}_{t'}$は，一次元の領域$\mathbb{R}^1$での局所重み付き平均

$$\hat{n}_{t'} = \frac{\sum_{t \in T} K(t, t') n_t}{\sum_{t \in T} K(t, t')} \tag{5.3}$$

として表される（Tilahun 2011；Tilahun *et al.* 2012）[21]。σ_tは素性によらず一定とする。また，地点の情報は無視する。た

[21] 定義域の両端付近において，カーネル関数は非対称となる。このため局所重み付き平均は，定義域の両端において望ましくないバイアスを生じる。局所線形回帰を用いることで，1 次までのバイアスを除去することができる。また，局所多項式回帰を用いることで，よりバイアスの少ない推定が可能となるが，計算量が増加してしまう（Hastie *et al.* 2009：Chapter 6）。このトレードオフを考慮し，本

だし，tがt'から大きく離れた場合，$K(t,t') \approx 0$となり，事実上$\hat{n}_{t'}$の計算には寄与しなくなる。よって，$|t-t'|$が閾値以上の場合，計算はスキップすることにする。本研究では，閾値は 20 とした。

たとえば，表 5.1 の年代分布が与えられたとする。1300 年に注目したときの重み$K(t,1300)$は，$\sigma_t = 2$として計算した。1298 年の頻度n_{1298}は欠損値である。

年代t	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305
頻度n_t	19	16	6	?	5	3	13	19	20	3	11
$K(t,1300)$ $(\sigma_t = 2)$	0.044	0.135	0.325	0.607	0.882	1.000	0.882	0.607	0.325	0.135	0.044

表 5.1　年代分布とカーネル関数 *K*(*t*, 1300)

このとき，1300 年の平滑化頻度$\hat{n}_{1300}$は，以下のように計算される。

$$\begin{aligned}
\hat{n}_{1300} &= \frac{\sum_{t\in\{1295,1296,\ldots,1305\}} K(t,1300)n_t}{\sum_{t\in\{1295,1296,\ldots,1305\}} K(t,1300)} \\
&= \frac{K(1295,1300)n_{1295} + \cdots + K(1300,1300)n_{1300} + \cdots + K(1305,1300)n_{1305}}{K(1295,1300) + \cdots + K(1300,1300) + \cdots K(1305,1300)} \\
&= \frac{0.044 \times 19 + \cdots + 1.000 \times 3 + \cdots + 0.044 \times 11}{0.044 + \cdots + 1.000 + \cdots + 0.044} \\
&\approx 8.57
\end{aligned}$$

すべての年代tについて平滑化頻度を計算すると，図 5.2 に示すように滑らかな分布になる。欠損値であった 1298 年の頻度も$\hat{n}_{1298} \approx 7.29$と補間されている。

研究では 0 次の局所多項回帰である局所重み付き平均を用いることにする。

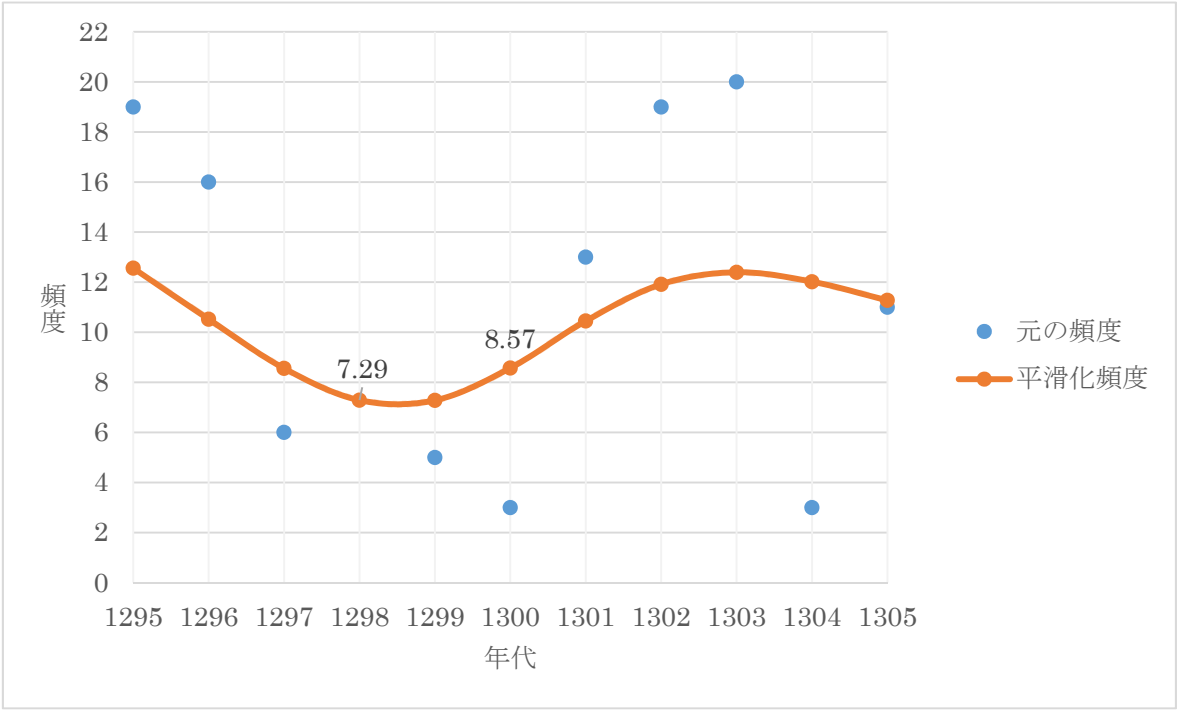


図 5.2　元の頻度と平滑化頻度

5.3　空間カーネル平滑化

地点の集合をL，着目する地点を$l' \in L$，観測地点を$l \in L$とする。着目地点l'に対する地点lの重みを表す空間カーネル関数$K(l, l')$を，ガウスカーネルを用いて，

$$K(l, l') = \exp\left(-\frac{Dis(l, l')^2}{2{\sigma_l}^2}\right) \tag{5.4}$$

と定義する。空間カーネル平滑化パラメータσ_lは，l'やlには依存せず一定とする。したがって，$K(l, l')$は二地点間の距離のみに依存し，l'からlへの距離が小さい（大きい）ほど，重みは大きく（小さく）なる。空間統計学における等方性（isotropy）を仮定しているので，南北方面にも東西方面にも同等である（瀬谷 堤 2014 : 46）。距離$Dis(l, l')$は，二地点の緯度・経度からヒュベニの公式を用いて求めた（付録 B の B1 を参照）[22]。距離の単位はkmとする。地点の粒度は，県（provincia）と自治州（comunidad autónoma）の二つのレベルを設けた。表 5.2 に県間距離行列の一部を示す。

[22] 二地点間の近さは，必ずしも地理的距離（直線距離）のみで説明できるわけではない。地点 A から同一距離にあるからといって，地点 B と地点 C が同程度に地点 A に類似しているとは限らない。たとえば，地理的障害，政治的対立，経済的活動，人口規模といった要因が影響している可能性がある。しかし，本研究では，データ収集の容易性から，地理的距離により二地点間の近さを定めた。

	Ávila	Burgos	León	Madrid	Sevilla	Toledo	Zaragoza
Ávila	0	238	251	106	370	114	353
Burgos	238	0	189	218	601	303	231
León	251	189	0	307	604	360	418
Madrid	106	218	307	0	395	90	262
Sevilla	370	601	604	395	0	305	632
Toledo	114	303	360	90	305	0	340
Zaragoza	353	231	418	262	632	340	0

表 5.2　県間距離行列

たとえば，$\sigma_l = 50$とした場合，Madrid と Ávila の距離$Dis(\text{Madrid}, \text{Ávila})$は 106km なので，$K(\text{Ávila}, \text{Madrid}) = \exp\left(-\frac{106^2}{2\times 50^2}\right) \approx 0.11$となる。一方, Madrid と Toledo の距離$Dis(\text{Madrid}, \text{Toledo})$は 90km なので, $K(\text{Toledo}, \text{Madrid}) = \exp\left(-\frac{90^2}{2\times 50^2}\right) \approx 0.20$となる。したがって，Madrid を中心とした場合，Toledo への重みのほうが Ávila への重みより約 2 倍大きくなる。

ある素性の地点lにおける出現頻度をn_lとすると，地点l'におけるカーネル平滑化後の出現頻度$\hat{n}_{l'}$は，一次元の領域$\mathbb{R}^1$での局所重み付き平均

$$\hat{n}_{l'} = \frac{\sum_{l\in L} K(l, l') n_l}{\sum_{l\in L} K(l, l')} \tag{5.5}$$

として表される（Serdyukov *et al.* 2009；Cheng *et al.* 2010；Lichman & Smyth 2014；Hulden *et al.* 2015）。σ_lは素性によらず一定とする。また，年代の情報は無視する。ただし，lがl'から大きく離れた場合，$K(l, l') \approx 0$となり，事実上$\hat{n}_{l'}$の計算には寄与しなくなる。よって，$Dis(l, l')$が閾値以上の場合，計算はスキップすることにする。本研究では，閾値は 200 とした。

たとえば，表 5.3 の地点分布と表 5.4 の地点間距離行列が与えられたとする。地点Fに注目したときの重み$K(l, F)$は，$\sigma_l = 10$として計算した。

地点l	A	B	C	D	E	F	G	H	I	J	K
頻度n_l	12	17	15	2	10	3	13	12	14	16	11
$K(l, F)(\sigma_l = 10)$	0.000	0.000	0.002	0.154	0.527	1.000	0.294	0.041	0.006	0.001	0.000

表 5.3　地点分布とカーネル関数 $K(l, F)$

地点	A	B	C	D	E	F	G	H	I	J	K
A	0	16	22	37	45	56	41	31	24	18	7
B	16	0	5	21	29	40	25	15	8	1	9
C	22	5	0	15	23	35	19	9	3	4	15
D	37	21	15	0	8	19	4	6	13	20	30
E	45	29	23	8	0	11	4	14	21	28	38
F	**56**	**40**	**35**	**19**	**11**	**0**	**16**	**25**	**32**	**39**	**50**
G	41	25	19	4	4	16	0	10	17	23	34
H	31	15	9	6	14	25	10	0	7	14	24
I	24	8	3	13	21	32	17	7	0	7	17
J	18	1	4	20	28	39	23	14	7	0	11
K	7	9	15	30	38	50	34	24	17	11	0

表 5.4　地点間距離行列

このとき，地点Fの平滑化頻度$\hat{n}_F$は以下のように計算される。

$$
\begin{aligned}
\hat{n}_F &= \frac{\sum_{l\in\{A,\ldots,K\}} K(l,F)n_l}{\sum_{l\in\{A,\ldots,K\}} K(l,F)} \\
&= \frac{K(A,F)n_A + \cdots + K(F,F)n_F + \cdots + K(K,F)n_K}{K(A,F) + \cdots + K(F,F) + \cdots + K(K,F)} \\
&= \frac{0.000 \times 12 + \cdots 1.000 \times 3 + \cdots + 0.000 \times 11}{0.000 + \cdots + 1.000 + \cdots + 0.000} \\
&\approx 6.43
\end{aligned}
$$

同様にすべての地点の平滑化頻度を計算すると，図 5.3 に示すように，大きい値は小さくなり，小さい値は大きくなり，凸凹な分布が滑らかになる。

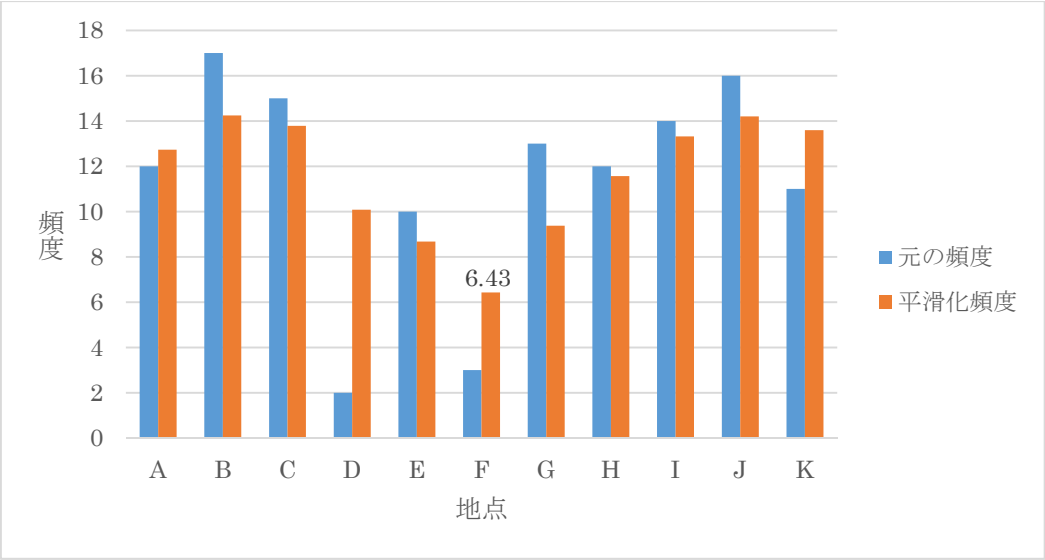


図 5.3　元の頻度と平滑化頻度の棒グラフ

5.4　時空間カーネル平滑化

ここで，Tilahun（2011）の r 次元多変数カーネル（r-dimensional multivariate kernel）に基づき，時間と空間の両方を考慮した二次元カーネル平滑化を考える（瀬谷 堤 2014：5.6；古谷 2011：9 章）。ある素性の年代t，地点lにおける頻度を$n_{t,l}$とする。このとき，年代t'，地点l'におけるカーネル平滑化頻度$\hat{n}_{t',l'}$は，二次元の領域$\mathbb{R}^2$での局所重み付き平均

$$\hat{n}_{t',l'} = \frac{\sum_{t\in T}\sum_{l\in L} K(t,t')K(l,l')n_{t,l}}{\sum_{t\in T}\sum_{l\in L} K(t,t')K(l,l')} \tag{5.6}$$

として表される。ただし，年代tと地点lには相関がないとする。時間カーネル関数$K(t,t')$と空間カーネル関数$K(l,l')$の積からなる$K(t,t') * K(l,l')$を，時空間カーネル関数と呼ぶことにする。時間カーネル平滑化パラメータσ_tと空間カーネル平滑化パラメータσ_lは，t'，t，l'，lや素性には依存せず一定とする[23]。年代tが着目年代t'に近いほど，また地点lが直目地点l'に近いほど，大きな重みが与えられる。たとえば，$\sigma_t = 2$，$\sigma_l = 10$として，$t' = 1300$，$l' = F$に着目したとき，表 5.1 と表 5.3 より，$K(1300,1300)K(F,F) = 1.000 \times 1.000 = 1.000$，$K(1298,1300)K(E,F) = 0.607 \times 0.527 = 0.320$，$K(1305,1300)K(D,F) = 0.044 \times 0.154 = 0.007$のように計算される。図 5.4 に，$K(t,1300) * K(l,F)$の分布を示す。

[23] 本研究では，カーネル平滑化のパラメータσ_tとσ_lを全クラスで同一としたが，クラスごとに異なるパラメータを用いることも可能である。たとえば，文書数が多いクラスでは小さいパラメータで平滑化範囲を狭くし，逆に文書数が少ないクラスでは大きいパラメータで平滑化範囲を広くすることなどが考えられる。また，年代tや地点lについて何らかの先験的な知識があれば，それに基づいてパラメータを調整することも可能であろう。

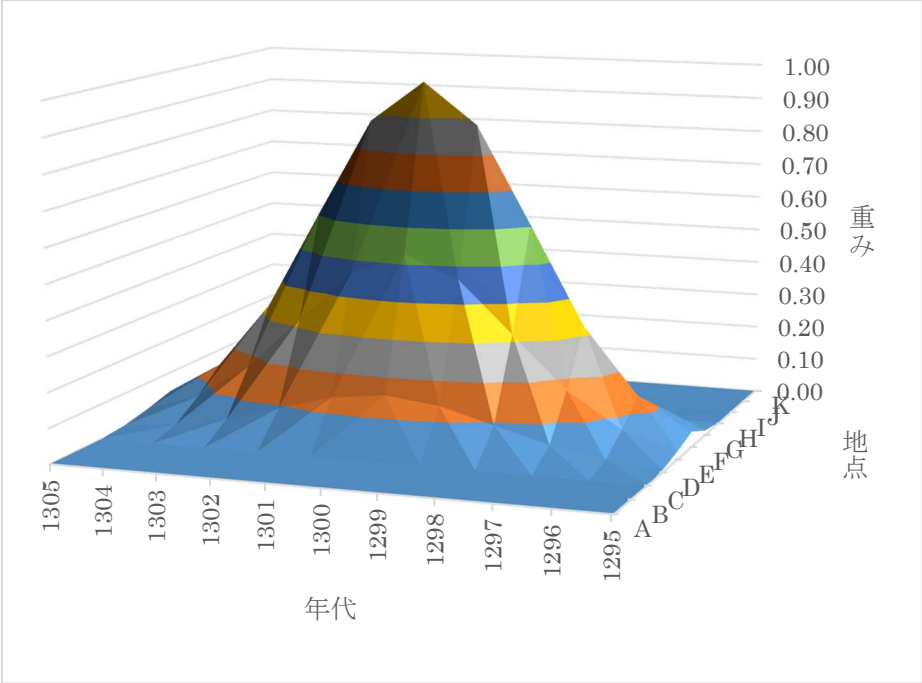


図 5.4　時空間カーネル関数 $K(t, 1300)*K(l, F)$（$\sigma_t = 2, \sigma_l = 10$）

たとえば, ある素性について, 表 5.5 のような時空間分布が与えられたとする。縦軸は年代$t \in \{1295, \ldots, 1305\}$を, 横軸は地点$l \in \{A, \ldots, K\}$を表している。

	A	B	C	D	E	F	G	H	I	J	K
1295	13	13	6	16	20	19	6	7	10	18	14
1296	16	12	10	12	15	16	3	2	19	8	13
1297	14	2	15	8	7	6	5	14	17	13	20
1298	10	2	20	12	9	0	4	16	17	18	0
1299	15	5	10	0	0	5	7	1	19	15	9
1300	12	17	15	2	10	3	13	12	14	16	11
1301	15	10	10	16	11	13	4	10	2	5	4
1302	2	6	15	20	18	19	10	13	10	17	0
1303	7	13	0	16	9	20	12	17	10	19	12
1304	7	19	13	3	15	3	8	10	16	8	18
1305	4	13	10	5	6	11	16	10	7	10	2

表 5.5　時空間分布

このとき，$t' = 1300$，$l' = F$における平滑化頻度$\hat{n}_{1300,F}$は，以下のように計算される。地点間距離は表 5.4 を用いた：

$$\hat{n}_{1300,F}$$

$$= \frac{\sum_{t\in\{1295,\ldots,1305\}}\sum_{l\in\{A,\ldots,K\}} K(t,1300)K(l,F)n_{t,l}}{\sum_{t\in\{1295,\ldots,1305\}}\sum_{l\in\{A,\ldots,K\}} K(t,1300)K(l,F)}$$

$$= \frac{K(1295,1300)K(A,F)n_{1295,A} + \cdots + K(1300,1300)K(F,F)n_{1300,F} + \cdots + K(1305,1300)K(K,F)n_{1305,F}}{K(1295,1300)K(A,F) + \cdots + K(1300,1300)K(F,F) + \cdots + K(1305,1300)K(K,F)}$$

$$= \frac{0.044 \times 0.000 \times 13 + \cdots + 1.000 \times 1.000 \times 3 + \cdots + 0.044 \times 0.000 \times 2}{0.044 \times 0.000 + \cdots + 1.000 \times 1.000 + \cdots + 0.044 \times 0.000}$$

$$\approx 8.79$$

図 5.5 は元の頻度の分布を，図 5.6 は時空間カーネル平滑化頻度の分布を示している。後者では，欠損値補間と頑健な推定が行われ，より滑らかな分布となっている。

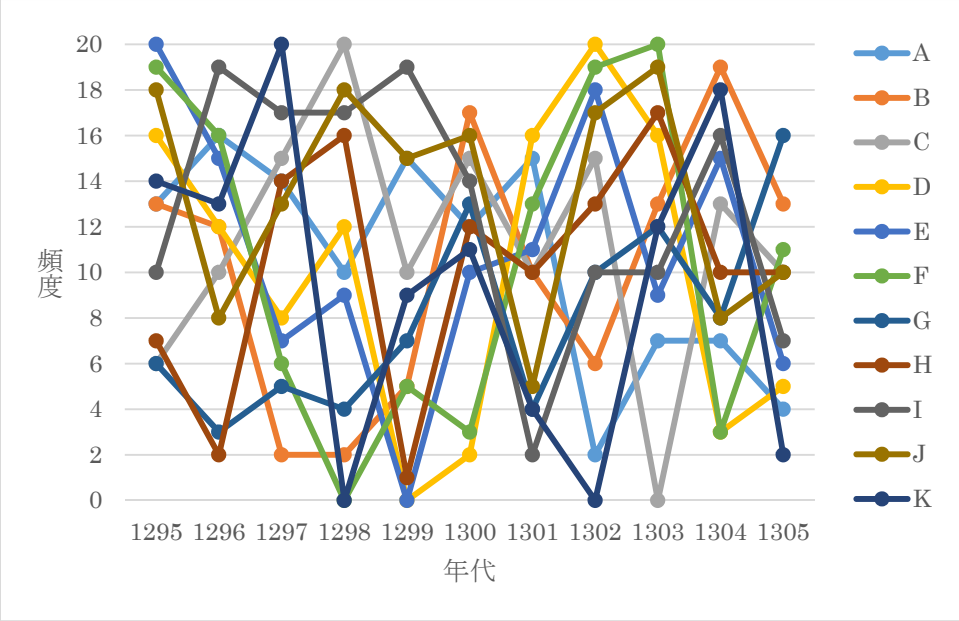


図 5.5　元の頻度の分布

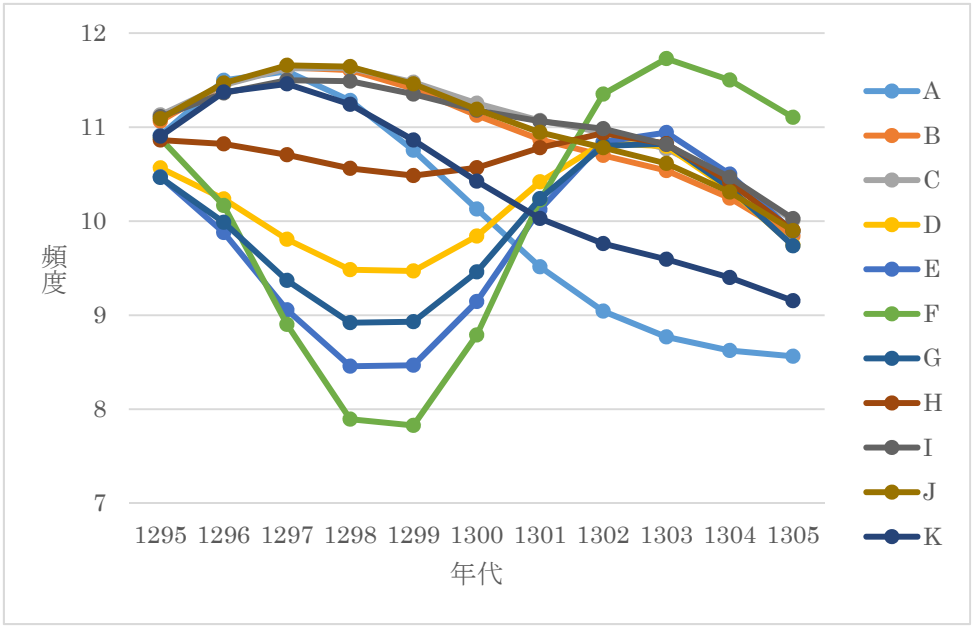


図 5.6　時空間カーネル平滑化頻度の分布

5.5　カーネル平滑化の言語学的解釈

上述のカーネル平滑化は，言語的連続体（linguistic continuum）という概念に基づいて，欠損値補間と頑健な推定を行う方法であると考えることができる。言語的連続体の概念によれば，言語は時間と空間の両方において漸進的に（滑らかに）変化する（Penny 2000）。たとえば，1520 年のスペイン語は 1519 年や 1521 年のスペイン語とほぼ同一であり，また 1500 年や 1480 年のスペイン語よりも 1510 年や 1490 年のスペイン語に似ているとみなすことができる。同様に，マドリードのスペイン語はアルカラ・デ・エナーレス（マドリードから約 30 ㎞）のスペイン語とほぼ同一であり，サラゴサ（マドリードから約 330 ㎞）のスペイン語よりもトレドのスペイン語（マドリードから約 70 ㎞）に似ているとみなすことができる。カーネル平滑化は，年代差や地理的距離が小さい（大きい）ほど類似度が大きい（小さい）という仮定をカーネル関数で表現し，隣接する年代や地点の頻度も考慮して滑らかな分布を作り出す方法である。図 5.7，図 5.8，図 5.9 は，それぞれ時間カーネル平滑化，空間カーネル平滑化，時空間カーネル平滑化のイメージ図である。矢印の太さは重みの大きさを表している。

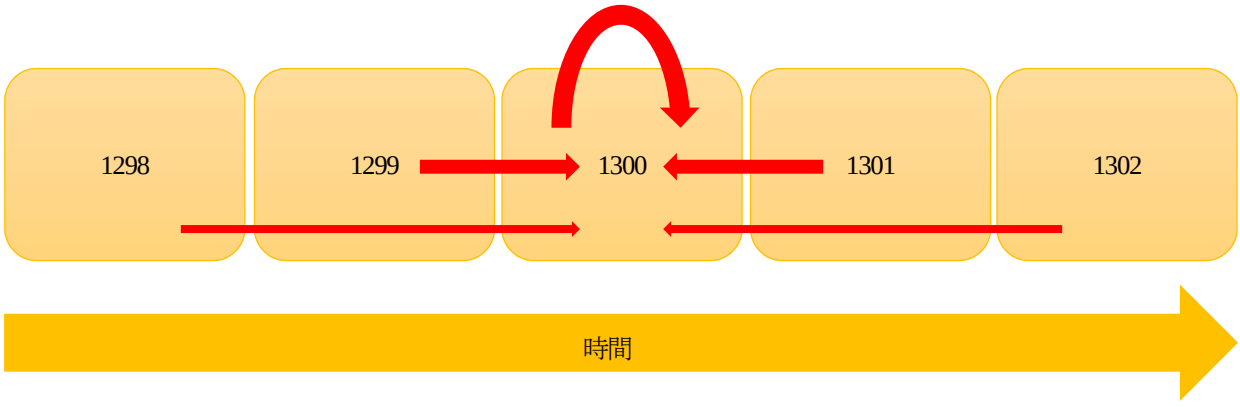


図 5.7　時間カーネル平滑化のイメージ

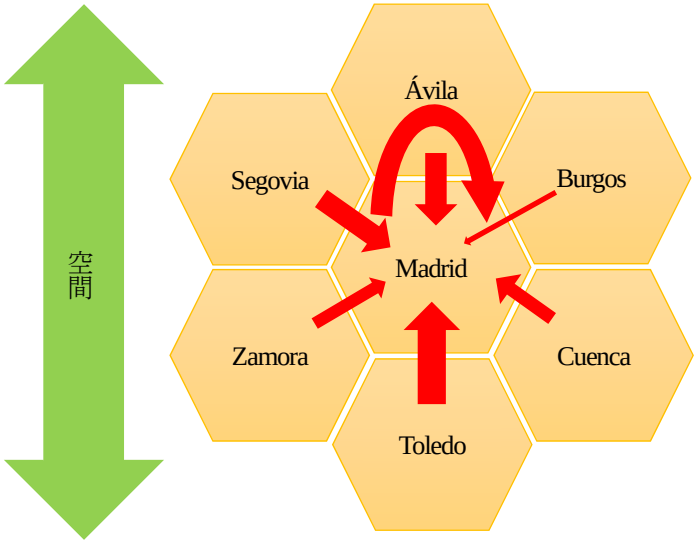


図 5.8　空間カーネル平滑化頻度のイメージ

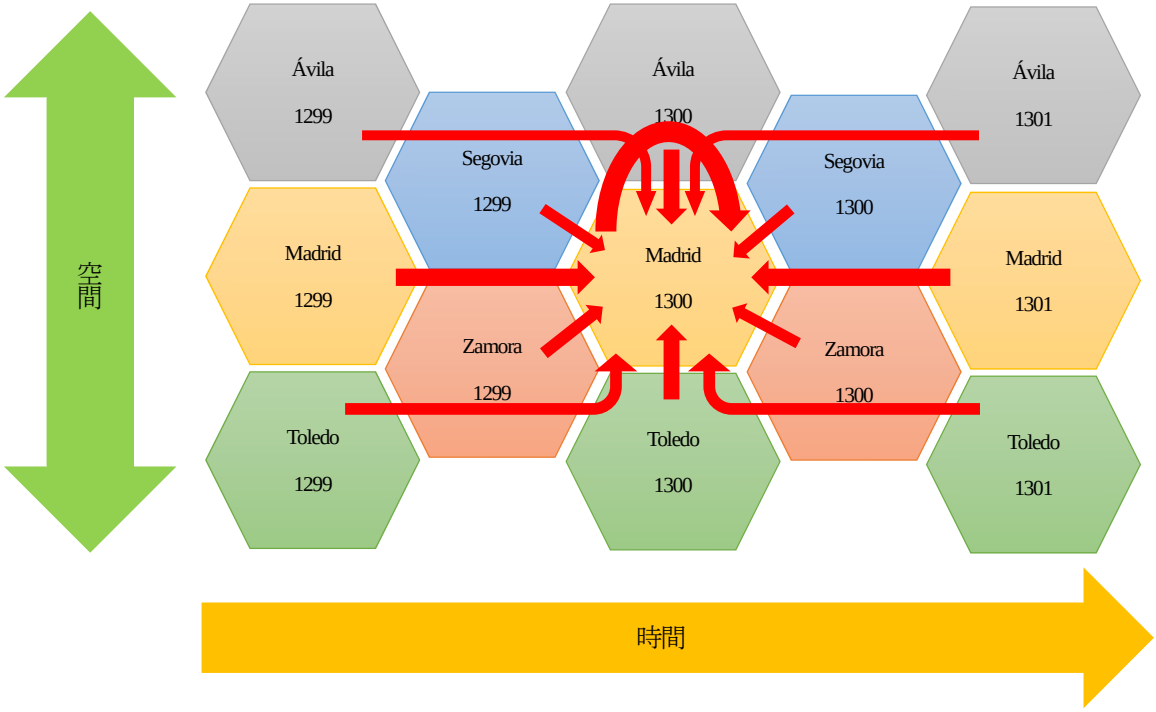


図 5.9　時空間カーネル平滑化のイメージ

5.6　応用例：文字 2-gram の条件付き確率

時空間カーネル平滑化の応用例として，文字 2-gram の条件付き確率の CL（Castilla y León）における年代推移を可視化する。CL は，スペイン中北部の自治州である。文字列aに文字列bが後続する条件付き確率を$P(b|a)$とする。$P(b|a)$が大きい（小さい）とき，文字列aの後ろに文字列bが続く確率が高く（低く）なる。条件付き確率の計算方法は，第 6 章を参照。ここでは，加算スムージングのパラメータ$\alpha = 0.001$とする。

時空間カーネル平滑化のパラメータは，$\sigma_l = 100$と固定したうえで，$\sigma_t = \{3, 5, 10\}$で変化させることにする。以下のグラフにおいて，その他の説明がない場合，中抜きの黒丸は実際に観測された値，曲線はカーネル平滑化後の値を

示している。青色は$\sigma_t = 3$，緑色は$\sigma_t = 5$，赤色は$\sigma_t = 10$の場合の曲線である（凡例の σ_t はσ_tを表している）。中抜きの黒丸が存在しない年代は，当該年代に文書が一つも存在しない年代である。グラフごとに，条件付き確率を表す縦軸の範囲が異なることに注意されたい。時空間カーネル平滑化のパラメータ値を変化させると，曲線の形も変わる。σ_lを固定した場合，σ_tを大きくするほど平滑化の効果が大きくなり，曲線はより滑らかになる。逆に，σ_tを小さくするほど局所的な特徴に敏感になり，曲線はより波打った形になる。

時空間カーネル平滑化では，時間カーネル平滑化に比べクラスがより細分化されているので，各クラスの文書数は少なくなる。そのため，年代推移を表すカーネル平滑化曲線は，より波打った形になる。この傾向は，以下のグラフに見られるように，特に文書数が少ない 1100 年代や 1600 年代で顕著になる。

5.6.1　*P*(*n*|*n*)の年代推移

図 5.10 は，文字 n に文字 n が後続する条件付き確率$P(n|n)$の CL における年代推移を表している。nn という文字列は，ラテン語では歯茎鼻音[n]の連続[nn]を（たとえば ANNU「年」），中近世スペイン語では一般的に硬口蓋鼻音[ɲ]を表している（たとえば **senn**or < SENIORE「主」）。$P(n|n)$は 13 世紀後半までは大きい値であるが，14 世紀以降は$P(n|n) \approx 0$となり，文字 n に文字 n が後続する現象は見られなくなる。これには二つの要因が考えられる。一つ目の要因は，14 世紀以降，ラテン語で文書を書くことが稀になったことである。二つ目の要因は，14 世紀以降，硬口蓋鼻音[ɲ]を表すために，nn に代わり n<n>や ñ が使用されるようになったことである（Ueda 2013a）。

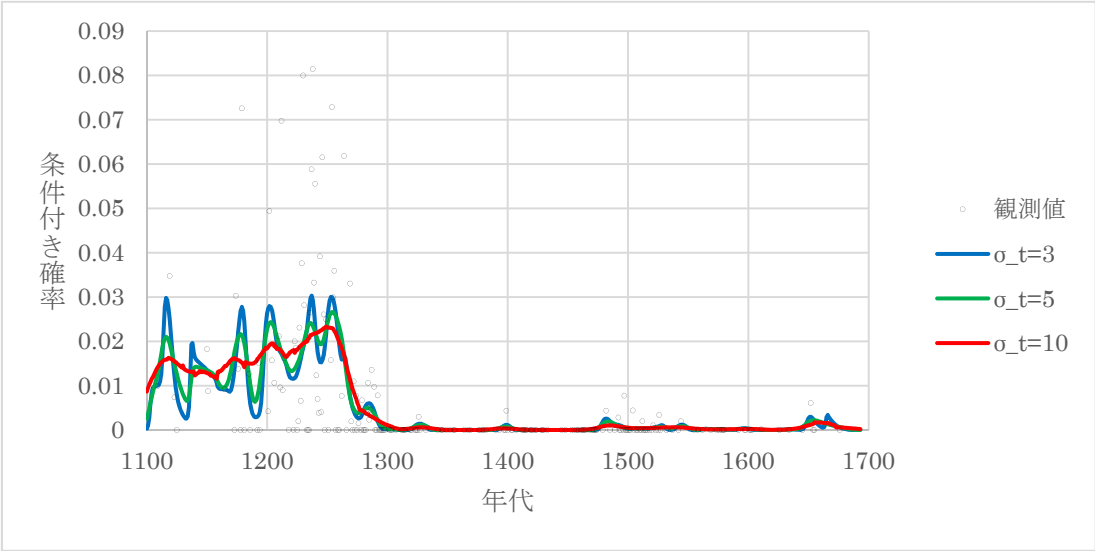


図 5.10　*P*(*n*|*n*)の CL における年代推移

5.6.2　*P*(_|*m*)の年代推移

図 5.11 は，文字 m にスペース（_）が後続する条件付き確率$P(_|m)$の CL における年代推移を表している。m_という文字列は，ラテン語には多く存在する（たとえば，名詞類の対格（ROSA「バラ」や DOMINU「主人」など）や動詞の活用形（SUM「ESSE「存在する」の直説法現在一人称単数」）など）。一方，中近世スペイン語では，語尾母音脱落（apócope）の場合を除き，通常見られない文字連続である。$P(_|m)$は 13 世紀後半までは大きい値であるが，14 世紀以降は$P(_|m) \approx 0$となり，文字 m にスペース（_）が後続する現象は見られなくなる。これは，14 世紀以降，ラテン語で文書を書くこと

が稀になったためだと考えられる。

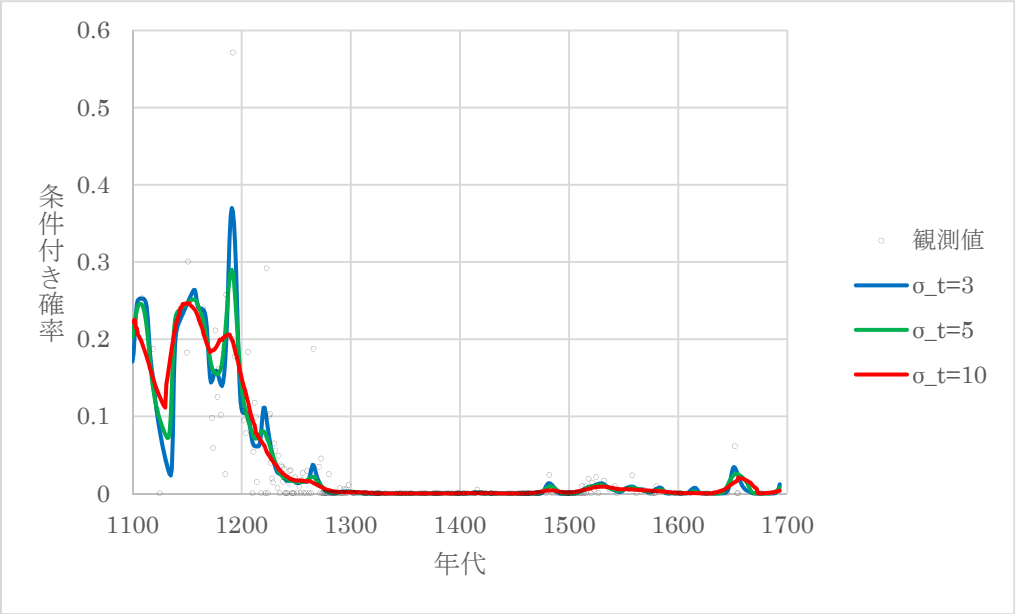


図 5.11　*P*(_|*m*)の CL における年代推移

5.6.3　*P*(*y*|_)の年代推移

図 5.12 は，スペース（_）に文字 y が後続する条件付き確率*P*(*y*|_)の CL における年代推移を表している。_y という文字列は，y で始まる単語とみなすことができる。13 世紀以前は，*P*(*y*|_) ≈ 0 である。これは，13 世紀以前の文書の大半がラテン語の文書であり，ラテン語では y で始まる単語がほぼ皆無であるためである。13 世紀から 15 世紀末までは，小さい値のまま推移しているが，16 世紀初から*P*(*y*|_)は急増する。これは，接続詞 e < ET「と」が y に変化し，y で始まる単語の割合が増加したためだと考えられる（Menéndez Pidal 1999：§130）。

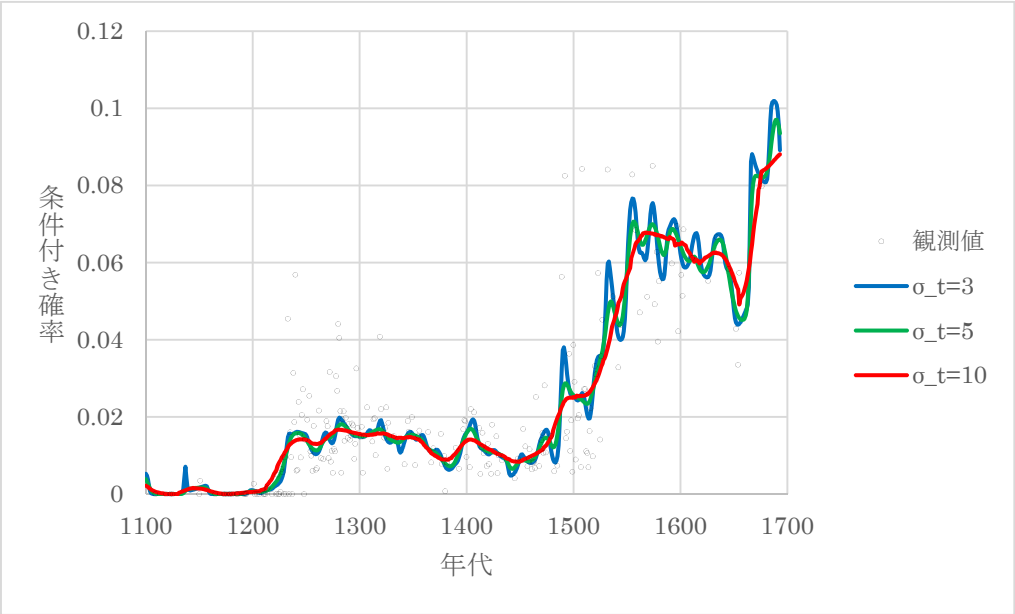


図 5.12　*P*(*y*|_)の CL における年代推移

5.6.4　*P*(*d*|*b*)の年代推移

図 5.13 は，文字 b に文字 d が後続する条件付き確率$P(d|b)$の CL における年代推移を表している。bd という文字列は，中近世スペイン語の ca**bd**al < CAPITALE，ci**bd**ad < CIVITATE，de**bd**a < DEBITA などに見られる（Sánchez-Prieto 1998：153；Sánchez-Prieto 2006：253）。$P(d|b)$は 15 世紀半に最大となり，それ以降は減少する。これは，cabdal > ca**ud**al，cibdad > ci**ud**ad，debda > de**ud**a のように，一部の単語の bd が ud となったためだと考えられる。

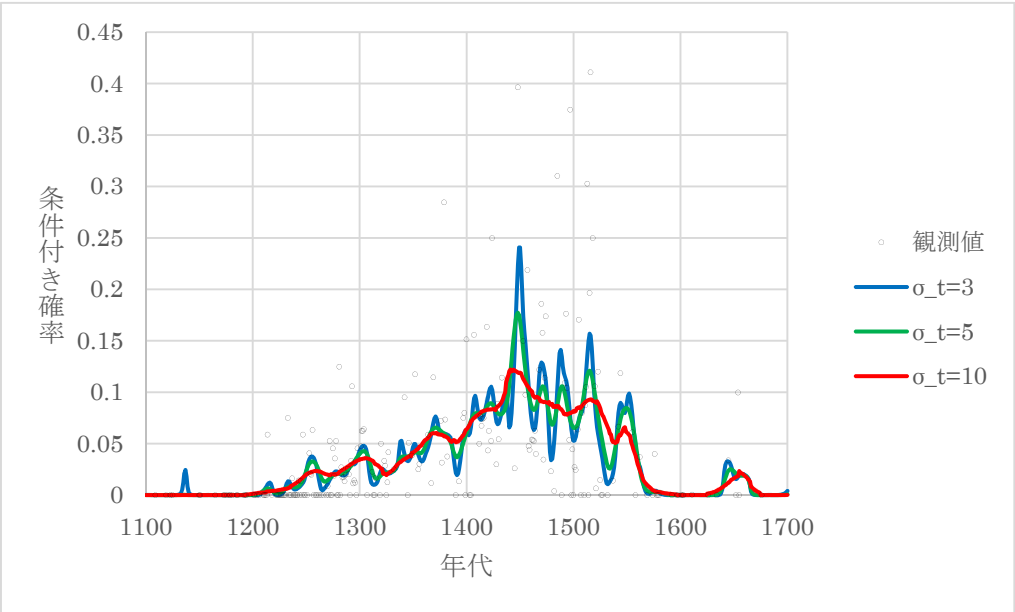


図 5.13　*P*(*d*|*b*)の CL における年代推移

5.6.5　*P*(*u*|*q*)と *P*(@|*q*)の年代推移

図 5.14 は，文字 q に文字 u が後続する条件付き確率$P(u|q)$の CL における年代推移を表している。スペイン語では，q に後続する文字は必ず u となり，qu は軟口蓋無声破裂音[k]を表す。$P(u|q)$の年代推移は，下に凸な放物線のような曲線となる。

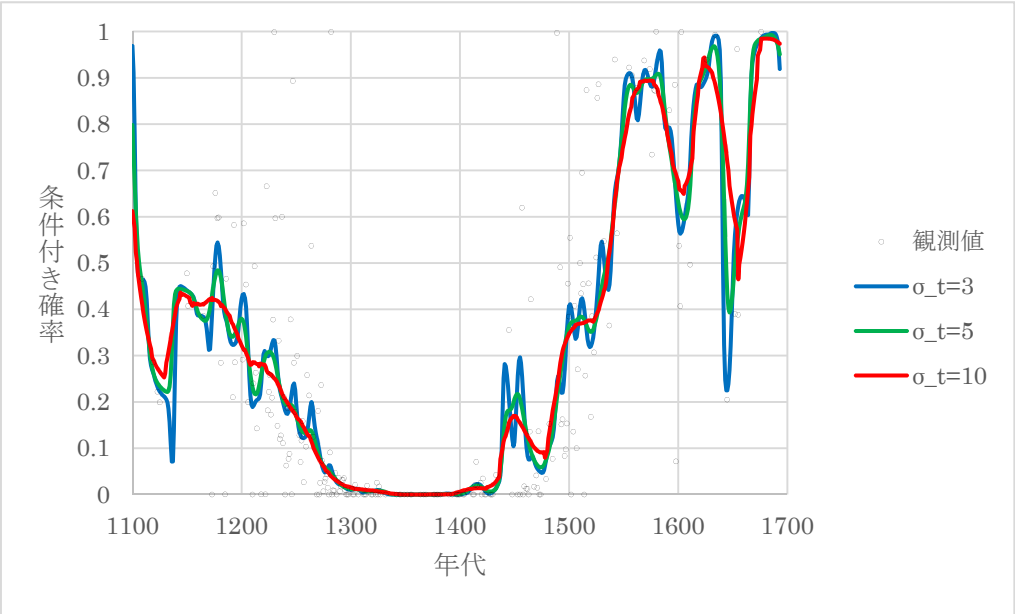


図 5.14　*P*(*u*|*q*)の CL における年代推移

$P(u|q)$とは反対に，文字 q に略記@が後続する条件付き確率$P(@|q)$の年代推移は，図 5.15 に示すように，上に凸な放物線のような曲線となる。文字列 q@は，q<ue>や q<ui>や q<ua>に相当する。14 世紀の間は，ほぼ常に略記法である q@が用いられていた（$P(@|q) \approx 1$）ため，$P(u|q) \approx 0$となる。

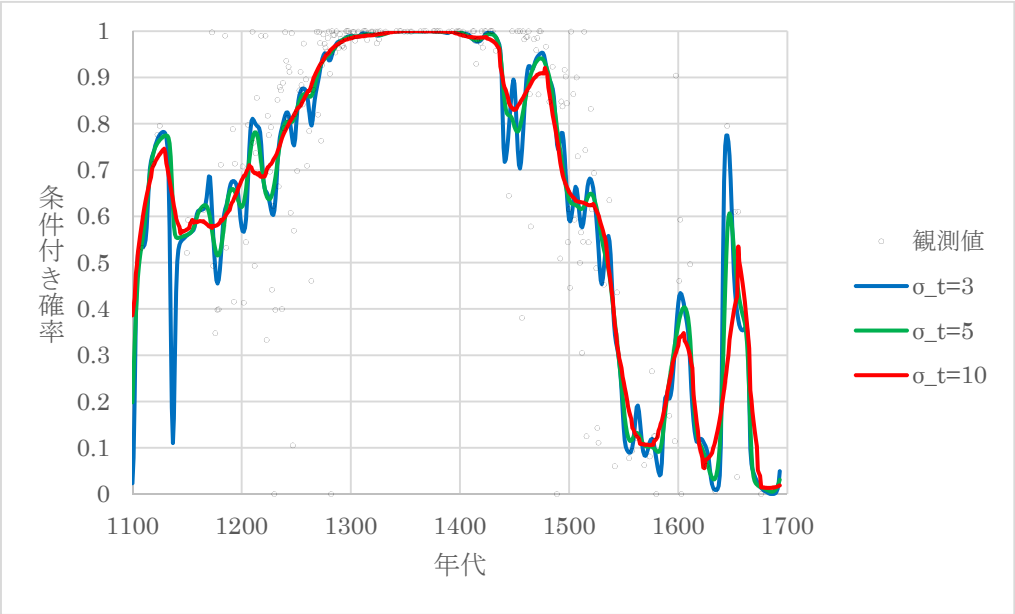


図 5.15　*P*(@|*q*)の CL における年代推移

5.6.6　*P*(s|_)の年代推移

図 5.16 は，スペース（_）に文字 s が後続する条件付き確率*P*(*s*|_)の CL における年代推移を表している。_s という文字列は，s で始まる単語とみなすことができる。*P*(*s*|_)は，13 世紀以降，大きな年代推移を示さない。これは，s で始まる単語の比率が年代により大きく変化しないためだと考えられる。

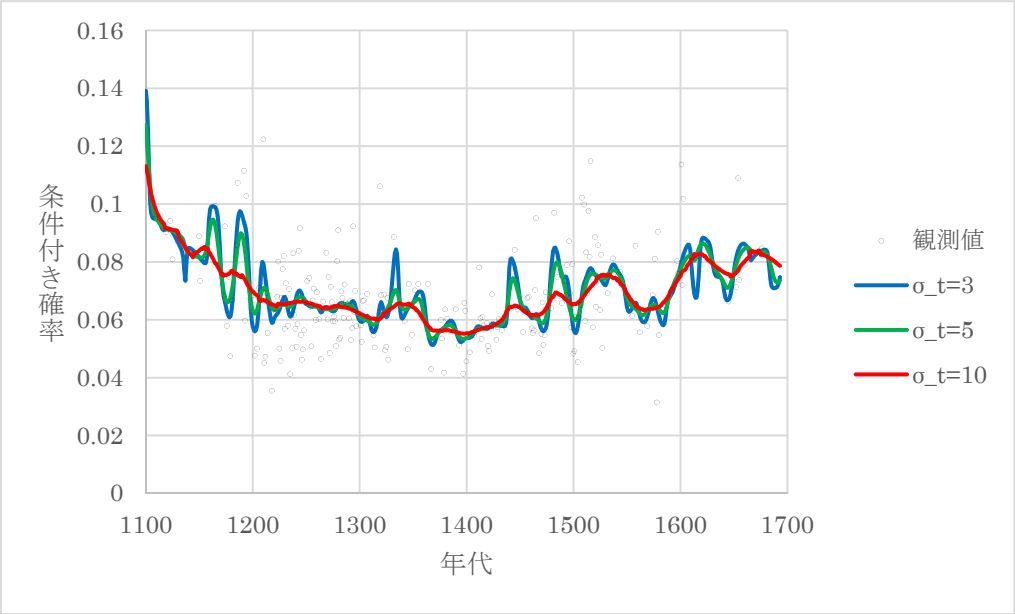


図 5.16　*P*(*s*|_)の CL における年代推移

5.6.7　*P*(c|#)，*P*(d|#)，*P*(i|#)，*P*(s|#)の年代推移

図 5.17 は，文頭記号#に文字 c，d，i，s が後続する条件付き確率$P(c|\#)$，$P(d|\#)$，$P(i|\#)$，$P(s|\#)$の CL における年代推移を表している。時空間カーネル平滑化のパラメータは，$\sigma_l = 100$，$\sigma_t = 10$である。$P(c|\#)$は文書が c で始まる確率を，$P(d|\#)$は文書が d で始まる確率を，$P(i|\#)$は文書が i で始まる確率を，$P(s|\#)$は文書が s で始まる確率を表している。$P(c|\#)$は，13 世紀後半にかけて増加している。これは **c**onocida cosa sea a cuantos esta carta vieren como …のように，c で始まる文書の増加を反映していると考えられる。$P(d|\#)$は，16 世紀半ばにかけて増加する。これは，**d**on enrique por la gracia de dios rey de castilla …や **d**oña juana por la gracia de dios reina de castilla …のように，d で始まる文書の増加を反映していると考えられる。$P(i|\#)$は，13 世紀前半にかけて増加している。これは，**i**n dei nomine …や **i**n nomine domini …のように，i で始まる文書の増加を反映していると考えられる。$P(s|\#)$は，14 世紀から 15 世紀半ばにかけて大きい値である。これは，**s**epan quantos esta carta vieren como …のように，s で始まる文書の多さを反映していると考えられる。

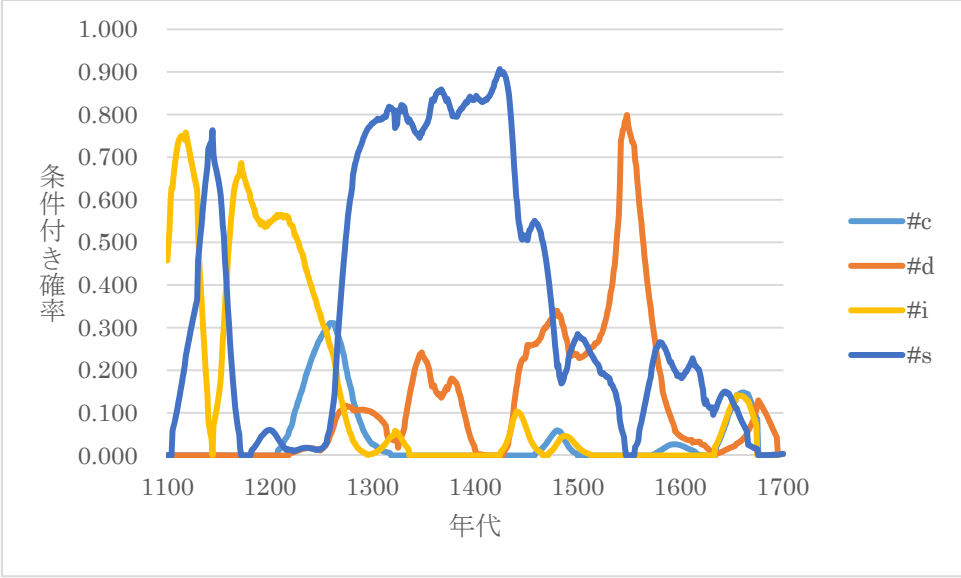


図 5.17　*P*(c|#)，*P*(d|#)，*P*(i|#)，*P*(s|#)の CL における年代推移

5.6.8　*P*(_|t)の年代推移

図 5.18 は，文字 t にスペース（_）が後続する条件付き確率$P(_|t)$の AR，CL，CM，MD，NA における年代推移と，地点を区別しない時間カーネル平滑化による年代推移を表している。時空間カーネル平滑化のパラメータは$\sigma_l = 100$と$\sigma_t = 10$，時間カーネル平滑化のパラメータは$\sigma_t = 10$とした。曲線が途切れている年代は，文書が存在しない年代である。t_という文字列は，t で終わる単語とみなすことができる。t で終わる単語は，ラテン語（たとえば，ET や AMANT）や中近世スペイン語（たとえば，ciudat や delant）には存在するが，現代スペイン語には外来語等を除き存在しない（Sánchez-Prieto Borja 1998；Torrens Álvarez 1998；Penny 2002：2.5.3.2.4；Kawasaki 2013a）。

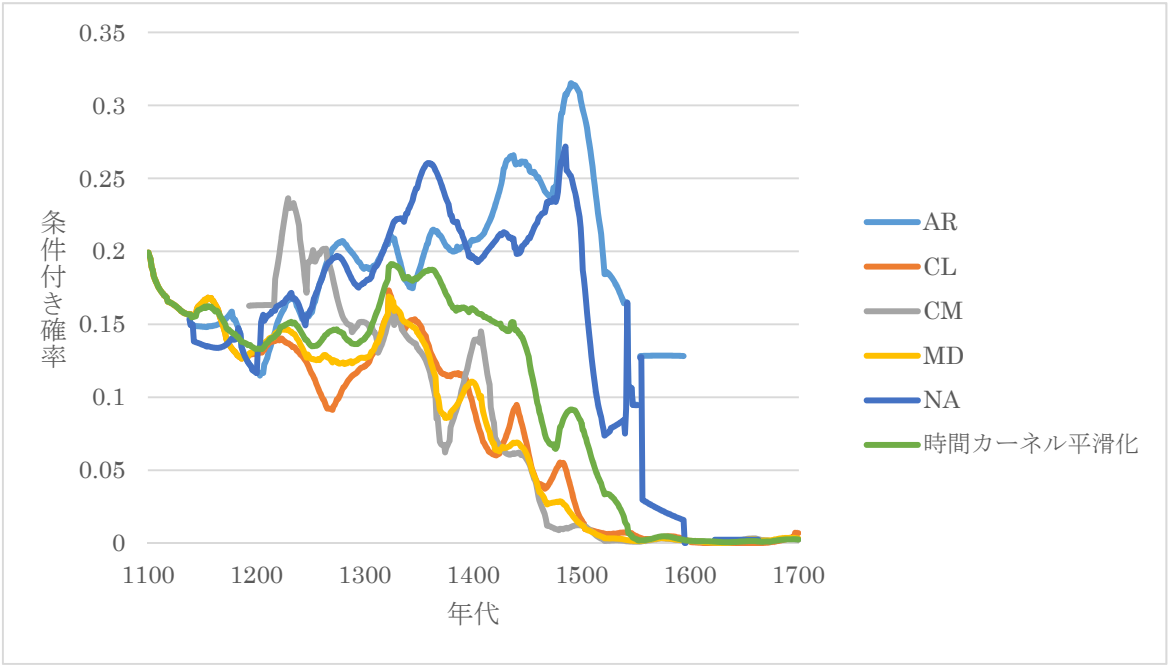


図 5.18　$P(_|t)$の各自治州における年代推移

$P(_|t)$は，スペイン中部・中北部の CL，CM，MD では，14 世紀初から低下し，16 世紀以降は$P(_|t) \approx 0$となる。一方，スペイン東部の AR と NA では，$P(_|t)$は 16 世紀まで大きいままである。地点の区別をしない時間カーネル平滑化を行うと，年代推移は全地点の平均的な曲線になってしまう。時空間カーネル平滑化を用いることで，各地点毎の年代推移を捉えることができる。

5.7　応用例：文献学的特徴の出現確率

つづいて，時空間カーネル平滑化の応用例として，文献学的素性セットのうち，いくつかの交替変数の CL（Castilla y León）における年代推移を可視化する。時間カーネル平滑化パラメータ$\sigma_t = 10$，空間カーネル平滑化のパラメータ$\sigma_l = 100$，ナイーブベイズ多変数ベルヌーイモデルのハイパーパラメータ$\alpha = 0.001$とした。確率の計算方法は，第 8 章を参照。

5.7.1　ove～uve の年代推移

図 5.19 は，andar，estar，tener，haber，placer，saber などの PYTA に関する交替変数 f37 の二つの変異形 ove（f37a）と uve（f37b）の CL における年代推移を表している。15 世紀半ばから，ove の出現確率が下がり，uve の出現確率が上昇するのが分かる。

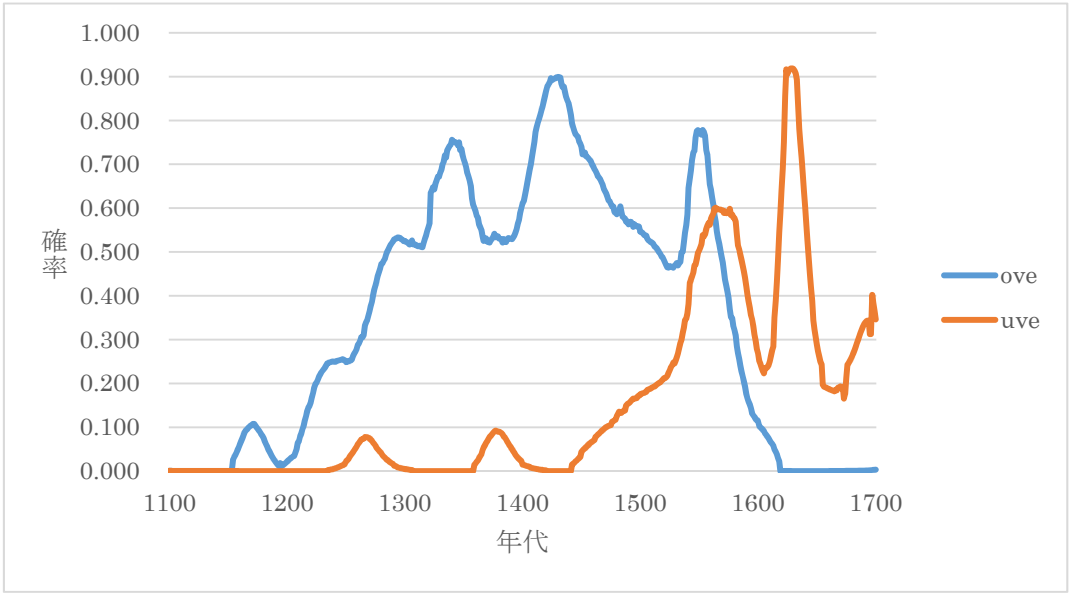


図 5.19　ove～uve の CL における年代推移

5.7.2　comiemos～comimos の年代推移

図 5.20 は，直説法点過去において弱変化する-er 動詞，-ir 動詞の 1 人称複数と 2 人称複数の語尾に関する交替変数 f46 の二つの変異形 comiemos（f46a）と comimos（f46b）の CL における年代推移を表している。15 世紀半ばには，comiemos の出現確率はほぼゼロとなり，comimos のみが用いられるのが分かる。

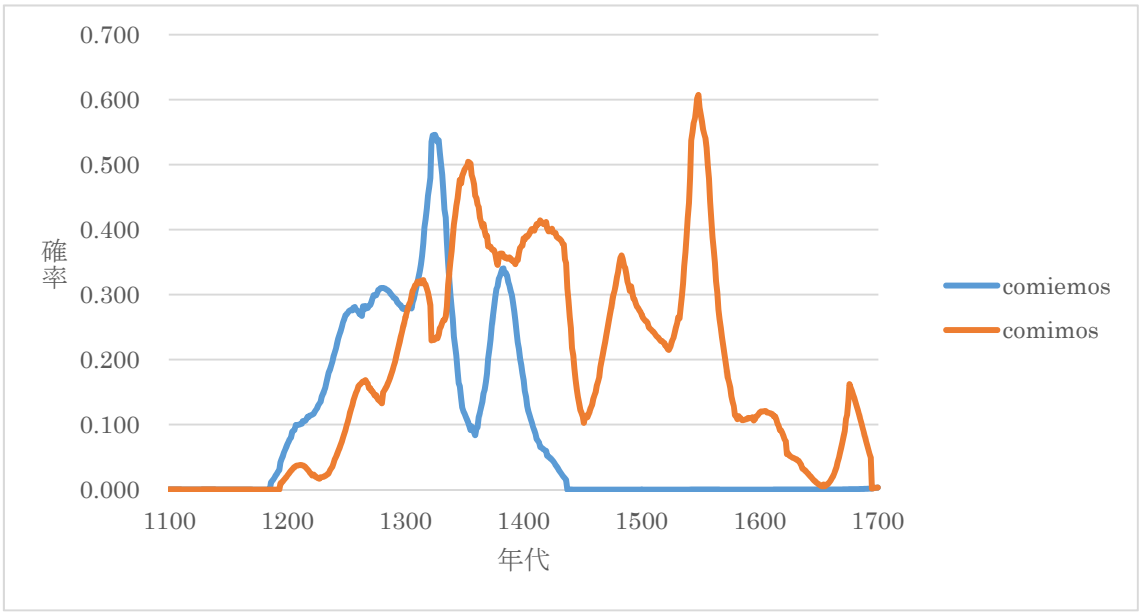


図 5.20　comiemos～comimos の CL における年代推移

5.7.3　agora～ahora の年代推移

図 5.21 は，副詞「今」を表す交替変数 f87 の二つの変異形 agora（f87a）と ahora（f87b）の CL における年代推移を表している。16 世紀半ば以降，ahora の出現確率が上昇し，agora の出現確率は低下するのが分かる。

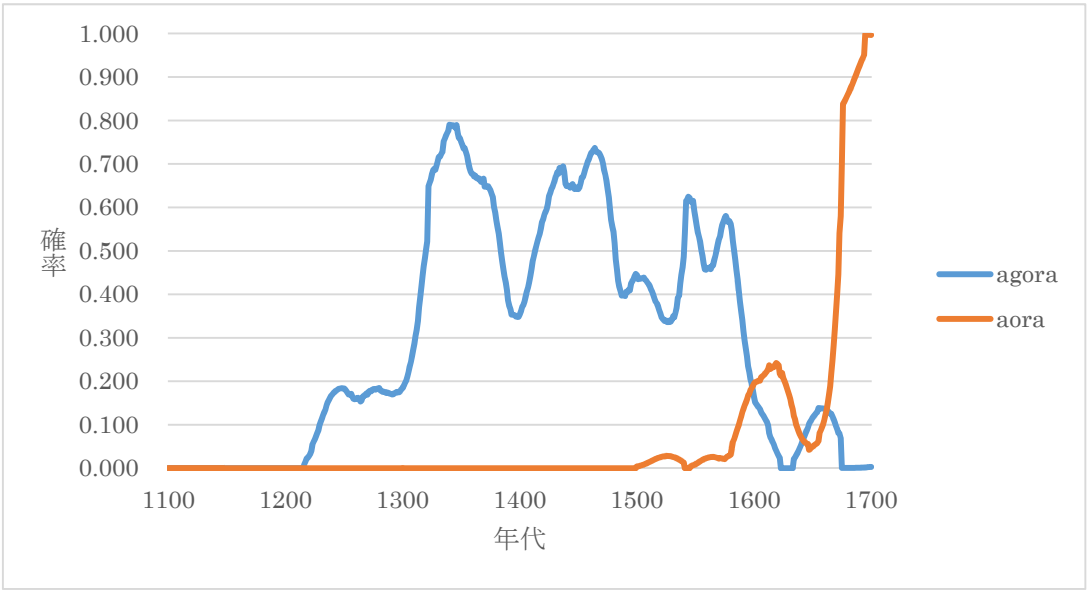


図 5.21　agora～ahora の CL における年代推移

5.7.4　vos～os の年代推移

図 5.22 は，2 人称複数の直接・間接目的格・再帰代名詞に関する交替変数 f12 の二つの変異形 vos（f12a）と os（f12b）の CL における年代推移を表している。16 世紀から os の出現頻度が上昇し，17 世紀以降は vos が使用されなくなるのが分かる。

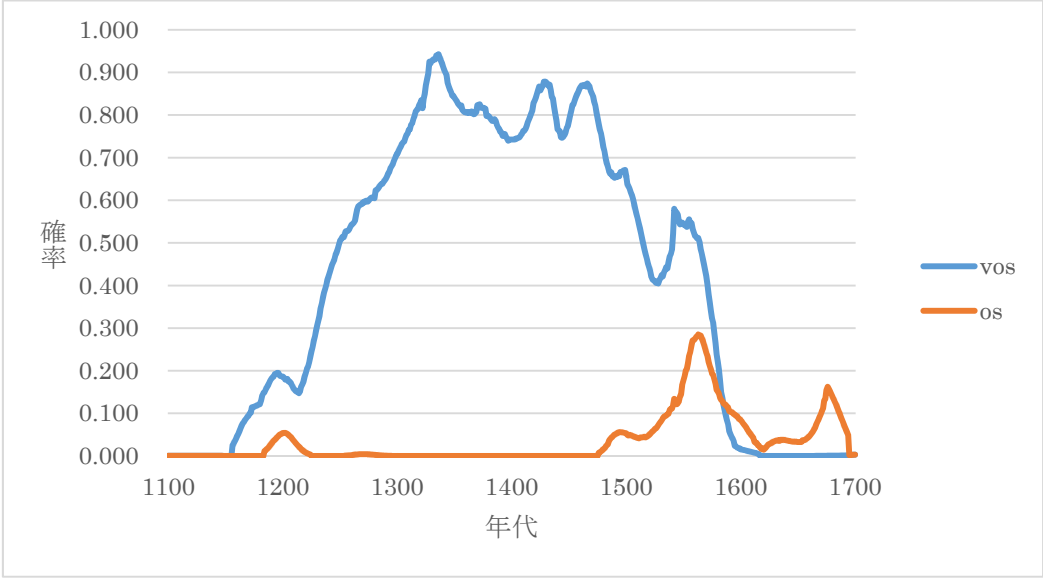


図 5.22　vos～os の CL における年代推移

5.7.5　gelo～selo の年代推移

図 5.23 は, 3 人称間接目的格代名詞と 3 人称直接目的格代名詞の連辞に関する交替変数 f15 の二つの変異形 gelo (f15a) と selo（f15b）の CL における年代推移を表している。16 世紀以降，selo の出現確率が上昇し，同世紀半ばには gelo は用いられなくなるのが分かる。

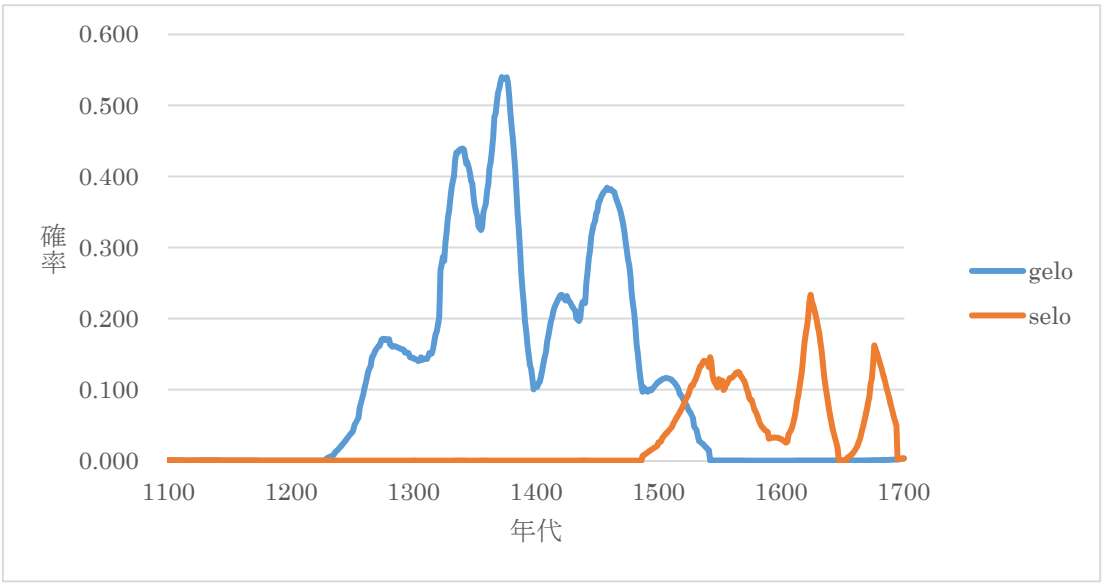


図 5.23　gelo～selo の CL における年代推移

5.7.6　que lo non mandó～que non lo mandó の年代推移

図 5.24 は，従属節における目的格代名詞と否定の副詞・主格代名詞の語順に関する交替変数 f31 の二つの変異形 que lo

non mandó（f31a）と que non lo mandó（f31b）の CL における年代推移を表している。16 世紀半ば以降は，que lo non mandó に代わり que non lo mandó のみが使用されるようになるのが分かる。

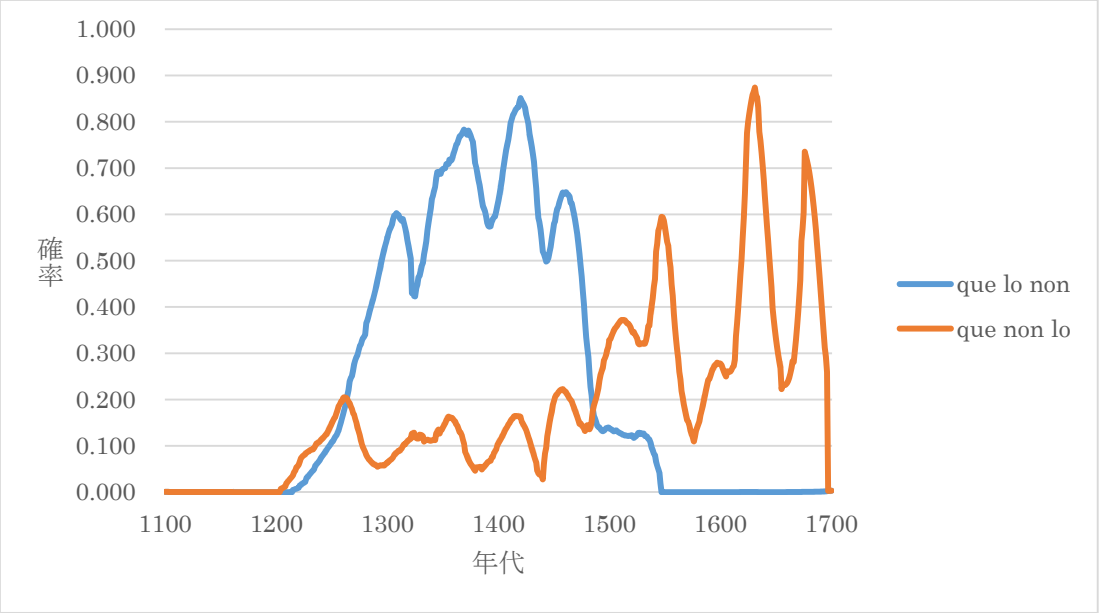


図 5.24　que lo non mandó～que non lo mandó の CL における年代推移

5.7.7　de lo hacer～de hacerlo の年代推移

図 5.25 は，前置詞句における目的格代名詞と不定詞の語順に関する交替変数 f32 の二つの変異形 de lo hacer（f32a）と de hacerlo（f32b）の CL における年代推移を表している。16 世紀以降，de lo hacer の出現確率が低下し，de hacerlo の出現確率が上昇するのが分かる。

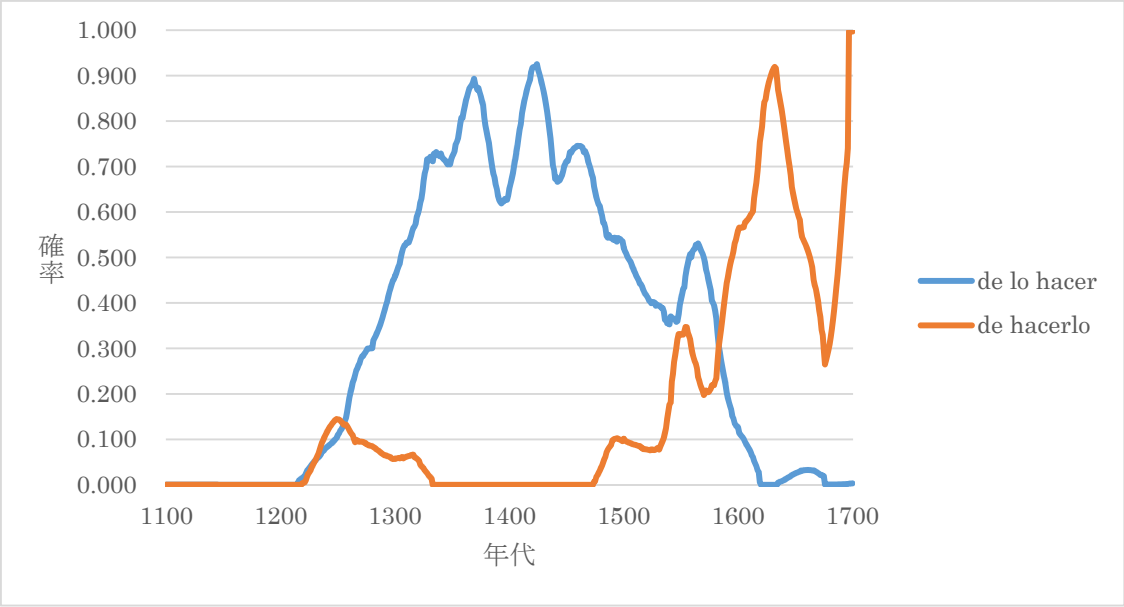


図 5.25　de lo hacer～de hacerlo の CL における年代推移

5.7.8　dó～doy の年代推移

図 5.26 は, 動詞 dar「与える」などの直説法現在 1 人称単数に関する交替変数 f33 の二つの変異形 dó (f33a) と doy (f33b) の CL における年代推移を表している。16 世紀以降, dó に代わり doy のみが使用されるのが分かる。

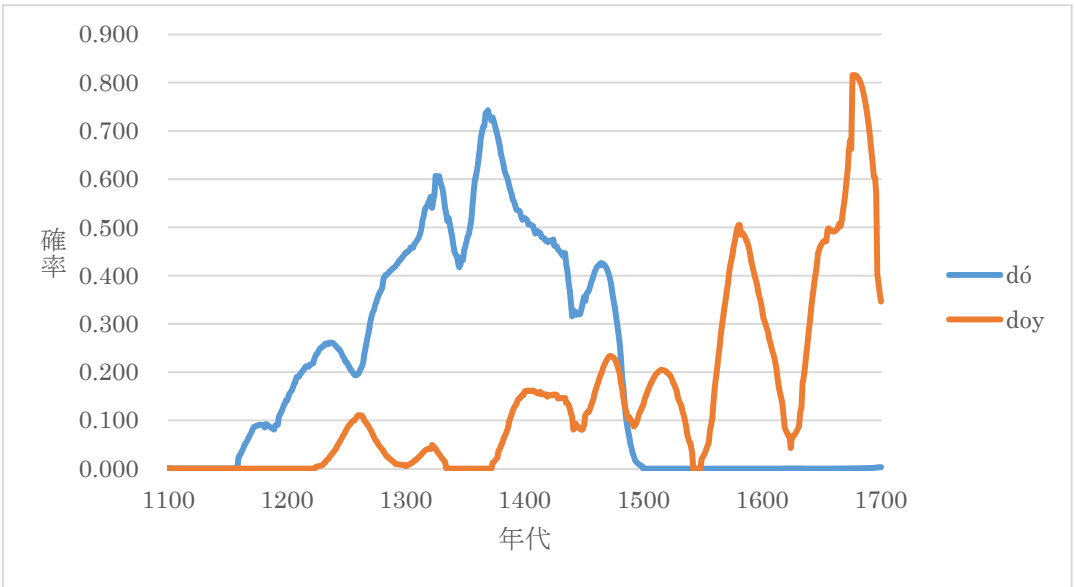


図 5.26　dó～doy の CL における年代推移

5.7.9　avía～avié の年代推移

図 5.27 は, -er 動詞と-ir 動詞の線過去と過去未来の語尾に関する交替変数 f36 の二つの変異形 avía (f36a) と avié (f36b)

の CL における年代推移を表している。14 世紀までは，avié の方がやや優勢であるが，以降は avía が優勢になるのが分かる。

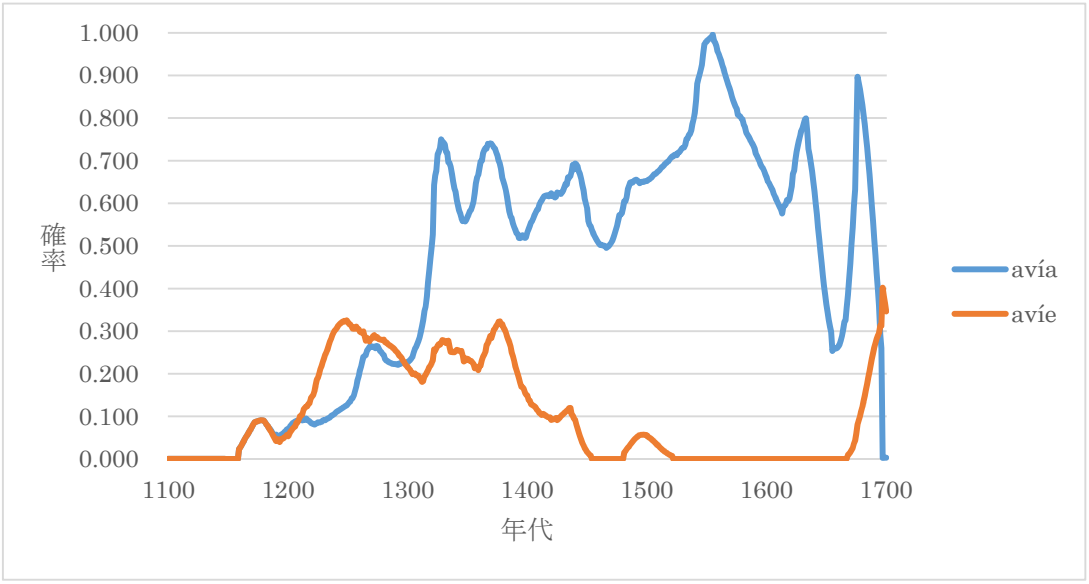


図 5.27　avía～avié の CL における年代推移

5.7.10　terné～tendré の年代推移

図 5.28 は，tener などの直説法未来と過去未来に関する交替変数 f52 の二つの変異形 terné（f52a）と tendré（f52c）の CL における年代推移を表している。16 世紀以降，tendré の出現確率が上がり，17 世紀には terné が用いられなくなるのが分かる。

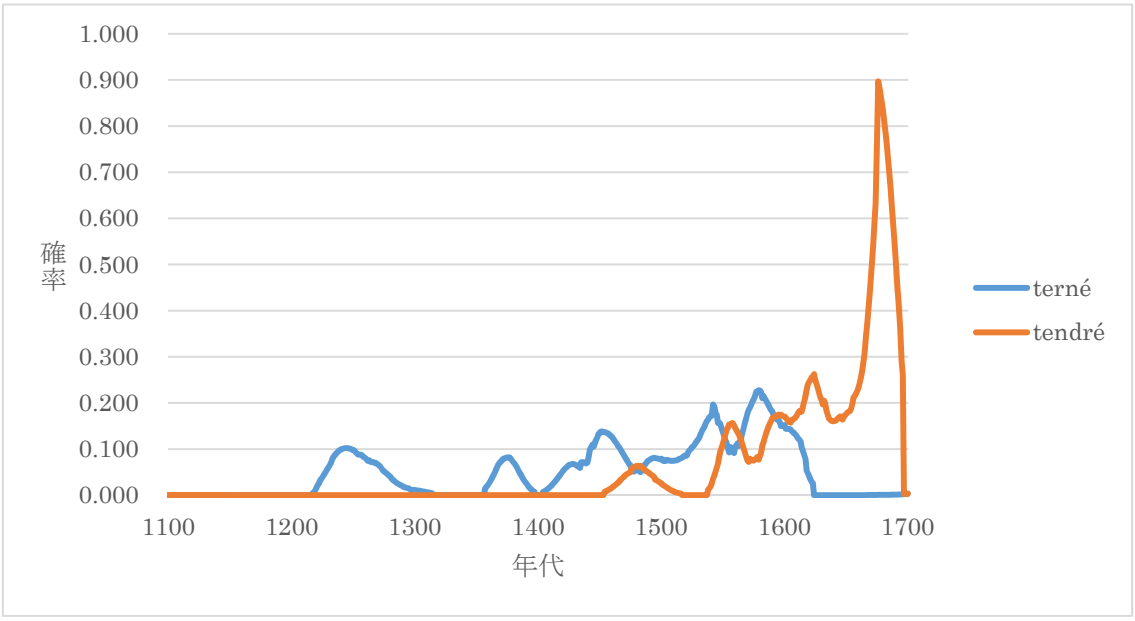


図 5.28　terné～tendré の CL における年代推移

5.7.11　seer～ser の年代推移

図 5.29 は，ser の語幹に関する交替変数 f59 の二つの変異形 seer（f59a）と ser（f59b）の CL における年代推移を表している。14 世紀以降，ser の出現確率が上がり，16 世紀以降は seer が用いられなくなるのが分かる。

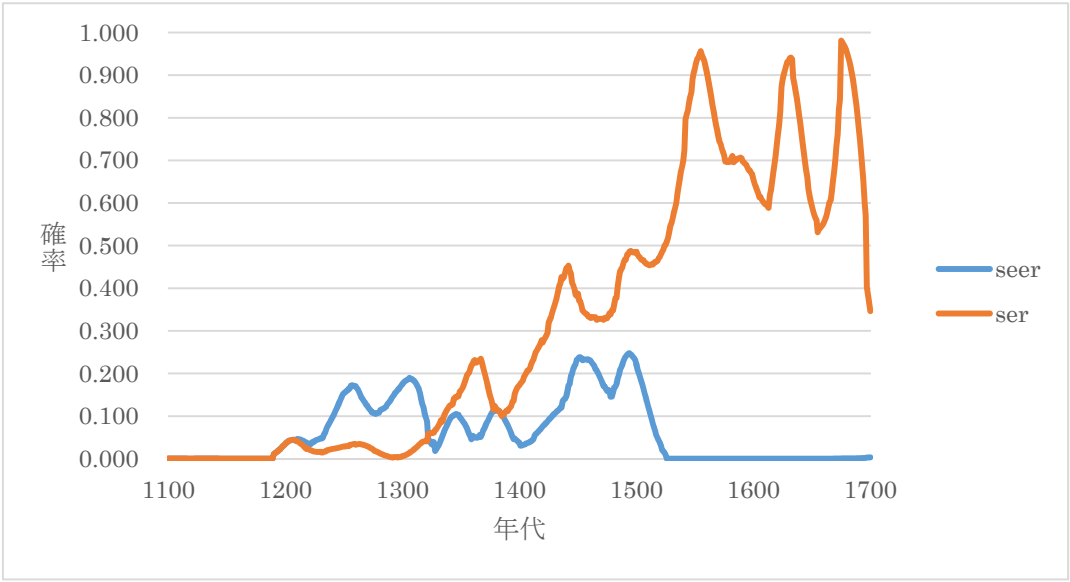


図 5.29　seer～ser の CL における年代推移

5.7.12　-ido～-udo の年代推移

図 5.30 は，-er 動詞の過去分詞に関する交替変数 f67 の二つの変異形-ido（f67a）と-udo（f67b）の CL における年代推移を表している。14 世紀後半以降，-udo の出現確率が下がり，-ido の出現確率が上昇するのが分かる。

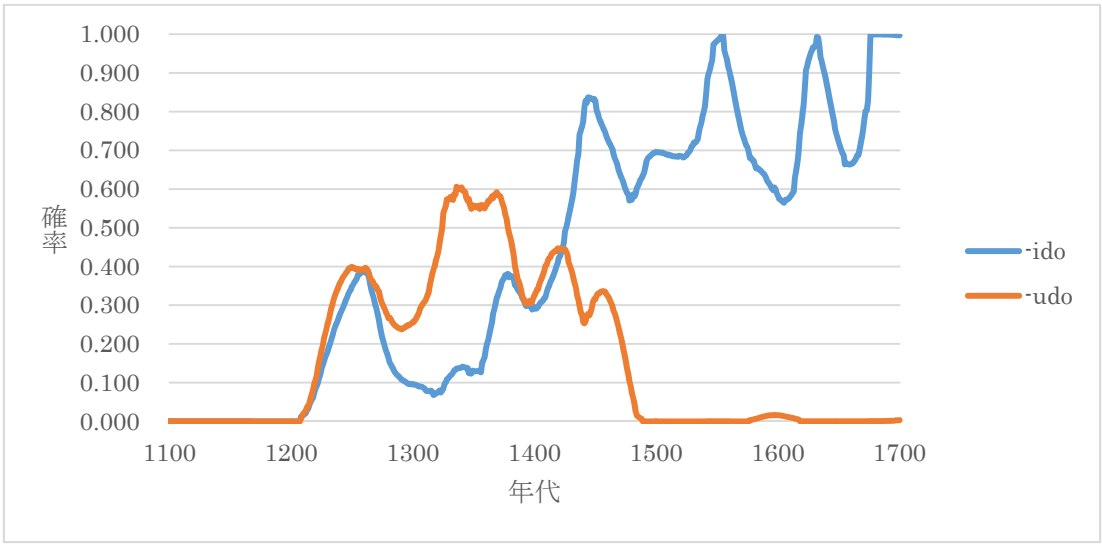


図 5.30　-ido～-udo の CL における年代推移

5.7.13　-mente～-miente～-mientre の年代推移

図 5.31 は，副詞を作る接尾辞に関する交替変数の三つの変異形-mente（f77a）と-miente（f77b）と-mientre（f77c）の CL における年代推移を表している。13 世紀半ばまでは-mientre が優勢で，13 世紀半ばから 15 世紀までは-miente が優勢となり，15 世紀以降は-mente のみが用いられるようになるのが分かる。

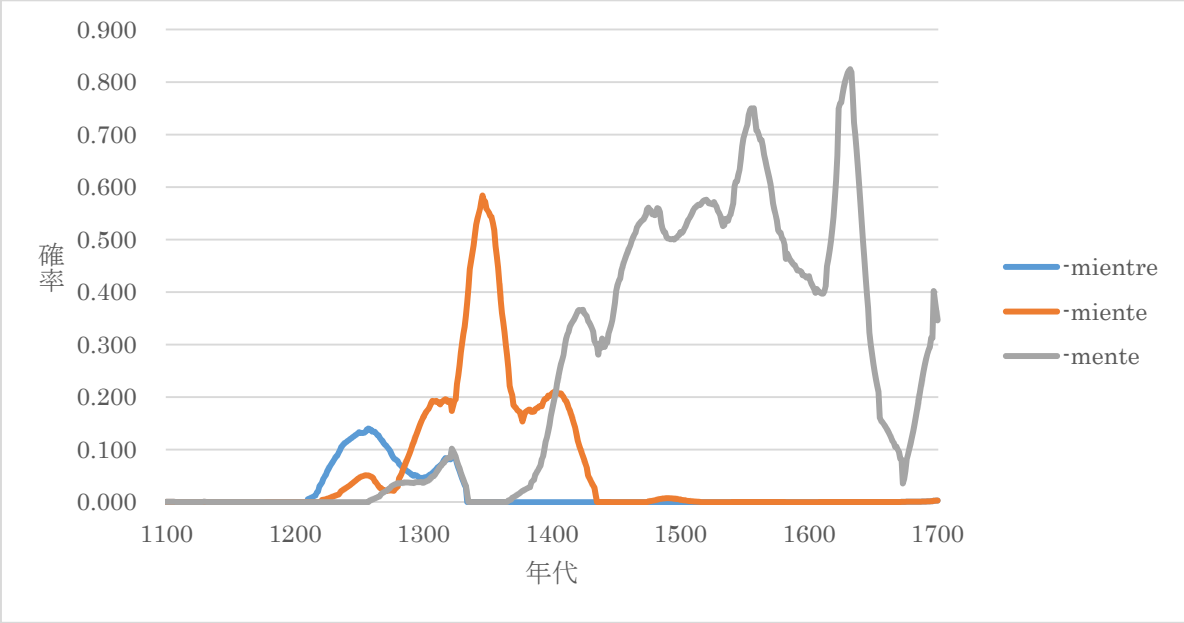


図 5.31　-mente～-miente～-mientre の CL における年代推移

5.7.14　para～pora の年代推移

図 5.32 は，前置詞「～のために」に関する交替変数 f79 の二つの変異形 para（f79a）と pora（f79b）の CL における年代推移を表している。13 世紀までは pora が優勢で，15 世紀半ば以降は para が優勢になるのが分かる。

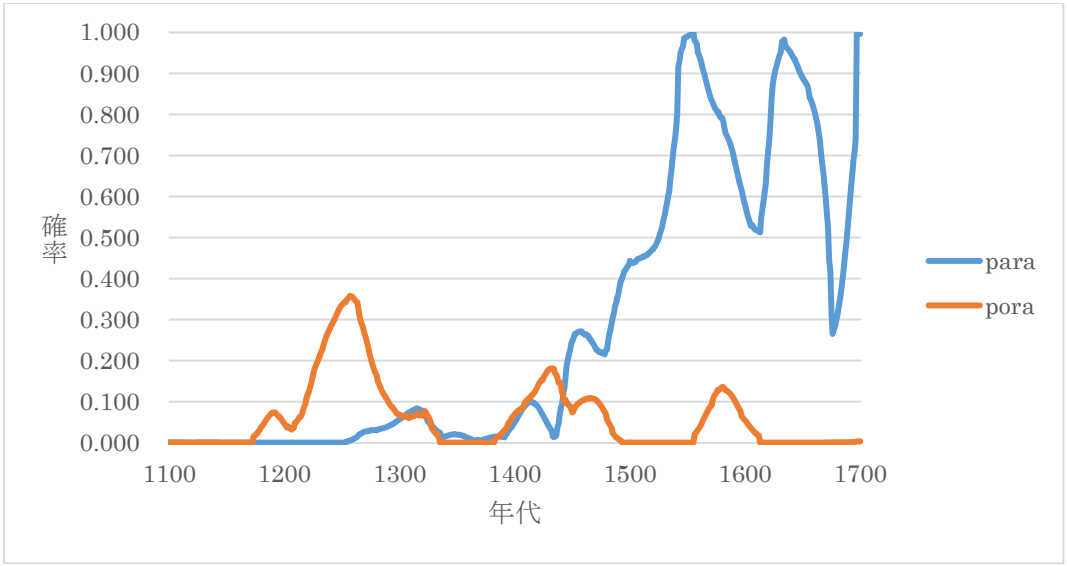


図 5.32　para～pora の CL における年代推移

5.7.15 non～no の年代推移

図 5.33 は，否定の副詞に関する交替変数 f82 の二つの変異形 non（f82a）と no（f82b）の CL における年代推移を表している。15 世紀半ば以降，non の出現確率が下がり，no の出現確率が上昇するのが分かる。

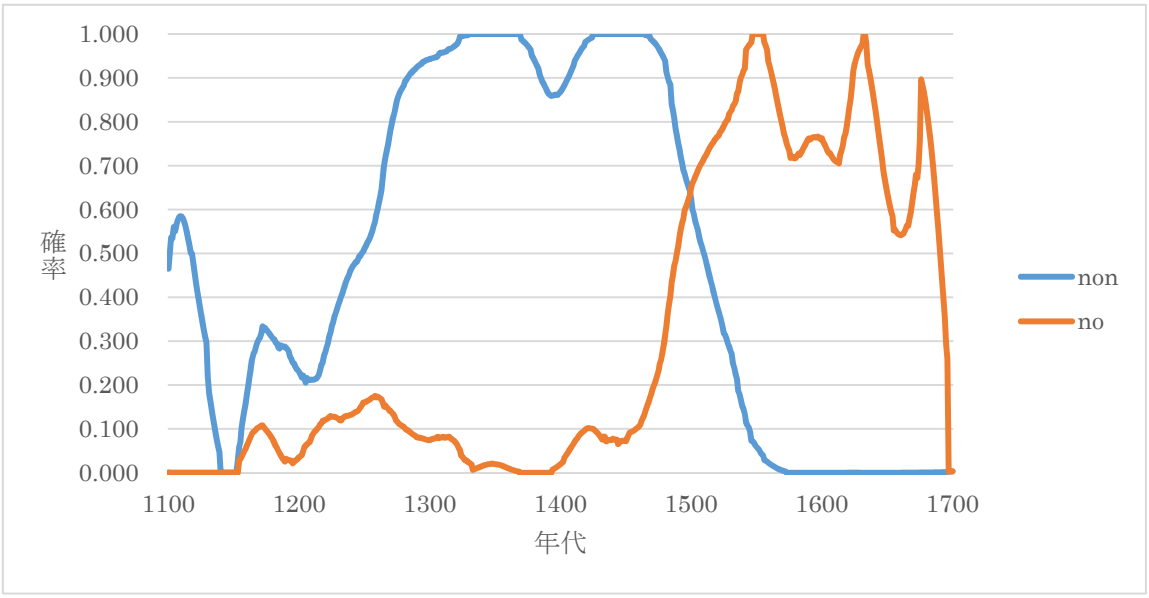


図 5.33　non～no の CL における年代推移

5.7.16 capiella～capilla の年代推移

図 5.34 は，ラテン語の指小辞-ĕllŭ に関する交替変数 f94 の二つの変異形 capiella（f94a）と capilla（f94b）の CL における年代推移を表している。15 世紀以降，capiella の出現確率が下がり，capilla の出現確率が上昇するのが分かる。

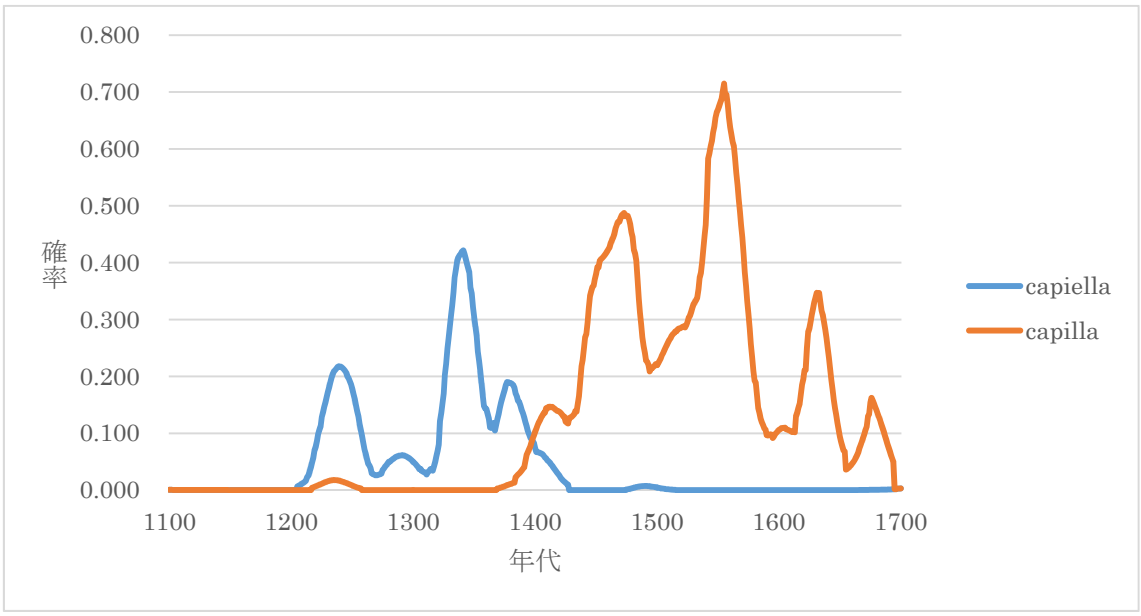


図 5.34　capiella～capilla の CL における年代推移

第6章 *n*-gram 言語モデルによる文書分類

6.1　分類

n-gram 言語モデル（*n*-gram language model）は，文書の生成確率（尤度）をモデル化した確率的言語モデルである（Chen & Goodman 1998；Jurafsky & Martin 2009：Chapter 4；北 1999）。*n*-gram 言語モデルの利点としては，理解しやすい単純なモデルであること，スケーラビリティーがあること，事前確率やその他の情報を組み込むことができること，事後確率を計算することができることが挙げられる（Han *et al.* 2014；Hulden *et al.* 2015）[24]。確かにサポートベクターマシンや対数線形モデルなどの識別モデルは汎化能力が高いが，本研究のようにクラス数が多い場合（最大で約 3 万クラス），実行時間が長大になってしまう（Manning *et al.* 2008：15.3）。また，本研究のように，クラス数に対して訓練データが少ない場合，生成モデルは識別モデルに勝るという報告もある（Ng & Jordan 2002）。

問題定義で述べたように，文書分類を行う文書qは，関数$\phi(q,c)$の値が最大となるクラス$c \in C$に分類される：

$$\hat{c_q} = \arg\max_{c \in C} \phi(q,c) \tag{6.1}$$

ここで，$\hat{c_q}$は文書qが属すると推定されたクラス，Cはクラスの集合である。*n*-gram 言語モデルにおいて，$\phi(q,c)$は，文書qがクラスcに属する事後確率（posterior probability）$P(c|q)$に対応する。事後確率$P(c|q)$とは，文書qが観測された時に，文書qがクラスcに属する確率である。したがって，文書qは事後確率$P(c|q)$が最大となるクラスcに分類される：

$$\hat{c_q} = \arg\max_{c \in C} P(c|q) \tag{6.2}$$

事後確率$P(c|q)$は，ベイズの定理（付録 A の A4）により，以下のように変形される：

$$P(c|q) = \frac{P(c)P(q|c)}{P(q)} = \frac{P(c)P(q|c)}{\sum_{c \in C} P(c)P(q|c)} \propto P(c)P(q|c) \tag{6.3}$$

ここで，$P(c)$はクラスcの事前確率（prior probability），$P(q|c)$はクラスcのもとで文書qが生成される確率（尤度 likelihood），$P(q)$は文書qの周辺確率（marginal probability）である。事後確率の値そのものではなく，事後確率が最大となるクラスcを知るだけならば$P(c)P(q|c)$を求めるだけでよい：

$$\hat{c_q} = \arg\max_{c \in C} P(c)P(q|c) \tag{6.4}$$

なぜならば，分母の$P(q) = \sum_{c \in C} P(c)P(q|c)$は全クラスで共通だからである。また，事前確率$P(c)$がすべてのクラスにおいて同一と考えるならば，尤度$P(q|c)$のみを求めればよい。

---

[24] 一方，*n*-gram 言語モデルの欠点として，*n* が大きくなると素性数が爆発的に増えることが挙げられる。語彙サイズを$|V|$とすると，総グラム数は$|V|^n$個になる。たとえば$|V| = 100$のとき，ユニグラムの総数は 100（$= 100^1$），バイグラムの総数は 10,000（$= 100^2$），トライグラムの総数は 1,000,000（$= 100^3$）となる。*n* が大きくなると，語順などより多くの情報捉えることができる一方，計算量やスパースネスが増加する。このため，タスクに応じて適切な *n* の値を選択する必要がある。

6.2　最尤推定

文字 *n*-gram 言語モデルによる尤度$P(q|c)$の求め方を説明する[25]。文書qが，長さlの文字列$w_1^l = w_1 w_2 \dots w_i \dots w_l$で表されているとする（$q = w_1^l = w_1 w_2 \dots w_i \dots w_l$）。ここで，$w_i$は$i$番目の文字を，$w_i^j$は$i$番目から$j$番目までの文字列を表している。たとえば，I am John という一文からなる文書は，$w_1 = \mathrm{I}, w_2 = _, w_3 = \mathrm{a}, w_4 = \mathrm{m}, w_5 = _, w_6 = \mathrm{J}, w_7 = \mathrm{o}, w_8 = \mathrm{h}, w_9 = \mathrm{n}$となり，文字数に基づく文書長$l = 9$である（スペースも一つの文字としてカウントしている）。このとき，クラスcにおける文書qの生成確率$P(q|c)$は，条件付き確率（conditional probability）を用いて，

$$\begin{aligned} P(q|c) &= P_c(w_0)P_c(w_1|w_0)P_c(w_2|w_0w_1)P_c(w_3|w_0w_1w_2)\cdots P_c(w_{l+1}|w_0w_1\cdots w_l) \\ &= \prod_{i=1}^{l+1} P_c\left(w_i|w_0^{i-1}\right) \end{aligned} \tag{6.5}$$

と表すことができる[26]。ここで，w_0は文頭記号，w_{l+1}は文末記号で，$P(w_0) = 1$である[27]。$P_c\left(w_i|w_0^{i-1}\right)$は，クラス$c$における$i$番目の文字$w_i$の条件付き確率である。$i$番目の文字$w_i$の出現は，それ以前の文字列$w_0^{i-1}(= w_0 \cdots w_{i-1})$に依存している。一般に，限られた量の訓練データから$P_c\left(w_i|w_0^{i-1}\right)$を正確に推定するのは困難である。そこで，*n*-gram 言語モデルでは，i番目の文字w_iの出現が直前の *n*-1 個の文字列w_{i-n+1}^{i-1}のみに依存する（*n*-1 重マルコフ過程とよばれる）と仮定して，文書qの生成確率をモデル化している：

$$\begin{aligned} P(q|c) &= \prod_{i=1}^{l+1} P_c\left(w_i|w_0^{i-1}\right) \\ &\approx \prod_{i=1}^{l+1} P_c\left(w_i|w_{i-n+1}^{i-1}\right) = \begin{cases} \prod_{i=1}^{l} P_c(w_i) & (n=1) \\ \prod_{i=1}^{l+1} P_c(w_i|w_{i-1}) & (n=2) \\ \prod_{i=1}^{l+1} P_c\left(w_i|w_{i-2}^{i-1}\right) & (n=3) \\ \vdots \end{cases} \end{aligned} \tag{6.6}$$

直前の *n*-1 個の文字列w_{i-n+1}^{i-1}は，履歴（hisotry）と呼ばれる。$n=1$ の時は，文字w_iの条件付き確率$P_c(w_i)$は直前の文字列に依存しない。これは，ナイーブベイズ多項モデルと等価である。$n = 2$ の時は，文字w_iの条件付き確率$P_c(w_i|w_{i-1})$は直前の文字w_{i-1}にのみ依存する。$n = 3$ の時は，文字w_iの条件付き確率$P_c\left(w_i|w_{i-2}^{i-1}\right)$は直前の二つの文字$w_{i-2}^{i-1} = w_{i-2}w_{i-1}$にのみ依存する。$n$ が大きいほど，部分的ではあるが，文字順や語順の情報を取り込むことが出来る。

式（6.6）において，クラスcにおける文字w_iの真の条件付き確率$P_c\left(w_i|w_{i-n+1}^{i-1}\right)$は分からないので，その推定値$\widehat{P_c}\left(w_i|w_{i-n+1}^{i-1}\right)$を訓練データから求める必要がある。最も単純な推定値は，

---

[25] 以下，「文字」を「単語」に置き換えれば，単語 *n*-gram 言語モデルの説明となる。

[26] 文書長lは，クラスcに依存しないとする。

[27] 文頭記号w_0は，$P(w_1|w_0)$を定義するために必要となる。文末記号w_{l+1}は，あらゆる文sに関して$\sum_s P(s) = 1$とするために必要となる。

$$\hat{P}_c\left(w_i \middle| w_{i-n+1}^{i-1}\right) = P_{c:ML}\left(w_i \middle| w_{i-n+1}^{i-1}\right) = \frac{n\left(w_{i-n+1}^{i-1} w_i, c\right)}{\sum_{w \in V} n\left(w_{i-n+1}^{i-1} w, c\right)}$$

$$= \begin{cases} \dfrac{n(w_i, c)}{\sum_{w \in V} n(w, c)} & (n = 1) \\ \dfrac{n(w_{i-1} w_i, c)}{\sum_{w \in V} n(w_{i-1} w, c)} & (n = 2) \\ \dfrac{n\left(w_{i-2}^{i-1} w_i, c\right)}{\sum_{w \in V} n\left(w_{i-2}^{i-1} w, c\right)} & (n = 3) \\ \vdots \end{cases} \tag{6.7}$$

で与えられる最尤推定値$P_{c:ML}\left(w_i \middle| w_{i-n+1}^{i-1}\right)$である。ここで，$c: ML$は，クラス$c$における最尤推定値（maximal likelihood (ML) estimate）の意味である。最尤推定値は，相対頻度と一致する。$n(w, c)$はクラスcにおける文字列wの頻度，Vはデータセット全体の文字集合である。式（6.7）より，最尤推定では訓練データに出現しない（つまり$n\left(w_{i-n+1}^{i-1} w_i, c\right) = 0$となる）文字$w_i$の条件付き確率はゼロとなってしまう。この状況は，ゼロ頻度問題（zero frequency problem）と呼ばれる。ゼロ頻度問題には，二つの問題点がある。一つ目は，最尤推定値がゼロだからといって，その文字w_iが現実に存在しないというわけではないという点である。訓練データの量は常に限られているので，偶然，その訓練データには文字w_iが出現しなかった可能性がある。二つ目は，少なくとも一つの文字w_iの条件付き確率$P_c\left(w_i \middle| w_{i-n+1}^{i-1}\right)$がゼロと推定された場合，式（6.6）より，クラス$c$における文書$q$の生成確率$P(q|c)$がゼロになってしまう点である。これは，計算上，望ましくない性質である。

6.3　スムージング

このゼロ頻度問題を回避するために，一般的にスムージング（smoothing）が行われる。スムージングとは，頻度がゼロもしくは低頻度の文字の出現頻度に下駄をはかせることで，訓練データに出現しない文字にもゼロより大きい出現確率を付与する方法である。結果的に，最尤推定値に比べ，高頻度の文字の出現確率は小さく，低頻度の文字の出現確率は大きくなり，確率分布の凸凹が滑らかになる。また一般的に，最尤推定よりもスムージングを用いた場合の方が，分類性能が高いとされる。

本研究では，最も単純なスムージングである加算スムージング（additive smoothing）を採用する[28]。加算スムージングによる，クラスcにおける文字w_iの条件付き確率の推定値$P_{c:AD}\left(w_i \middle| w_{i-n+1}^{i-1}\right)$は，次式で与えられる：

[28] 一般的に，他のスムージング法に比べ，加算スムージングの性能は低いことが知られている（Chen & Goodman 1998）。しかし，予備実験において，実験精度が最高といわれる Kneser-Ney 法よりも，加算スムージングの方が性能が高かったので，加算スムージングを採用した。本研究では，カーネル平滑化により頻度が非整数値となる。Kneser-Ney 法は整数値しか扱うことができないので，値を丸めて整数にする必要がある。一方，加算スムージングでは，整数化の必要はない。性能の違いは，この整数化の有無によるものだと考えられる。Kneser-Ney 法における非整数頻度の扱いについては，Zhang & Chiang（2014）を参照。

$$\hat{P}_c\left(w_i \middle| w_{i-n+1}^{i-1}\right) = P_{c:AD}\left(w_i \middle| w_{i-n+1}^{i-1}\right) = \frac{n\left(w_{i-n+1}^{i-1} w_i, c\right) + \alpha}{\sum_{w \in V} n\left(w_{i-n+1}^{i-1} w, c\right) + \alpha|V|}$$

$$= \begin{cases} \dfrac{n(w_i, c) + \alpha}{\sum_{w \in V} n(w, c) + \alpha|V|} & (n = 1) \\ \dfrac{n(w_{i-1} w_i, c) + \alpha}{\sum_{w \in V} n(w_{i-1} w, c) + \alpha|V|} & (n = 2) \\ \dfrac{n\left(w_{i-2}^{i-1} w_i, c\right) + \alpha}{\sum_{w \in V} n\left(w_{i-2}^{i-1} w, c\right) + \alpha|V|} & (n = 3) \\ \vdots \end{cases} \tag{6.8}$$

ここで，$|V|$はデータセット全体の文字の種類数である。加算スムージングでは，出現頻度$n\left(w_{i-n+1}^{i-1} w_i, c\right)$に定数$\alpha$を加え，頻度に下駄を履かせている。一般的に，$\alpha \in (0, 1]$が用いられる。これにより，すべての文字$w_i$の条件付き確率がゼロより大きくなる。$\alpha = 0$の時，$P_{c:AD}\left(w_i \middle| w_{i-n+1}^{i-1}\right)$は最尤推定値$P_{ML}\left(w_i \middle| w_{i-n+1}^{i-1}\right)$と一致する。上記より，クラス$c$のもとで文書$q$が生成される確率$P(q|c)$の推定値$\hat{P}(q|c)$は，次式のように計算される：

$$P(q|c) \approx \hat{P}(q|c) = \prod_{i=1}^{l+1} \hat{P}_c\left(w_i \middle| w_0^{i-1}\right)$$

$$= \prod_{i=1}^{l+1} P_{c:AD}\left(w_i \middle| w_{i-n+1}^{i-1}\right)$$

$$= \prod_{i=1}^{l+1} \frac{n\left(w_{i-n+1}^{i-1} w_i, c\right) + \alpha}{\sum_{w \in V} n\left(w_{i-n+1}^{i-1} w, c\right) + \alpha|V|}$$

$$= \begin{cases} \dfrac{n(w_i, c) + \alpha}{\sum_{w \in V} n(w, c) + \alpha|V|} & (n = 1) \\ \dfrac{n(w_{i-1} w_i, c) + \alpha}{\sum_{w \in V} n(w_{i-1} w, c) + \alpha|V|} & (n = 2) \\ \dfrac{n\left(w_{i-2}^{i-1} w_i, c\right) + \alpha}{\sum_{w \in V} n\left(w_{i-2}^{i-1} w, c\right) + \alpha|V|} & (n = 3) \\ \vdots \end{cases} \tag{6.9}$$

次に，クラスcの事前確率$P(c)$の推定値$\hat{P}(c)$の求め方を説明する。クラスcに属する文書数をN_c，クラスの集合をCとすると，最尤推定値$P_{ML}(c)$はクラスcに属する文書の相対頻度として与えられる：

$$\hat{P}(c) = P_{ML}(c) = \frac{N_c}{\sum_{c \in C} N_c} \tag{6.10}$$

事前確率$P(c)$に関する頻度ゼロ問題を回避するために，スムージングを行うことが考えられる。加算スムージングによる事前確率$P(c)$の推定値$P_{AD}(c)$は，次式で与えられる：

$$\hat{P}(c) = P_{AD}(c) = \frac{N_c + \beta}{\sum_{c \in C} N_c + \beta|C|} \tag{6.11}$$

ここで，$|C|$はデータセット全体におけるクラス数である。加算スムージングでは，すべてのクラスcに対して定数βを加

え，頻度に下駄を履かせている。一般的に，$\beta \in (0,1]$が用いられる。$\beta = 0$の時，$P_{AD}(c)$は最尤推定値$P_{ML}(c)$と一致する。

上記より，式（6.2）は，式（6.4），式（6.9），式（6.11）を用いて，以下のように変形できる：

$$\begin{aligned}\widehat{c_q} &= \arg\max_{c\in C} P(c|q) \\ &= \arg\max_{c\in C} P(c)P(q|c) \\ &= \arg\max_{c\in C} \hat{P}(c)\hat{P}(q|c) \\ &= \arg\max_{c\in C} P_{AD}(c)\prod_{i=1}^{l+1} P_{c:AD}\left(w_i|w_{i-n+1}^{i-1}\right)\end{aligned} \tag{6.12}$$

さて，1 以下の値である確率を多数掛け合わせると，通常の計算機では扱えないほど小さい値になってしまうことがある。アンダーフロー（underflow）と呼ばれるこの問題を回避するために，確率を対数に変換して計算を行うことが一般的である[29]。対数関数は単調増加するので，正の実数を対数に変換しても大小関係は変化しない（$0 < \mathrm{a} < \mathrm{b} \Leftrightarrow \log a < \log b$）。よって，式（6.12）は$\log ab = \log a + \log b$を用いて，

$$\begin{aligned}\widehat{c_q} &= \arg\max_{c\in C} P_{AD}(c)\prod_{i=1}^{l+1} P_{c:AD}\left(w_i|w_{i-n+1}^{i-1}\right) \\ &= \arg\max_{c\in C} \log P_{AD}(c)\prod_{i=1}^{l+1} P_{c:AD}\left(w_i|w_{i-n+1}^{i-1}\right) \\ &= \arg\max_{c\in C} \log P_{AD}(c) + \log\prod_{i=1}^{l+1} P_{c:AD}\left(w_i|w_{i-n+1}^{i-1}\right) \\ &= \arg\max_{c\in C} \log \hat{P}(c) + \sum_{i=1}^{l+1} \log P_{c:AD}\left(w_i|w_{i-n+1}^{i-1}\right)\end{aligned} \tag{6.13}$$

と変形できる。対数の底には自然数$e = 2.71828\ldots$を用いる。

本研究では，事前確率$P(c)$はすべてのクラスcにおいて同一とする。したがって，文書qの属するクラス$\widehat{c_q}$は，対数尤度$\log P(q|c)$が最大になるクラスcとなる：

$$\widehat{c_q} = \arg\max_{c\in C} \sum_{i=1}^{l+1} \log P_{c:AD}\left(w_i|w_{i-n+1}^{i-1}\right) \tag{6.14}$$

本研究では，対数の底に自然数eを用いるので，尤度$P(q|c)$は$\exp(\log P(q|c))$で求められる。

6.4　例

たとえば，文字 2-gram 言語モデルを考えてみる。訓練データとして，クラスc_1には$\{abbca, bacab, ccaab\}$という 3 文書が，クラスc_2には$\{abca, bacab\}$という 2 文書が与えられているとする。各クラスの 2-gram の頻度は表 6.1 のように

[29] たとえば，明らかに$10^{-1000} \gg 10^{-2000}$であるにもかかわらず，通常の計算機では$10^{-1000}$も$10^{-2000}$もどちらもゼロだとみなされてしまう。底を 10 として対数を取れば，$\log_{10} 10^{-1000} = -1000 \gg \log_{10} 10^{-2000} = -2000$となり，通常の計算機で扱える値になる。

なる。ここで，#は文頭記号，$は文末記号である。文字集合$V = \{a, b, c, \$\}$より，文字サイズ$|V| = 4$となる。文頭記号は文字集合には含まれないが，文末記号は文字集合に含まれる。$P_{c_1}(\#) = P_{c_2}(\#) = 1$とする。

	#a	#b	#c	#$	aa	ab	ac	a$	ba	bb	bc	b$	ca	cb	cc	c$
c_1	1	1	1	0	1	3	1	1	1	1	1	2	3	0	1	0
c_2	1	1	0	0	0	2	1	1	1	0	1	1	2	0	0	0

表 6.1　各クラスの 2-gram の頻度

このとき，文書$q = bcca$の属するクラス$\hat{c_q}$は，

$$\begin{aligned}\hat{c_q} &= \arg\max_{c \in C} P(c)P(q|c) \\ &= \arg\max_{c \in C} P(q|c)\end{aligned}$$

と予測される。ただし，各クラスの事前確率は同一とする。

クラスc_1における文書$q = bcca$の最尤推定による尤度$P(q|c_1)$は，以下のように計算される。

$$\begin{aligned}P(q|c_1) &= P_{c_1}(\#)P_{c_1}(b|\#)P_{c_1}(c|b)P_{c_1}(c|c)P_{c_1}(a|c)P_{c_1}(\$|a) \\ &= 1 * \frac{1}{3} * \frac{1}{5} * \frac{1}{4} * \frac{3}{4} * \frac{1}{6} = 0.00283\end{aligned}$$

たとえば，$P_{c_1:ML}(b|\#)$は，

$$\begin{aligned}P_{c_1:ML}(b|\#) &= \frac{n(\#b, c_1)}{\sum_{w \in V} n(\#w, c_1)} \\ &= \frac{n(\#b, c_1)}{n(\#a, c_1) + n(\#b, c_1) + n(\#c, c_1) + n(\#\$, c_1)} \\ &= \frac{1}{1 + 1 + 1 + 0} = \frac{1}{3}\end{aligned}$$

と計算される。$P(b|\#)$は，文書が b で始まる確率を表している。

同様に，クラスc_2における文書$q = bcca$の最尤推定による尤度$P(q|c_2)$は，

$$\begin{aligned}P(q|c_2) &= P_{c_2}(\#)P_{c_2}(b|\#)P_{c_2}(c|b)P_{c_2}(c|c)P_{c_2}(a|c)P_{c_2}(\$|a) \\ &= 1 * \frac{1}{2} * \frac{1}{3} * \frac{0}{2} * \frac{2}{2} * \frac{1}{4} = 0\end{aligned}$$

と計算される。しかし，$n(cc, c_2) = 0$のゼロ頻度問題により$P(q|c_2) = 0$となってしまう。

最尤推定の場合，$P(q|c_1) > P(q|c_2)$となるので，文書qはクラスc_1に属すると判断される。各クラスの事後確率は，

$$P(c_1|q) = \frac{P(q|c_1)}{P(q|c_1) + P(q|c_2)} = \frac{0.00283}{0.00283 + 0} = 1$$

$$P(c_2|q) = \frac{P(q|c_2)}{P(q|c_1) + P(q|c_2)} = \frac{0}{0.00283 + 0} = 0$$

と求められる。各クラスの事前確率は同一なので，省略できる。

一方，$\alpha = 1$として加算スムージングを行った場合の，各クラスにおける文書$q = bcca$の尤度は,

$$\begin{aligned} P(q|c_1) &= P_{c_1}(\#)P_{c_1}(b|\#)P_{c_1}(c|b)P_{c_1}(c|c)P_{c_1}(a|c)P_{c_1}(\$|a) \\ &= 1 * \frac{1+1}{3+4} * \frac{1+1}{5+4} * \frac{1+1}{4+4} * \frac{3+1}{4+4} * \frac{1+1}{6+4} = 0.001587 \end{aligned}$$

$$\begin{aligned} P(q|c_2) &= P_{c_2}(\#)P_{c_2}(b|\#)P_{c_2}(c|b)P_{c_2}(c|c)P_{c_2}(a|c)P_{c_2}(\$|a) \\ &= 1 * \frac{1+1}{2+4} * \frac{1+1}{3+4} * \frac{0+1}{2+4} * \frac{2+1}{2+4} * \frac{1+1}{4+4} = 0.001984 \end{aligned}$$

となる。このとき，最尤推定の場合とは異なり，$P(q|c_1) < P(q|c_2)$となるので，文書qはクラスc_2に属すると予測される。各クラスの事後確率は,

$$P(c_1|q) = \frac{P(q|c_1)}{P(q|c_1) + P(q|c_2)} = \frac{0.001587}{0.001587 + 0.001984} \approx 0.44$$

$$P(c_2|q) = \frac{P(q|c_2)}{P(q|c_1) + P(q|c_2)} = \frac{0.001984}{0.001587 + 0.001984} \approx 0.56$$

と求められる。最尤推定の場合と比べ，事後確率の差は小さくなっている。

第7章 JS 情報量による文書分類

7.1　分類

問題定義で述べたように，文書分類を行う文書qは，関数$\phi(q,c)$の値が最大となるクラス$c \in C$に分類される：

$$\widehat{c_q} = \arg\max_{c \in C} \phi(q,c) \tag{7.1}$$

ここで，$\widehat{c_q}$は文書qが属すると推定されたクラス，Cはクラスの集合である。類似度に基づいた分類において，$\phi(q,c)$は，文書qとクラスcとの類似度$\mathrm{Sim}(q,c)$に対応する。したがって，文書qは類似度$\mathrm{Sim}(q,c)$が最大となるクラスcに分類される：

$$\widehat{c_q} = \arg\max_{c \in C} \mathrm{Sim}(q,c) \tag{7.2}$$

本研究では，類似度として JS 情報量（Jensen-Shanon divergence）を用いる[30]。JS 情報量は，同じ事象空間上の二つの確率分布間の差異を情報理論に基づき測る尺度である（高村 2010 1.6）。素性の集合を$F = \{f_1, \ldots, f_{|F|}\}$，文書$q$の確率分布を$P_q$，クラス$c$の確率分布を$P_c$とする（$\sum_{f \in F} P_q(f) = \sum_{f \in F} P_c(f) = 1$）。このとき，二つの確率分布$P_q$と$P_c$の JS 情報量$D_{JS}(P_q||P_c)$は，次式で与えられる。対数の底には，自然数$e = 2.71828\ldots$を用いる：

$$\begin{aligned} D_{JS}(P_q||P_c) &= \frac{1}{2}\left(D_{KL}(P_q||R) + D_{KL}(P_c||R)\right) \\ &= \frac{1}{2}\left(\sum_{f \in F} P_q(f) \log \frac{P_q(f)}{R(f)} + \sum_{f \in F} P_c(f) \log \frac{P_c(f)}{R(f)}\right) \\ &= \frac{1}{2}\left(\sum_{f \in F} P_q(f) \log \frac{P_q(f)}{\frac{P_q(f)+P_c(f)}{2}} + \sum_{f \in F} P_c(f) \log \frac{P_c(f)}{\frac{P_q(f)+P_c(f)}{2}}\right) \end{aligned} \tag{7.3}$$

ここで，Rは二つの確率分布P_qとP_cの平均として定義される確率分布$R(f) = \frac{P_q(f)+P_c(f)}{2}$である。$D_{KL}(P_q||R)$は$P_q$と$R$の KL 情報量（Kullback-Leibler divergence），$D_{KL}(P_c||R)$はP_cとRの KL 情報量である（KL 情報量については，付録 A の A10 を参照）。したがって，JS 情報量は，二つの確率分布の平均である確率分布Rまでの KL 情報量の平均とみなすことができる。確率分布の値は最尤推定で求めた：

$$P_q(f) = \frac{n(f,q)}{\sum_{f \in F} n(f,q)} \tag{7.4}$$

[30] 予備実験では，コサイン類似度による実験も行った。しかし，推定精度が低かったので，説明は省略する。

$$P_c(f) = \frac{n(f,c)}{\sum_{f \in F} n(f,c)}$$

ここで, $n(f,q)$は文書qにおける素性$f \in F$の頻度, $n(f,c)$はクラスcにおける素性$f \in F$の頻度である。$P_q(f) = P_c(f) = 0$の場合，$P_q(f) \log \frac{P_q(f)}{\frac{P_q(f)+P_c(f)}{2}} = P_c(f) \log \frac{P_c(f)}{\frac{P_q(f)+P_c(f)}{2}} = 0 \log \frac{0}{0} = 0$となる。

JS 情報量には，非負性，同一性，対称性の性質がある：

$$D_{JS}(P_q||P_c) \geq 0 \tag{7.5}$$

$$D_{JS}(P_q||P_c) = 0 \Leftrightarrow P_q = P_c \tag{7.6}$$

$$D_{JS}(P_q||P_c) = D_{JS}(P_c||P_q) \tag{7.7}$$

JS 情報量が小さい（大きい）ほど二つの確率分布の類似度は高い（低い）ので，類似度としては JS 情報量の逆数を用いることにする：

$$\mathrm{Sim}(P_q, P_c) = D_{JS}(P_q||P_c)^{-1} \tag{7.8}$$

したがって，文書qの属するクラス$\hat{c_q}$は,

$$\begin{aligned} \hat{c_q} &= \arg\max_{c \in C} D_{JS}(P_q||P_c)^{-1} \\ &= \arg\min_{c \in C} D_{JS}(P_q||P_c) \end{aligned} \tag{7.9}$$

と予測される。$\arg\min_{c \in C} D_{JS}(P_q||P_c)$は，$D_{JS}(P_q||P_c)$が最小（min）になるクラス$c$に文書$q$を分類することを意味している。arg min は argument of the minimum の意味である。

7.2　例

たとえば，訓練データとして，クラスc_1には$\{abbca,\ bacab,\ ccaab\}$という 3 文書が，クラス$c_2$には$\{abca,\ bacab\}$という 2 文書が与えられているとする。ここで，文書$q = bcca$の属するクラス$\hat{c_q}$を JS 情報量に基づいて予測する。素性には 2-gram を用いるとする。表 7.1 に文書q，クラスc_1，クラスc_2の 2-gram の頻度を示す。ここで，#は文頭記号，$は文末記号である。

	#a	#b	#c	#$	aa	ab	ac	a$	ba	bb	bc	b$	ca	cb	cc	c$
q	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0
c_1	1	1	1	0	1	3	1	1	1	1	1	2	3	0	1	0
c_2	1	1	0	0	0	2	1	1	1	0	1	1	2	0	0	0

表 7.1　2-gram の頻度

表 7.1 より，最尤推定による確率分布は表 7.2 のようになる。

	#a	#b	#c	#$	aa	ab	ac	a$	ba	bb	bc	b$	ca	cb	cc	c$
P_q	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.20	0.00	0.20	0.00	0.20	0.00
P_{c_1}	0.06	0.06	0.06	0.00	0.06	0.17	0.06	0.06	0.06	0.06	0.06	0.11	0.17	0.00	0.06	0.00
P_{c_2}	0.09	0.09	0.00	0.00	0.00	0.18	0.09	0.09	0.09	0.00	0.09	0.09	0.18	0.00	0.00	0.00

表 7.2　2-gram の確率分布

このとき，文書qとクラスc_1の JS 情報量$D_{JS}(P_q||P_{c_1})$は，以下のように計算される：

$$
\begin{aligned}
D_{JS}(P_q||P_{c_1}) &= \frac{1}{2}\left(D_{KL}(P_q||R) + D_{KL}(P_{c_1}||R)\right) \\
&= \frac{1}{2}\left(\sum_{f\in F} P_q(f)\log\frac{P_q(f)}{R(f)} + \sum_{f\in F} P_{c_1}(f)\log\frac{P_{c_1}(f)}{R(f)}\right) \\
&= \frac{1}{2}\left(\sum_{f\in F} P_q(f)\log\frac{P_q(f)}{\frac{P_q(f)+P_{c_1}(f)}{2}} + \sum_{f\in F} P_{c_1}(f)\log\frac{P_{c_1}(f)}{\frac{P_q(f)+P_{c_1}(f)}{2}}\right) \\
&= \frac{1}{2}\left(\left(0.00 * \log\frac{0.00}{\frac{0.00+0.06}{2}} + \cdots + 0.00 * \log\frac{0.00}{\frac{0.00+0.00}{2}}\right)\right. \\
&\quad \left. + \left(0.06 * \log\frac{0.06}{\frac{0.00+0.06}{2}} + \cdots + 0.00 * \log\frac{0.00}{\frac{0.00+0.00}{2}}\right)\right) \\
&\approx 0.060
\end{aligned}
$$

文書qとクラスc_2の JS 情報量$D_{JS}(P_q||P_{c_2})$は，以下のように計算される：

$$D_{JS}(P_q||P_{c_2}) = \frac{1}{2}\left(D_{KL}(P_q||R) + D_{KL}(P_{c_2}||R)\right)$$

$$= \frac{1}{2}\left(\sum_{f\in F} P_q(f)\log\frac{P_q(f)}{R(f)} + \sum_{f\in F} P_{c_2}(f)\log\frac{P_{c_2}(f)}{R(f)}\right)$$

$$= \frac{1}{2}\left(\sum_{f\in F} P_q(f)\log\frac{P_q(f)}{\frac{P_q(f)+P_{c_2}(f)}{2}} + \sum_{f\in F} P_{c_2}(f)\log\frac{P_{c_2}(f)}{\frac{P_q(f)+P_{c_2}(f)}{2}}\right)$$

$$= \frac{1}{2}\left(\left(0.00 * \log\frac{0.00}{\frac{0.00+0.09}{2}} + \cdots + 0.00 * \log\frac{0.00}{\frac{0.00+0.00}{2}}\right)\right.$$

$$\left. + \left(0.09 * \log\frac{0.09}{\frac{0.00+0.09}{2}} + \cdots + 0.00 * \log\frac{0.00}{\frac{0.00+0.00}{2}}\right)\right)$$

$$\approx 0.042$$

よって$D_{JS}(P_q||P_{c_1}) > D_{JS}(P_q||P_{c_2})$となるので，文書$q$はクラス$c_2$に属すると予測される。

第8章 ナイーブベイズ多変数ベルヌーイモデルによる文書分類

8.1　分類

ナイーブベイズは, ベイズの定理に基づき, 文書の事後確率をモデル化した確率的言語モデルである（Domingos & Pazzani 1996；McCallum & Nigam 1998；高村 2010：第 4 章）。ナイーブベイズは，SVM や対数線形モデルなどの識別モデルと比べると，精度が劣る傾向にある。しかし，本研究のようにクラス数が多いタスクにおいては，学習の速いナイーブベイズを用いる意義はある。また，各クラスにおける条件付き確率を計算することができるという利点もある。言語変異の年代推移や地理変異を可視化することができるからである。事後確率をモデル化する対数線形モデルでは，条件付き確率の計算はできない。

問題定義で述べたように，文書分類を行う文書qは，関数$\phi(q,c)$の値が最大となるクラス$c \in C$に分類される：

$$\hat{c_q} = \arg\max_{c \in C} \phi(q,c) \tag{8.1}$$

ここで，$\hat{c_q}$は文書qが属すると推定されたクラス，Cはクラスの集合である。ナイーブベイズにおいて，$\phi(q,c)$は，文書qがクラスcに属する事後確率（posterior probability）$P(c|q)$に対応する。事後確率$P(c|q)$とは，文書qが観測された時に，文書qがクラスcに属する確率である。したがって，文書qは事後確率$P(c|q)$が最大となるクラスcに分類される：

$$\hat{c_q} = \arg\max_{c \in C} P(c|q) \tag{8.2}$$

事後確率$P(c|q)$は，ベイズの定理により，以下のように変形される：

$$P(c|q) = \frac{P(c)P(q|c)}{P(q)} = \frac{P(c)P(q|c)}{\sum_{c \in C} P(c)P(q|c)} \propto P(c)P(q|c) \tag{8.3}$$

ここで, $P(c)$はクラスcの事前確率（prior probability）, $P(q|c)$はクラスcのもとで文書qが生成される確率（尤度 likelihood）, $P(q)$は文書qの周辺確率（marginal probability）である。事後確率の値そのものではなく，事後確率が最大となるクラスcを知るだけならば$P(c)P(q|c)$を求めるだけでよい：

$$\hat{c_q} = \arg\max_{c \in C} P(c)P(q|c) \tag{8.4}$$

なぜならば，分母の$P(q) = \sum_{c \in C} P(c)P(q|c)$は全クラスで共通だからである。また，事前確率$P(c)$がすべてのクラスにおいて同一と考えるならば，$P(q|c)$のみを求めればよい。

本研究では，文献学的特徴を素性として，ナイーブベイズ多変数ベルヌーイモデル（Naive Bayes Multivariate Bernoulli Model）により分類を行う（川崎 2015a）。このモデルでは，素性が出現するかしないかのみに注目する。したがって，出現回数が 1 回でも 100 回でも同様に扱われる。この点が，出現回数を考慮するナイーブベイズ多項モデル（Naive Bayes Multinomial Model）とは異なる点である。文書をトピック別に分類する一般的な文書分類のタスクでは，ナイーブベイ

ズ多項モデルの方が性能が高いことが示されている（McCallum & Nigam 1998）。しかし，文献学では，素性の出現の有無に注目するのが自然な方法なので，ナイーブベイズ多変数ベルヌーイモデルを採用する。

ナイーブベイズ多変数ベルヌーイモデルでは，尤度$P(q|c)$を以下のように求める（高村 2010：第 4 章）。素性の集合を$F = \{f_1, f_2, \ldots, f_{|F|}\}$とする。クラス$c$において，素性の出現確率が，それ以外の素性の出現と独立（条件付き独立）だと仮定すると（McCallum & Nigam 1998），事後確率は素性の出現確率の積として，

$$P(q|c) = P(f_1, f_2, \ldots, f_{|F|}|c) = \prod_{i=1}^{|F|} P(f_i|c) \tag{8.5}$$

で与えられる。「ナイーブ」と呼ばれるのは，各クラスにおいて，ある素性の生起が他の素性の生起と独立だと仮定しているためである。もちろん，実際には，素性の生起は独立ではないが，独立性を仮定することで，パラメータの学習を個別に行うことが可能となる（Domingos & Pazzani 1996；McCallum & Nigam 1998）。独立性を仮定しない場合，パラメータの学習を行うときに，自分以外の素性の生起も考慮することが必要となり，学習が困難となる。この傾向は，特に素性の数が多い場合，顕著になる。また，素性の出現順番は影響しないと仮定している。

クラスcの素性fに関して，$X_{f,c}$をベルヌーイ分布（付録 A の A7）に従う確率変数とする。つまり，クラスcにおいて素性fが出現する場合，$X_{w,c} = 1$となり，出現しない場合$X_{w,c} = 0$となる。$X_{w,c} = 1$となる確率$P(X_{f,c} = 1) = p_{w,c}$，$X_{w,c} = 0$となる確率$P(X_{f,c} = 0) = 1 - p_{f,c}$とする。したがって，$P(X_{f,c} = 1) + P(X_{f,c} = 0) = p_{f,c} + 1 - p_{f,c} = 1$が成り立つ。$p_{f,c}$は，各ベルヌーイ分布を規定するパラメータである。クラスcにおいて，各ベルヌーイ分布は独立と仮定している。

クラスcにおいて，素性fが出現するかどうかを表す確率は，

$$p_{f,c}^{\delta_{f,q}}(1 - p_{f,c})^{1-\delta_{f,q}} \tag{8.6}$$

となる。ここで，$\delta_{f,q}$は，素性fが文書qに出現したとき$\delta_{f,q} = 1$となり，出現しないときに$\delta_{f,q} = 0$となる関数である。式（8.6）は，素性fが出現するという情報と出現しないという情報の両者を考慮している：

$$p_{f,c}^{\delta_{f,q}}(1 - p_{f,c})^{1-\delta_{f,q}} = \begin{cases} p_{f,c}^0(1 - p_{f,c})^{1-0} = 1 - p_{f,c} & (\delta_{f,q} = 0) \\ p_{f,c}^1(1 - p_{f,c})^{1-1} = p_{f,c} & (\delta_{f,q} = 1) \end{cases} \tag{8.7}$$

したがって，素性の集合Fが与えられたとき，文書qのクラスcにおける尤度$P(q|c)$は，

$$P(q|c) = \prod_{f \in F} p_{f,c}^{\delta_{f,q}}(1 - p_{f,c})^{1-\delta_{f,q}} \tag{8.8}$$

と表される。文書qの事前確率$P(c) = p_c$とすると，多変数ベルヌーイモデルを採用した場合，文書qは，

$$P(c)P(q|c) = p_c \prod_{f \in F} p_{f,c}^{\delta_{f,q}} \left(1 - p_{f,c}\right)^{1-\delta_{f,q}} \tag{8.9}$$

が最大となるクラスcに分類される。

8.2　最尤推定

多変数ベルヌーイモデルのパラメータであるp_cと$p_{f,c}$を，訓練データ$D = \{(d^{(1)}, c^{(1)}), \ldots, (d^{(i)}, c^{(i)}), \ldots, (d^{(|D|)}, c^{(|D|)})\}$から最尤推定（付録 A の A9）により求めることを考える。$P(D)$を訓練データ内の文書全体の生成確率とすると，最尤推定では,

$$\begin{aligned}
P(D) &\propto \log P(D) \\
&= \log \prod_{(d,c) \in D} P(d,c) \\
&= \log \prod_{(d,c) \in D} P(c)P(d|c) \\
&= \sum_{(d,c) \in D} \log \left(p_c \prod_{f \in F} p_{f,c}^{\delta_{f,d}} \left(1 - p_{f,c}\right)^{1-\delta_{f,d}} \right) \\
&= \sum_{(d,c) \in D} \left(\log p_c + \sum_{f \in F} \left(\delta_{f,d} \log p_{f,c} + \left(1 - \delta_{f,d}\right) \log\left(1 - p_{f,c}\right) \right) \right) \\
&= \sum_c N_c \log p_c + \sum_c \sum_{f \in F} N_{f,c} \log p_{f,c} + \sum_c \sum_{f \in F} \left(N_c - N_{f,c}\right) \log\left(1 - p_{f,c}\right)
\end{aligned} \tag{8.10}$$

を最大化することになる。ここで，N_cはクラスcの文書数，$N_{f,c} (\leq N_c)$はクラスcの文書の中で素性fを含む文書の数である。対数関数は単調増加関数なので，$P(D)$の最大化は，$P(D)$の対数の最大化と等価である。

$\sum_c p_c = 1$という等式制約のもとでの$\log P(D)$の最大化は，等式制約付き凸計画問題なので，ラグランジュの未定乗数法による行うことができる（Bishop 2006 : 707-710 ; 高村 2010 : 1.2.3）。未定乗数λを用いて，次のようにラグランジュ関数を定義する：

$$L(\theta, \lambda) = \log P(D) + \lambda \left(\sum_c p_c - 1 \right) \tag{8.11}$$

ここで，θは，求めようとしている$|F||C| + |C|$個のパラメータ集合$\left(\{p_{f,c}\}_{f \in F, c \in C}, \{p_c\}_{c \in C}\right)$である。式（8.11）を各パラメータについて偏微分（付録 A の A6 を参照）すると,

$$\frac{\partial L(\theta, \lambda)}{\partial p_{f,c}} = \frac{N_{f,c}}{p_{f,c}} - \frac{N_c - N_{f,c}}{1 - p_{f,c}}, \qquad \frac{\partial L(\theta, \lambda)}{\partial p_c} = \frac{N_c}{p_c} + \lambda \tag{8.12}$$

となる。各式を 0 とおき，$\sum_c p_c = 1$という制約を用いて変形すると，

$$p_{f,c} = \frac{N_{f,c}}{N_c}, \qquad p_c = \frac{N_c}{\sum_c N_c} \tag{8.13}$$

となる。つまり，$p_{f,c}$は，クラスcに属する文書数のうち，素性fを含む文書の数である。p_cは，訓練データの文書のうち，クラスcに属する文書数の割合となっている。したがって，クラスが同一の場合，多くの文書で現われる素性fほど$p_{f,c}$が大きくなり，大きな重みを持つことになる。

8.3　MAP 推定

最尤推定では，訓練データに出現しない素性fの出現確率$p_{f,c}$がゼロと推定されてしまうゼロ頻度問題が生じる（$N_{f,c} = 0$より$p_{f,c} = \frac{N_{f,c}}{N_c} = 0$）。ゼロ頻度問題を回避するには，すべての素性$f$の出現確率$p_{f,c} > 0$となるように調整する必要がある。常套手段は，MAP（Maximum A Posteriori）推定によるパラメータ推定である（付録 A の A9 を参照）。MAP 推定では，パラメータの事前分布としてディリクレ分布（付録 A の A8）を導入することで，すべてのパラメータがゼロより大きな値を取るようにすることができる。

MAP 推定では，

$$\begin{aligned}
\log P(\theta|D) &= \log \frac{P(\theta)P(D|\theta)}{P(D)} \\
&\propto \log P(\theta) + \log P(D) \\
&= \log\left(\left(\prod_c p_c^{\alpha_c - 1}\right) \times \left(\prod_{f,c} p_{f,c}^{\alpha_{f,c} - 1}\right) \times \left(\prod_{f,c} (1 - p_{f,c})^{\alpha_{f,c} - 1}\right)\right) \\
&\quad + \sum_{(d,c) \in D} \log P(d,c) + (\text{定数}) \\
&= \log\left(\left(\prod_c p_c^{\alpha_c - 1}\right) \times \left(\prod_{f,c} \left(p_{f,c}^{\alpha_{f,c} - 1}(1 - p_{f,c})^{\alpha_{f,c} - 1}\right)\right)\right) + \sum_{(d,c) \in D} \log P(d,c) \\
&\quad + (\text{定数}) \\
&= \sum_c (\alpha_c - 1) * \log p_c + \sum_{f,c} (\alpha_{f,c} - 1) * \left(\log p_{f,c} + \log(1 - p_{f,c})\right) \\
&\quad + \sum_{(d,c) \in D} \log\left(p_c \prod_{f \in F} p_{f,c}^{\delta_{f,d}} (1 - p_{f,c})^{1 - \delta_{f,d}}\right) + (\text{定数})
\end{aligned} \tag{8.14}$$

で与えられる目的関数を最大化することになる。ここで，$\alpha_{f,c} - 1$は，クラスcの素性fに対して事前に与えられる観測数（出現回数）である。$\alpha_c - 1$は，クラスcに対して事前に与えられる観測数（文書数）である。$\alpha_{f,c}$（$|F||C|$個）とα_c

（$|C|$個）は，それぞれ，$p_{w,c}$とp_cの分布を規定するハイパーパラメータである。$P(\theta)$はディリクレ分布によるパラメータの事前確率である。

最尤推定と同様に，$\sum_c p_c = 1$という等式制約のもとで式（8.14）を最大化する。未定乗数λを用いて，次のようにラグランジュ関数を定義する：

$$\mathrm{L}(\theta, \lambda) = \log P(\theta) + \log P(D) + \lambda \sum_c (p_c - 1) \tag{8.15}$$

式（8.15）を各パラメータについて偏微分すると，

$$\frac{\partial L(\theta,\lambda)}{\partial p_{f,c}} = \frac{\alpha_{f,c}-1}{p_{f,c}} - \frac{\alpha_{f,c}-1}{1-p_{f,c}} + \frac{N_{f,c}}{p_{f,c}} - \frac{N_c - N_{f,c}}{1-p_{f,c}}, \qquad \frac{\partial L(\theta,\lambda)}{\partial p_c} = \frac{\alpha_c - 1}{p_c} + \frac{N_c}{p_c} + \lambda \tag{8.16}$$

となる。各式を 0 とおき，$\sum_c p_c = 1$という制約を用いて変形すると，

$$p_{f,c} = \frac{N_{f,c} + (\alpha_{f,c} - 1)}{N_c + 2(\alpha_{f,c} - 1)}, \qquad p_c = \frac{N_c + (\alpha_c - 1)}{\sum_c N_c + \sum_c (\alpha_c - 1)} \tag{8.17}$$

が得られる。MAP 推定では，クラスcに対して$\alpha_c - 1$を加え，文書数を水増ししている。また，クラスcの素性fに対しては$\alpha_{f,c} - 1$を加え，出現回数を水増ししている。$\alpha_c = 1$，$\alpha_{f,c} = 1$の時，MAP 推定値は最尤推定値と一致する。

本研究では，事前確率p_cはすべてのクラスcにおいて同一とする。したがって，文書qの属するクラス$\hat{c_q}$は，対数尤度$\log P(q|c)$が最大になるクラスcとなる。

8.4　例

たとえば，訓練データとして，クラスc_1には$\{f_1 f_1, f_3 f_3, f_2 f_3\}$という 3 文書が，クラス$c_2$には$\{f_1 f_3 f_2, f_3 f_1, f_2 f_3 f_1\}$という 2 文書が与えられているとする。素性集合$F = \{f_1, f_2, f_3\}$，$N_{c_1} = N_{c_2} = 3$となる。

文書$q = f_1 f_2 f_2$の属するクラス$\hat{c_q}$は，

$$\hat{c_q} = \arg\max_{c \in C} P(c) P(q|c)$$

と予測される。

クラスc_1のパラメータは，

$$p_{c_1} = \frac{N_{c_1}}{\sum_c N_c} = \frac{N_{c_1}}{N_{c_1} + N_{c_2}} = \frac{3}{3+3} = \frac{1}{2}$$

$$p_{f_1,c_1} = \frac{N_{f_1,c_1}}{N_{c_1}} = \frac{1}{3}$$

$$p_{f_2,c_1} = \frac{N_{f_2,c_1}}{N_{c_1}} = \frac{1}{3}$$

$$p_{f_3,c_1} = \frac{N_{f_3,c_1}}{N_{c_1}} = \frac{2}{3}$$

となる。したがって，クラスc_1における文書$q = f_1 f_2 f_2 f_1$の最尤推定による同時確率$P(c_1)P(q|c_1)$は,

$$\begin{aligned} P(c_1)P(q|c_1) &= p_{c_1} \prod_{f \in F} p_{f,c_1}^{\delta_{f,q}} \left(1 - p_{f,c_1}\right)^{1-\delta_{f,q}} \\ &= p_{c_1} \times p_{f_1,c_1} \times p_{f_2,c_1} \times \left(1 - p_{f_3,c_1}\right) \\ &= \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times \left(1 - \frac{2}{3}\right) \\ &\approx 0.019 \end{aligned}$$

と計算される。同様に，クラスc_2のパラメータは,

$$\begin{aligned} p_{c_2} &= \frac{N_{c_2}}{\sum_c N_c} = \frac{N_{c_2}}{N_{c_1} + N_{c_2}} = \frac{3}{3+3} = \frac{1}{2} \\ p_{f_1,c_2} &= \frac{N_{f_1,c_2}}{N_{c_2}} = \frac{3}{3} \\ p_{f_2,c_2} &= \frac{N_{f_2,c_2}}{N_{c_2}} = \frac{2}{3} \\ p_{f_3,c_2} &= \frac{N_{f_3,c_2}}{N_{c_2}} = \frac{3}{3} \end{aligned}$$

となるので，クラスc_2における文書$q = f_1 f_2 f_2 f_1$の最尤推定による同時確率$P(c_2)P(q|c_2)$は,

$$\begin{aligned} P(c_2)P(q|c_2) &= p_{c_2} \prod_{f \in F} p_{f,c_2}^{\delta_{f,q}} \left(1 - p_{f,c_2}\right)^{1-\delta_{f,q}} \\ &= p_{c_2} \times p_{f_1,c_2} \times p_{f_2,c_2} \times \left(1 - p_{f_3,c_2}\right) \\ &= \frac{1}{2} \times \frac{3}{3} \times \frac{2}{3} \times \left(1 - \frac{3}{3}\right) \\ &= 0 \end{aligned}$$

と計算される。最尤推定の場合，$P(c_1)P(q|c_1) > P(c_2)P(q|c_2)$となるので，文書$q$はクラス$c_1$に属すると判断される。

次に，$\alpha_{c_1} = \alpha_{c_2} = \alpha_{f_1,c_1} = \alpha_{f_2,c_1} = \alpha_{f_3,c_1} = \alpha_{f_1,c_2} = \alpha_{f_2,c_2} = \alpha_{f_3,c_2} = 2$として，MAP 推定によりパラメータを求めることにする。このとき，クラスc_1のパラメータは，以下のようになる：

$$p_{c_1} = \frac{N_{c_1} + \left(\alpha_{c_1} - 1\right)}{\sum_c N_c + \sum_c (\alpha_c - 1)} = \frac{N_{c_1} + \left(\alpha_{c_1} - 1\right)}{N_{c_1} + N_{c_2} + \left(\alpha_{c_1} - 1\right) + \left(\alpha_{c_2} - 1\right)} = \frac{3+1}{3+3+1+1} = \frac{1}{2}$$

$$p_{f_1,c_1} = \frac{N_{f_1,c_1} + (\alpha_{f_1,c_1} - 1)}{N_{c_1} + 2(\alpha_{f_1,c_1} - 1)} = \frac{1+1}{3+2} = \frac{2}{5}$$
$$p_{f_2,c_1} = \frac{N_{f_2,c_1} + (\alpha_{f_2,c_1} - 1)}{N_{c_1} + 2(\alpha_{f_2,c_1} - 1)} = \frac{1+1}{3+2} = \frac{2}{5}$$
$$p_{f_3,c_1} = \frac{N_{f_3,c_1} + (\alpha_{f_3,c_1} - 1)}{N_{c_1} + 2(\alpha_{f_3,c_1} - 1)} = \frac{2+1}{3+2} = \frac{3}{5}$$

したがって，クラスc_1における文書$q = f_1 f_2 f_2 f_1$の最尤推定による同時確率$P(c_1)P(q|c_1)$は，

$$\begin{aligned} P(c_1)P(q|c_1) &= p_{c_1} \prod_{f \in F} p_{f,c_1}^{\delta_{f,q}} (1 - p_{f,c_1})^{1-\delta_{f,q}} \\ &= p_{c_1} \times p_{f_1,c_1} \times p_{f_2,c_1} \times (1 - p_{f_3,c_1}) \\ &= \frac{1}{2} \times \frac{2}{5} \times \frac{2}{5} \times \left(1 - \frac{3}{5}\right) \\ &= 0.032 \end{aligned}$$

と計算される。同様に，クラスc_2のパラメータは，以下のようになる：

$$p_{c_2} = \frac{N_2 + (\alpha_2 - 1)}{\sum_c N_c + \sum_c (\alpha_c - 1)} = \frac{N_{c_2} + (\alpha_{c_2} - 1)}{N_{c_1} + N_{c_2} + (\alpha_{c_1} - 1) + (\alpha_{c_2} - 1)} = \frac{3+1}{3+3+1+1} = \frac{1}{2}$$
$$p_{f_1,c_2} = \frac{N_{f_1,c_2} + (\alpha_{f_1,c_2} - 1)}{N_{c_2} + 2(\alpha_{f_1,c_2} - 1)} = \frac{3+1}{3+2} = \frac{4}{5}$$
$$p_{f_2,c_2} = \frac{N_{f_2,c_2} + (\alpha_{f_2,c_2} - 1)}{N_{c_2} + 2(\alpha_{f_2,c_2} - 1)} = \frac{2+1}{3+2} = \frac{3}{5}$$
$$p_{f_3,c_2} = \frac{N_{f_3,c_2} + (\alpha_{f_3,c_2} - 1)}{N_{c_1} + 2(\alpha_{f_3,c_2} - 1)} = \frac{3+1}{3+2} = \frac{4}{5}$$

したがって，クラスc_2における文書$q = f_1 f_2 f_2 f_1$の最尤推定による同時確率$P(c_2)P(q|c_2)$は，

$$\begin{aligned} P(c_2)P(q|c_2) &= p_{c_2} \prod_{f \in F} p_{f,c_2}^{\delta_{f,q}} (1 - p_{f,c_2})^{1-\delta_{f,q}} \\ &= p_{c_2} \times p_{f_1,c_2} \times p_{f_2,c_2} \times (1 - p_{f_3,c_2}) \\ &= \frac{1}{2} \times \frac{4}{5} \times \frac{3}{5} \times \left(1 - \frac{4}{5}\right) \\ &= 0.048 \end{aligned}$$

と計算される。MAP 推定では，最尤推定の場合とは異なり，$P(c_1)P(q|c_1) < P(c_2)P(q|c_2)$となるので，文書$q$はクラス$c_2$に属すると予測される。

第9章 実験

9.1　実験設定

データセットは，作成年代と作成場所が既知の 1237 文書である。データセットを約 10 対 1 で訓練データとテストデータに分割し，10 分割交差検定（10-fold cross-validation）を行った（交差検定については，付録 A の A12 を参照）。したがって，訓練データは約 1100 文書，テストデータは約 120 文書である。分割基準は，文書 ID を 10 で割ったときの余りとした。サブデータセットは，質的に大きく異ならないことを確認した。文書数が少ないため，データセットを訓練データ，開発データ，テストデータに分割して実験を行うことはしない。

本研究では，テストデータに属する文書qの作成年代t_qと作成場所l_qがともに不明だと仮定し，推定年代$\hat{t_q}$と推定場所$\hat{l_q}$を求める実験を行った[31]。推定方法は，個別推定と同時推定の二種類である。個別推定では，作成年代と作成場所を独立に推定する。同時推定では，両者を同時に推定する。*n*-gram 言語モデルと JS 情報量に基づくモデルでは，素性として文字 2-gram を使用した。計算時間が長大になるため，3-gram 以上での実験は行わなかった。また，文字順を考慮しない 1-gram での実験も行わなかった。ナイーブベイズ多変数ベルヌーイモデルでは，素性として文献学的特徴を使用した。

最適なパラメータは，グリッドサーチで求めた。グリッドサーチとは，すべてのパラメータの組み合わせで実験を行い，最良の結果を生むパラメータの組み合わせを決定する作業のことである。本研究の場合，最良の結果とは，式(9.5)の誤差が最小になることを意味する。予備実験の結果を踏まえ，パラメータ空間は，時間カーネル平滑化パラメータ$\sigma_t \in \{0, 3, 5, 10\}$，空間カーネル平滑化パラメータ$\sigma_l \in \{0, 25, 50, 75, 100\}$，*n*-gram 言語モデルの加算スムージングのパラメータ$\alpha \in \{0.001, 0.01, 0.1, 1\}$，ナイーブベイズ多変数ベルヌーイモデルのハイパーパラメータ$\{\alpha_{f,c}\}_{f\in F, c\in C} \in \{1.001, 1.01, 1.1, 2\}$とした。$\sigma_t = 0$と$\sigma_l = 0$は，カーネル平滑化しないことを意味している。$\alpha_{f,c}$は，すべての素性で同一とした。時間カーネル平滑化パラメータと空間カーネル平滑化パラメータの組み合わせは，$(\sigma_t, \sigma_l) = \{(0,0), (3,25), (3,50), (3,75), (3,100), (5,25), (5,50), (5,75), (5,100), (10,25), (10,50), (10,75), (10,100)\}$の 13 通りに限定した。*n*-gram 言語モデルでは，これらの各組み合わせに対し，加算スムージングのパラメータαが四種類，推定方法が個別推定・同時推定の二種類あるので，合計で 104 通り（=13×4×2）の組み合わせが存在する。JS 情報量に基づくモデルでは，各組み合わせに対し，推定方法が個別推定・同時推定の二種類あるので，合計で 26 通り（=13×2）の組み合わせが存在する。*n*-gram 言語モデルナイーブベイズ多変数ベルヌーイモデルでは，各組み合わせに対し，ハイパーパラメータ$\alpha_{f,c}$が四種類，推定方法が個別推定・同時推定の二種類あるので，合計で 104 通り（=13×4×2）の組み合わせが存在する。各クラスの事前確率は，全クラスで同一としたので，無視する。実験は，すべて筆者が Excel VBA で実装して行った（付録 C を参照）。

年代推定・場所推定の評価指標は，いずれも絶対値誤差平均（Mean Absolute Error：以下 MAE），二乗平均平方根誤差（Root Mean Squared Error：以下 RMSE），絶対値誤差中央値（Median Absolute Error：以下 MedAE）である。これらの値が小さいほど，予測精度が高いといえる。推定精度が 100%のとき，いずれの指標も 0 となる。年代推定の絶対値誤差平均をMAE_t，二乗平均平方根誤差をRMSE_t，絶対値誤差中央値をMedAE_t，場所推定の絶対値誤差平均をMAE_l，二乗平均平方根誤差をRMSE_l，絶対値誤差中央値をMedAE_lとする。

データセットをD，データセットの文書数を$|D|$，文書$d \in D$の推定年代を$\hat{t_d}$，実年代をt_dとする。このとき，MAE_t

[31] 予備実験において，作成年代もしくは作成場所を既知として，もう一つの属性を推定する実験を行ったが，推定精度の大きな向上は見られなかった。

とRMSE_tは，それぞれ，次のように定義される：

$$\mathrm{MAE}_t = \frac{1}{|D|}\sum_{d\in D}|\widehat{t_d} - t_d| \tag{9.1}$$

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{|D|}\sum_{d\in D}|\widehat{t_d} - t_d|^2} \tag{9.2}$$

同様に文書$d \in D$の推定場所を$\widehat{l_d}$，実作成場所をl_dとすると，MAE_lとRMSE_lは，それぞれ，次のように定義される：

$$\mathrm{MAE}_l = \frac{1}{|D|}\sum_{d\in D} Dis\left(\widehat{l_d}, l_d\right) \tag{9.3}$$

$$\mathrm{RMSE}_l = \sqrt{\frac{1}{|D|}\sum_{d\in D} Dis\left(\widehat{l_d}, l_d\right)^2} \tag{9.4}$$

ここで，$Dis\left(\widehat{l_d}, l_d\right)$は推定場所$\widehat{l_d}$と実作成場所$l_d$との距離（単位はkm）を表している。$Dis\left(\widehat{l_d}, l_d\right)$は，推定場所$\widehat{l_d}$と実作成場所$l_d$の緯度・経度からヒュベニの公式（Hubeny's distance formula）を用いて計算する（付録 B の B1 を参照）。

年代推定・場所推定の結果をまとめて評価するために，RMSE_tとRMSE_lの積で与えられる評価指標を定義する：

$$\mathrm{STEP} = \mathrm{RMSE}_t \times \mathrm{RMSE}_l \tag{9.5}$$

これを，時空間誤差積（Spatio-Temporal Error Product：以下 STEP）と呼ぶ。STEP が小さいほど，推定精度が高いと評価する[32]。STEP の単位は年kmである。したがって，グリッドサーチにより求めるパラメータの組み合わせ（$\sigma_t, \sigma_l, \alpha, \alpha_{f,c}$）は，次式で与えられる：

$$\left(\sigma_t, \sigma_l, \alpha, \alpha_{f,c}\right) = \arg\min_{\left(\sigma_t, \sigma_l, \alpha, \alpha_{f,c}\right)} \mathrm{STEP} \tag{9.6}$$

ランダムな個別推定として，全文書の作成年代を平均作成年代 1393 年で推定すると，MAE_tは 119.00 年，RMSE_tは 135.43 年，MedAE_tは 117.00 年となる。また，作成場所を文書数が最大の地点（県レベルでは Valladolid，自治州レベルだと CL）で推定すると，県レベルではMAE_lは 227.82 km，RMSE_lは 291.91 km，MedAE_lは 195.30 km，STEP は 39532.68 年km，自治州レベルではMAE_lは 194.14 km，RMSE_lは 286.32 km，MedAE_lは 200.94 km，STEP は 38775.73 年kmとなる。同時推定の推定精度は，個別推定より低くなるので省略する。

[32] RMSE の代わりに MAE を用いても，以下の議論に大きな影響はない。

9.2　実験結果

実験は，*n*-gram 言語モデル，JS 情報量，ナイーブベイズ多変数ベルヌーイモデルの各モデルにおいて，個別推定か同時推定か（2 通り），またカーネル平滑化をするかしないか（2 通り）の合計 4 通り（= 2×2）の組み合わせで行う。つまり，①カーネル平滑化なしの個別推定，②カーネル平滑化なしの同時推定，③時間カーネル平滑化・空間カーネル平滑化ありの個別推定，④時空間カーネル平滑化ありの同時推定の 4 通りである。各々の設定において，パラメータの組み合わせを変化させ，STEP の値を比較する。

表 9.1 に，場所推定の粒度を県としたときの，4 つの設定における各モデルの最良の実験結果を示す。実験番号の LM（Language Model）は *n*-gram 言語モデル，JSD（Jensen-Shanon Divergence）は JS 情報量，NB（Naive Bayes）はナイーブベイズ多変数ベルヌーイモデルによる推定であることを表している。Prv は場所推定の粒度が県であることを表している。比較のために，ランダムな推定の精度も示す。すべての実験結果は，付録 B の B2 を参照。

実験番号		σ_t	σ_l	α	$\alpha_{f,c}$	年代推定			場所推定（県）			STEP
						MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
ランダム	個別	*	*	*	*	119.00	135.43	117.00	227.82	291.91	195.30	39532.68
①LM_Prv_1	個別	0	0	0.001	*	18.63	31.70	10.00	157.54	268.20	100.98	8503.12
②LM_Prv_54	同時	0	0	0.1	*	17.63	32.82	9.00	134.22	237.14	0.00	7782.90
③LM_Prv_11	個別	5	25	0.001	*	18.67	31.19	11.00	167.06	276.16	104.17	8612.13
④LM_Prv_76	同時	5	50	0.1	*	**13.99**	**25.51**	**7.00**	**109.98**	**208.43**	**0.00**	**5317.94**
①JSD_Prv_1	個別	0	0	*	*	16.48	27.35	9.00	224.11	380.93	120.54	10417.97
②JSD_Prv_2	同時	0	0	*	*	16.06	28.31	8.00	124.88	234.75	0.00	6646.79
③JSD_Prv_3	個別	3	25	*	*	20.21	33.01	12.00	236.73	388.56	144.25	12826.54
④JSD_Prv_6	同時	3	50	*	*	**14.61**	**23.68**	**8.00**	**144.36**	**215.13**	**107.80**	**5093.38**
①NB_Prv_1	個別	0	0	*	1.001	29.42	48.33	16.00	166.30	286.66	105.08	13853.46
②NB_Prv_28	同時	0	0	*	1.01	29.29	49.18	15.00	158.79	262.05	104.17	12888.27
③NB_Prv_5	個別	3	50	*	1.001	28.55	49.21	16.00	199.50	293.39	146.06	14437.90
④NB_Prv_34	同時	3	75	*	1.01	**24.33**	**42.00**	**12.00**	**168.67**	**244.98**	**124.07**	**10289.82**

表 9.1　年代推定・場所推定（県）の最良の推定精度

STEP が最小となるのは，*n*-gram 言語モデル，JS 情報量，ナイーブベイズ多変数ベルヌーイモデルの各モデルにおいて，いずれも，④時空間カーネル平滑化ありの同時推定の場合である。したがって，本研究の提案する手法の有効性が確認された。以下，②時空間カーネル平滑化なしの同時推定，①カーネル平滑化なしの個別推定，③カーネル平滑化ありの個別推定の順で，STEP は大きくなる。STEP が最大となる③カーネル平滑化ありの個別推定でも，ランダムな推定より推定精度は良い。ただし，JS 情報量による個別推定（①カーネル平滑化なしの個別推定，③カーネル平滑化ありの個別推定）では，カーネル平滑化の有無にかかわらず，ランダムな推定より，RMSE_lがかなり大きくなってしまう。

つづいて，表 9.2 に，場所推定の粒度を自治州としたときの，4 つの設定における各モデルの最良の実験結果を示す。実験番号の CA は，場所推定の粒度が自治州であることを表している。すべての実験結果は，付録 B の B3 を参照。

実験番号		σ_t	σ_l	α	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
						MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
ランダム	個別	*	*	*	*	119.00	135.43	117.00	194.14	286.32	200.94	38775.73
①LM_CA_1	個別	0	0	0.001	*	18.63	31.70	10.00	170.59	287.36	123.00	9110.40
②LM_CA_54	同時	0	0	0.1	*	16.28	30.65	8.00	105.68	214.24	0.00	6565.74
③LM_CA_3	個別	3	25	0.001	*	17.87	30.41	10.00	170.78	286.66	157.88	8717.37
④LM_CA_76	同時	10	75	0.1	*	**15.32**	**26.56**	**9.00**	**106.63**	**198.50**	**0.00**	**5272.13**
①JSD_CA_1	個別	0	0	*	*	16.48	27.35	9.00	287.33	476.08	206.65	13020.21
②JSD_CA_2	同時	0	0	*	*	15.68	27.36	8.00	107.61	224.13	0.00	6132.54
③JSD_CA_3	個別	3	25	*	*	20.21	33.01	12.00	294.24	481.12	206.65	15882.07
④JSD_CA_4	同時	3	25	*	*	**15.43**	**26.70**	**9.00**	**106.71**	**206.43**	**0.00**	**5511.08**
①NB_CA_1	個別	0	0	*	1.001	29.42	48.33	16.00	188.19	331.95	123.00	16041.97
②NB_CA_54	同時	0	0	*	1.01	27.96	46.98	14.00	141.79	256.26	0.00	12039.68
③NB_CA_5	個別	3	50	*	1.001	28.55	49.21	16.00	195.33	337.21	172.85	16594.19
④NB_CA_40	同時	5	50	*	1.01	**24.84**	**41.87**	**13.00**	**126.07**	**231.59**	**0.00**	**9697.67**

表 9.2　年代推定・場所推定（自治州）の最良の推定精度

STEP が最小となるのは，*n*-gram 言語モデル，JS 情報量，ナイーブベイズ多変数ベルヌーイモデルの各モデルにおいて，いずれも，④時空間カーネル平滑化ありの同時推定の場合である。したがって，本研究の提案する手法の有効性が確認された。*n*-gram 言語モデルでは，以下，②時空間カーネル平滑化なしの同時推定，③カーネル平滑化ありの個別推定，①カーネル平滑化なしの個別推定の順で，STEP が大きくなる。JS 情報量とナイーブベイズ多変数ベルヌーイモデルでは，以下，②時空間カーネル平滑化なしの同時推定，①カーネル平滑化なしの個別推定，③カーネル平滑化ありの個別推定の順で，STEP が大きくなる。STEP が最大になる場合でも，ランダムな推定より推定精度は良い。ただし，JS 情報量による個別推定（①カーネル平滑化なしの個別推定，③カーネル平滑化ありの個別推定）では，カーネル平滑化の有無にかかわらず，ランダムな推定より，MAE_l と RMSE_l がかなり大きくなってしまう。

場所推定の粒度を自治州としたときの，最良の推定結果を以下に図示する。場所推定の粒度が県の場合も同様の結果となるので，図示は省略する。図 9.1，図 9.2，図 9.3 は，*n*-gram 言語モデル，JS 情報量，ナイーブベイズ多変数ベルヌーイモデルの各モデルにおいて STEP が最小となる④LM_CA_76，④JSD_CA_4，④NB_CA_40 の実年代と推定年代の散布図である。

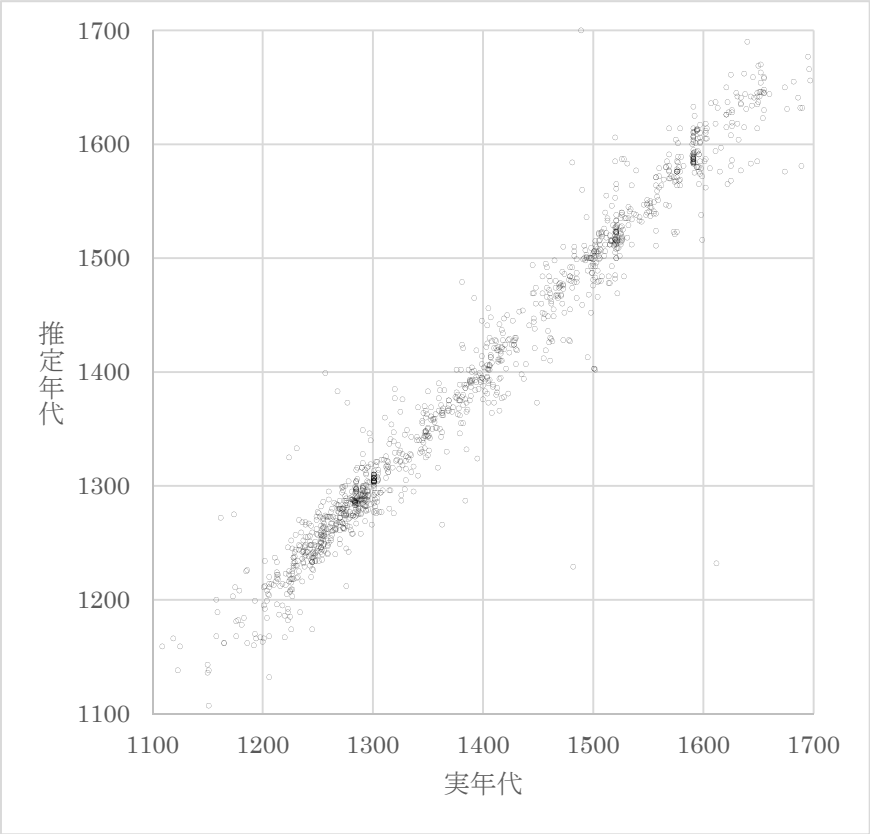


図 9.1　実年代と推定年代の散布図（④LM_CA_76）

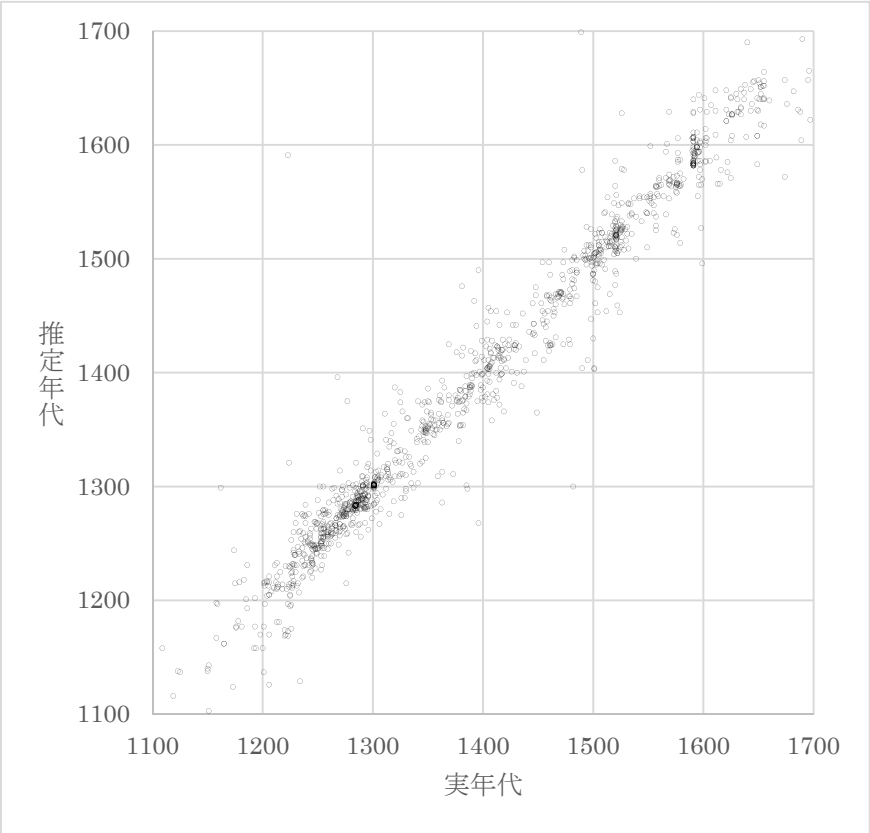


図 9.2　実年代と推定年代の散布図（④JSD_CA_4）

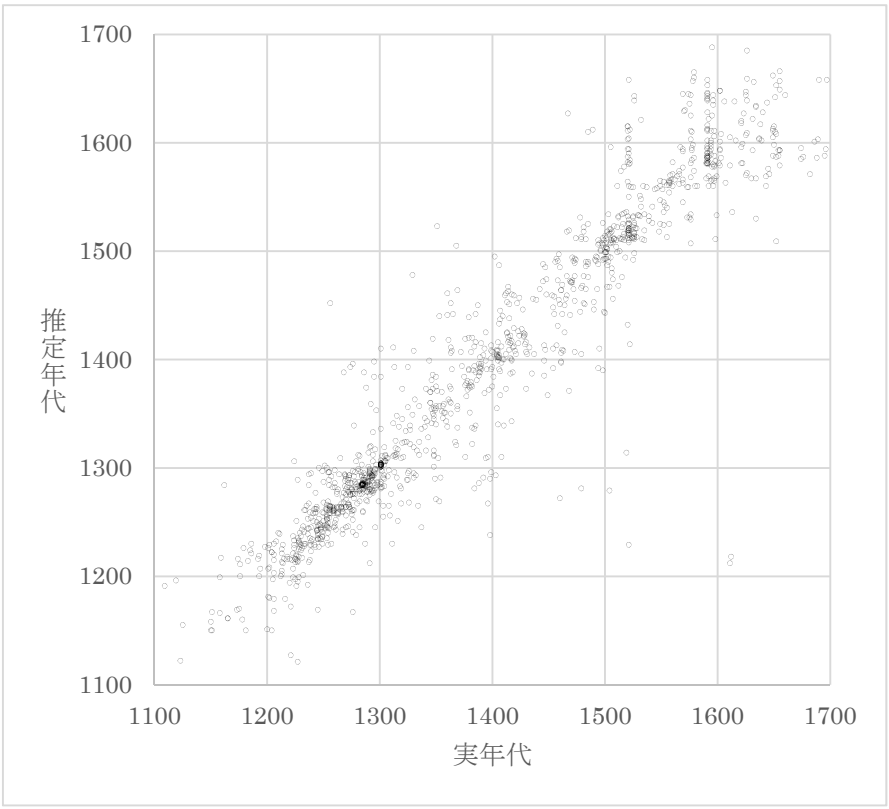


図 9.3　実年代と推定年代の散布図（④NB_CA_40）

全体的に，点が対角線上付近に集中している。対角線上の点もしくはその付近の点は，年代推定が正確な文書である。一方，対角線上から遠いところに位置する点ほど，年代推定が不正確な文書である。推定年代が実年代から大きく異なる外れ値も散見される。文書数が少ない両端の1100年代と1600年代後半は，年代推定の誤差が大きくなる。④LM_CA_76（図 9.1）と④JSD_CA_4（図 9.2）に比べ推定精度が低い④NB_CA_40（図 9.3）には，対角線から離れたところに多くの点が存在する。

表 9.3，表 9.4，表 9.5 はそれぞれ，④LM_CA_76 の場所推定の対応表，④JSD_CA_4 の場所推定の対応表，④NB_CA_40 の場所推定の対応表である。一列目が実際の自治州，一行目の*が付いているのが推定自治州である。対角線上のセルは場所推定が正確な文書数を表している。精度とは，当該クラスに分類された文書のうち，正しく分類されている文書の割合である。再現率とは，当該クラスに分類されるべき文書のうち，正しく分類されている文書の割合である。右下の値（それぞれ，59%，63%，57%）は，全体における正解率（accuracy）である。

	*AN	*AR	*AS	*CB	*CL	*CM	*CT	*EX	*GA	*IB	*IT	*LR	*MD	*MU	*NA	*PT	*PV	*VC	合計	再現率
AN	71				18	13						1	13				2		118	60%
AR		136		1	5							2	2		4		1		151	90%
AS			32	5	4				2			1	1						45	71%
CB	2			36	10							1	2				1		52	69%
CL	28		3	41	308	43		1	1			9	75		2		8		519	59%
CM	9			4	23	55		2				1	30				1		125	44%
CT																			0	*
EX	1				5	1		12					3						22	55%
GA					1				2										3	67%
IB																			0	*
IT					2	1					8		1						12	67%
LR	2			11	14	1		1				13			1		4		47	28%
MD	9			2	17	15					1		31			1	1		77	40%
MU	1					1													2	0%
NA		6		1	2							14	1		46		7		77	60%
PT													3						3	0%
PV	1			2	3	4						2	1				5		18	28%
VC					1								1						2	0%
合計	124	142	35	103	413	134	0	16	5	0	9	44	164	0	53	1	30	0	1273	
精度	57%	96%	91%	35%	75%	41%	*	75%	40%	*	89%	30%	19%	*	87%	0%	17%	*		59%

表 9.3　場所推定の対応表（④LM_CA_76）

	*AN	*AR	*AS	*CB	*CL	*CM	*CT	*EX	*GA	*IB	*IT	*LR	*MD	*MU	*NA	*PT	*PV	*VC	合計	再現率
AN	63				33	12		1					7				2		118	53%
AR		140			6	1									3		1		151	93%
AS			32	1	11								1						45	71%
CB				31	14	1						2					4		52	60%
CL	30		6	2	373	41		2				7	28		2		28		519	72%
CM	16				40	47		2			1	2	16				1		125	38%
CT																			0	*
EX	1				5			13					3						22	59%
GA					1				2										3	67%
IB																			0	*
IT					1	1					9		1						12	75%
LR	1				15	1		2				21			1		6		47	45%
MD	7			1	24	13					2	2	25			1	2		77	32%
MU	1					1													2	0%
NA	1	2			3	1						6	1		41		22		77	53%
PT					1	1							1						3	0%
PV	2			2	5	4						1			2		2		18	11%
VC					1								1						2	0%
合計	122	142	38	37	533	124	0	20	2	0	12	41	84	0	49	1	68	0	1273	
精度	52%	99%	84%	84%	70%	38%	*	65%	100%	*	75%	51%	30%	*	84%	0%	3%	*		63%

表 9.4　場所推定の対応表（④JSD_CA_4）

	*AN	*AR	*AS	*CB	*CL	*CM	*CT	*EX	*GA	*IB	*IT	*LR	*MD	*MU	*NA	*PT	*PV	*VC	合計	再現率
AN	50				38	11						1	15				3		118	42%
AR		132		1	5	1			1				1		6		4		151	87%
AS			35	1	9														45	78%
CB		2		23	21	1						2	1				2		52	44%
CL	25	2	5	11	354	31		1	5		2	15	51		4		13		519	68%
CM	14			1	45	37					2		20		2		4		125	30%
CT																			0	*
EX	2			1	6	1		9					2				1		22	41%
GA		2			1														3	0%
IB																			0	*
IT	1					2					7		2						12	58%
LR	3			5	14	4		1				14	3		1		2		47	30%
MD	13				24	9						2	27		1		1		77	35%
MU					1												1		2	0%
NA	3	3			3							8	3		41		16		77	53%
PT					1								2						3	0%
PV	1				8	2						3	2				2		18	11%
VC	1												1						2	0%
合計	113	141	40	43	530	99	0	11	6	0	11	45	130	0	55	0	49	0	1273	
精度	44%	94%	88%	53%	67%	37%	*	82%	0%	*	64%	31%	21%	*	75%	*	4%	*		57%

表 9.5　場所推定の対応表（④NB_CA_40）

いずれの場合も，AR は，精度，再現率ともに高い。これは，AR には，他の自治州とは異なる特異なパターンがあることを示唆している。

図 9.4 は，④LM_CA_76 による文書 ID1（作成年代：1251 年，県：Sevilla，自治州：AN）の推定年代・推定場所の上位 10 候補を表している[33]。図 9.5 は，その対数尤度の時空間分布を表している。対数尤度が小さい自治州の分布は省略した。カーネル平滑化パラメータが大きいので，滑らかな曲線になっている。

	対数尤度	事後確率
1276:MD	-14196.8	20.0%
1275:MD	-14196.8	19.4%
1274:MD	-14197.0	15.6%
1280:MD	-14197.3	12.0%
1273:MD	-14197.6	9.1%
1279:MD	-14197.9	6.2%
1272:MD	-14198.0	6.0%
1271:MD	-14198.2	4.7%
1278:MD	-14198.6	3.4%
1277:MD	-14198.7	3.0%

図 9.4　④LM_CA_76 による文書 ID1（作成年代：1251 年，自治州：AN）の推定年代・推定場所の上位 10 候補

[33] 事後確率や JS 情報量の逆数を重みとした重み付き平均により年代推定・場所推定を行ったが，ほとんどの場合，推定精度は統計的に有意には改善しなかった。

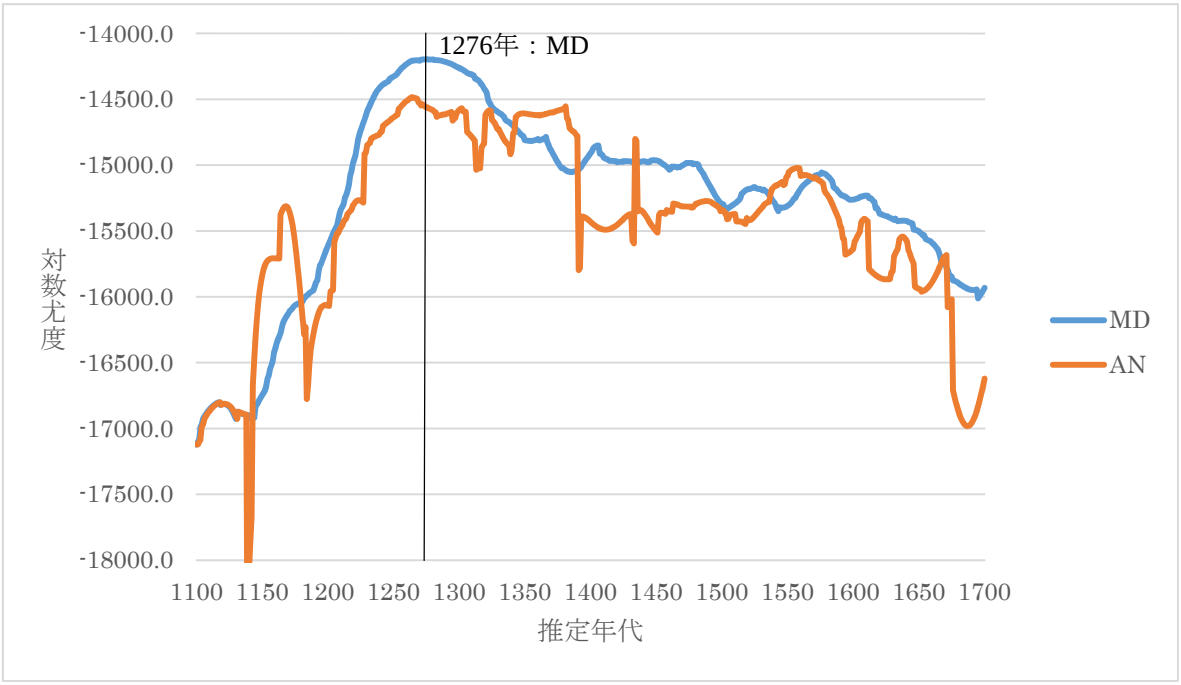


図 9.5　④LM_CA_76 による，文書 ID1（作成年代：1251 年，自治州：AN）の対数尤度の時空間分布

推定年代の 1276 年は，実年代の 1251 年に比較的近い。一方，推定場所の MD（Madrid）は，実際の作成場所である AN（Andalucía）からは，大きく離れている。上位 10 候補は，推定年代・推定場所の近くに集中している。これは，推定の信頼性を高める望ましい性質である。ただし，一定の年代や場所に集中しているからといって，推定精度が高くなるという傾向は見られなかった。

図 9.6 は，④JSD_CA_4 による文書 ID1（作成年代：1251 年，県：Sevilla，自治州：AN）の推定年代・推定場所の上位 10 候補を表している。図 9.7 は，JS 情報量の逆数の時空間分布を表している。JS 情報量の逆数が小さい自治州の分布は省略した。カーネル平滑化パラメータが小さいので，凸凹な曲線になっている。

	JS情報量	JS情報量の逆数
1277:CL	0.0321	31.16
1278:CL	0.0322	31.10
1276:CL	0.0323	31.00
1279:CL	0.0326	30.68
1275:CL	0.0326	30.66
1282:PV	0.0330	30.28
1274:CL	0.0332	30.14
1280:CL	0.0336	29.80
1273:CL	0.0340	29.43
1293:CL	0.0346	28.93

図 9.6　④JSD_CA_4 による文書 ID1（作成年代：1251 年，自治州：AN）の推定年代・推定場所の上位 10 候補

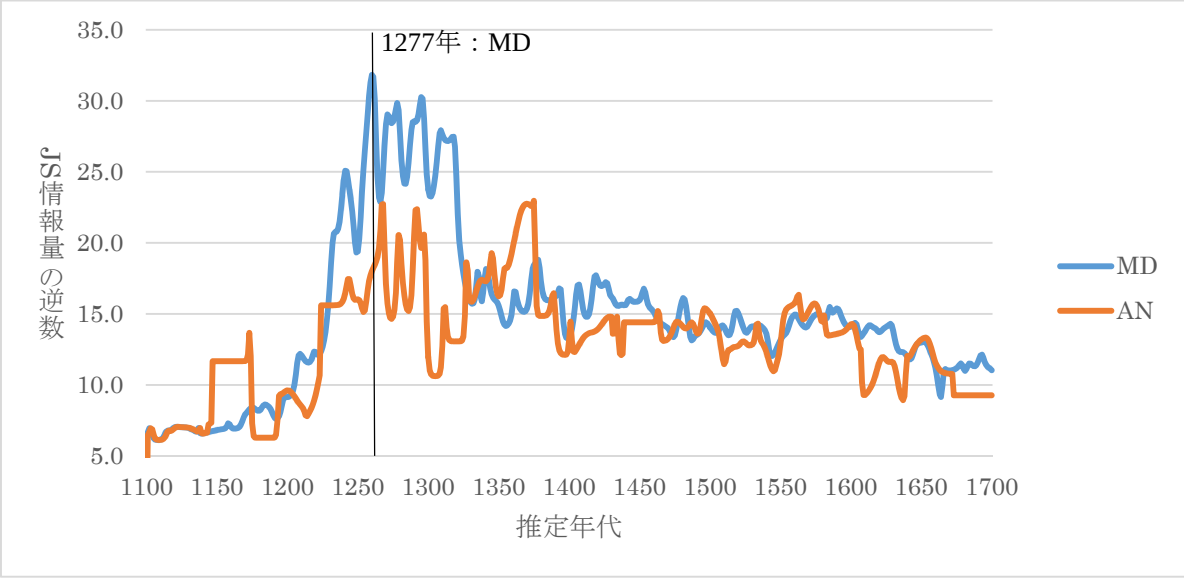


図 9.7　④JSD_CA_4 による，文書 ID1（作成年代：1251 年，自治州：AN）の JS 情報量の逆数の時空間分布

推定年代の 1277 年は，実年代の 1251 年に比較的近い。一方，推定場所の MD（Madrid）は，実際の作成場所である AN（Andalucía）からは，大きく離れている。上位 10 候補は，推定年代・推定場所の近くに集中している。

図 9.8 は，④NB_CA_40 による文書 ID1（作成年代：1251 年，県：Sevilla，自治州：AN）の推定年代・推定場所の上位 10 候補を示している。図 9.9 は，その対数尤度の時空間分布を表している。対数尤度が小さい自治州の分布は省略した。カーネル平滑化パラメータは中程度の値だが，素性の出現頻度が小さいため，波打った曲線になっている。

	対数尤度	事後確率
1302:MD	-103.5	23.9%
1301:MD	-103.6	23.0%
1303:MD	-103.9	16.9%
1300:MD	-104.1	13.0%
1304:MD	-104.5	9.4%
1305:MD	-105.2	4.4%
1299:MD	-105.3	3.9%
1306:MD	-106.2	1.7%
1229:AN	-106.3	1.4%
1230:AN	-106.8	0.9%

図 9.8　④NB_CA_40 による文書 ID1（作成年代：1251 年，自治州：AN）の推定年代・推定場所の上位 10 候補

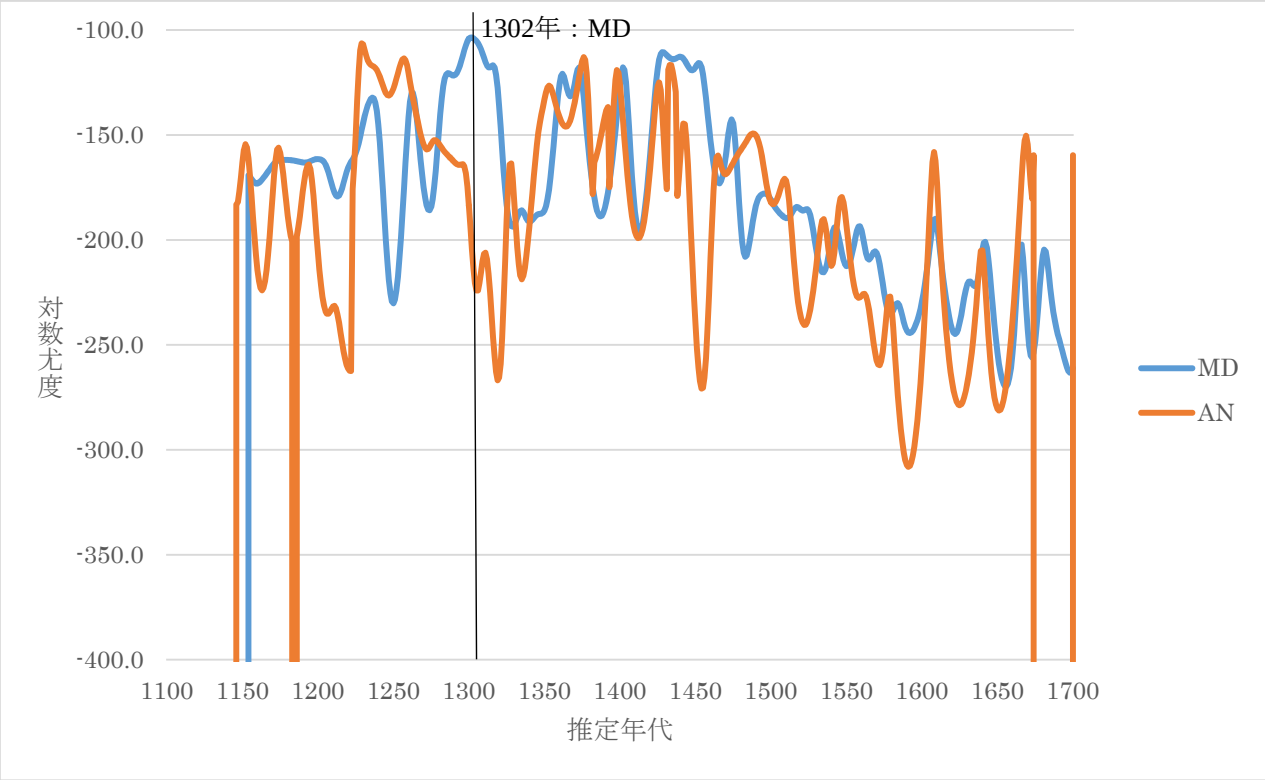


図 9.9　④NB_CA_40 による，文書 ID1（作成年代：1251 年，自治州：AN）の対数尤度の時空間分布

推定年代の 1302 年は，実年代の 1251 年から大きく離れている。一方，推定場所の MD（Madrid）も，実際の作成場所である AN（Andalucía）から大きく離れている。図 9.9 から，複数の峰が存在することが分かる。これは，推定の信頼性を下げる望ましくない性質である。

9.3　考察

9.3.1　分類器，素性，平滑化の有無による推定精度の違い

まずは，文字 *n*-gram を素性として用いた *n*-gram 言語モデルと JS 情報量による推定を比較する。同時推定では，カーネル平滑化の有無にかかわらず，両モデルの STEP に大きな差異は見られない。個別推定では，*n*-gram 言語モデルの方が，JS 情報量による推定よりも STEP が小さくなる傾向にある。これは，JS 情報量による推定において，場所推定の精度が極端に低下するためである。ただし，場所推定の精度が低下する理由は不明である。

次に，文字 *n*-gram を素性とした用いた *n*-gram 言語モデルと JS 情報量と，文献学的特徴を素性として用いたナイーブベイズ多変数ベルヌーイモデルを比較する。表 9.1 と表 9.2 より，前者よりも後者の方が，推定精度が低いのが分かる。推定精度の低さが，素性セットと分類器のどちら（もしくは両者）に起因するかは分からない。ここでは，素性セットのみに注目することにする。考えられる理由の一つは，文献学的特徴が純粋に形式的な特徴だけを捉えているのに対し，文字 *n*-gram はより多くの情報を捉えていることである。また，文字 *n*-gram の素性数（900 弱）の方が，文献学的特徴（300 弱）よりも多いので，より多くの情報を捉えていると考えられる。推定精度を基準にした場合，文書分類のタスクには，機械的に抽出した文字 *n*-gram の方が適していると考えられる（Rosenfeld 2000）。ただし，文献学的特徴を素性として用いることには，文献学的に解釈できるモデルの構築という方法論的な意義がある。

次に，個別推定・同時推定による推定精度の差異を考察する。パラメータ（加算スムージングのパラメータα，ナイーブベイズ多変数ベルヌーイモデルのハイパーパラメータ$\alpha_{f,c}$）が同一の場合，ほとんど常に，個別推定よりも同時推定の方が推定精度が高くなる。年代推定ではMAE_tがマイナス 2〜5 年，場所推定ではMAE_lがマイナス 80 ㎞〜180 ㎞である。この差は，統計的に有意である（有意水準 0.05）。検定は，対応のある t 検定により行った。推定精度の向上は，同時推定により，言語の時空間変異を上手く捉えられているためだと考えられる。

次に，カーネル平滑化の効果を考察する。表 9.6 は，パラメータ（加算スムージングのパラメータα，ナイーブベイズ多変数ベルヌーイモデルのハイパーパラメータ$\alpha_{f,c}$）を同一としたとき，分類器の種類，個別推定か同時推定か，年代推定か場所推定かに注目して，カーネル平滑化の効果を示している。LM（Language Model）は *n*-gram 言語モデル，JSD（Jensen-Shanon Divergence）は JS 情報量，NB（Naive Bayes）はナイーブベイズ多変数ベルヌーイモデルによる推定であることを表している。「○」は，カーネル平滑化の効果が，平滑化パラメータの値に関わらず，多くの場合ポジティブであることを表している。「△」は，カーネル平滑化の効果が，平滑化パラメータの値に関わらず，多くの場合ゼロであることを表している。「×」は，カーネル平滑化の効果が，平滑化パラメータの値に関わらず，多くの場合ネガティブであることを表している。

	個別推定		同時推定	
	年代推定	場所推定	年代推定	場所推定
LM	△	△	○	△
JSD	×	×	△	×
NB	△	×	○	△

表 9.6　カーネル平滑化の効果

上表より，平滑化の効果が見られるのは，同時推定による年代推定のみである。個別推定の場合と同時推定による場所推定では，平滑化の効果はゼロもしくはネガティブである。

次に，カーネル平滑化パラメータの値の大小と推定精度の関係を考察する。推定精度が高いのは，表 9.1 と表 9.2 より，時間カーネル平滑化パラメータσ_tと空間カーネル平滑化パラメータσ_lが，$\sigma_t = 3,5$や$\sigma_l = 25,50$のように，比較的小さい場合が多い。したがって，平滑化の幅は，小さい方が効果的なようである。これは，現時点での文書量が少ないため，各年代や地点において類似した文書が集中し，付近の年代や地点の情報を考慮しなくても，ある程度の分類精度が出てしまうためかもしれない。また，空間カーネル平滑化を行うことで，推定精度が落ちることもある。ただし，推定精度への貢献がないとしても，平滑化には欠損値補間と頑健な推定という方法論的な意義がある。

最後に，加算スムージングパラメータα，ナイーブベイズ多変数ベルヌーイモデルのハイパーパラメータ$\alpha_{f,c}$の大小と推定精度の関係を考察する。表 9.1 と表 9.2 からは，特異なパターンは見つからない。

9.3.2　個別推定と同時推定の比較

前節で，同一のパラメータのもとでは，個別推定より同時推定の方が推定精度が高いことを確認した。ここでは，同時推定の精度向上は，全体的な傾向なのか，それとも特定の年代や地点のみで見られる現象なのかを考察する。一例として，*n*-gram 言語モデルによる個別推定（③LM_CA_75）と同時推定（④LM_CA_76）を比較する。④LM_CA_76 は，*n*-gram 言語モデルにおいて，STEP が最小となる場合である（表 9.2）。パラメータは，どちらも，$\sigma_t = 10$，σ_l =75，$\alpha_{f,c}$ = 0.10 である。

表 9.7 は同時推定による年代推定の誤差増減を，表 9.8 は同時推定による場所推定の誤差増減を示している。各セルの値は，該当の年代と地点における誤差増減の合計を表している。同時推定により，全体として，年代推定ではマイナス 2.5 年（17.87 年→15.32 年），場所推定ではマイナス 64 km（171 km→107 km），誤差が減少している。

	AN	AR	AS	CB	CL	CM	EX	GA	IB	IT	LR	MD	MU	NA	PT	PV	VC	合計	文書数	平均
1100					5													5	1	5.0
1125					4													4	4	1.0
1150					6													6	2	3.0
1175		-30		1	35						-11							-5	11	-0.5
1200		3		-5	18	-6		-9										1	15	0.1
1225		-20		-3	75	1					7			-9				51	48	1.1
1250	49	25	1	10	71	3		-57			-4	-4		-14				80	99	0.8
1275	-20	-107	27	-69	-137	14					-36	-2		-23				-353	122	-2.9
1300	-8	-510	6	3	229	60	-4				-21			-26		-16		-287	156	-1.8
1325	-21	-355	-71	8	88	-42					-24			1				-416	91	-4.6
1350	-8	-386	-74	1	-125	1	24				13	-17		27		-9		-553	53	-10.4
1375	-20	-204	-12	-5	-78	-24	38				-5			-61				-371	41	-9.0
1400		38	-6	-35	-75	-81	-53					-86		-63	17	2		-242	65	-3.7
1425	13	235		-58	-131	8	-12				-95	-97		60		7		-120	69	-1.7
1450		-42	-67		-38	-1					9			-16				-205	27	-7.6
1475	-45	-31		-23	40	45						-19						-83	51	-1.6
1500	33	-95	0	32	-36	23					-7	-13		-5			3	-115	53	-2.2
1525	77	-220		-146	329	-47					21	-54	0	20		-9		-39	114	-0.3
1550	-4	-32			-38	-33	8					-16						-115	37	-3.1
1575	-117	-1			8	16						-43						-137	42	-3.3
1600	109				-6	-46	8			112		-89	8		-6	-4		136	87	1.6
1625	-9				-39	-50				-16		83						-31	29	-1.1
1650	-31	12			-55	-41						-15						-130	30	-4.3
1675	15				19	-102						-51						-119	17	-7.0
1700					-35	-28						-73					-71	-207	9	-23.0
合計	13	-1770	-196	-289	-16	-280	9	-66		96	-153	-406	8	-109	11	-29	-68	-3245	1273	-2.5
文書数	118	151	45	52	519	125	22	3	0	12	47	77	2	77	3	18	2	1273		
平均	0.1	-11.7	-4.4	-5.6	-0.0	-2.2	0.4	-22.0		8.0	-3.3	-5.3	4.0	-1.4	3.7	-1.6	-34.0	-2.5		

表 9.7　同時推定による年代推定の誤差増減

表 9.7 では，三つの点が注目に値する。一つ目は，AR の誤差が大きく減少していることである。全体の誤差減少分の半分以上を占めている。二つ目は，AR の文書が多い 14 世紀の誤差が大きく減少していることである。三つ目は，文書数が少ない年代や地点では，誤差が増加していることである。これは，同時推定によりクラスが細分化され，文書数が少ないクラスでは，パラメータの推定値がさらに不正確になるためだと考えられる。

	AN	AR	AS	CB	CL	CM	EX	GA	IB	IT	LR	MD	MU	NA	PT	PV	VC	合計	文書数	平均
1100					0													0	1	0
1125					-862													-862	4	-216
1150					0													0	2	0
1175		-351		0	-458						201							-608	11	-55
1200		-276		-145	-745	133		0										-1031	15	-69
1225		-385		-969	-1457	135					-112			264				-2524	48	-53
1250	-1714	-329	0	0	-865	-207		0			281	0		166				-2663	99	-27
1275	-1853	-589	667	-1055	-2667	337					-39	-59		361				-4896	122	-40
1300	-1083	-667	0	-495	-1699	-2726	-155				154			352		-67		-6375	156	-41
1325	-139	0	-423	-166	-3025	-592					769			646				-2930	91	-32
1350	-457	-7	313	14	276	17	321				310	-310		264		146		887	53	17
1375	-1331	0	-317	181	-990	-468	641				0			0				-2284	41	-56
1400		158	-272	-501	-3674	-291	279					-815		0	0	0		-5117	65	-79
1425	84	0		0	-3699	17	521				178	-194		-316		117		-3292	69	-48
1450		0	-685		-1470	-345					271			0				-2229	27	-83
1475	-735	0		-431	-3811	248						144						-4585	51	-90
1500	-2137	316	-296	-668	-1739	-36					-116	-369		0			241	-4804	53	-91
1525	101	-145		50	-5648	1826					-235	-689	0	-1418		-1207		-7361	114	-65
1550	396	-240			-173	748	-516					663						878	37	24
1575	-3446	-287			-223	-163						706						-3413	42	-81
1600	-9493				-23	591	0			-1359		-1283	-122		-1389	-528		-13606	87	-156
1625	-581				-1031	-1034				68		1253						-1324	29	-46
1650	-1735	0			-2288	-2899						-1220						-8142	30	-271
1675	-82				-2785	-369						173						-3064	17	-180
1700					123	-1322						-1097					0	-2295	9	-255
合計	-24203	-2801	-1013	-4185	-38929	-6397	1090	0		-1290	1663	-3088	-122	323	-1389	-1538	241	-81639	1273	-64
文書数	118	151	45	52	519	125	22	3	0	12	47	77	2	77	3	18	2	1273		
平均	-205	-19	-23	-80	-75	-51	50	0		-108	35	-40	-61	4	-463	-85	121	-64		

表 9.8　同時推定による場所推定の誤差増減

表 9.8 では，AN と CL の誤差が大きく減少していることが注目に値する。全体の誤差減少分の大半を占めている。それ以外は，特別なパターンは見られない。

第10章　結論

10.1 まとめ

本研究では，中近世スペイン語古文書の作成年代と作成場所を言語的特徴に基づき統計的に推定する方法を提案した。第 1 章では，研究背景の説明と問題定義を行った。第 2 章では，年代推定・場所推定に関連する先行研究を紹介した。第 3 章では，本研究で用いるコーパスの概要と記述的統計を示した。第 4 章では，素性として用いる文字 *n*-gram と文献学的特徴について説明した。第 5 章では，欠損値補間と頑健な推定を可能にするカーネル平滑化について説明した。第 6 章では *n*-gram 言語モデルによる年代推定法・場所推定法を，第 7 章では JS 情報量に基づく年代推定法・場所推定法を，第 8 章ではナイーブベイズ多変数ベルヌーイモデルによる年代推定法・場所推定法を説明した。第 9 章では，年代推定・場所推定の実験結果を示した。

本論文の貢献は，以下の四点である。

一つ目は，作成年代と作成場所を同時に推定する方法の提案である。管見の限り，同時推定を提案した研究は存在しない。先行研究では，年代推定・場所推定は個別に行われている。年代推定の研究では，言語の空間的変異を無視している。同様に，場所推定の研究では，言語の時間的変異を無視している。しかし，同年代における空間的変異や同地域における時間的変異が存在するので，言語の変異を扱う際には，時間軸と空間軸を同時に考慮する必要がある。実験により，作成年代と作成場所の個別推定に比べ，同時推定の方が常に予測精度が高くなることを示した。ただし，トレードオフとして，同時推定では計算量が増加する。

二つ目は，時空間カーネル平滑化の応用である。カーネル平滑化とは，カーネル関数を用いて，ある関数からより滑らかな関数を推定する方法である。本研究では，素性の出現頻度に対して，カーネル平滑化を適用した。カーネル平滑化では，関数に関して線形性や S 字カーブなど特殊な性質を仮定する必要がない。カーネル平滑化により，データセットに点在する欠損値補間と頑健な推定が可能になる。時間軸と空間軸の各々においてカーネル平滑化を行う先行研究は存在するが，両者を組み合わせた時空間カーネル平滑化を文書分類のタスクに応用した研究は，管見の限り，存在しない。ただし，上述の利点がある一方で，本研究の実験では，カーネル平滑化の年代推定・場所推定への効果は限定的だった。

三つ目は，言語に依存しない年代推定法・場所推定法の提案である。素性として文字 *n*-gram を用いることで，単語毎に分かち書きされない言語（日本語や中国語など）の文書や，正書法が確立されていない時代の文書もそのまま扱うことができる。素性として単語 *n*-gram を用いる場合は，単語分割やステミングなどの技術開発が必要となる。また文字 *n*-gram は，単語 *n*-gram に比べ，素性数を大幅に削減できるという利点がある。

四つ目は，文献学的特徴に基づく計量的な年代推定法・場所推定法の提案である。スペイン語で書かれた文書の年代推定・場所推定は，スペイン語文献学の大きな目標の一つである。今日まで，記述的研究には大きな蓄積があるが，年代推定・場所推定を正面から扱った研究は存在しない。先行研究により，各年代や地点に特有の言語的特徴は，ある程度，判明している。しかし，どれだけ細かな記述をしても，記述は記述に過ぎない。スペイン語史の記述から年代推定・場所推定という予測に移るには，計量的なアプローチが必要となる。重要性の異なる複数の証拠から総合的に判断するには，専門家の「勘」よりも，計量的手法の方が信頼性・実証性に勝るからである。「塵も積もれば山となる」というように，小さな証拠でも複数集まれば，大きな差異を生むことになる。本研究では，各々の証拠，つまり文献学的特徴の重みをデータから決定した。多くの文書に現れる特徴ほど，大きな重みが与えられる。この重みをプロットすることで，各特徴の年代推移・地理的変異を可視化することができる。これは，年代推定・場所推定の副産物として，スペ

イン語文献学への大きな貢献となる。

10.2 今後の展開

今後の展開として，以下の四点が挙げられる。

一つ目は，社会言語学的属性を考慮することである。本研究では，言語的特徴が文書の作成年代と作成場所のみに依存すると仮定し，文書の内容，作成者や発行機関（王室，教会，裁判所，地方政府，個人），差出人，受取人等の属性は無視した。しかし，これらの社会言語学的要因も言語的特徴の出現頻度に影響していると考えられる。したがって，これらの変数も考慮することが望ましい。ただし，現時点の文書数では多数の要因をコントロールした上で頑健な推定を行うことは困難である。今後，文書数の増加が望まれる。

二つ目は，新たな素性の導入である。たとえば，文字 *n*-gram を拡張した文字列カーネル（Hastie *et al.* 2009：18.5；Ionescu *et al.* 2014）が考えられる。文字列カーネルは文字 *n*-gram より多くの情報を表現できるので，推定精度の改善が期待される。ただし，計算量が増えるというトレードオフがある。文献学的特徴に関しては，語順の情報を用いることなどが考えられる。ただし，これには，品詞やレンマのタグ付けなどの作業が必要となる。

三つ目は，平滑化パラメータを適応的に調整させることである。本研究では，すべての年代，場所，素性でカーネル平滑化パラメータは同一としたが，これらの変数に従って，カーネル平滑化パラメータを調整させることが望ましい。たとえば，文書数の多い年代や場所では，平滑化パラメータを小さくし，局所的な情報を重視して捉え，逆に文書数が少ない年代や場所では，平滑化パラメータを大きくとり，平滑化の効果を大きくすることが考えられる。ただし，モデルの自由度が大きくなる半面，適切なパラメータを求めるための計算量が増えるというトレードオフがある。

四つ目は，素性間の相関を考慮できるような分類器を用いることである。本研究で用いた分類器は，各素性の独立性を仮定している。しかし，実際には，素性間には相関が存在する。素性間の相関は，たとえば，対数線形モデルで扱うことができる。対数線形モデルの特長としては，このほか，事後確率も計算することができること，異なる種類の素性を組み込めることが挙げられる。たとえば，異なるオーダーの文字 *n*-gram や文字の種類や，文書の種類と文献学的特徴を組み込むことができる。ただし，計算量が増えるというトレードオフがある。

参考文献

Abbasi, A., & Hsinchun, C. (2008). Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems, 26*(2), 7:1-7:29.

Alvar, M. (Ed.). (1996). *Manual de dialectología hispánica: El español de España.* Barcelona: Editorial Planeta.

Azofra, M. E. (2009). *Morfosintaxis histórica del español: de la teoría a la práctica.* Madrid: Universidad Nacional de Educación a Distancia.

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: Improving geographical prediction with social and spatial proximity. *Proceeding of the 19th International Conference on World Wide Web*, 61-70.

Binongo, J. N. (2003). Who wrote the 15th Book of Oz? An application of multivariate analysis to authorship attribution. *Chance, 16*(2), 9-17.

Binongo, J. N., & Smith, M. (1999). The application of principal component analysis to stylometry. *Literary and Linguistic Computing, 14*(4), 445-465.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York: Springer.

Brinegar, C. S. (1963). Mark Twain and the Quintus Curtius Snodgrass letters: A statistical test of authorship. *Journal of the American Statistical Association, 58*, 85-96.

Bruster, D., & Smith, G. (2014). A new chronology for Shakespeare's plays. *Digital Scholarship in the Humanities*.

Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating gender on Twitter. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1301-1309.

Burrows, J. (2002). Delta: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing, 17*(3), 267-287.

Bybee, J. (2015). *Language change.* Cambridge: Cambridge University Press.

Chalmers, T. (2000). Beyond DEEDS: A role for personal names? In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 177-189). Suffolk: Boydell & Brewer.

Chambers, N. (2012). Labeling documents with timestamps: Learning from their time expressions. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 98-106.

Chen, S., & Goodman, J. (1998). An empirical study of smoothing techniques for language modeling. *Technical Report TR-10-98,Harvard University*.

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: A content-based approach to geo-locating Twitter users. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 759-768.

Chibnal, M. (2000). Dating the charters of the smaller religious houses in Suffolk in the twelfth and thirteenth centuries. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 51-59). Suffolk: Boydell & Brewer.

Clement, R., & Sharp, D. (2003). Ngram and Bayesian classification of documents for topic and authorship. *Literary and Linguistic Computing, 18*(4), 423-447.

Company Company, C. (2006). *Sintaxis histórica de la lengua española. Primera parte: La frase verbal* (1a ed.). México: Fondo de Cultura Económica y Universidad Nacional Autónoma de México.

Conde Silvestre, J. C. (2007). *Sociolingüística histórica.* Madrid: Gredos.

Cortina Gómez, R. (1977). On dating the Lazarillo. *Hispanic Review, 45*(1), 61-66.

Coseriu, E. (1978). *Sincronía, diacronía e historia* (tercera ed.). Madrid: Gredos.

Daelemans, W. (2013). Explanation in computational stylometry. *Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing, 2*, 451-462 .

Dalli, A., & Wilks, Y. (2006). Automatic dating of documents and temporal text classification. *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, 17-22.

De Jong, F., Rode, H., & Hiemstra, D. (2005). Temporal language models for the disclosure of historical text. *Proceedings of the 16th International Conference of the Association for History and Computing*, 161-168.

Díaz Moreno, R., Rocío, M. S., Ramírez Luengo, J. L., & Sánchez-Prieto Borja, P. (2015). Hacia una cronología evolutiva del español. *Actas del IX Congreso Internacional de Historia de la Lengua Española (Cádiz, 10-14 de septiembre de 2012).*, 435-447.

Diederich, J., Kindermann, J., Leopold, E., & Paass, G. (2003). Authorship attribution with support vector machines. *Applied Intelligence*, *19*(1), 109-123.

Douvier, É. (1995). L'alternance des graphies MP-MB et NP-NB dans les manuscrits médiévaux. *Cahiers de linguistique hispanique médiévale*(20), 235-256.

Eberenz, R. (1991). Castellano antiguo y español moderno: reflexiones sobre la periodización en la historia de la lengua. *Revista de Filología Española*(71), 79-106.

Eberenz, R. (2009). La periodización de la historia morfosintáctica del español: propuestas y aportaciones recientes. *Cahiers d'études hispaniques médiévales*(32), 181-201.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2010). A latent variable model for geographical lexical variation. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1277-1287.

Ellegård, A. (1962). *A statistical method for determining authorship: The Junius letters, 1769-1772 (Gothenburg studies in English 13).* Göteborg: Acta Universitatis Gothoburgensis.

Enrique-Arias, A. (2012). Dos problemas en el uso de corpus diacrónicos del español: perspectiva y comparabilidad. *Scriptum Digital*(1), 85-106.

Escalante, H. J., Solorio, T., & Montes-y-Gómez, M. (2011). Local histograms of character n-grams for authorship attribution. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 288-298.

Evert, S., Proisl, T., Jannidis, F., Pielström, S., Schöch, C., & Vitt, T. (2015). Towards a better understanding of Burrow's Delta in literary authorship attribution. *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, 79-88.

Faigenbaum-Golovin, S., Shaus, A., Sober, B., Levin, D., Na'aman, N., Sass, B., . . . Finkelstein, I. (2016). Algorithmic handwriting analysis of Judah's military correspondence sheds light on composition of biblical texts. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1522200113

Feuerverger, A., Hall, P., Tilahun, G., & Gervers, M. (2005). Distance measures and smoothing methodology for imputing features of documents. *Journal of Computational and Graphical Statistics, 14*(2), 255-262.

Feuerverger, A., Hall, P., Tilahun, G., & Gervers, M. (2008). Using statistical smoothing to date medieval manuscripts. (N. Balakrishnan, E. Pena, & M. J. Silvapulle, Eds.) *Beyond parametrics in interdisciplinary research: Festshrift in honor of professor Pranab K. Sen, 1*, 321-331.

Fiallos, R. (1997). Procedure for dating undated documents using a relational database. In J. Brown, & W. P. Stoneman (Eds.), *A distinct voice: medieval studies in honor of Leonard E. Boyle, O.P.* (pp. 480-504). Notre Dame (Indiana).

Fiallos, R. (2000). An overview of the process of dating undated medieval charters: latest results and future developments. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 37-48). Suffolk: Boydell & Brewer.

Gazeau, V. (2000). Recherches autour de la datation des actes normands aux Xe-XIIe siècles. In M. Gervers (Ed.), *Dating undated*

medieval charters (pp. 61-79). Suffolk: Boydell & Brewe.

Gervers, M. (1997). The dating of medieval English private charters of the twelfth and thirteenth centuries. In J. Brown, & W. P. Stoneman (Eds.), *A distinct voice: medieval studies in honor of Leonard E. Boyle, O.P.* (pp. 455-480). Notre Dame (Indiana).

Gervers, M. (Ed.). (2000a). *Dating undated medieval charters.* Suffolk: Boydell & Brewer.

Gervers, M. (2000b). The DEEDS project and the development of a computerised methodology for dating undated English private charters of the twelfth and thirteenth centuries. In M. Gervers, *Dating undated medieval charters* (pp. 13-35). Suffolk: Boydell & Brewer.

Gervers, M., & Hamonic, N. (2011). Pro Amore Dei: diplomatic evidence of social conflict. In K. Pennington, & M. Harris Eichbauer (Eds.), *Law as profession and practice in medieval Europe : essays in honor of James A. Brundage* (pp. 231-261). Surrey, England: Ashgate.

Gervers, M., & Tilahun, G. (2013). Statistical approaches to the diplomatics of institutional topography. *Presentation at Digital Diplomatics 2013*. Retrieved from http://www.cei.lmu.de/digdipl13/wp-content/uploads/Tilahun.pdf

Global Administrative Areas. (2015). *GADM database (ver. 2.8).* http://www.gadm.org/

González Ollé, F. (1996). Navarro. In M. Alvar (Ed.), *Manual de dialectología hispánica: El español de España* (pp. 305-316). Barcelona: Editorial Planeta.

Granvik, A., & Sánchez Lancis, C. (2015). Un acercamiento cuantitativo a la periodización en la historia del español. *Libro de resúmenes del X Congreso Internacional de Historia de la Lengua Española (Zaragoza, 7-11 de septiembre de 2015)*, 118-119.

Grieve, J. (2016). *Regional variation in written American English.* Cambridge: Cambridge University Press.

Grupo de Investigación de Textos para la Historia del Español (GITHE). (2010-). *Corpus de Documentos Españoles Anteriores a 1700 (CODEA).* (P. Sánchez-Prieto Borja, Ed.) http://demos.bitext.com/codea

Han, B., Cook, P., & Baldwin, T. (2014). Text-based Twitter user geolocation prediction. *Journal of Artificial Intelligence Research, 49*, 451-500.

Harvey, P. (2000). Seals and the dating of documents. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 207-210). Suffolk: Boydell & Brewer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Second ed.). Springer New York.

He, S., & Schomaker, L. (2014). Delta-n Hinge: rotation-invariant features for writer identification. *Proceedings of the International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, 2014*.

He, S., Samara, P., Burgers, J., & Schomaker, L. (2014). Towards style-based dating of historical documents. *Proceedings of the International Conference on Frontiers in Handwriting Recognition(ICFHR) (Crete, Greece, 2014).*

Hillebrandt, M. (2000). Social groups as recognition patterns: A means of dating medieval charters. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 163-175). Suffolk: Boydell & Brewer.

Hoover, D. L. (2003a). Another perspective on vocabulary richness. *Computers and the Humanities, 37*, 151-178.

Hoover, D. L. (2003b). Multivariate analysis and the study of style variation. *Literary and Linguistic Computing, 18*(4), 341-360.

Hope, J. (1994). *The authorship of Shakespeare's plays.* Cambridge: Cambridge University Press.

Hulden, M., Silfverberg, M., & Francom, J. (2015). Kernel density estimation for text-based geolocation. *Proceedings of the 29th AAAI Conference on Artificial Intelligence* , 145-150.

Ionescu, R. T., Popescu, M., & Cahill, A. (2014). Can characters reveal your native language? A language-independent approach to native language identification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1363-1373.

Iqbal, F., Hadjidi, R., Fung, B. C., & Debbadi, M. (2008). A novel approach of mining write-prints for authorship attribution in e-mail forensics. *Digital Investigation, 5*, S42-S51.

Jin, M., & Jiang, M. (2012). Text clustering on authorship attribution based on the features of punctuation usage. *Proceedings of the 11th International Conference on Signal Processing (ICSP), 3*, 2175-2178.

Jin, M., & Murakami, M. (1993). Authors' characteristic writing styles as seen through their use of commas. *Behaviormetrika, 20*(1), 63-76.

Jockers, M. L., Witten, D. M., & Criddle, C. S. (2008). Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification. *Literary and Linguistic Computing, 23*(4), 465-491.

Juola, P. (2006). Authorship attribution. *Foundations and Trends in Information Retrieval, 1*(3), 233-334.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing (Second edition).* Pearson Education.

Jurgens, D., Finethy, T., McCorriston, J., Xu, Y. T., & Ruths, D. (2015). Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice. *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media*, 188–197.

Kabatek, J. (Ed.). (2008). *Sintaxis histórica del español y cambio lingüístico: Nuevas perspectivas desde las Tradiciones Discursivas.* Madrid/Frankfurt: Iberoamericana/Vervuert.

Kanhabua, N., & Nørvåg, K. (2008). Improving temporal language models for determining time of non-timestamped documents. *Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries 2008 (ECDL'2008)*, 358-370.

Kanhabua, N., & Nørvåg, K. (2009). Using temporal language models for document dating. *Proceedings of ECML PKDD'2009*.

Kanhabua, N., & Nørvåg, K. (2010). Determining time of queries for re-ranking search results. *Proceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries 2008 (ECDL'2010)*, 261-272.

Kawasaki, Y. (2013a). Apuntes para la datación de documentos dialectales: los casos de -nt~-nd, -t~-d final. *Hispánica*(57), 111-117.

Kawasaki, Y. (2013b). Datación de documentos medievales castellanos según la 'coocurrencia' de parámetros. 『ロマンス語研究』*(Studia Romanica), 46*, 57-66.

Kawasaki, Y. (2014a). Datación por coeficientes de asociación. *Actas del Congreso Internacional sobre el Español y la Cultura Hispánica en Japón (Instituto Cervantes, Tokio, 3 de octubre de 2013)*, 265-276.

Kawasaki, Y. (2014b). Datación crono-geográfica de documentos notariales medievales. *Scriptum Digital (Revista de corpus diacrónicos y edición digital en lenguas iberorrománicas), 3*, 29-63.

Kawasaki, Y. (2015b). Datación de documentos castellanos medievales. *Actas del IX Congreso Internacional de Historia de la Lengua Española (Cádiz, 10-14 de septiembre de 2012)*, 477-488.

Kawasaki, Y. (2015c). La determinación cronológica de cambios gráfico-fonéticos y la datación de documentos no fechados en el CODEA. (J. Sánchez Méndez, M. De la Torre, & V. Codita, Eds.) *Temas, Problemas y métodos para la edición y el estudio de documentos hispánicos antiguos*, 491-515.

Kešelj, V., Peng, F., Cercone, N., & Thomas, C. (2003). N-gram-based author profiles for authorship attribution. *Proceedings of the Pacific Association for Computational Linguistics*, 255-264.

Kestemont, M. (2014). Function words in authorship attribution: From black magic to theory? *Proceedings of the 3rd Workshop on Computational Linguistics for Literature*, 59-66.

Kestemont, M., Moens, S., & Deploige, J. (2015). Collaborative authorship in the twelfth century: A stylometric study of Hildegard of Bingen and Guibert of Gembloux. *Digital Scholarship in the Humanities, 30*(2), 199-224.

Kneser, R., & Ney, H. (1995). Improved backing-off for m-gram language modeling. *Proceedings of the IEEE International Conference*

on Acoustics, Speech and Signal Processing, 1, 181-184.

Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing, 17*(4), 401-412.

Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information and Science and Technology, 60*(1), 9-26.

Koppel, M., Schler, J., & Argamon, S. (2011). Authorship attribution in the wild. *Language Resources and Evaluation*, *45*(1), 83-94.

Kotsakos, D., Lappas, T., Kotzias, D., Gunopulos, D., Kanhabua, N., & Nørvåg, K. (2014). A burstiness-aware approach for document dating. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1003-1006.

Kraaij, W. (2004). *Variations on language modeling for information retrieval*. Ph.D. thesis, University of Twente, Netherlands.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79-86.

Kumar, A. (2013). *Supervised language models for temporal resolution of text in absence of explicit temporal cues*. Department of Computer Science. The University of Texas at Austin.

Kumar, A., Baldridge, J., Lease, M., & Ghosh, J. (2012). Dating texts without explicit temporal cues. Retrieved from http://arxiv.org/pdf/1211.2290.pdf

Kumar, A., Lease, M., & Baldridge, J. (2011). Supervised language modeling for temporal resolution of texts. *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM)*, 2069-2072.

Lafferty, J., & Zhai, C. (2001). Document language models, query models, and risk minimization for information retrieval. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 111-119.

Lass, R. (1980). *On explaining language change*. Cambridge: Cambridge University Press.

Li, J., Zheng, R., & Chen, H. (2006). From fingerprint to writeprint. *Communications of the ACM*, *49*(4), 76-82.

Li, X., & Croft, W. B. (2003). Time-based language models. *Proceedings of the 12th ACM Conference on Information and Knowledge Management (CIKM)*.

Lichman, M., & Smyth, P. (2014). Modeling human location data with mixtures of kernel densities. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 35-44.

Love, H. (2002). *Attributing authorship: An introduction*. Cambridge: Cambridge University Press.

Luyckx, K., & Daelemans, W. (2011). The effect of author set size and data size in authorship attribution. *Literary and Linguistic Computing, 26*(1), 35-55.

Maiden, M. (2001). A strange affinity: 'Perfecto y tiempos afines'. *Bulletin of Hispanic Studies, 78*, 441-464.

Malkiel, Y. (1968). Range of variation as a clue to dating (I). *Romance Philology, 21*(4), 463-501.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. New York: Cambridge University Press.

Mei, Q., Liu, C., Su, H., & Zhai, C. (2006). A probabilistic approach to spatiotemporal theme pattern mining on weblogs. *Proceedings of the 15th International Conference on World Wide Web*, 533-542 .

Mendenhall, T. C. (1887). The characteristic curves of composition. *Science, IX*(214), 237–249.

Menéndez Pidal, R. (1908). *Cantar de mío cid: texto, gramática y vocabulario*. Madrid: Bailly-Baillière é hijos.

Menéndez Pidal, R. (1999). *Manual de Gramática Histórica Española* (Vigésima tercera ed.). Madrid: Espasa-Calpe.

Mikawa, K., Ishida, T., & Goto, M. (2011). A proposal of extended cosine measure for distance metric learning in text classification. *Systems, Man, and Cybernetics (SMC)*, 1741-1746.

Mochihashi, D., Kikui, G., & Kita, K. (2004). Learning Nonstructural Distance Metric by Minimum Cluster Distortions. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 341-348.

Moreno Bernal, J., & Horcajada, B. (1997). Sobre no y non en español medieval. *Revista de Filología Románica, I*(14), 345-361.

Morton, A. (1965). The authorship of Greek prose. *Journal of the Royal Statistical Society, 128*(2), 169-233.

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. *Journal of the American Statistical Association, 58*(302), 275-309.

Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. *Advances in Neural Information Processing Systems 14*, 841-848.

Nieuwenhuijsen, D. (2006). Cambios en la colocación de los pronombres átonos. En C. Company Company, *Sintaxis histórica de la lengua española. Primera parte: La frase verbal* (págs. 1337-1404). México: Fondo de Cultura Económica y Universidad Nacional Autónoma de México.

Oda, H., & Ikeda, K. (2010a). Radiocarbon dating of kohitsugire calligraphies attributed to Asukai Masatsune and the periods of origin of Genji Monogatari Emaki and Ban Dainagon Ekotoba. *Radiocarbon, 52*(2), 520-525.

Oda, H., & Ikeda, K. (2010b). Radiocarbon dating of kohitsugire calligraphies attributed to Fujiwara Shunzei: Akihiro-gire, Oie-gire, and Ryosa-gire. *Nuclear instruments and methods in phisics reserach B, 268*, 1041-1044.

Oda, H., Ikeda, K., & Nakamura, T. (2007). Radiocarbon age of the kohitsugire calligraphy and the kiwamefuda certificate. *Nuclear instruments and methods in physics research B, 259*, 374-377.

Oda, H., Ikeda, K., Masuda, T., & Nakamura, T. (2004). Radiocarbon dating of kohitsugire (paper fragments) attributed to Japanese calligraphists in the Heian-Kamakura period. *Radiocarbon, 46*(1), 369-375.

Oda, H., Yasu, H., Ikeda, K., Sakamoto, M., & Yoshizawa, Y. (2011a). Radiocarbon dating of ancient Japanese calligraphy sheets and the discovery of 45 letters of a lost manuscript. *Proceedings in Radiochemistry A Supplement to Radiochimica Act, 1*(1), 331-334.

Omi, T., Ogata, Y., Hirata, Y., & Aihara, K. (2013). Forecasting large aftershocks within one day after the main shock. *Scientific Reports, 3*(2218).

Paredes García, F. (2015). Factores condicionantes de la variación <otro/otri/otre/otrie> en español medieval. In J. P. Sánchez Méndez, M. De la Torre, & V. Codita (Eds.), *Temas, Problemas y métodos para la edición y el estudio de documentos hispánicos antiguos* (pp. 227-260). Valencia: Tirant Humanidades.

Pato, E. (2012). Qual manera quier: la «interposición» en los indefinidos compuestos del español medieval. *Revista de Filología Española, XCII*(2), 273-310.

Pato, E., & O'Neill, P. (2013). Los gerundios 'analógicos' en la historia del español (e iberorromance). *Nueva Revista de Filología Hispánica, 61*(1), 1-27.

Peersman, C., Daelemans, W., & van Vaerenbergh, L. (2011). Predicting age and gender in online social networks. *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents*, 37-44.

Peng, F., Shuurmans, D., Kešelj, V., & Wang, S. (2003). Language independent authorship attribution using character level language models. *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, 267-274.

Penny, R. (2000). *Variation and change in Spanish.* Cambridge: Cambridge University Press.

Penny, R. (2002). *A History of the Spanish Language* (2nd ed.). Cambridge: Cambridge University Press.

Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. *Proceeding of SIGIR'1998.*

Priedhorsky, R., Culotta, A., & Del Valle, S. Y. (2014). Inferring the origin locations of tweets with quantitative confidence. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1523-1536 .

R Core Team. (2015). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

https://www.R-project.org/

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in Twitter. *Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents*, 37-44.

Real Academia Española. (2016). *Diccionario de la lengua española (vigesimotercera edición)*. http://dle.rae.es/?w=diccionario (15/05/2015)

Reynolds, N. B., Schaalje, G. B., & Hilton, J. L. (2012). Who wrote Bacon? Assessing the respective roles of Francis Bacon and his secretaries in the production of his English works. *Literary and Linguistic Computing, 27*(4), 409-425.

Riesco Terrero, Á. (Ed.). (2004). *Introducción a la paleografía y la diplomática general.* Madrid: Editorial Sintesis.

Roller, S., Speriosu, M., Rallapalli, S., Wing, B., & Baldridge, J. (2012). Supervised text-based geolocation using language models on an adaptative grid. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1500-1510.

Romaine, S. (1982). *Socio-historical linguistics.* Cambridge: Cambridge University Press.

Rosenfeld, R. (2000). Two decades of statistical language modeling: where do we go from here? *Proceedings of IEEE, 88*(8).

Rybicki, J. (2015). Vive la différence:Tracing the (authorial) gender signal by multivariate analysis of word frequencies. *Digital Scholarship in the Humanities.*

Sánchez Lancis, C. (2009). Corpus diacrónicos y periodización del español. *Cahiers d'études hispaniques médiévales*(32), 159-180.

Sánchez-Prieto Borja, P. (1998). *Cómo editar los textos medievales: criterios para su presentación gráfica.* Madrid: Arco/Libros.

Sánchez-Prieto Borja, P. (2006). Interpretación fonemática de las grafías medievales. (J. J. Tovar, & J. L. Alconchel, Eds.) *Actas del VI Congreso Internacional de Historia de la Lengua Española: Madrid, 29 de septiembre 3 de octubre de 2003*, 219-260.

Sánchez-Prieto Borja, P. (2012). Desarrollo y explotación del Corpus de Documentos Españoles Anteriores a 1700 (CODEA). *Scriptum Digital*(1), 5-35.

Sánchez-Prieto Borja, P., Díaz Moreno, R., Martínez Sánchez, R., & Vázquez Balonga, D. (2012). El CODEA, un corpus primario de fuentes documentales del español peninsular. *Actas del XVI Congreso Internacional de la ALFAL*, 2629-2638.

Sánchez-Prieto Borja, Pedro (coord.). (2010-). *Corpus Hispánico y Americano en la Red: Textos Antiguos (CHARTA).* http://www.bibliateca.es/charta/index.html

Sapkota, U., Bethard, S., Montes-y-Gómez, M., & Solorio, T. (2015). Not all character n-grams are created equal: A study in authorship attribution. *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 93-102.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*(1), 1-47.

Serdyukov, P., Murdock, V., & van Zwol, R. (2009). Placing Flickr photos on a map. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 484-491.

Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas, 18*(3), 491-504.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology, 60*(3), 538-556.

Stamatatos, E. (2013). On the robustness of authorship attribution based on character n-gram features. *Journal of Law & Policy, 21*(2), 421-439.

Tao, T., Wang, X., Mei, Q., & Zhai, C. (2006). Language model information retrieval with document expansion. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 407-414.

Thinniyam, R. (2014). *On statistical sequencing of document collections.* Department of Statistical Sciences. University of Toronto.

Tilahun, G. (2011). *Statistical methods for dating collections of historical documents.* Doctoral thesis, University of Toronto, Graduate Department of Statistics.

Tilahun, G., Feuerverger, A., & Gervers, M. (2012). Dating medieval English charters. *The Annnals of Applied Statistics, 6*(4), 1615-1640.

Tock, B.-M. (2000). L'étude du vocabulaire et la datation des actes: l'apport des bases de données informatisées. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 81-96). Suffolk: Boydell & Brewer.

Torrens Álvarez, M. J. (1995). La paleografía como instrumento de datación. La escritura denominada «littera textualis». *Cahiers de linguistique hispanique médiévale*(20), 345-380.

Torrens Álvarez, M. J. (1998). ¿Ensordecimiento de las consonantes finales? El caso de -t y -d. (C. García Turza, F. González Bachiller, & J. J. Mangado Martínez, Edits.) *Actas del IV Congreso Internacional de Historia de la Lengua Española (La Rioja, 1-5 de abril 1997)*, 303-317.

Tuten, D. N. (2003). *Koineization in medieval Spanish.* Berlin-New York: Mouton de Gruyter.

Ueda, H. (2011). Razones de 'gelo' medieval y 'selo' moderno: Un estudio filológico y enseñanza-aprendizaje de ELE. *Actas del VII Congreso Internacional de la Asociación Asiática de Hispanistas (Universidad de Estudios Extranjeros de Beijing, 26-28 de agosto de 2010)*, 60-69.

Ueda, H. (2013a). La función de la tilde en la grafía abreviada n<n> del español medieval: Evidencias en los documentos notariales castellanos del siglo XIII al XV. *Cuadernos del Instituto Historia de la Lengua*(8), 343-360.

Ueda, H. (2013b). Pautas y frecuencias grafotácticas de formas abreviadas: Su utilización para la datación de los documentos notariales del siglo XIII al XVII. *Comunicación oral en el III Congreso Internacional Tradición e Innovación: nuevas perspectivas para la edición, la investigación y el estudio de documentos antiguos (5-7 de junio de 2013, Salamanca).*

Ueda, H. (2013c). Una nota sobre el método de taxonomía cuantitativa de grandes datos: Coeficientes de asociación aplicados a los variantes del Diccionario de americanismos. *Dialectologia Special issue, IV*, 221-235.

Ueda, H. (2015). La vocal débil en la apócope extrema medieval: Observaciones sobre el «Corpus de Documentos Españoles Anteriores a 1700». In J. P. Sánchez Méndez, M. De La Torre, & V. Codita, *Temas, problemas y métodos para la edición y el estudio de documentos hispánicos antiguos* (pp. 585-607). Valencia: Tirant Humanidades.

van de Velden, M., Groenen, P. J., & Poblome, J. (2009). Seriation by constrained correspondence analysis: A simulation study. *Computational Statistics and Data Analysis*(53), 3129-3138.

Veszprémy, L. (2000). On the border of book and charter paleography: The dating of some Hungarian manuscripts from the eleventh to the thirteenth century. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 193-205). Suffolk: Boydell & Brewer.

Vincent, N. (2000). The charters of King Henry II: The introduction of the royal inspeximus revisited. In M. Gervers (Ed.), *Dating undated medieval charters* (pp. 97-120). Suffolk: Boydell & Brewer.

Wake, W. C. (1957). Sentence-length distribution of Greek authors. *Journal of the Royal Statistical Society, 120*(3), 331-346.

Wing, B., & Baldridge, J. (2011). Simple supervised document geolocation with geodesic grids. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 955-964.

Wing, B., & Baldridge, J. (2014). Hierarchical discriminative classification for text-based geolocation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 336-348.

Yule, G. U. (1939). On sentence-length as a statistical characteristics of style in prose: With application to two cases of disputed authorship. *Biometrika, 30*(3), 363-390.

Zamora Vicente, A. (1967). *Dialectología española(2a edición muy aumentada).* Madrid: Gredos.

Zhai, C., & Lafferty, J. (2004). A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems, 22*(2), 179-214.

Zhao, Y., & Zobel, J. (2007). Searching with style: Authorship attribution in classic literature. *Proceedings of the 30th Australasian Conference on Computer Science, 62*, 59-68.

Zheng, R., Li, J., Huang, Z., & Chen, H. (2006). A framework of authorship identification for online messages: writing style features and classification techniques. *The Journal of the American Society for Information Science and Technology(JASIST), 57*(3), 378-393.

五神真. (2016). 平成 28 年度東京大学大学院入学式　総長式辞. http://www.u-tokyo.ac.jp/gen01/b_message28_02_j.html (2016 年 4 月 12 日)

エイデンエレツ, ミシェルジャン＝バティースト. (2016). カルチャロミクス : 文化をビッグデータで計測する. (阪本芳久, 訳) 東京: 草思社.

小田寛貴. (2011). 歴史時代資料の 14C 年代測定 : 古文書・古筆切の測定を中心に. 著: 中條利一郎, 酒井英男, 石田肇 (共同編集). 京都: 臨川書店.

川崎義史. (2014). 中世スペイン語公証文書における従属節の否定語（主格代名詞）と斜格代名詞の語順の通時的変異 : que lo non mandó ~ que non lo mandó. 日本ロマンス語学会第 52 回大会（口頭発表）.

川崎義史. (2015a). 機械学習による中世スペイン語古文書の作成年代推定法. 言語処理学会第 21 回年次大会発表論文集, 333-336.

川崎義史. (2016). 中近世スペイン語古文書の統計的年代推定・場所推定. 言語処理学会第 22 回年次大会発表論文集, 1153-1156.

北研二. (1999). 確率的言語モデル. 東京: 東京大学出版会.

金明哲. (2009). テキストデータの統計科学入門. 東京: 岩波書店.

金明哲. (2013). 文節パターンに基づいた文章の書き手の識別. 行動計量学, 40(1), 17-28.

金明哲, 村上征勝. (2007). ランダムフォレスト法による文章の書き手の同定. 統計数理, 55(2), 255-268.

財津, 亘., & 金, 明. (2015). テキストマイニングを用いた犯罪に関わる文書の筆者識別. *法科学技術, 20*(1), 1-14.

杉本智彦. (2016). カシミール 3D マニュアルページ ヒュベニの距離計算式. http://www.kashmir3d.com/ (2016 年 5 月 26 日)

瀬谷創, 堤盛人. (2014). 空間統計学 : 自然科学から人文・社会科学まで. 東京: 朝倉書店.

高村大也. (2010). 言語処理のための機械学習入門（Introduction to machine learning for natural language processing）（奥村学監修）. 東京: コロナ社.

古谷知之. (2011). R による空間データの統計分析. 東京: 朝倉書店.

水谷智洋 (編). (2009). 改訂版　羅和辞典. 東京: 研究社.

村上征勝. (1994). 真贋の科学　計量文献学入門. 東京: 朝倉書店.

村上征勝. (2002). 文化を計る　文化計量学序説. 東京: 朝倉書店.

村上征勝. (2004). シェークスピアは誰ですか？計量文献学の世界. 東京: 文藝春秋.

やまだらけ. (2015 年 5 月 26 日). 日本は山だらけ〜. http://yamadarake.jp/trdi/report000001.html

付録A数学的基礎

A1　記号

$\propto$　：$A \propto B$は，AがBに比例することを意味する。

$\hat{x}$　：$\hat{x}$は，xの推定値であることを意味する。

$\equiv$　：$A \equiv B$は，AをBと定義するという意味である。

$\mathbb{R}$　：$\mathbb{R}$は，実数全体の集合を表している。$\mathbb{R}^n$は，n次元の実数の集合を表している。

$(a, b]$：$(a, b]$は，aより大きくb以下の実数の集合（半閉区間）を表している（ただし，$a < b$）。

$\|a - b\|$：$\|a - b\|$は，点aと点bのユークリッド距離（ノルム）を表している。

A2　集合

n個の値$x_1, \ldots, x_i, \ldots x_n$を要素（元）とする集合$X$を，

$$X = \{x_1, \ldots, x_i, \ldots x_n\} \tag{A2.1}$$

と表す。各要素は，実数，記号，文字列，文書，集合などである。要素は小文字で，集合は大文字で表記するのが一般的である。

$x \in X$は，要素xが集合Xに属することを，$x \notin X$は，要素xが集合Xに属さないことを表している。$|X|$は集合Xの要素数を表している。たとえば，$X = \{1, 2, 3\}$という集合が与えられとき，$1 \in X$，$2 \in X$，$3 \in X$，$4 \notin X$，$|X| = 3$である。

$\subseteq$や$\supseteq$は，部分集合を表す記号である。たとえば，集合Yが集合Zの部分集合である（集合Zが集合Yを包含する）ことを，$Y \subseteq Z$，$Z \supseteq Y$のように表す。

$\cup$（カップ）は，和集合を表す記号である。たとえば，$x_i \in y$であるようなx_iの集合の和をX'とすることを，

$$X' = \bigcup_{x_i \in y} x_i \tag{A2.2}$$

と表す。

A3　総和記号・総乗記号

シグマ記号Σは総和を表している：

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \tag{A3.1}$$

また，集合$X = \{x_1, x_2, \ldots, x_{|X|}\}$のすべての要素に関する関数$f(x)$の和は，$\sum_{x \in X} f(x)$もしくは$\sum_x f(x)$のように表す。

パイ記号Πは総乗を表している：

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times \cdots \times x_n \tag{A3.2}$$

また，集合$X = \{x_1, x_2, \ldots, x_{|X|}\}$のすべての要素に関する関数$f(x)$の積は，$\prod_{x \in X} f(x)$もしくは$\prod_x f(x)$のように表す。

たとえば，集合$X = \{2, 4, 6\}$のすべての要素に関する総和は12，総乗は48となる。

$$\sum_{x \in X} x = \sum_{x} x = \sum_{i=1}^{3} x_i = x_1 + x_2 + x_3 = 2 + 4 + 6 = 12$$

$$\prod_{x \in X} x = \prod_{x} x = \prod_{i=1}^{3} x_i = x_1 \times x_2 \times x_3 = 2 \times 4 \times 6 = 48$$

A4　確率

1から6の各目が等確率（つまり1/6）で出るサイコロを考える。このサイコロを一回投げた時，偶数の目が出る事象を事象A，事象Aの起きる確率を$P(A)$とする。すべての目の出方の場合の数は6通り，偶数の目の出る場合の数は2，4，6の3通りである。したがって，$P(A) = 3/6$である：

$$P(A) = \frac{1 + 1 + 1}{6} = \frac{3}{6}$$

また，このサイコロを一回投げた時，3以上の目が出る事象を事象B，事象Bの起きる確率を$P(B)$とする。すべての目の出方の場合の数は6通り，3以上の目の出る場合の数は3，4，5，6の4通りである。したがって，$P(B) = 4/6$である：

$$P(B) = \frac{1 + 1 + 1 + 1}{6} = \frac{4}{6}$$

このサイコロを一回投げた時，偶数でかつ3以上の目が出る，つまり$A \cap B$（AかつB）となる場合の数は，4か6が出る2通りである。したがって，事象$A \cap B$が起きる確率$P(A \cap B) = 2/6$である：

$$P(A \cap B) = \frac{1 + 1}{6} = \frac{2}{6}$$

$P(\mathrm{A} \cap \mathrm{B})$は$P(\mathrm{A}, \mathrm{B})$と書かれることも多い。本論文では，後者の表記を用いる。$P(A, B)$は事象Aと事象Bの同時確率（joint probability）と呼ばれる。また，$A \cap B$と$B \cap A$は同一の事象を表しているので，$P(A, B) = P(B, A)$となる。

さて，このサイコロを一回投げ偶数の目が出たとき，その目が3以上である，つまり事象Aが起きたという条件の下で事象Bの起こる確率$P(B|A)$を考える。$P(B|A)$は事象Aが起きたという条件の下で事象Bの起きる条件付き確率（conditional probability）と呼ばれる。定義より$P(B|A)$は，

$$P(B|A) = \frac{P(A,B)}{P(A)} \tag{A4.1}$$

となる。したがって，$P(B|A)$は以下のように計算される：

$$P(B|A) = \frac{P(A,B)}{P(A)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$$

次に，このサイコロを一回投げ 3 以上の目が出たとき，その目が偶数である，つまり事象Bが起きたという条件の下で事象Aの起こる確率$P(A|B)$を考える。$P(A|B)$は事象Bが起きたという条件での事象Aの起きる条件付き確率である。定義より$P(B|A)$は,

$$P(A|B) = \frac{P(A,B)}{P(B)} \tag{A4.2}$$

と表すことができる。したがって，$P(A|B)$は以下のように計算される：

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{\frac{2}{6}}{\frac{4}{6}} = \frac{2}{4}$$

ここで同時確率$P(A,B)$に関して，式（A4.1）と式（A4.2）を整理すると，

$$P(A,B) = P(A)P(B|A) = P(B)P(A|B) \tag{A4.3}$$

となる。上式より条件付き確率$P(A|B)$は,

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} \tag{A4.4}$$

と表すことができる。上式は，ベイズの定理（Bayes' theorem）と呼ばれる。ベイズの定理を用いると，$P(A|B)$は以下のように計算される：

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{\frac{3}{6} \times \frac{2}{3}}{\frac{4}{6}} = \frac{2}{4}$$

同様に，式（A4.3）より条件付き確率$P(B|A)$は,

$$P(B|A) = \frac{P(A,B)}{P(A)} = \frac{P(B)P(A|B)}{P(A)} \tag{A4.5}$$

と表すことができる。上式より，$P(B|A)$は以下のように計算される：

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{\frac{4}{6}\times\frac{2}{4}}{\frac{3}{6}} = \frac{2}{3}$$

A5　指数関数・対数関数

まず，指数関数について説明する。$a \in \mathbb{R}$，$x \in \mathbb{R}$とする。このとき，

$$y = a^x \tag{A5.1}$$

を，aを底（てい）とするxの指数関数とよぶ。上式は，aのx乗がyになることを表している。xをaを底としたときの指数とよぶ。指数関数は，$a > 1$のとき単調増加，$a < 1$のとき単調減少する。$b \in \mathbb{R}$，$c \in \mathbb{R}$のとき，指数には以下の性質がある：

$$a^0 = 1 \tag{A5.2}$$
$$a^b a^c = a^{b+c} \tag{A5.3}$$
$$(a^b)^c = a^{bc} \tag{A5.4}$$

また，$n > 0$のとき，

$$a^{\frac{1}{n}} = \sqrt[n]{a} \tag{A5.5}$$
$$a^{-n} = \frac{1}{a^n} \tag{A5.6}$$

となる。

次に，対数関数について説明する。$a > 0$，$a \neq 0$，$x > 0$とする。このとき，

$$y = \log_a x \tag{A5.7}$$

は，底（てい）をaとしたときのxの対数関数とよばれる。上式は，aのy乗がxになることを表している。対数とは，aを何乗するとxになるかを表す値（y）である。対数関数は，$a > 1$のとき単調増加，$a < 1$のとき単調減少する。$b > 0$，$c > 0$のとき，対数には以下の性質がある：

$$\log_a a = 1 \tag{A5.8}$$

$$\log_a 1 = 0 \tag{A5.9}$$

$$\log_a bc = \log b + \log c \tag{A5.10}$$

$$\log_a \frac{b}{c} = \log_a b - \log_a c \tag{A5.11}$$

$$\log_a b^c = c \log_a b \tag{A5.12}$$

また，対数関数の微分は,

$$(\log_a x)' = \frac{1}{x \log_e a} \tag{A5.13}$$

となる。ここで，eは,

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = 2.71828 \ldots \tag{A5.14}$$

で定義される値（自然数もしくはネイピア数と呼ばれる）である。eを底とした対数関数は自然対数（natural logarithm）とよばれ，$\ln x$（$= \log_e x$）とも表記される。ln は natural logarithm の意味である。eを底としたとき，式（A5.13）は,

$$(\ln x)' = \frac{1}{x} \tag{A5.15}$$

となる。

eのx乗は，e^xのほか,

$$\exp(x) = e^x \tag{A5.16}$$

とも表記される。exp は exponential（指数）の意である。指数関数の微分は,

$$(a^x)' = a^x \ln a \tag{A5.17}$$

となる。eを底としたとき，式（A5.17）は,

$$(e^x)' = e^x \tag{A5.18}$$

となる。

底が同一のとき，指数関数と対数関数は互いに逆関数なので,

$$\exp(\ln x) = \ln(\exp x) = x \tag{A5.19}$$

が成り立つ。

A6　微分・偏微分

微分（differentiation）とは，変数を少しだけ動かしたときの，関数の値の変化量を求める操作である。関数$f(x)$の変数xによる微分$f'(x)$は，記号dを用いて，

$$f'(x) = \frac{df(x)}{dx} \tag{A6.1}$$

と表され，

$$f'(x) = \frac{df(x)}{dx} = \lim_{h\to 0}\frac{f(x+h)-f(x)}{(x+h)-x} \tag{A6.2}$$

と定義される。たとえば，aをxに無関係な定数だとすると，変数xによる$f(x) = ax^n$の微分は，

$$\begin{aligned}
&f'(x)\\
&= \lim_{h\to 0}\frac{f(x+h)-f(x)}{(x+h)-x}\\
&= \lim_{h\to 0}\frac{a(x+h)^n - ax^n}{(x+h)-x}\\
&= \lim_{h\to 0}\frac{a\left(\sum_{k=0}^{n}\binom{n}{k}x^{n-k}h^k\right) - ax^n}{(x+h)-x}\\
&= \lim_{h\to 0}\frac{a\left(\binom{n}{0}x^nh^0 + a\binom{n}{1}x^{n-1}h^1 + a\binom{n}{2}x^{n-2}h^2 + \cdots + a\binom{n}{n-1}x^1h^{n-1} + a\binom{n}{n}x^0h^n\right) - ax^n}{h}\\
&= \lim_{h\to 0}\frac{\left(ax^n + anx^{n-1}h + a\frac{n(n-1)}{2}x^{n-2}h^2 + \cdots + anxh^{n-1} + ah^n\right) - ax^n}{h}\\
&= \lim_{h\to 0}\frac{anx^{n-1}h + a\frac{n(n-1)}{2}x^{n-2}h^2 + \cdots + anxh^{n-1} + ah^n}{h}\\
&= \lim_{h\to 0}\left(anx^{n-1} + a\frac{n(n-1)}{2}x^{n-2}h \ldots + anxh^{n-2} + ah^{n-1}\right)\\
&= anx^{n-1} + \lim_{h\to 0}\left(a\frac{n(n-1)}{2}x^{n-2}h + \cdots + anxh^{n-2} + ah^{n-1}\right)\\
&= anx^{n-1}
\end{aligned} \tag{A6.3}$$

となる。ここで，$\binom{n}{k}$は二項係数である：

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \tag{A6.4}$$

微分には線形性がある。aとbをxに無関係な定数，$f(x)$と$g(x)$をxの関数とすると，

$$\big(af(x)+bg(x)\big)' = a\frac{df(x)}{dx}+b\frac{dg(x)}{dx} = af'(x)+bg'(x) \tag{A6.5}$$

が成り立つ。和の微分は，微分の和に等しい。

たとえば，$f(x)$=x^3+x^2+x+5を，xで微分すると，

$$f'(x) = \frac{df(x)}{dx} = \frac{dx^3}{dx}+\frac{dx^2}{dx}+\frac{dx}{dx}+\frac{d5}{dx} = 3x^2+2x+1+0 = 3x^2+2x+1 \tag{A6.6}$$

となる。

偏微分（partial differentiation）とは，多変数関数を，特定の変数以外は定数だとみなして微分したものである。n個の変数$x_1, x_2, \dots, x_i, \dots, x_n$を持つ多変数関数$f(x_1, x_2, \dots, x_i, \dots, x_n)$が与えられたとする。$i$番目の変数$x_i$による$f(x_1, x_2, \dots, x_i, \dots, x_n)$の偏微分は，記号$\partial$（ラウンドディー）を用いて，

$$\frac{\partial f(x_1, x_2, \dots, x_i, \dots, x_n)}{\partial x_i} \tag{A6.7}$$

と表され，

$$\frac{\partial f(x_1, x_2, \dots, x_i, \dots, x_n)}{\partial x_i} = \lim_{h\to 0}\frac{f(x_1, x_2, \dots, x_i+h, \dots, x_n)-f(x_1, x_2, \dots, x_i, \dots, x_n)}{(x_i+h)-x_i} \tag{A6.8}$$

と定義される。

たとえば，二変数関数$f(x_1, x_2)$=${x_1}^3+{x_2}^2+x_1x_2+x_1$を，$x_1$で偏微分すると，以下のようになる：

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{\partial {x_1}^3}{\partial x_1}+\frac{\partial {x_2}^2}{\partial x_1}+\frac{\partial x_1x_2}{\partial x_1}+\frac{\partial x_1}{\partial x_1} = 3{x_1}^2+0+x_2+1 = 3{x_1}^2+x_2+1 \tag{A6.9}$$

同様に，$f(x_1, x_2)$=${x_1}^3+{x_2}^2+x_1x_2+x_1$を，$x_2$で偏微分すると，以下のようになる：

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{\partial {x_1}^3}{\partial x_2}+\frac{\partial {x_2}^2}{\partial x_2}+\frac{\partial x_1x_2}{\partial x_2}+\frac{\partial x_1}{\partial x_2} = 0+2x_2+x_1+0 = 2x_2+x_1 \tag{A6.10}$$

A7　ベルヌーイ分布

ベルヌーイ分布（Bernoulli distribution）は，取り得る値が二つであるような確率変数を記述する分布である（Bishop 2006 : 685 ; 高村 2010 : 1.3.4）。ベルヌーイ分布に従う確率変数Xが確率pでaを，確率$1-p$でbを取るとする。このとき，確率

$P(X=x;p)$は,

$$\begin{aligned} P(X=x;p) &= \delta(x,a)p + \delta(x,b)(1-p) \\ &= \delta(x,a)p + \big(1-\delta(x,a)\big)(1-p) \\ &= p^{\delta(x,a)}(1-p)^{1-\delta(x,a)} \end{aligned} \tag{A7.1}$$

で与えられる。「$;p$」は，pがこの確率変数のパラメータであることを表している。ここで，xは確率変数Xの取る値である。また，$\delta(x,a)$は，$x=a$のときに1を，$x\neq a$のときに0を取る関数である。$x=a$のときに$P(X=a;p)=p^1(1-p)^0=p$を，$x\neq a$つまり$x=b$のときに$P(X=b;p)=p^0(1-p)^1=1-p$となる。

さて，ベルヌーイ分布を多変数に拡張することを考える。互いに独立なm個の確率変数$X_1,X_2,\dots,X_m$が，それぞれ確率$p_1,p_2,\dots,p_m$で1を，確率$(1-p_1),(1-p_2),\dots,(1-p_m)$で0を取るとする。したがって，$X_1,X_2,\dots,X_m$はベルヌーイ分布に従う確率分布である。このとき，$X_1,X_2,\dots,X_m$を要素とする確率変数ベクトル$\boldsymbol{X}=(X_1,X_2,\dots,X_m)$が従う分布は，多変数ベルヌーイ分布（multivariatte Bernoulli distribution）となる。その確率関数は,

$$P(\boldsymbol{X}=\boldsymbol{x};p_1,p_2,\dots,p_m)=\prod_{i=1}^{m}\big(\delta(x_i,1)p_i+\delta(x_i,0)(1-p_i)\big) \tag{A7.2}$$

で与えられ，$p_1,p_2,\dots,p_m$がパラメータある。ここで，$\boldsymbol{x}=(x_1,x_2,\dots,x_m)$は各確率変数の値を要素とするベクトルである。また，$\delta(x_i,a)$は，$x_i=a$のときに1を，$x_i\neq a$のときに0を取る関数である。

A8　ディリクレ分布

ディリクレ分布（Dirichlet distribution）は，$0\leq x_i\leq 1$，$\sum_i x_i=1$を満たすような変数x_iを要素とするベクトル$\boldsymbol{x}=(x_1,x_2,\dots,x_n)$に対して確率を与える連続確率分布である（Bishop 2006：687；高村 2010：1.4.2）。確率密度関数は,

$$\begin{aligned} p(\boldsymbol{x};\boldsymbol{\alpha}) &= \frac{1}{\int_0^1\prod_{i=1}^n x_i^{\alpha_i-1}\,d\boldsymbol{x}}\prod_{i=1}^{n}x_i^{\alpha_i-1} \\ &\propto \prod_{i=1}^{n}x_i^{\alpha_i-1} \end{aligned} \tag{A8.1}$$

で与えられる。$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\dots,\alpha_n)$がパラメータである。分母の$\int_0^1\prod_{i=1}^n x_i^{\alpha_i-1}\,d\boldsymbol{x}$は積分が1になるようにするための正規化定数であり，確率密度は$\prod_{i=1}^n x_i^{\alpha_i-1}$に比例する。ディリクレ分布により，$\boldsymbol{x}=(x_1,x_2,\dots,x_n)$の各要素$x_i$は0や1に近い値を取りにくくなる。

A9　パラメータ推定法

あるデータが与えられたとき，このデータが何らかの確率分布から生成されたと仮定する。しかし，確率分布のパラメータθ（たとえば，正規分布における平均や標準偏差）の値は未知だとする。このとき，与えられたデータからパラメータθの値を推定することを考える。ここでは，最尤推定（maximum likelihood estimation）と最大事後確率推定（maximum a posteriori estimation；以下，MAP 推定）と呼ばれる二つのパラメータの推定法を説明する（高村 2010：1.5）。

独立に同一の確率分布に従うデータ$D=\{x^{(1)},x^{(2)},\dots,x^{(N)}\}$が与えられたとする。このとき，あるパラメータ$\theta$のも

とでのデータDの生成確率$P(D|\theta)$は,

$$P(D|\theta) = \prod_{x^{(i)} \in D} p\left(x^{(i)}|\theta\right) \tag{A9.1}$$

で与えられる。ここで，$p\left(x^{(i)}|\theta\right)$は，仮定した確率分布からサンプル$x^{(i)}$が生成される確率である。サンプル$x^{(i)}$は他のサンプルから独立なので積で表すことができる。また，全サンプルは同一の確率分布から生成されたと仮定しているので同一の確率関数$p(\cdot|\theta)$で表すことができる。$P(D|\theta)$はパラメータθのもとでの尤度（likelihood）と呼ばれる。最尤推定は，尤度$P(D|\theta)$を最大化する，つまり，モデルをデータに最大限フィットさせることでパラメータθを決定する方法である。ただし，尤度$P(D|\theta)$はアンダーフローの問題などがあるため，尤度の対数

$$\begin{aligned} \log P(D|\theta) &= \log \prod_{x^{(i)} \in D} p\left(x^{(i)}|\theta\right) \\ &= \sum_{x^{(i)} \in D} \log p\left(x^{(i)}|\theta\right) \end{aligned} \tag{A9.2}$$

を最大化するのが一般的である。対数関数は単調増加関数なので，尤度の最大化と対数尤度の最大化は等価である。したがって，尤度$P(D|\theta)$の最大化は，次のように変形できる：

$$\begin{aligned} \arg\max_{\theta} P(D|\theta) &= \arg\max_{\theta} \log P(D|\theta) \\ &= \arg\max_{\theta} \sum_{x^{(i)} \in D} \log p\left(x^{(i)}|\theta\right) \end{aligned} \tag{A9.3}$$

次に，MAP 推定を説明する。上述の最尤推定と同一の状況において，値の推定を行いたいパラメータθに関して何らかの確率分布$P(\theta)$が仮定できるとする。$P(\theta)$はパラメータθの事前確率分布（prior distribution）とよばれる。また，データDが与えられた時のパラメータθの確率分布$P(\theta|D)$を事後確率分布（posterior distribution）とよばれる。MAP 推定では，事後確率$P(\theta|D)$の最大化によりパラメータθの値を決定する。事後確率$P(\theta|D)$の最大化は，ベイズの定理を用いて次のように変形できる：

$$\begin{aligned} \arg\max_{\theta} P(\theta|D) &= \arg\max_{\theta} \frac{P(\theta)P(D|\theta)}{P(D)} \\ &= \arg\max_{\theta} P(\theta)P(D|\theta) \\ &= \arg\max_{\theta} \log P(\theta)P(D|\theta) \\ &= \arg\max_{\theta} \log P(\theta) + \log P(D|\theta) \\ &= \arg\max_{\theta} \log P(\theta) + \sum_{x^{(i)} \in D} \log P\left(x^{(i)}|\theta\right) \end{aligned} \tag{A9.4}$$

上式において，$P(\theta)P(D|\theta)$の最大化と$\log P(\theta)P(D|\theta)$の最大化が等価であることを用いた。MAP 推定では，パラメータθを確率変数とみなし，その事前確率分布$P(\theta)$も考慮している。パラメータθに一様分布を仮定すると，MAP 推定は最尤推定と一致する。

A10 KL 情報量

同じ事象空間上で，二つの離散確率分布PとQが与えられたとする（$\sum_x P(x) = \sum_x Q(x) = 1$）。このとき，$P$からみた$Q$の KL 情報量$D_{KL}(P||Q)$は，

$$D_{KL}(P||Q) = \sum_x P(X = x) \log \frac{P(X = x)}{Q(X = x)} \tag{A10.1}$$

と定義される（Kullback & Leibler 1951；Lafferty & Zhai 2001；Manning *et al.* 2008：12.4；高村 2010：1.6）。KL 情報量が小さい（大きい）ほど，二つの確率分布は類似している（異なっている）と見なされる。対数の底には 2 や自然数$e = 2.71828 \ldots$が用いられることが多い。

KL 情報量には，非負性と同一性の性質がある：

$$D_{KL}(P||Q) \geq 0 \tag{A10.2}$$

$$D_{KL}(P||Q) = 0 \Leftrightarrow P = Q \tag{A10.3}$$

しかし，KL 情報量は，JS 情報量と異なり，一般的には対称性を持たない（$D_{KL}(P||Q) \neq D_{KL}(Q||P)$）。

たとえば，下表のように，x_1，x_2，x_3のいずれかの値を取る確率変数Xに関して，三つの確率分布P，Q_1，Q_2が与えられたとする。

	x_1	x_2	x_3
P	0.5	0.3	0.2
Q_1	0.4	0.5	0.1
Q_2	0.6	0.1	0.3

対数の底を自然数$e = 2.71828 \ldots$としたとき，PからみたQ_1の KL 情報量$D_{KL}(P||Q_1)$は，以下のように計算される：

$$\begin{aligned} D_{KL}(P||Q_1) &= \sum_x P(X = x) \log \frac{P(X = x)}{Q_1(X = x)} \\ &= P(X = x_1) \log \frac{P(X = x_1)}{Q_1(X = x_1)} + P(X = x_2) \log \frac{P(X = x_2)}{Q_1(X = x_2)} \\ &\quad + P(X = x_3) \log \frac{P(X = x_3)}{Q_1(X = x_3)} \\ &= 0.5 \log \frac{0.5}{0.4} + 0.3 \log \frac{0.3}{0.5} + 0.2 \log \frac{0.2}{0.1} \\ &\approx 0.097 \end{aligned}$$

同様に，PからみたQ_2の KL 情報量$D_{KL}(P||Q_2)$は，以下のように計算される：

$$
\begin{aligned}
D_{KL}(P||Q_2) &= \sum_x P(X=x)\log\frac{P(X=x)}{Q_2(X=x)} \\
&= P(X=x_1)\log\frac{P(X=x_1)}{Q_2(X=x_1)} + P(X=x_2)\log\frac{P(X=x_2)}{Q_2(X=x_2)} \\
&\quad + P(X=x_3)\log\frac{P(X=x_3)}{Q_2(X=x_3)} \\
&= 0.5\log\frac{0.5}{0.6} + 0.3\log\frac{0.3}{0.1} + 0.2\log\frac{0.2}{0.3} \\
&\approx 0.157
\end{aligned}
$$

$D_{KL}(P||Q_1) < D_{KL}(P||Q_2)$となるので，$P$は$Q_2$より$Q_1$に近いとみなされる。ちなみに，ユークリッド距離を用いた場合，Q_1とQ_2はPから等距離にあるとみなされる。

A11 bag-of-words 表現

bag-of-words 表現（bag-of-words model）とは，単語（素性）の頻度を要素として文書をベクトル表現する方法である（Manning *et al.* 2008：6.3；高村 2010）。ここで単語とは，タイプやレンマと呼ばれる単位に当たる。語彙集合をV，語彙サイズを$|V|$とすると，文書は$|V|$次元のベクトルとして表現される。つまり，各文書は$|V|$次元上の一点として表現されることになる。各単語が一次元を張り，次元間には相関がないと仮定している。ベクトルには，単語の頻度を要素とする頻度ベクトル（frequency vector）と，単語の有無（1 か 0）を要素とする二値ベクトル（binary vector）がある。

単語を素性としたとき，文の構造や語順の情報は無視される。たとえば，$d_1 = \text{Mary loves John}$と$d_2 = \text{John loves Mary}$という二つの文を bag-of-words 表現すると以下のようになる。

$$
\begin{aligned}
d_1 &= \big(n(Mary, d_1), n(loves, d_1), n(John, d_1)\big) = (1,1,1) \\
d_2 &= \big(n(Mary, d_2), n(loves, d_2), n(John, d_2)\big) = (1,1,1)
\end{aligned}
$$

ここで，$n(w,d)$は，文書dにおける単語wの頻度，語彙集合は$V = \{Mary, loves, John\}$とする。語順は考慮されないので，二つの文書は同じベクトルとなる。*n*-gram（n≧2）を用いると部分的に語順の情報を取り込むことができる。単語の代わりに，*n*-gram を素性として用いる場合は，bag-of-*n*-grams 表現と呼ばれることもある。bag-of-words 表現は，多くの情報を捨てている一方，簡潔に文を表現することができるので，自然言語処理において広く用いられている。

A12 交差検定

交差検定（cross-validation）とは，ある評価指標値（正解率や誤差など）の値を推定するために行う実験のことである（Bishop 2006：1.3；Hastie *et al.* 2009：7.10；高村 2010：6.2.1）。まず，データセット D をほぼ同サイズ（文書数）の k 個のサブデータセット（$D_1, D_2, \ldots, D_k$）に分割する。サブデータセットは可能な限り量的・質的に均質になるように分割することが望ましい。分割方法としては文書 ID を k で割った余りを基準としたり，データセットを k 等分するなどの方法が考えられる。たとえば，データセット$D = \{d_1, d_2, \ldots, d_{100}\}$を 10 分割する場合，文書 ID を k = 10 で割った余りを基準にすれば$D_1 = \{d_1, d_{11}, \ldots, d_{91}\}, D_2 = \{d_2, d_{12}, \ldots, d_{92}\}, \ldots, D_{10} = \{d_{10}, d_{20}, \ldots, d_{100}\}$となり，10 等分すれば$D_1 = \{d_1, d_2, \ldots, d_{10}\}, D_2 = \{d_{11}, d_{12}, \ldots, d_{20}\}, \ldots, D_{10} = \{d_{91}, d_{92}, \ldots, d_{100}\}$となる。次に，$D_i$ 以外（つまり，$D_1, D_2, \ldots, D_{i-1}, D_{i+1}, \ldots, D_k$）を用いてモデルを作り（この過程は訓練もしくは学習と呼ばれる），D_iでモデルの評価（テ

スト）を行う。D_iを除いたデータセット（つまり，$D_1, D_2, \ldots, D_{i-1}, D_{i+1}, \ldots, D_k$）は訓練データ，$D_i$はテストデータと呼ばれる。これを i = 1, 2, …, k まで k 回行い，各D_iの評価指標値を算出する。各文書は 1 回ずつテストセットに入り，残りの k-1 回は訓練セットに入ることになる。k 個の評価指標値の平均を，データセットDの評価指標値とする。このような方法は，k 分割交差検定（k-fold cross-validation）と呼ばれる。$k = |D|$のとき，つまり自分自身以外を訓練データとして用いる交差検定は，一個抜き交差検定（leave-one-out cross-validation）と呼ばれる。

評価指標値に影響するパラメータが存在する場合には，パラメータの全組み合わせに対して交差検定を行い，評価指標値が最良となるパラメータの組み合わせを決定する。このような手法は，グリッドサーチ（grid search）と呼ばれる。

付録B距離計算

B1　ヒュベニの公式

ヒュベニの公式（Hubeny's distance formula）は，二地点の緯度・経度から距離を求める式である（やまだらけ 2015；杉本 2016）。地点 1 の緯度・経度を(lon_1, lat_1)，地点 2 の緯度・経度を(lon_2, lat_2)とする。ただし，緯度・経度は 10 進数で表されているとする。このとき，ヒュベニの公式による二地点間の距離D（単位は㎞）は，

$$D = \sqrt{(d_{lon}M)^2 + (d_{lat}N\cos\mu_{lon})^2} \tag{B1.1}$$

で与えられる。ここで

$$d_{lon} = lon_1 - lon_2 \quad \text{緯度の差（ラジアン）} \tag{B1.2}$$

$$d_{lat} = lat_1 - lat_2 \quad \text{経度の差（ラジアン）} \tag{B1.3}$$

$$\mu_{lon} = \frac{lon_1 + lon_2}{2} \quad \text{緯度の平均値（ラジアン）} \tag{B1.4}$$

$$M = \frac{a(1-e^2)}{W^3} \quad \text{子午線曲率半径} \tag{B1.5}$$

$$N = \frac{a}{W} \quad \text{卯酉（ぼうゆう）線曲率半径} \tag{B1.6}$$

$$W = \sqrt{1-(e\sin\mu_{lon})^2} \tag{B1.7}$$

$$e = \sqrt{\frac{a^2-b^2}{a^2}} \quad \text{第一離心率} \tag{B1.8}$$

$$a = 6378.137 \text{（㎞）} \quad \text{長半径（赤道半径）} \tag{B1.9}$$

$$b = 6356.752 \text{（㎞）} \quad \text{短半径（極半径）} \tag{B1.10}$$

である。

ヒュベニの公式を用いて，県間距離と自治州間距離を計算した。各県・各自治州の緯度・経度には，Global Administrative Areas の GADM データベース（ver. 2.8）のデータを用いた。このデータの緯度・経度は，各領域の重心に対応しているようである。イタリアとポルトガルは国を一つの県・自治州とみなし，それぞれローマとリスボンで代表させた。この二都市の緯度・経度は，上記のデータに含まれていないので，Google Maps から取得した。表 B1.1 に各県の緯度・経度を，表 B1.2 に県間距離行列を，表 B1.3 に各県の緯度・経度を，表 B1.4 に自治州間距離行列を示す：

県	経度	緯度	県	経度	緯度
A Coruña	-8.46421	43.12600	La Rioja	-2.51718	42.27508
Álava	-3.03137	43.00949	Las Palmas	-14.03649	28.40611
Albacete	-1.98033	38.82552	León	-5.83937	42.61967
Alicante	-0.56829	38.47881	Lleida	1.04775	42.04374
Almería	-2.34477	37.19601	Lugo	-7.44599	43.01168
Asturias	-5.99299	43.29231	Madrid	-3.71606	40.49496
Ávila	-4.94536	40.57105	Málaga	-4.72577	36.81398
Badajoz	-6.14164	38.70973	Murcia	-1.48522	38.00207
Baleares	2.95661	39.61345	Navarra	-1.64766	42.66755
Barcelona	1.98367	41.73112	Ourense	-7.59231	42.19637
Burgos	-3.60380	42.36142	Palencia	-4.53780	42.37071
Cáceres	-6.16048	39.71189	Pontevedra	-8.45960	42.43538
Cádiz	-5.75992	36.55419	Portugal	-9.13934	38.72232
Cantabria	-4.03440	43.19633	Salamanca	-6.06530	40.80494
Castellón	-0.14669	40.24142	Santa Cruz de Tenerife	-16.58348	28.27241
Ciudad Real	-3.81594	38.91935	Segovia	-4.05429	41.17107
Córdoba	-4.80903	37.99286	Sevilla	-5.68231	37.43580
Cuenca	-2.19551	39.89610	Soria	-2.58867	41.62076
Girona	2.67559	42.12694	Tarragona	0.81794	41.08762
Granada	-3.26777	37.31251	Teruel	-0.81546	40.66130
Guadalajara	-2.62364	40.81351	Toledo	-4.14794	39.79377
Guipúzcoa	-2.19458	43.14358	Valencia	-0.78432	39.34497
Huelva	-6.82820	37.57948	Valladolid	-4.84353	41.63149
Huesca	-0.07303	42.20285	Vizcaya	-2.85361	43.24141
Italia	12.49634	41.90276	Zamora	-5.98045	41.72715
Jaén	-3.44154	38.01655	Zaragoza	-1.06468	41.62174

表 B1.1　各県の緯度・経度

付録 B

	A Coruña	Álava	Albacete	Alicante	Almería	Asturias	Ávila	Badajoz	Baleares	Barcelona	Burgos	Cáceres	Cádiz	Cantabria	Castellón	Ciudad Real	Córdoba	Cuenca	Girona	Granada	Guadalajara	Guipúzcoa	Huelva	Huesca	Italia	Jaén	La Rioja	Las Palmas	León	Lleida	Lugo	Madrid	Málaga	Murcia	Navarra	Ourense	Palencia	Pontevedra	Portugal	Salamanca	Santa Cruz de Tenerife	Segovia	Sevilla	Soria	Tarragona	Teruel	Toledo	Valencia	Valladolid	Vizcaya	Zamora	Zaragoza
A Coruña	0	451	748	867	870	205	421	552	1055	890	415	444	800	367	781	632	676	652	937	813	562	519	662	708	1759	736	506	1773	226	805	85	506	803	847	569	131	339	81	517	339	1868	436	705	523	817	707	535	790	350	465	265	645
Álava	451	0	496	569	680	247	328	568	646	446	90	468	788	86	405	482	604	371	488	664	259	71	713	265	1306	583	96	1964	238	359	366	300	736	599	122	393	146	457	724	363	2106	232	689	167	394	332	387	469	222	31	289	231
Albacete	748	496	0	131	192	624	329	367	442	482	435	381	428	539	230	162	269	126	556	210	238	504	454	427	1300	160	405	1647	553	457	676	246	338	105	449	623	468	698	632	423	1831	327	367	330	358	236	221	121	410	521	483	335
Alicante	867	569	131	0	217	729	453	494	337	439	523	512	516	625	208	291	381	218	509	276	325	562	567	436	1199	261	473	1715	662	438	792	361	419	98	497	745	568	822	758	550	1909	436	472	406	327	255	349	102	522	589	602	368
Almería	870	680	192	217	0	777	454	381	547	649	611	447	317	714	403	239	239	314	722	84	421	693	405	615	1406	137	591	1497	701	637	809	402	219	121	640	740	632	809	631	532	1694	486	301	515	531	424	341	285	560	706	615	527
Asturias	205	247	624	729	777	0	330	535	873	692	227	418	785	162	610	543	626	511	736	735	406	315	670	510	1553	653	313	1869	80	605	125	379	762	728	368	185	162	228	597	291	1986	297	683	346	628	534	438	640	217	260	183	457
Ávila	421	328	329	453	454	330	0	240	694	607	238	145	472	316	416	216	300	250	674	406	201	380	385	456	1495	325	287	1635	251	539	355	106	437	426	372	294	213	369	424	100	1782	104	370	236	498	355	114	388	124	359	161	353
Badajoz	552	568	367	494	381	535	240	0	805	786	478	116	252	554	553	206	145	372	858	303	392	618	145	663	1652	253	520	1397	457	732	514	296	254	421	602	425	449	478	265	244	1549	339	153	458	665	517	215	477	358	599	352	555
Baleares	1055	646	442	337	547	873	694	805	0	261	646	795	853	727	279	599	710	450	294	613	503	602	897	398	861	594	562	2045	829	327	971	587	756	434	531	952	717	1030	1066	793	2245	632	805	531	252	349	619	328	711	651	807	418
Barcelona	890	446	482	439	649	692	607	786	261	0	476	738	908	532	252	598	733	417	74	689	407	387	907	182	889	640	385	2132	666	87	804	508	820	528	324	809	554	883	1024	694	2313	517	835	387	124	269	573	366	578	442	674	258
Burgos	415	90	435	523	611	227	238	478	646	476	0	379	701	104	385	402	519	312	528	588	199	149	623	297	1355	506	92	1874	189	393	329	218	653	540	167	335	78	407	638	277	2017	144	601	122	403	309	303	427	135	121	213	231
Cáceres	444	468	381	512	447	418	145	116	795	738	379	0	368	445	526	225	232	346	807	378	332	524	254	597	1620	310	432	1494	341	675	401	231	359	457	517	315	340	374	285	128	1637	249	268	379	623	476	175	471	251	499	236	491
Cádiz	800	788	428	516	317	785	472	252	853	908	701	368	0	788	656	325	187	501	982	242	567	827	153	824	1717	269	723	1221	706	873	766	492	98	419	796	675	685	724	393	494	1400	557	102	651	783	647	402	549	595	818	601	719
Cantabria	367	86	539	625	714	162	316	554	727	532	104	445	788	0	477	499	610	416	574	689	303	152	697	350	1390	606	166	1937	165	445	284	317	746	644	208	319	105	379	681	328	2071	237	686	221	477	402	398	528	195	98	237	310
Castellón	781	405	230	208	403	610	416	553	279	252	385	526	656	477	0	355	486	182	326	438	223	381	665	229	1098	388	312	1878	558	234	698	309	568	286	311	675	449	752	805	514	2060	353	587	264	129	76	351	118	434	419	528	179
Ciudad Real	632	482	162	291	239	543	216	206	599	598	402	225	325	499	355	0	139	182	673	193	244	512	309	500	1449	110	407	1542	465	555	570	184	257	232	475	502	407	572	470	294	1713	263	239	332	475	331	106	271	328	511	377	395
Córdoba	676	604	269	381	239	626	300	145	710	733	519	232	187	610	486	139	0	319	808	159	380	641	187	639	1570	122	537	1403	546	693	628	306	137	296	609	546	510	606	393	345	1575	376	101	464	609	467	217	389	423	635	446	533
Cuenca	652	371	126	218	314	511	250	372	450	417	312	346	501	416	182	182	319	0	491	315	113	379	489	325	1280	244	279	1723	444	374	576	149	422	229	327	534	351	611	623	350	1895	218	420	204	294	149	170	139	304	395	389	223
Girona	937	488	556	509	722	736	674	858	294	74	528	807	982	574	326	673	808	491	0	763	476	424	981	231	828	714	437	2207	716	137	852	577	894	601	368	863	606	935	1093	758	2388	581	909	448	199	343	644	440	638	479	732	321
Granada	813	664	210	276	84	735	406	303	613	689	588	378	242	689	438	193	159	315	763	0	411	686	321	633	1475	83	581	1443	657	667	756	372	144	178	639	682	599	748	548	474	1633	454	217	505	567	445	298	323	521	692	565	537
Guadalajara	562	259	238	325	421	406	201	392	503	407	199	332	567	303	223	244	380	113	476	411	0	275	527	271	1293	333	171	1778	344	344	481	101	500	342	232	452	244	530	617	295	1941	129	476	94	296	156	177	234	212	285	305	163
Guipúzcoa	519	71	504	562	693	315	380	618	602	387	149	524	827	152	381	512	641	379	424	686	275	0	762	208	1237	607	105	2017	309	300	436	335	769	602	72	464	215	528	787	426	2164	279	730	181	349	313	425	459	284	56	358	202
Huelva	662	713	454	567	405	670	385	145	897	907	623	254	153	697	665	309	187	489	981	321	527	762	0	796	1756	307	663	1254	593	860	635	435	209	480	743	541	592	583	245	381	1411	483	104	598	784	638	349	573	502	744	488	688
Huesca	708	265	427	436	615	510	456	663	398	182	297	597	824	350	229	500	639	325	231	633	271	208	796	0	1059	568	205	2045	485	96	623	369	747	504	143	632	375	704	880	534	2212	358	739	223	151	191	448	339	408	262	501	108
Italia	1759	1306	1300	1199	1406	1553	1495	1652	861	889	1355	1620	1717	1390	1098	1449	1570	1280	828	1475	1293	1237	1756	1059	0	1456	1264	2901	1542	965	1674	1392	1619	1296	1190	1692	1434	1763	1905	1585	3111	1407	1667	1276	996	1143	1448	1180	1467	1292	1562	1148
Jaén	736	583	160	261	137	653	325	253	594	640	506	310	269	606	388	110	122	244	714	83	333	607	307	568	1456	0	502	1488	575	609	677	289	181	174	564	605	516	673	512	397	1670	371	211	426	516	384	216	281	438	611	485	468
La Rioja	506	96	405	473	591	313	287	520	562	385	92	432	723	166	312	407	537	279	437	581	171	105	663	205	1264	502	0	1919	281	301	420	232	664	506	86	426	170	499	705	347	2070	183	627	77	315	238	321	373	210	117	299	144
Las Palmas	1773	1964	1647	1715	1497	1869	1635	1397	2045	2132	1874	1494	1221	1937	1878	1542	1403	1723	2207	1443	1778	2017	1254	2045	2901	1488	1919	0	1804	2099	1790	1690	1308	1617	2000	1697	1832	1697	1271	1608	252	1737	1303	1854	2007	1871	1602	1766	1744	1995	1706	1938
León	226	238	553	662	701	80	251	457	829	666	189	341	706	165	558	465	546	444	716	657	344	309	593	485	1542	575	281	1804	0	582	141	307	682	656	350	155	113	220	536	213	1927	227	604	297	590	483	360	577	143	259	105	418
Lleida	805	359	457	438	637	605	539	732	327	87	393	675	873	445	234	555	693	374	137	667	344	300	860	96	965	609	301	2099	582	0	719	444	790	519	237	727	471	800	960	622	2273	444	794	311	113	226	517	352	500	355	595	185
Lugo	85	366	676	792	809	125	355	514	971	804	329	401	766	284	698	570	628	576	852	756	481	436	635	623	1674	677	420	1790	141	719	0	431	759	778	484	96	254	108	521	283	1896	358	667	439	731	624	469	711	272	381	194	559
Madrid	506	300	246	361	402	379	106	296	587	508	218	231	492	317	309	184	306	149	577	372	101	335	435	369	1392	289	232	1690	307	444	431	0	437	350	309	385	230	462	516	205	1847	84	395	163	395	250	90	288	164	329	241	262
Málaga	803	736	338	419	219	762	437	254	756	820	653	359	98	746	568	257	137	422	894	144	500	769	209	747	1619	181	664	1308	682	790	759	437	0	322	732	674	647	731	452	478	1492	510	112	589	696	564	349	457	560	766	582	644
Murcia	847	599	105	98	121	728	426	421	434	528	540	457	419	644	286	232	296	229	601	178	342	602	480	504	1296	174	506	1617	656	519	778	350	322	0	544	720	573	794	684	516	1810	431	381	432	411	315	314	168	514	622	584	423
Navarra	569	122	449	497	640	368	372	602	531	324	167	517	796	208	311	475	609	327	368	639	232	72	743	143	1190	564	86	2000	350	237	484	309	732	544	0	501	244	570	791	433	2154	268	702	146	278	245	397	395	295	121	380	132
Ourense	131	393	623	745	740	185	294	425	952	809	335	315	675	319	675	502	546	534	863	682	452	464	541	632	1692	605	426	1697	155	727	96	385	674	720	501	0	257	78	426	208	1806	323	578	428	724	603	407	672	241	414	147	555
Palencia	339	146	468	568	632	162	213	449	717	554	78	340	685	105	449	407	510	351	606	599	244	215	592	375	1434	516	170	1832	113	471	254	230	647	573	244	257	0	329	581	224	1968	146	583	186	477	374	303	478	90	173	143	306
Pontevedra	81	457	698	822	809	228	369	478	1030	883	407	374	724	379	752	572	606	611	935	748	530	528	583	704	1763	673	499	1697	220	800	108	462	731	794	570	78	329	0	437	278	1796	401	630	504	800	681	480	750	319	476	225	630
Portugual	517	724	632	758	631	597	424	265	1066	1024	638	285	393	681	805	470	393	623	1093	548	617	787	245	880	1905	512	705	1271	536	960	521	516	452	684	791	426	581	437	0	361	1389	526	342	660	908	760	455	738	503	753	444	776
Salamanca	339	363	423	550	532	291	100	244	793	694	277	128	494	328	514	294	345	350	758	474	295	426	381	534	1585	397	347	1608	213	622	283	205	478	516	433	208	224	278	361	0	1742	177	393	311	590	451	203	488	142	393	108	437
Santa Cruz de Tenerife	1868	2106	1831	1909	1694	1986	1782	1549	2245	2313	2017	1637	1400	2071	2060	1713	1575	1895	2388	1633	1941	2164	1411	2212	3111	1670	2070	252	1927	2273	1896	1847	1492	1810	2154	1806	1968	1796	1389	1742	0	1886	1473	2010	2187	2045	1762	1952	1883	2137	1834	2103
Segovia	436	232	327	436	486	297	104	339	632	517	144	249	557	237	353	263	376	218	581	454	129	279	483	358	1407	371	183	1737	227	444	358	84	510	431	268	323	146	401	526	177	1886	0	457	135	416	284	161	354	86	262	176	260
Sevilla	705	689	367	472	301	683	370	153	805	835	601	268	102	686	587	239	101	420	909	217	476	730	104	739	1667	211	627	1303	604	794	667	395	112	381	702	578	583	630	342	393	1473	457	0	557	710	569	305	487	493	719	500	632
Soria	523	167	330	406	515	346	236	458	531	387	122	379	651	221	264	332	464	204	448	505	94	181	598	223	1276	426	77	1854	297	311	439	163	589	432	146	428	186	504	660	311	2010	135	557	0	296	188	252	307	191	191	288	129
Tarragona	817	394	358	327	531	628	498	665	252	124	403	623	783	477	129	475	609	294	199	567	296	349	784	151	996	516	315	2007	590	113	731	395	696	411	278	724	477	800	908	590	2187	416	710	296	0	149	454	246	486	398	583	172
Teruel	707	332	236	255	424	534	355	517	349	269	309	476	647	402	76	331	467	149	343	445	156	313	638	191	1143	384	238	1871	483	226	624	250	564	315	245	603	374	681	760	451	2045	284	569	188	149	0	305	153	362	347	458	114
Toledo	535	387	221	349	341	438	114	215	619	573	303	175	402	398	351	106	217	170	644	298	177	425	349	448	1448	216	321	1602	360	517	469	90	349	314	397	407	303	480	455	203	1762	161	305	252	454	305	0	298	222	417	275	340
Valencia	790	469	121	102	285	640	388	477	328	366	427	471	549	528	118	271	389	139	440	323	234	459	573	339	1180	281	373	1766	577	352	711	288	457	168	395	672	478	750	738	488	1952	354	487	307	246	153	298	0	440	488	527	266
Valladolid	350	222	410	522	560	217	124	358	711	578	135	251	595	195	434	328	423	304	638	521	212	284	502	408	1467	438	210	1744	143	500	272	164	560	514	295	241	90	319	503	142	1883	86	493	191	486	362	222	440	0	252	97	320
Vizcaya	465	31	521	589	706	260	359	599	651	442	121	499	818	98	419	511	635	395	479	692	285	56	744	262	1292	611	117	1995	259	355	381	329	766	622	121	414	173	476	753	393	2137	262	719	191	398	347	417	488	252	0	316	242
Zamora	265	289	483	602	615	183	161	352	807	674	213	236	601	237	528	377	446	389	732	565	305	358	488	501	1562	485	299	1706	105	595	194	241	582	584	380	147	143	225	444	108	1834	176	500	288	583	458	275	527	97	316	0	417
Zaragoza	645	231	335	368	527	457	353	555	418	258	231	491	719	310	179	395	533	223	321	537	163	202	688	108	1148	468	144	1938	418	185	559	262	644	423	132	555	306	630	776	437	2103	260	632	129	172	114	340	266	320	242	417	0

表 B1.2　県間距離行列

自治州	経度	緯度
AN	-4.57355	37.46363
AR	-0.66030	41.51988
AS	-5.99299	43.29231
CB	-4.03440	43.19633
CL	-4.78819	41.75120
CM	-3.00458	39.58108
CN	-16.55631	28.29106
CT	1.52853	41.79800
EX	-6.15069	39.19141
GA	-7.91038	42.75708
IB	2.95661	39.61345
IT	12.49634	41.90276
LR	-2.51718	42.27508
MD	-3.71606	40.49496
MU	-1.48522	38.00207
NA	-1.64766	42.66755
PT	-9.13934	38.72232
PV	-2.62166	43.03215
VC	-0.54301	39.39044

表 B1.3　各自治州の緯度・経度

	AN	AR	AS	CB	CL	CM	CN	CT	EX	GA	IB	IT	LR	MD	MU	NA	PT	PV	VC
AN	0	583	690	670	499	282	1547	733	244	681	712	1575	588	360	283	657	432	670	421
AR	583	0	493	344	351	303	2118	188	546	627	382	1115	180	287	416	158	803	242	248
AS	690	493	0	162	207	502	1983	653	479	171	873	1553	313	379	728	368	597	281	650
CB	670	344	162	0	181	431	2067	493	501	326	727	1390	166	317	644	208	681	119	535
CL	499	351	207	181	0	296	1894	534	321	287	711	1461	201	173	522	285	516	235	457
CM	282	303	502	431	296	0	1814	467	279	559	520	1358	317	123	227	378	548	404	216
CN	1547	2118	1983	2067	1894	1814	0	2283	1588	1848	2242	3107	2067	1844	1807	2151	1386	2127	1969
CT	733	188	653	493	534	467	2283	0	728	800	283	927	345	473	514	286	989	376	332
EX	244	546	479	501	321	279	1588	728	0	442	798	1636	476	260	435	560	269	540	492
GA	681	627	171	326	287	559	1848	800	442	0	997	1714	455	443	784	522	482	441	742
IB	712	382	873	727	711	520	2242	283	798	997	0	861	562	587	434	531	1066	620	307
IT	1575	1115	1553	1390	1461	1358	3107	927	1636	1714	861	0	1264	1392	1296	1190	1905	1272	1159
LR	588	180	313	166	201	317	2067	345	476	455	562	1264	0	232	506	86	705	89	377
MD	360	287	379	317	173	123	1844	473	260	443	587	1392	232	0	350	309	516	311	304
MU	283	416	728	644	522	227	1807	514	435	784	434	1296	506	350	0	544	684	594	181
NA	657	158	368	208	285	378	2151	286	560	522	531	1190	86	309	544	0	791	92	394
PT	432	803	597	681	516	548	1386	989	269	482	1066	1905	705	516	684	791	0	751	759
PV	670	242	281	119	235	404	2127	376	540	441	620	1272	89	311	594	92	751	0	461
VC	421	248	650	535	457	216	1969	332	492	742	307	1159	377	304	181	394	759	461	0

表 B1.4　自治州間距離行列

B2　年代推定・場所推定（県レベル）の結果

表 B2.5 に *n*-gram 言語モデルによる年代推定・場所推定（県）の結果を，表 B2.6 に JS 情報量による年代推定・場所推定（県）の結果を，表 B2.7 にナイーブベイズ多変数ベルヌーイモデルによる年代推定・場所推定（県）の結果を示す。

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（県）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
①LM_Prv_1	個別	0	0	0.001	18.63	31.70	10.00	157.54	268.20	100.98	8503.12
②LM_Prv_2	同時	0	0	0.001	21.83	37.83	11.00	153.83	253.98	102.22	9608.76
③LM_Prv_3	個別	3	25	0.001	17.87	30.41	10.00	167.06	276.16	104.17	8397.98
④LM_Prv_4	同時	3	25	0.001	16.01	28.66	8.00	119.57	212.50	76.75	6089.97
③LM_Prv_5	個別	3	50	0.001	17.87	30.41	10.00	191.34	287.65	141.41	8747.46
④LM_Prv_6	同時	3	50	0.001	13.91	25.16	7.00	143.55	218.59	107.80	5500.78
③LM_Prv_7	個別	3	75	0.001	17.87	30.41	10.00	222.83	317.42	175.46	9652.63
④LM_Prv_8	同時	3	75	0.001	13.91	25.41	7.00	147.98	212.87	116.61	5408.72
③LM_Prv_9	個別	3	100	0.001	17.87	30.41	10.00	244.93	332.76	181.99	10119.14
④LM_Prv_10	同時	3	100	0.001	14.04	25.78	7.00	155.94	213.77	131.80	5510.00
③LM_Prv_11	個別	5	25	0.001	18.67	31.19	11.00	167.06	276.16	104.17	8612.13
④LM_Prv_12	同時	5	25	0.001	16.99	29.05	11.00	117.20	211.14	75.62	6133.04
③LM_Prv_13	個別	5	50	0.001	18.67	31.19	11.00	191.34	287.65	141.41	8970.52
④LM_Prv_14	同時	5	50	0.001	14.86	25.55	9.00	139.61	219.39	105.41	5604.54
③LM_Prv_15	個別	5	75	0.001	18.67	31.19	11.00	222.83	317.42	175.46	9898.77
④LM_Prv_16	同時	5	75	0.001	14.78	26.00	9.00	150.57	223.17	113.83	5802.80
③LM_Prv_17	個別	5	100	0.001	18.67	31.19	11.00	244.93	332.76	181.99	10377.18
④LM_Prv_18	同時	5	100	0.001	14.49	23.98	8.00	158.53	223.89	127.60	5368.40
③LM_Prv_19	個別	10	25	0.001	19.16	31.35	12.00	167.06	276.16	104.17	8656.23
④LM_Prv_20	同時	10	25	0.001	17.46	28.62	12.00	116.33	208.44	75.62	5965.65
③LM_Prv_21	個別	10	50	0.001	19.16	31.35	12.00	191.34	287.65	141.41	9016.46
④LM_Prv_22	同時	10	50	0.001	15.51	25.61	10.00	133.46	216.04	103.58	5532.15
③LM_Prv_23	個別	10	75	0.001	19.16	31.35	12.00	222.83	317.42	175.46	9949.46
④LM_Prv_24	同時	10	75	0.001	15.06	25.68	10.00	145.74	222.32	107.80	5708.82
③LM_Prv_25	個別	10	100	0.001	19.16	31.35	12.00	244.93	332.76	181.99	10430.32
④LM_Prv_26	同時	10	100	0.001	15.11	25.63	10.00	157.47	225.51	124.74	5779.63
①LM_Prv_27	個別	0	0	0.01	17.86	31.06	9.00	164.88	287.26	103.58	8921.27
②LM_Prv_28	同時	0	0	0.01	19.15	32.76	9.00	142.03	240.32	89.82	7872.72
③LM_Prv_29	個別	3	25	0.01	18.08	30.55	10.00	166.05	288.41	103.58	8811.90
④LM_Prv_30	同時	3	25	0.01	15.94	28.63	7.00	112.78	217.14	0.00	6216.16
③LM_Prv_31	個別	3	50	0.01	18.08	30.55	10.00	197.54	304.40	141.41	9300.39
④LM_Prv_32	同時	3	50	0.01	13.56	25.07	6.00	129.18	214.46	96.93	5376.41
③LM_Prv_33	個別	3	75	0.01	18.08	30.55	10.00	229.51	330.80	179.19	10107.04
④LM_Prv_34	同時	3	75	0.01	13.89	25.94	7.00	141.99	209.02	108.15	5421.83
③LM_Prv_35	個別	3	100	0.01	18.08	30.55	10.00	255.64	352.33	183.53	10764.75
④LM_Prv_36	同時	3	100	0.01	13.91	25.79	7.00	156.79	215.39	129.30	5555.96

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（県）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
③LM_Prv_37	個別	5	25	0.01	18.89	31.31	11.00	166.05	288.41	103.58	9029.57
④LM_Prv_38	同時	5	25	0.01	16.01	27.54	9.00	107.24	208.63	0.00	5745.95
③LM_Prv_39	個別	5	50	0.01	18.89	31.31	11.00	197.54	304.40	141.41	9530.13
④LM_Prv_40	同時	5	50	0.01	14.41	25.71	8.00	124.33	211.37	90.20	5434.17
③LM_Prv_41	個別	5	75	0.01	18.89	31.31	11.00	229.51	330.80	179.19	10356.70
④LM_Prv_42	同時	5	75	0.01	14.23	25.66	8.00	146.70	219.16	112.72	5623.89
③LM_Prv_43	個別	5	100	0.01	18.89	31.31	11.00	255.64	352.33	183.53	11030.66
④LM_Prv_44	同時	5	100	0.01	14.43	25.68	8.00	159.42	221.25	129.30	5682.32
③LM_Prv_45	個別	10	25	0.01	19.06	31.10	12.00	166.05	288.41	103.58	8968.79
④LM_Prv_46	同時	10	25	0.01	17.16	28.25	12.00	107.66	207.28	0.00	5855.25
③LM_Prv_47	個別	10	50	0.01	19.06	31.10	12.00	197.54	304.40	141.41	9465.99
④LM_Prv_48	同時	10	50	0.01	15.38	26.25	10.00	124.15	211.98	90.20	5564.97
③LM_Prv_49	個別	10	75	0.01	19.06	31.10	12.00	229.51	330.80	179.19	10286.99
④LM_Prv_50	同時	10	75	0.01	14.91	25.44	10.00	144.12	219.65	107.80	5588.56
③LM_Prv_51	個別	10	100	0.01	19.06	31.10	12.00	255.64	352.33	183.53	10956.42
④LM_Prv_52	同時	10	100	0.01	15.08	25.69	10.00	159.50	226.48	127.60	5819.33
①LM_Prv_53	個別	0	0	0.1	17.08	30.00	9.00	170.41	297.14	104.10	8915.13
②LM_Prv_54	同時	0	0	0.1	17.63	32.82	9.00	134.22	237.14	0.00	7782.90
③LM_Prv_55	個別	3	25	0.1	18.18	30.70	10.00	171.49	298.93	104.17	9176.32
④LM_Prv_56	同時	3	25	0.1	15.59	28.43	7.00	113.36	219.77	0.00	6247.88
③LM_Prv_57	個別	3	50	0.1	18.18	30.70	10.00	203.33	325.66	124.07	9996.87
④LM_Prv_58	同時	3	50	0.1	14.19	26.12	6.00	110.22	208.81	0.00	5454.85
③LM_Prv_59	個別	3	75	0.1	18.18	30.70	10.00	237.10	348.03	179.19	10683.30
④LM_Prv_60	同時	3	75	0.1	13.47	25.46	6.00	123.90	211.96	89.82	5397.30
③LM_Prv_61	個別	3	100	0.1	18.18	30.70	10.00	269.65	379.99	183.53	11664.52
④LM_Prv_62	同時	3	100	0.1	13.76	25.91	6.00	140.09	211.33	106.22	5475.33
③LM_Prv_63	個別	5	25	0.1	19.00	31.65	11.00	171.49	298.93	104.17	9461.31
④LM_Prv_64	同時	5	25	0.1	15.52	27.90	7.00	105.79	203.63	0.00	5681.98
③LM_Prv_65	個別	5	50	0.1	19.00	31.65	11.00	203.33	325.66	124.07	10307.34
④LM_Prv_66	同時	5	50	0.1	13.99	25.51	7.00	109.98	208.43	0.00	5317.94
③LM_Prv_67	個別	5	75	0.1	19.00	31.65	11.00	237.10	348.03	179.19	11015.09
④LM_Prv_68	同時	5	75	0.1	13.91	25.69	7.00	127.10	212.24	91.71	5451.74
③LM_Prv_69	個別	5	100	0.1	19.00	31.65	11.00	269.65	379.99	183.53	12026.79
④LM_Prv_70	同時	5	100	0.1	13.96	25.71	7.00	147.27	216.76	112.72	5572.29
③LM_Prv_71	個別	10	25	0.1	19.34	31.56	12.00	171.49	298.93	104.17	9434.92
④LM_Prv_72	同時	10	25	0.1	16.18	27.72	10.00	104.77	200.75	0.00	5564.85

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（県）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
③LM_Prv_73	個別	10	50	0.1	19.34	31.56	12.00	203.33	325.66	124.07	10278.60
④LM_Prv_74	同時	10	50	0.1	15.16	26.13	9.00	110.28	205.85	0.00	5378.97
③LM_Prv_75	個別	10	75	0.1	19.34	31.56	12.00	237.10	348.03	179.19	10984.37
④LM_Prv_76	同時	10	75	0.1	14.42	25.38	9.00	130.26	213.27	96.11	5412.41
③LM_Prv_77	個別	10	100	0.1	19.34	31.56	12.00	269.65	379.99	183.53	11993.25
④LM_Prv_78	同時	10	100	0.1	14.88	25.81	9.00	149.11	219.34	113.10	5660.77
①LM_Prv_79	個別	0	0	1.0	18.86	31.32	10.00	166.89	288.32	105.08	9031.19
②LM_Prv_80	同時	0	0	1.0	22.31	38.81	12.00	157.40	254.98	105.08	9896.72
③LM_Prv_81	個別	3	25	1.0	18.15	31.01	11.00	167.66	288.88	105.08	8959.30
④LM_Prv_82	同時	3	25	1.0	17.49	29.72	9.00	125.72	223.91	0.00	6655.43
③LM_Prv_83	個別	3	50	1.0	18.15	31.01	11.00	188.56	321.39	112.72	9967.49
④LM_Prv_84	同時	3	50	1.0	17.11	29.80	9.00	123.03	223.90	0.00	6672.54
③LM_Prv_85	個別	3	75	1.0	18.15	31.01	11.00	233.68	365.42	162.35	11332.99
④LM_Prv_86	同時	3	75	1.0	16.36	29.10	9.00	124.17	221.91	78.30	6457.53
③LM_Prv_87	個別	3	100	1.0	18.15	31.01	11.00	264.12	386.63	178.43	11990.99
④LM_Prv_88	同時	3	100	1.0	15.97	28.60	8.00	124.48	215.02	86.00	6150.43
③LM_Prv_89	個別	5	25	1.0	18.90	31.60	11.00	167.66	288.88	105.08	9128.25
④LM_Prv_90	同時	5	25	1.0	16.95	29.22	9.00	122.15	221.00	0.00	6457.64
③LM_Prv_91	個別	5	50	1.0	18.90	31.60	11.00	188.56	321.39	112.72	10155.45
④LM_Prv_92	同時	5	50	1.0	16.64	29.34	9.00	118.98	219.71	0.00	6445.84
③LM_Prv_93	個別	5	75	1.0	18.90	31.60	11.00	233.68	365.42	162.35	11546.70
④LM_Prv_94	同時	5	75	1.0	15.49	27.86	8.00	119.50	214.26	78.30	5968.31
③LM_Prv_95	個別	5	100	1.0	18.90	31.60	11.00	264.12	386.63	178.43	12217.11
④LM_Prv_96	同時	5	100	1.0	15.06	27.24	8.00	122.29	211.69	86.00	5767.55
③LM_Prv_97	個別	10	25	1.0	19.14	31.33	11.00	167.66	288.88	105.08	9051.15
④LM_Prv_98	同時	10	25	1.0	16.25	27.96	9.00	116.30	211.76	0.00	5920.58
③LM_Prv_99	個別	10	50	1.0	19.14	31.33	11.00	188.56	321.39	112.72	10069.68
④LM_Prv_100	同時	10	50	1.0	15.67	27.87	9.00	112.36	208.99	0.00	5823.73
③LM_Prv_101	個別	10	75	1.0	19.14	31.33	11.00	233.68	365.42	162.35	11449.18
④LM_Prv_102	同時	10	75	1.0	15.06	26.70	8.00	119.10	209.91	78.30	5604.02
③LM_Prv_103	個別	10	100	1.0	19.14	31.33	11.00	264.12	386.63	178.43	12113.92
④LM_Prv_104	同時	10	100	1.0	14.92	26.45	8.00	124.79	210.74	90.20	5574.45

表 B2.5　*n*-gram 言語モデルによる年代推定・場所推定（県）の結果

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（県）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
①JSD_Prv_1	個別	0	0	*	16.48	27.35	9.00	224.11	380.93	120.54	10417.97
②JSD_Prv_2	同時	0	0	*	16.06	28.31	8.00	124.88	234.75	0.00	6646.79
③JSD_Prv_3	個別	3	25	*	20.21	33.01	12.00	236.73	388.56	144.25	12826.54
④JSD_Prv_4	同時	3	25	*	15.52	24.82	10.00	135.31	219.23	100.01	5441.36
③JSD_Prv_5	個別	3	50	*	20.21	33.01	12.00	267.74	415.77	189.46	13724.80
④JSD_Prv_6	同時	3	50	*	14.61	23.68	8.00	144.36	215.13	107.80	5093.38
③JSD_Prv_7	個別	3	75	*	20.21	33.01	12.00	311.12	466.66	223.43	15404.64
④JSD_Prv_8	同時	3	75	*	14.63	24.31	8.00	160.19	221.53	131.80	5384.86
③JSD_Prv_9	個別	3	100	*	20.21	33.01	12.00	338.98	495.37	226.23	16352.36
④JSD_Prv_10	同時	3	100	*	14.83	24.22	9.00	171.28	227.37	142.93	5506.27
③JSD_Prv_11	個別	5	25	*	21.18	35.49	13.00	236.73	388.56	144.25	13788.26
④JSD_Prv_12	同時	5	25	*	16.19	25.12	11.00	132.07	220.30	96.11	5532.98
③JSD_Prv_13	個別	5	50	*	21.18	35.49	13.00	267.74	415.77	189.46	14753.87
④JSD_Prv_14	同時	5	50	*	15.19	24.28	9.00	144.75	223.27	106.22	5421.95
③JSD_Prv_15	個別	5	75	*	21.18	35.49	13.00	311.12	466.66	223.43	16559.67
④JSD_Prv_16	同時	5	75	*	15.59	24.90	10.00	164.51	234.25	127.60	5832.55
③JSD_Prv_17	個別	5	100	*	21.18	35.49	13.00	338.98	495.37	226.23	17578.44
④JSD_Prv_18	同時	5	100	*	15.55	24.76	10.00	177.63	239.17	145.19	5921.90
③JSD_Prv_19	個別	10	25	*	21.48	35.50	13.00	236.73	388.56	144.25	13795.12
④JSD_Prv_20	同時	10	25	*	17.02	26.05	12.00	132.81	218.44	96.25	5691.40
③JSD_Prv_21	個別	10	50	*	21.48	35.50	13.00	267.74	415.77	189.46	14761.20
④JSD_Prv_22	同時	10	50	*	16.25	24.55	11.00	145.85	230.35	105.08	5656.21
③JSD_Prv_23	個別	10	75	*	21.48	35.50	13.00	311.12	466.66	223.43	16567.90
④JSD_Prv_24	同時	10	75	*	16.10	25.21	10.00	167.25	239.18	127.60	6030.77
③JSD_Prv_25	個別	10	100	*	21.48	35.50	13.00	338.98	495.37	226.23	17587.18
④JSD_Prv_26	同時	10	100	*	16.16	25.07	11.00	181.55	247.41	142.93	6202.62

表 B2.6　JS 情報量による年代推定・場所推定（県）の結果

実験番号	推定方法	σ_t	σ_l	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
①NB_Prv_1	個別	0	0	1.001	29.42	48.33	16.00	166.30	286.66	105.08	13853.46
②NB_Prv_2	同時	0	0	1.001	33.92	57.05	18.00	166.61	267.51	112.72	15262.20
③NB_Prv_3	個別	3	25	1.001	28.55	49.21	16.00	175.72	293.66	107.91	14451.24
④NB_Prv_4	同時	3	25	1.001	25.61	44.07	13.00	150.03	256.98	86.00	11325.67
③NB_Prv_5	個別	3	50	1.001	28.55	49.21	16.00	199.50	293.39	146.06	14437.90
④NB_Prv_6	同時	3	50	1.001	24.57	42.71	12.00	170.47	247.42	124.07	10566.73
③NB_Prv_7	個別	3	75	1.001	28.55	49.21	16.00	227.91	328.23	179.19	16152.26
④NB_Prv_8	同時	3	75	1.001	24.47	42.57	12.00	179.49	251.00	142.76	10685.20
③NB_Prv_9	個別	3	100	1.001	28.55	49.21	16.00	244.05	334.78	179.19	16474.96
④NB_Prv_10	同時	3	100	1.001	24.30	42.00	12.00	184.91	252.13	144.38	10589.56
③NB_Prv_11	個別	5	25	1.001	29.62	49.75	16.00	175.72	293.66	107.91	14608.48
④NB_Prv_12	同時	5	25	1.001	26.11	44.03	14.00	146.18	248.49	86.00	10940.31
③NB_Prv_13	個別	5	50	1.001	29.62	49.75	16.00	199.50	293.39	146.06	14594.99
④NB_Prv_14	同時	5	50	1.001	25.19	43.73	13.00	164.77	249.41	113.83	10907.74
③NB_Prv_15	個別	5	75	1.001	29.62	49.75	16.00	227.91	328.23	179.19	16328.01
④NB_Prv_16	同時	5	75	1.001	25.35	43.31	13.00	181.87	257.66	142.76	11160.19
③NB_Prv_17	個別	5	100	1.001	29.62	49.75	16.00	244.05	334.78	179.19	16654.22
④NB_Prv_18	同時	5	100	1.001	25.32	42.35	13.00	189.34	259.02	145.19	10968.82
③NB_Prv_19	個別	10	25	1.001	31.73	53.76	18.00	175.72	293.66	107.91	15786.81
④NB_Prv_20	同時	10	25	1.001	27.70	46.33	16.00	148.66	252.95	89.67	11719.28
③NB_Prv_21	個別	10	50	1.001	31.73	53.76	18.00	199.50	293.39	146.06	15772.24
④NB_Prv_22	同時	10	50	1.001	25.14	41.42	14.00	165.46	249.49	113.83	10333.59
③NB_Prv_23	個別	10	75	1.001	31.73	53.76	18.00	227.91	328.23	179.19	17645.04
④NB_Prv_24	同時	10	75	1.001	25.43	43.01	14.00	180.95	259.10	141.41	11142.93
③NB_Prv_25	個別	10	100	1.001	31.73	53.76	18.00	244.05	334.78	179.19	17997.56
④NB_Prv_26	同時	10	100	1.001	26.07	45.48	14.00	191.57	263.16	145.19	11967.67
①NB_Prv_27	個別	0	0	1.01	28.91	50.24	15.00	170.64	294.43	105.41	14792.82
②NB_Prv_28	同時	0	0	1.01	29.29	49.18	15.00	158.79	262.05	104.17	12888.27
③NB_Prv_29	個別	3	25	1.01	29.06	49.81	16.00	173.25	295.69	107.91	14729.11
④NB_Prv_30	同時	3	25	1.01	24.96	45.56	12.00	142.37	263.07	78.30	11985.50
③NB_Prv_31	個別	3	50	1.01	29.06	49.81	16.00	204.11	302.61	145.19	15073.74
④NB_Prv_32	同時	3	50	1.01	24.59	43.47	12.00	149.33	243.61	103.58	10590.58
③NB_Prv_33	個別	3	75	1.01	29.06	49.81	16.00	233.36	337.47	181.99	16810.36
④NB_Prv_34	同時	3	75	1.01	24.33	42.00	12.00	168.67	244.98	124.07	10289.82
③NB_Prv_35	個別	3	100	1.01	29.06	49.81	16.00	248.16	343.26	181.99	17098.73
④NB_Prv_36	同時	3	100	1.01	24.62	42.96	12.00	184.03	254.23	142.93	10922.31

実験番号	推定方法	σ_t	σ_l	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
③NB_Prv_37	個別	5	25	1.01	30.15	51.04	17.00	173.25	295.69	107.91	15092.18
④NB_Prv_38	同時	5	25	1.01	25.53	45.74	13.00	136.13	247.34	71.96	11313.32
③NB_Prv_39	個別	5	50	1.01	30.15	51.04	17.00	204.11	302.61	145.19	15445.31
④NB_Prv_40	同時	5	50	1.01	24.41	42.20	12.00	152.45	249.19	102.22	10515.93
③NB_Prv_41	個別	5	75	1.01	30.15	51.04	17.00	233.36	337.47	181.99	17224.73
④NB_Prv_42	同時	5	75	1.01	25.68	45.48	13.00	172.39	250.00	127.60	11370.96
③NB_Prv_43	個別	5	100	1.01	30.15	51.04	17.00	248.16	343.26	181.99	17520.21
④NB_Prv_44	同時	5	100	1.01	25.36	43.07	13.00	185.73	256.41	145.19	11043.69
③NB_Prv_45	個別	10	25	1.01	31.87	54.09	19.00	173.25	295.69	107.91	15992.27
④NB_Prv_46	同時	10	25	1.01	26.82	45.81	16.00	139.35	257.02	78.30	11774.40
③NB_Prv_47	個別	10	50	1.01	31.87	54.09	19.00	204.11	302.61	145.19	16366.47
④NB_Prv_48	同時	10	50	1.01	25.39	42.49	14.00	152.36	246.79	104.10	10486.84
③NB_Prv_49	個別	10	75	1.01	31.87	54.09	19.00	233.36	337.47	181.99	18252.01
④NB_Prv_50	同時	10	75	1.01	25.59	44.31	14.00	177.94	258.16	131.80	11438.54
③NB_Prv_51	個別	10	100	1.01	31.87	54.09	19.00	248.16	343.26	181.99	18565.11
④NB_Prv_52	同時	10	100	1.01	26.14	45.79	14.00	194.94	266.81	145.19	12216.18
①NB_Prv_53	個別	0	0	1.1	28.66	51.82	15.00	178.18	304.63	114.10	15786.71
②NB_Prv_54	同時	0	0	1.1	28.27	50.43	14.00	152.31	261.06	91.71	13164.42
③NB_Prv_55	個別	3	25	1.1	29.34	52.60	15.00	178.34	304.70	114.10	16027.67
④NB_Prv_56	同時	3	25	1.1	24.73	45.40	11.00	135.18	256.16	0.00	11629.65
③NB_Prv_57	個別	3	50	1.1	29.34	52.60	15.00	203.42	315.88	142.63	16615.80
④NB_Prv_58	同時	3	50	1.1	24.71	46.22	12.00	137.40	250.87	79.85	11595.63
③NB_Prv_59	個別	3	75	1.1	29.34	52.60	15.00	241.74	359.49	179.19	18909.72
④NB_Prv_60	同時	3	75	1.1	24.81	46.12	12.00	143.15	248.90	90.20	11478.99
③NB_Prv_61	個別	3	100	1.1	29.34	52.60	15.00	259.94	362.87	183.24	19087.49
④NB_Prv_62	同時	3	100	1.1	24.37	45.14	12.00	157.00	251.10	105.74	11335.78
③NB_Prv_63	個別	5	25	1.1	30.63	53.31	17.00	178.34	304.70	114.10	16243.89
④NB_Prv_64	同時	5	25	1.1	23.88	43.58	11.00	134.41	251.26	0.00	10949.68
③NB_Prv_65	個別	5	50	1.1	30.63	53.31	17.00	203.42	315.88	142.63	16839.95
④NB_Prv_66	同時	5	50	1.1	24.77	46.06	11.00	138.77	255.55	84.03	11771.78
③NB_Prv_67	個別	5	75	1.1	30.63	53.31	17.00	241.74	359.49	179.19	19164.81
④NB_Prv_68	同時	5	75	1.1	24.99	46.12	12.00	146.23	250.25	91.71	11540.47
③NB_Prv_69	個別	5	100	1.1	30.63	53.31	17.00	259.94	362.87	183.24	19344.98
④NB_Prv_70	同時	5	100	1.1	24.81	45.26	12.00	162.59	252.44	112.72	11425.51
③NB_Prv_71	個別	10	25	1.1	32.56	58.19	18.00	178.34	304.70	114.10	17731.05
④NB_Prv_72	同時	10	25	1.1	25.55	46.02	13.00	133.06	246.39	0.00	11338.84

実験番号	推定方法	σ_t	σ_l	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
③NB_Prv_73	個別	10	50	1.1	32.56	58.19	18.00	203.42	315.88	142.63	18381.68
④NB_Prv_74	同時	10	50	1.1	24.95	44.32	12.00	138.15	245.85	84.03	10897.20
③NB_Prv_75	個別	10	75	1.1	32.56	58.19	18.00	241.74	359.49	179.19	20919.39
④NB_Prv_76	同時	10	75	1.1	25.17	45.24	12.00	148.99	243.40	101.32	11010.97
③NB_Prv_77	個別	10	100	1.1	32.56	58.19	18.00	259.94	362.87	183.24	21116.05
④NB_Prv_78	同時	10	100	1.1	26.07	46.71	13.00	171.95	252.83	124.74	11809.75
①NB_Prv_79	個別	0	0	2	36.14	57.18	21.00	156.59	255.45	105.41	14606.79
②NB_Prv_80	同時	0	0	2	69.05	118.94	31.00	241.15	350.49	163.93	41686.27
③NB_Prv_81	個別	3	25	2	33.72	61.38	17.00	156.87	255.75	105.41	15698.45
④NB_Prv_82	同時	3	25	2	37.35	65.08	19.00	156.82	257.94	107.91	16787.44
③NB_Prv_83	個別	3	50	2	33.72	61.38	17.00	162.73	261.77	107.91	16067.96
④NB_Prv_84	同時	3	50	2	37.60	65.79	18.00	156.67	260.53	105.41	17139.68
③NB_Prv_85	個別	3	75	2	33.72	61.38	17.00	185.97	277.33	134.93	17023.06
④NB_Prv_86	同時	3	75	2	36.07	64.02	17.00	158.65	266.48	105.41	17061.31
③NB_Prv_87	個別	3	100	2	33.72	61.38	17.00	216.72	303.84	162.25	18650.23
④NB_Prv_88	同時	3	100	2	33.36	59.57	16.00	159.26	264.69	107.80	15768.53
③NB_Prv_89	個別	5	25	2	33.61	61.07	16.00	156.87	255.75	105.41	15618.30
④NB_Prv_90	同時	5	25	2	35.27	63.60	17.00	150.53	255.10	103.58	16225.36
③NB_Prv_91	個別	5	50	2	33.61	61.07	16.00	162.73	261.77	107.91	15985.93
④NB_Prv_92	同時	5	50	2	34.17	60.92	16.00	150.23	254.80	102.22	15520.96
③NB_Prv_93	個別	5	75	2	33.61	61.07	16.00	185.97	277.33	134.93	16936.16
④NB_Prv_94	同時	5	75	2	32.95	60.61	15.00	151.66	255.96	104.17	15512.54
③NB_Prv_95	個別	5	100	2	33.61	61.07	16.00	216.72	303.84	162.25	18555.02
④NB_Prv_96	同時	5	100	2	31.47	58.54	15.00	155.04	261.40	105.08	15303.66
③NB_Prv_97	個別	10	25	2	33.39	60.49	17.00	156.87	255.75	105.41	15468.97
④NB_Prv_98	同時	10	25	2	32.91	61.12	15.00	145.48	258.13	100.98	15776.12
③NB_Prv_99	個別	10	50	2	33.39	60.49	17.00	162.73	261.77	107.91	15833.08
④NB_Prv_100	同時	10	50	2	31.74	59.30	14.00	142.94	252.16	100.01	14952.67
③NB_Prv_101	個別	10	75	2	33.39	60.49	17.00	185.97	277.33	134.93	16774.22
④NB_Prv_102	同時	10	75	2	30.59	58.26	14.00	143.81	249.57	100.01	14540.99
③NB_Prv_103	個別	10	100	2	33.39	60.49	17.00	216.72	303.84	162.25	18377.61
④NB_Prv_104	同時	10	100	2	29.58	57.17	13.00	147.59	252.55	100.98	14438.24

表 B2.7　ナイーブベイズ多変数ベルヌーイモデルによる年代推定・場所推定（県）の結果

B3　年代推定・場所推定（自治州）の結果

表 B3.8 に *n*-gram 言語モデルによる年代推定・場所推定（自治州）の結果を，表 B3.9 に JS 情報量による年代推定・場所推定（自治州）の結果を，表 B3.10 にナイーブベイズ多変数ベルヌーイモデルによる年代推定・場所推定（自治州）の結果を示す。

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
①LM_CA_1	個別	0	0	0.001	18.63	31.70	10.00	170.59	287.36	123.00	9110.40
②LM_CA_2	同時	0	0	0.001	19.15	33.00	10.00	116.71	220.73	0.00	7283.92
③LM_CA_3	個別	3	25	0.001	17.87	30.41	10.00	170.78	286.66	157.88	8717.37
④LM_CA_4	同時	3	25	0.001	15.88	26.80	9.00	103.45	211.03	0.00	5655.21
③LM_CA_5	個別	3	50	0.001	17.87	30.41	10.00	180.16	294.07	172.85	8942.69
④LM_CA_6	同時	3	50	0.001	15.46	27.49	9.00	121.83	213.99	0.00	5882.67
③LM_CA_7	個別	3	75	0.001	17.87	30.41	10.00	214.18	333.01	188.25	10126.71
④LM_CA_8	同時	3	75	0.001	14.77	26.94	8.00	147.07	224.37	166.24	6045.26
③LM_CA_9	個別	3	100	0.001	17.87	30.41	10.00	238.59	364.78	188.25	11093.06
④LM_CA_10	同時	3	100	0.001	14.43	24.43	8.00	145.88	221.00	166.24	5399.50
③LM_CA_11	個別	5	25	0.001	18.67	31.19	11.00	170.78	286.66	157.88	8939.66
④LM_CA_12	同時	5	25	0.001	16.07	26.04	9.00	104.38	213.84	0.00	5568.37
③LM_CA_13	個別	5	50	0.001	18.67	31.19	11.00	180.16	294.07	172.85	9170.73
④LM_CA_14	同時	5	50	0.001	15.64	26.64	9.00	116.47	215.10	0.00	5730.37
③LM_CA_15	個別	5	75	0.001	18.67	31.19	11.00	214.18	333.01	188.25	10384.94
④LM_CA_16	同時	5	75	0.001	15.31	26.64	9.00	137.13	219.81	157.88	5856.83
③LM_CA_17	個別	5	100	0.001	18.67	31.19	11.00	238.59	364.78	188.25	11375.94
④LM_CA_18	同時	5	100	0.001	14.82	24.34	9.00	145.93	226.07	166.24	5503.11
③LM_CA_19	個別	10	25	0.001	19.16	31.35	12.00	170.78	286.66	157.88	8985.44
④LM_CA_20	同時	10	25	0.001	16.45	27.27	10.00	101.87	206.29	0.00	5626.38
③LM_CA_21	個別	10	50	0.001	19.16	31.35	12.00	180.16	294.07	172.85	9217.69
④LM_CA_22	同時	10	50	0.001	15.84	26.44	10.00	114.91	210.52	0.00	5565.58
③LM_CA_23	個別	10	75	0.001	19.16	31.35	12.00	214.18	333.01	188.25	10438.12
④LM_CA_24	同時	10	75	0.001	15.71	26.40	10.00	136.32	215.38	157.88	5685.93
③LM_CA_25	個別	10	100	0.001	19.16	31.35	12.00	238.59	364.78	188.25	11434.19
④LM_CA_26	同時	10	100	0.001	15.38	26.30	9.00	141.27	218.43	166.24	5744.18
①LM_CA_27	個別	0	0	0.01	17.86	31.06	9.00	179.69	300.39	172.85	9329.30
②LM_CA_28	同時	0	0	0.01	17.65	31.66	9.00	110.87	211.78	0.00	6704.68
③LM_CA_29	個別	3	25	0.01	18.08	30.55	10.00	181.38	302.21	172.85	9233.41
④LM_CA_30	同時	3	25	0.01	15.24	25.99	8.00	102.00	208.64	0.00	5422.78
③LM_CA_31	個別	3	50	0.01	18.08	30.55	10.00	184.81	311.45	172.85	9515.88

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
④LM_CA_32	同時	3	50	0.01	15.00	27.09	8.00	106.73	207.24	0.00	5614.76
③LM_CA_33	個別	3	75	0.01	18.08	30.55	10.00	227.74	361.76	188.25	11053.02
④LM_CA_34	同時	3	75	0.01	14.67	27.02	7.00	129.88	214.01	123.00	5782.24
③LM_CA_35	個別	3	100	0.01	18.08	30.55	10.00	260.53	395.90	206.65	12096.18
④LM_CA_36	同時	3	100	0.01	14.34	26.24	7.00	142.80	217.59	162.35	5710.23
③LM_CA_37	個別	5	25	0.01	18.89	31.31	11.00	181.38	302.21	172.85	9461.50
④LM_CA_38	同時	5	25	0.01	15.49	25.37	9.00	100.15	208.06	0.00	5277.60
③LM_CA_39	個別	5	50	0.01	18.89	31.31	11.00	184.81	311.45	172.85	9750.94
④LM_CA_40	同時	5	50	0.01	15.48	26.78	9.00	105.63	206.11	0.00	5520.22
③LM_CA_41	個別	5	75	0.01	18.89	31.31	11.00	227.74	361.76	188.25	11326.05
④LM_CA_42	同時	5	75	0.01	14.85	26.16	9.00	126.87	213.63	91.61	5589.33
③LM_CA_43	個別	5	100	0.01	18.89	31.31	11.00	260.53	395.90	206.65	12394.98
④LM_CA_44	同時	5	100	0.01	14.70	26.15	8.00	141.29	217.96	162.35	5700.07
③LM_CA_45	個別	10	25	0.01	19.06	31.10	12.00	181.38	302.21	172.85	9397.81
④LM_CA_46	同時	10	25	0.01	16.20	26.94	10.00	98.99	204.22	0.00	5502.53
③LM_CA_47	個別	10	50	0.01	19.06	31.10	12.00	184.81	311.45	172.85	9685.31
④LM_CA_48	同時	10	50	0.01	15.80	26.52	10.00	103.20	200.61	0.00	5319.98
③LM_CA_49	個別	10	75	0.01	19.06	31.10	12.00	227.74	361.76	188.25	11249.81
④LM_CA_50	同時	10	75	0.01	15.38	25.85	10.00	126.66	208.41	118.63	5386.85
③LM_CA_51	個別	10	100	0.01	19.06	31.10	12.00	260.53	395.90	206.65	12311.55
④LM_CA_52	同時	10	100	0.01	15.31	26.40	9.00	140.99	218.82	162.35	5775.85
①LM_CA_53	個別	0	0	0.1	17.08	30.00	9.00	193.77	326.30	172.85	9790.26
②LM_CA_54	同時	0	0	0.1	16.28	30.65	8.00	105.68	214.24	0.00	6565.74
③LM_CA_55	個別	3	25	0.1	18.18	30.70	10.00	193.23	326.02	172.85	10007.91
④LM_CA_56	同時	3	25	0.1	14.80	27.29	7.00	96.83	203.51	0.00	5554.66
③LM_CA_57	個別	3	50	0.1	18.18	30.70	10.00	195.07	333.35	172.85	10232.96
④LM_CA_58	同時	3	50	0.1	14.85	27.56	7.00	98.09	203.10	0.00	5598.22
③LM_CA_59	個別	3	75	0.1	18.18	30.70	10.00	235.97	385.30	188.25	11827.61
④LM_CA_60	同時	3	75	0.1	14.31	26.44	7.00	101.12	201.98	0.00	5339.37
③LM_CA_61	個別	3	100	0.1	18.18	30.70	10.00	269.17	419.18	206.65	12867.49
④LM_CA_62	同時	3	100	0.1	14.11	26.08	7.00	117.19	209.19	0.00	5455.87
③LM_CA_63	個別	5	25	0.1	19.00	31.65	11.00	193.23	326.02	172.85	10318.73
④LM_CA_64	同時	5	25	0.1	15.18	26.69	8.00	99.43	202.56	0.00	5405.25
③LM_CA_65	個別	5	50	0.1	19.00	31.65	11.00	195.07	333.35	172.85	10550.77
④LM_CA_66	同時	5	50	0.1	15.18	26.89	8.00	102.31	206.08	0.00	5541.12
③LM_CA_67	個別	5	75	0.1	19.00	31.65	11.00	235.97	385.30	188.25	12194.94

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
④LM_CA_68	同時	5	75	0.1	14.72	26.42	8.00	105.77	203.95	0.00	5387.93
③LM_CA_69	個別	5	100	0.1	19.00	31.65	11.00	269.17	419.18	206.65	13267.12
④LM_CA_70	同時	5	100	0.1	14.54	26.27	8.00	117.56	208.62	0.00	5479.88
③LM_CA_71	個別	10	25	0.1	19.34	31.56	12.00	193.23	326.02	172.85	10289.95
④LM_CA_72	同時	10	25	0.1	15.81	26.74	9.00	97.52	198.42	0.00	5304.97
③LM_CA_73	個別	10	50	0.1	19.34	31.56	12.00	195.07	333.35	172.85	10521.34
④LM_CA_74	同時	10	50	0.1	15.66	26.56	9.00	101.30	199.94	0.00	5309.66
③LM_CA_75	個別	10	75	0.1	19.34	31.56	12.00	235.97	385.30	188.25	12160.93
④LM_CA_76	同時	10	75	0.1	15.32	26.56	9.00	106.63	198.50	0.00	5272.13
③LM_CA_77	個別	10	100	0.1	19.34	31.56	12.00	269.17	419.18	206.65	13230.11
④LM_CA_78	同時	10	100	0.1	15.23	26.55	9.00	127.45	211.64	91.61	5619.18
①LM_CA_79	個別	0	0	1	18.86	31.32	10.00	191.73	325.15	172.85	10184.85
②LM_CA_80	同時	0	0	1	19.71	34.83	11.00	120.47	227.68	0.00	7929.17
③LM_CA_81	個別	3	25	1	18.15	31.01	11.00	191.74	325.16	172.85	10084.45
④LM_CA_82	同時	3	25	1	16.19	29.24	9.00	106.49	211.45	0.00	6182.13
③LM_CA_83	個別	3	50	1	18.15	31.01	11.00	192.27	331.04	172.85	10266.85
④LM_CA_84	同時	3	50	1	16.18	29.24	9.00	103.64	208.43	0.00	6095.52
③LM_CA_85	個別	3	75	1	18.15	31.01	11.00	225.27	386.16	180.55	11976.41
④LM_CA_86	同時	3	75	1	15.95	28.66	9.00	101.63	202.57	0.00	5805.63
③LM_CA_87	個別	3	100	1	18.15	31.01	11.00	260.87	424.06	206.65	13151.63
④LM_CA_88	同時	3	100	1	15.62	27.85	8.00	105.18	204.69	0.00	5701.44
③LM_CA_89	個別	5	25	1	18.90	31.60	11.00	191.74	325.16	172.85	10274.61
④LM_CA_90	同時	5	25	1	15.55	27.75	8.00	105.00	207.97	0.00	5770.99
③LM_CA_91	個別	5	50	1	18.90	31.60	11.00	192.27	331.04	172.85	10460.46
④LM_CA_92	同時	5	50	1	15.61	27.89	8.00	102.51	203.76	0.00	5682.88
③LM_CA_93	個別	5	75	1	18.90	31.60	11.00	225.27	386.16	180.55	12202.26
④LM_CA_94	同時	5	75	1	15.62	27.83	8.00	103.00	203.61	0.00	5665.97
③LM_CA_95	個別	5	100	1	18.90	31.60	11.00	260.87	424.06	206.65	13399.63
④LM_CA_96	同時	5	100	1	15.53	27.82	8.00	105.75	206.01	0.00	5730.56
③LM_CA_97	個別	10	25	1	19.14	31.33	11.00	191.74	325.16	172.85	10187.84
④LM_CA_98	同時	10	25	1	16.09	27.67	9.00	105.38	207.45	0.00	5740.76
③LM_CA_99	個別	10	50	1	19.14	31.33	11.00	192.27	331.04	172.85	10372.11
④LM_CA_100	同時	10	50	1	16.14	27.66	9.00	104.34	204.55	0.00	5658.50
③LM_CA_101	個別	10	75	1	19.14	31.33	11.00	225.27	386.16	180.55	12099.20
④LM_CA_102	同時	10	75	1	15.79	27.17	9.00	106.69	208.13	0.00	5654.31
③LM_CA_103	個別	10	100	1	19.14	31.33	11.00	260.87	424.06	206.65	13286.46

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
④LM_CA_104	同時	10	100	1	15.66	27.10	8.00	110.39	207.59	0.00	5624.99

表 B3.8　*n*-gram 言語モデルによる年代推定・場所推定（自治州）の結果

実験番号	推定方法	σ_t	σ_l	α	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
①JSD_CA_1	個別	0	0	*	16.48	27.35	9.00	287.33	476.08	206.65	13020.21
②JSD_CA_2	同時	0	0	*	15.68	27.36	8.00	107.61	224.13	0.00	6132.54
③JSD_CA_3	個別	3	25	*	20.21	33.01	12.00	294.24	481.12	206.65	15882.07
④JSD_CA_4	同時	3	25	*	15.43	26.70	9.00	106.71	206.43	0.00	5511.08
③JSD_CA_5	個別	3	50	*	20.21	33.01	12.00	308.10	486.07	206.65	16045.54
④JSD_CA_6	同時	3	50	*	15.33	25.82	9.00	135.26	220.90	91.61	5703.14
③JSD_CA_7	個別	3	75	*	20.21	33.01	12.00	336.90	520.84	206.65	17193.11
④JSD_CA_8	同時	3	75	*	15.51	27.36	9.00	142.94	219.99	162.35	6018.41
③JSD_CA_9	個別	3	100	*	20.21	33.01	12.00	352.73	525.08	282.26	17333.17
④JSD_CA_10	同時	3	100	*	15.56	27.36	9.00	152.58	224.69	172.85	6147.36
③JSD_CA_11	個別	5	25	*	21.18	35.49	13.00	294.24	481.12	206.65	17072.89
④JSD_CA_12	同時	5	25	*	16.54	27.88	10.00	106.73	205.71	0.00	5735.69
③JSD_CA_13	個別	5	50	*	21.18	35.49	13.00	308.10	486.07	206.65	17248.62
④JSD_CA_14	同時	5	50	*	16.20	26.00	10.00	129.70	220.56	0.00	5734.89
③JSD_CA_15	個別	5	75	*	21.18	35.49	13.00	336.90	520.84	206.65	18482.23
④JSD_CA_16	同時	5	75	*	15.80	25.24	9.00	142.39	220.63	157.88	5569.76
③JSD_CA_17	個別	5	100	*	21.18	35.49	13.00	352.73	525.08	282.26	18632.79
④JSD_CA_18	同時	5	100	*	16.33	27.76	10.00	153.28	225.47	171.02	6259.99
③JSD_CA_19	個別	10	25	*	21.48	35.50	13.00	294.24	481.12	206.65	17081.38
④JSD_CA_20	同時	10	25	*	17.42	28.29	11.00	111.22	210.62	0.00	5957.88
③JSD_CA_21	個別	10	50	*	21.48	35.50	13.00	308.10	486.07	206.65	17257.20
④JSD_CA_22	同時	10	50	*	17.42	28.16	11.00	129.71	218.56	0.00	6154.46
③JSD_CA_23	個別	10	75	*	21.48	35.50	13.00	336.90	520.84	206.65	18491.42
④JSD_CA_24	同時	10	75	*	17.36	28.45	11.00	146.73	226.91	166.24	6455.31
③JSD_CA_25	個別	10	100	*	21.48	35.50	13.00	352.73	525.08	282.26	18642.06
④JSD_CA_26	同時	10	100	*	17.35	28.63	10.00	161.14	236.36	172.85	6767.77

表 B3.9　JS 情報量による年代推定・場所推定（自治州）の結果

実験番号	推定方法	σ_t	σ_l	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
①NB_CA_1	個別	0	0	1.001	29.42	48.33	16.00	188.19	331.95	123.00	16041.97
②NB_CA_2	同時	0	0	1.001	30.40	50.74	16.00	148.17	262.04	0.00	13296.70
③NB_CA_3	個別	3	25	1.001	28.55	49.21	16.00	209.32	351.17	172.85	17281.37
④NB_CA_4	同時	3	25	1.001	24.74	42.04	13.00	129.53	241.78	0.00	10164.91
③NB_CA_5	個別	3	50	1.001	28.55	49.21	16.00	195.33	337.21	172.85	16594.19
④NB_CA_6	同時	3	50	1.001	25.63	43.87	13.00	141.01	238.05	89.09	10442.89
③NB_CA_7	個別	3	75	1.001	28.55	49.21	16.00	222.39	346.68	188.25	17060.30
④NB_CA_8	同時	3	75	1.001	25.10	43.34	13.00	157.38	240.37	171.02	10418.38
③NB_CA_9	個別	3	100	1.001	28.55	49.21	16.00	226.91	346.70	188.25	17061.43
④NB_CA_10	同時	3	100	1.001	25.58	43.87	14.00	162.24	240.87	172.85	10566.22
③NB_CA_11	個別	5	25	1.001	29.62	49.75	16.00	209.32	351.17	172.85	17469.41
④NB_CA_12	同時	5	25	1.001	24.82	41.76	13.00	125.31	234.32	0.00	9784.09
③NB_CA_13	個別	5	50	1.001	29.62	49.75	16.00	195.33	337.21	172.85	16774.74
④NB_CA_14	同時	5	50	1.001	26.06	45.20	14.00	138.90	237.30	0.00	10724.66
③NB_CA_15	個別	5	75	1.001	29.62	49.75	16.00	222.39	346.68	188.25	17245.93
④NB_CA_16	同時	5	75	1.001	26.28	45.11	14.00	154.02	239.89	166.24	10822.20
③NB_CA_17	個別	5	100	1.001	29.62	49.75	16.00	226.91	346.70	188.25	17247.07
④NB_CA_18	同時	5	100	1.001	26.38	45.51	14.00	160.99	241.16	172.85	10976.55
③NB_CA_19	個別	10	25	1.001	31.73	53.76	18.00	209.32	351.17	172.85	18878.50
④NB_CA_20	同時	10	25	1.001	27.41	46.81	15.00	132.17	240.90	0.00	11277.71
③NB_CA_21	個別	10	50	1.001	31.73	53.76	18.00	195.33	337.21	172.85	18127.81
④NB_CA_22	同時	10	50	1.001	27.85	48.23	15.00	144.39	244.97	0.00	11815.92
③NB_CA_23	個別	10	75	1.001	31.73	53.76	18.00	222.39	346.68	188.25	18637.00
④NB_CA_24	同時	10	75	1.001	26.51	46.00	15.00	160.50	250.83	171.02	11539.48
③NB_CA_25	個別	10	100	1.001	31.73	53.76	18.00	226.91	346.70	188.25	18638.23
④NB_CA_26	同時	10	100	1.001	26.74	46.15	15.00	170.54	254.34	172.85	11738.30
①NB_CA_27	個別	0	0	1.01	28.91	50.24	15.00	201.65	349.12	172.85	17540.73
②NB_CA_28	同時	0	0	1.01	27.96	46.98	14.00	141.79	256.26	0.00	12039.68
③NB_CA_29	個別	3	25	1.01	29.06	49.81	16.00	201.72	349.12	172.85	17390.70
④NB_CA_30	同時	3	25	1.01	24.74	44.56	12.00	125.44	238.88	0.00	10643.91
③NB_CA_31	個別	3	50	1.01	29.06	49.81	16.00	204.32	349.43	172.85	17406.36
④NB_CA_32	同時	3	50	1.01	24.96	44.31	13.00	127.06	230.25	0.00	10202.02
③NB_CA_33	個別	3	75	1.01	29.06	49.81	16.00	233.15	368.91	188.25	18376.55
④NB_CA_34	同時	3	75	1.01	25.16	44.65	12.00	140.15	230.10	123.00	10274.14
③NB_CA_35	個別	3	100	1.01	29.06	49.81	16.00	241.92	371.20	188.25	18490.89
④NB_CA_36	同時	3	100	1.01	25.45	43.92	14.00	155.31	238.85	166.24	10490.05

実験番号	推定方法	σ_t	σ_l	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
③NB_CA_37	個別	5	25	1.01	30.15	51.04	17.00	201.72	349.12	172.85	17819.38
④NB_CA_38	同時	5	25	1.01	24.78	42.51	13.00	122.20	232.07	0.00	9865.01
③NB_CA_39	個別	5	50	1.01	30.15	51.04	17.00	204.32	349.43	172.85	17835.43
④NB_CA_40	同時	5	50	1.01	24.84	41.87	13.00	126.07	231.59	0.00	9697.67
③NB_CA_41	個別	5	75	1.01	30.15	51.04	17.00	233.15	368.91	188.25	18829.53
④NB_CA_42	同時	5	75	1.01	25.64	44.54	14.00	139.31	229.53	123.00	10222.50
③NB_CA_43	個別	5	100	1.01	30.15	51.04	17.00	241.92	371.20	188.25	18946.68
④NB_CA_44	同時	5	100	1.01	26.54	47.32	14.00	155.83	239.25	166.24	11321.10
③NB_CA_45	個別	10	25	1.01	31.87	54.09	19.00	201.72	349.12	172.85	18882.12
④NB_CA_46	同時	10	25	1.01	26.84	46.56	14.00	129.85	241.17	0.00	11230.29
③NB_CA_47	個別	10	50	1.01	31.87	54.09	19.00	204.32	349.43	172.85	18899.13
④NB_CA_48	同時	10	50	1.01	27.67	47.84	15.00	133.96	239.98	0.00	11481.57
③NB_CA_49	個別	10	75	1.01	31.87	54.09	19.00	233.15	368.91	188.25	19952.52
④NB_CA_50	同時	10	75	1.01	27.27	47.60	14.00	150.30	244.25	123.00	11627.35
③NB_CA_51	個別	10	100	1.01	31.87	54.09	19.00	241.92	371.20	188.25	20076.66
④NB_CA_52	同時	10	100	1.01	27.18	47.11	14.00	167.87	256.88	172.85	12101.29
①NB_CA_53	個別	0	0	1.1	28.66	51.82	15.00	224.62	383.62	180.55	19880.00
②NB_CA_54	同時	0	0	1.1	27.05	48.44	14.00	137.49	254.27	0.00	12316.73
③NB_CA_55	個別	3	25	1.1	29.34	52.60	15.00	224.47	383.55	180.55	20175.09
④NB_CA_56	同時	3	25	1.1	25.27	47.32	12.00	120.46	240.34	0.00	11372.55
③NB_CA_57	個別	3	50	1.1	29.34	52.60	15.00	224.28	384.04	180.55	20201.23
④NB_CA_58	同時	3	50	1.1	25.07	47.07	12.00	120.13	240.20	0.00	11306.67
③NB_CA_59	個別	3	75	1.1	29.34	52.60	15.00	245.74	402.14	188.25	21153.02
④NB_CA_60	同時	3	75	1.1	25.29	46.72	13.00	120.53	238.09	0.00	11123.88
③NB_CA_61	個別	3	100	1.1	29.34	52.60	15.00	264.86	408.37	200.94	21481.03
④NB_CA_62	同時	3	100	1.1	25.16	46.77	12.00	129.14	234.85	0.00	10983.37
③NB_CA_63	個別	5	25	1.1	30.63	53.31	17.00	224.47	383.55	180.55	20447.25
④NB_CA_64	同時	5	25	1.1	26.08	47.66	13.00	123.14	237.91	0.00	11339.38
③NB_CA_65	個別	5	50	1.1	30.63	53.31	17.00	224.28	384.04	180.55	20473.75
④NB_CA_66	同時	5	50	1.1	25.75	47.01	13.00	123.93	238.21	0.00	11197.07
③NB_CA_67	個別	5	75	1.1	30.63	53.31	17.00	245.74	402.14	188.25	21438.37
④NB_CA_68	同時	5	75	1.1	26.26	48.41	13.00	125.73	236.35	0.00	11442.42
③NB_CA_69	個別	5	100	1.1	30.63	53.31	17.00	264.86	408.37	200.94	21770.81
④NB_CA_70	同時	5	100	1.1	26.15	47.65	13.00	132.96	237.81	0.00	11332.05
③NB_CA_71	個別	10	25	1.1	32.56	58.19	18.00	224.47	383.55	180.55	22319.24
④NB_CA_72	同時	10	25	1.1	27.63	48.51	14.00	123.88	234.68	0.00	11384.98

実験番号	推定方法	σ_t	σ_l	$\alpha_{f,c}$	年代推定			場所推定（自治州）			STEP
					MAE_t	RMSE_t	MedAE_t	MAE_l	RMSE_l	MedAE_l	
③NB_CA_73	個別	10	50	1.1	32.56	58.19	18.00	224.28	384.04	180.55	22348.16
④NB_CA_74	同時	10	50	1.1	27.63	48.33	14.00	124.53	232.90	0.00	11254.89
③NB_CA_75	個別	10	75	1.1	32.56	58.19	18.00	245.74	402.14	188.25	23401.10
④NB_CA_76	同時	10	75	1.1	26.63	47.37	14.00	130.66	234.52	0.00	11110.08
③NB_CA_77	個別	10	100	1.1	32.56	58.19	18.00	264.86	408.37	200.94	23763.98
④NB_CA_78	同時	10	100	1.1	26.81	47.62	14.00	140.26	236.10	86.06	11241.92
①NB_CA_79	個別	0	0	2	36.14	57.18	21.00	172.92	280.98	172.85	16066.63
②NB_CA_80	同時	0	0	2	54.19	98.96	24.00	183.25	299.91	0.00	29679.03
③NB_CA_81	個別	3	25	2	33.72	61.38	17.00	172.92	280.98	172.85	17246.98
④NB_CA_82	同時	3	25	2	32.49	56.51	16.00	137.55	254.56	0.00	14384.06
③NB_CA_83	個別	3	50	2	33.72	61.38	17.00	170.40	280.06	166.24	17190.95
④NB_CA_84	同時	3	50	2	32.43	56.50	16.00	136.71	253.97	0.00	14348.17
③NB_CA_85	個別	3	75	2	33.72	61.38	17.00	180.58	289.69	172.85	17782.06
④NB_CA_86	同時	3	75	2	32.38	56.38	16.00	134.24	250.73	0.00	14136.83
③NB_CA_87	個別	3	100	2	33.72	61.38	17.00	197.86	299.56	180.55	18387.84
④NB_CA_88	同時	3	100	2	31.49	54.75	16.00	131.72	248.04	0.00	13580.47
③NB_CA_89	個別	5	25	2	33.61	61.07	16.00	172.92	280.98	172.85	17158.93
④NB_CA_90	同時	5	25	2	31.31	57.60	15.00	132.05	249.07	0.00	14347.41
③NB_CA_91	個別	5	50	2	33.61	61.07	16.00	170.40	280.06	166.24	17103.19
④NB_CA_92	同時	5	50	2	31.34	57.63	15.00	131.78	248.88	0.00	14343.64
③NB_CA_93	個別	5	75	2	33.61	61.07	16.00	180.58	289.69	172.85	17691.28
④NB_CA_94	同時	5	75	2	31.17	57.54	15.00	129.53	246.26	0.00	14170.81
③NB_CA_95	個別	5	100	2	33.61	61.07	16.00	197.86	299.56	180.55	18293.96
④NB_CA_96	同時	5	100	2	31.03	57.55	15.00	128.63	245.13	0.00	14106.78
③NB_CA_97	個別	10	25	2	33.39	60.49	17.00	172.92	280.98	172.85	16994.87
④NB_CA_98	同時	10	25	2	31.10	58.60	14.00	132.68	250.47	0.00	14678.65
③NB_CA_99	個別	10	50	2	33.39	60.49	17.00	170.40	280.06	166.24	16939.66
④NB_CA_100	同時	10	50	2	30.83	57.77	14.00	132.04	249.80	0.00	14430.07
③NB_CA_101	個別	10	75	2	33.39	60.49	17.00	180.58	289.69	172.85	17522.13
④NB_CA_102	同時	10	75	2	30.34	57.31	14.00	128.87	247.22	0.00	14167.57
③NB_CA_103	個別	10	100	2	33.39	60.49	17.00	197.86	299.56	180.55	18119.05
④NB_CA_104	同時	10	100	2	29.79	56.72	14.00	129.06	245.88	0.00	13946.59

表 B3.10　ナイーブベイズ多変数ベルヌーイモデルによる年代推定・場所推定（自治州）の結果

付録C自作プログラム

本研究では，コーパス作成から文書分類まですべての作業と実験を Excel VBA（Visual Basic for Application）の自作プログラムにより実施した[34]。実験環境の OS は, Windows 8.1（64bit）, CPU は Intel（R）Core（TM）i7-4700MQ CPU @ 2.40GHz, メモリは 8.00GB である。筆者が作成したプログラムは大きく分けて三つある。一つ目は Word で作成されたデータを Excel へ読み込むプログラム，二つ目は文字 *n*-gram 抽出を行うプログラム，三つ目はカーネル平滑化と文書分類を行うプログラムである。

以下，本研究で用いたデータセット（コーパス）とこれら三つのプログラムについて説明する。

C1　データセット

CODEA のデータセットは，スペインのアルカラ大学の Sánchez-Prieto Borja 教授から提供していただいた Word ファイルに，筆者が手を加えて作成した。各文書は 20 個の属性を持ち，構造化されている。#1 は文書の ID，#2 は文書の属するコーパスの作成グループ名，#3 は文書の属するコーパス名，#4 は文書の保管されている公文書館，#5 は文書の作成年代，#6 は文書の作成場所，#7 は文書の作成県，#8 は文書の作成自治州，#9 は文書の作成国，#10 と#11 は文書の種類，#12 は文書が原本か写しか，#13 は文書で使用されている文字の種類，#14 は文書の写字生，#15 は文書の送り主や受け取り主に女性が含まれるかどうか，#16 は文書内容の要約，#17 は文書のトランスクリプションを行った研究者，#18 はキーワード，#19 はパレオグラフィカ（versión paleográfica），#20 はクリティカ（presentación crítica）である。欠損値は「*」で示した。これらの属性のうち，本研究では，#1（文書の ID），#5（文書の作成年代），#6（文書の作成場所），#7（文書の作成県），#8（文書の作成自治州），#19（パレオグラフィカ），#20（クリティカ）のみに注目した。

例として，以下に，文書 ID1（作成年代 : 1251 年，県 : Sevilla，自治州 : AN）を示す。

#1:0001
#2:GITHE
#3:CODEA+ 2015
#4:AMGU, 1H1.1
#5:1251 abril 13
#6:Sevilla
#7:Sevilla
#8:*
#9:España
#10:Cancilleresco
#11:Textos legislativos: carta plomada
#12:*
#13:Letra de privilegios
#14:[Juán Pérez de Berlanga] (FACERE: *[it:Johannes Petri de Berlanga fecit:it]*)
#15:[Mujer: No]

[34] Word と Excel は米国 Microsoft 社の製品である。Microsoft, Windows, Word, Excel は米国 Microsoft 社の登録商標である。

#16:Carta plomada de Fernando III por la que confirma los fueros de Guadalajara, devuelve las aldeas que había segregado de su jurisdición y establece normas sobre los hombres buenos enviados ante el rey, juez que lleva la “seña”, las cofradías, los alcaldes y los matrimonios.

#17:Pedro Sánchez-Prieto Borja

#18:*[it:concejo de Guadalajara, hombres buenos, caveros, aldeas, fuero, seña, cofradías, alcaldes, bodas:it]*

#19:TRANSCRIPCIÓN PALEOGRÁFICA

{h 1r} {1} Connoscida cosa sea a todos los q<ue> esta carta uieren como yo don Fferrando por la gr<aci>a de dios Rey de Castiella de Toledo de Leon de Gallizia de Seuilla de Cordoua de Murc<ia> & de Jahen enbie {2} mis cartas auos el Conceio de Guadalfaiara q<ue> enbiassedes u<uest>ros om<n>es buenos de u<uest>ro Conceio a mj. por cosas q<ue> auya de ueer & de fablar con uusco por bue<n> paramie<n>to de u<uest>ra villa. Et uos enbiastes u<uest>ros {3} om<n>es buenos ante mi. & yo fable conellos aq<ue>llas cosas q<ue> ente<n>di q<ue> eran bue<n> paramie<n>to dela tierra. Et ellos saliero<n> me bie<n> & recudiero<n> me bie<n> a todas las cosas q<ue> les yo dix. deguisa q<ue> les yo fuy so pagado. {4} Et esto passado rogaro<n> me & pidiero<n> me merçet por su villa q<ue> les touyesse aq<ue>llos fueros & aq<ue>lla uida et aq<ue>llos usos q<ue> ouyeran [*[it:mano 2: :it]*ouyeron] en tie<n>po del Rey do<n> Alfonso mio Auuelo & assu muerte assi como gelos yo {5} p<ro>meti & gelos otorgue q<ua>ndo fuy Rey de Casti<e>lla. q<ue> gelo ternia & gelos guardaria ante mi Madre. & ante mios Ricos om<n>es. & antel Arçobispo & ante los Obispos. & ante Caueros d<e> Casti<e>lla & {6} de Estremadura. & ante toda mi corte. Et bie<n> connosco & es uerdat q<ue> q<ua>ndo yo era mas nin<n>o q<ue> aparte las Aldeas delas villas en algunos logares. & ala sazon q<ue> fiz esto. era me mas nin<n>o. & no<n> pare hy {7} tanto mie<n>tes. Et por q<ue> tenia q<ue> era cosa q<ue> deuya a eme<n>dar. oue mio co<n>seio co<n> don Alfonso mio fijo. & co<n> don Alfonso mio h<er>mano. & co<n> don diago lop<e>z. & co<n> don Nunno gonçaluez. & co<n> don Rodrigo Alfonso. & {8} co<n> el Obispo de Palençia. & co<n> el Obispo de Segouya. & co<n> el Maestro de Calatraua. & co<n> el Maestro de vcles. & con el Maestro del Temple. & con el gra<n>t Come<n>dador del Hospital. & co<n> otros Rycos {9} om<n>es & Caueros & om<n>es buenos de Casti<e>lla & de Leon. Et toue por d<er>echo & por razon de tornar las Aldeas alas villas. Assi como eran en dias de mio Auuelo & a su muerte. & q<ue> esse ffuero & esse d<er>echo {10} & essa uida ouyessen los delas Aldeas con los delas villas. & los delas villas con los de las Aldeas. q<ue> ouyero<n> en dias de mio Auuelo el Rey do<n> Alfonso & a su muerte. Et pues q<ue> esto les fiz & este {11} amor. & toue por d<er>echo de tornar las Aldeas alas villas. Mando otro si alos delas villas & deffiendo les so pena de mio amor & de mi gr<aci>a & delos cuerpos & de q<ua>nto q<ue> an. q<ue> ni<n>guno tanbie<n> Jurado {12} como Alcalde como ot<ro> Cauall<er>o ni<n>guno poderoso ni<n> otro q<ua>l q<ui>ere. de mala cuenta nj<n> de mal despechamie<n>to nj<n> mala p<re>mia nj<n> mala terreria nj<n> mal ffuero fiziesse alos pueblos. tanbie<n> dela villa como delas Aldeas {13} nj<n> les tomasse conducho a tuerto nj<n> a ffuerça. q<ue> yo q<ue> me tornasse a ellos a fazer les Justicia en los cuerpos & en los aueres & en q<ua>nto han. como om<n>es q<ue> tal yerro & tal tuerto & tal atreuimie<n>to {14} fazen a sennor. Et maguer yo entie<n>do q<ue> todo esto deuo a fazer & a uedar por mio debdo & por mio d<er>echo como sennor. plogo a ellos & otorgaro<n> melo. & touyero<n> q<ue> era d<er>echo q<ue> yo q<ue> diesse aq<ue>lla pena {15} q<ue> sobredicha es en los cuerpos & en los aueres. a aq<ue>llos q<ue> me errassen & tuerto me fiziessen a mios pueblos assi como sobredicho es en esta carta. Et ma<n>do & te<n>go por bie<n> q<ue> q<ua>ndo yo enbiare por om<n>es {16} de u<uest>ro Conceio q<ue> uengan a mj por cosas q<ue> ouyere de ffablar co<n> ellos. o q<ua>ndo q<ui>sieredes uos amj enbiar u<uest>ros om<n>es buenos de p<ro> de u<uest>ro Conceio. q<ue> uos catedes en u<uest>ro Conceio Caueros atales q<ua>les {17} touieredes por guisados de enbiar a mj. Et aq<ue>llos Caueros q<ue> en esta guisa tomaredes pora enbiar a mj. q<ue>les dedes despesa de Conceio en esta guisa. Que q<ua>ndo uiniere<n> fata Toledo q<ue> dedes a cada {18} Cauero medio m<o>r<abedi> cadadia. & non mas. Et de Toledo cont<ra> la ffrontera q<ue> dedes a cada Cauero un m<o>r<abedi> cadadia & no<n> mas. Et ma<n>do & deffiendo q<ue> estos q<ue> amj enbiaredes q<ue> non sean mas de tres fata

q<ua>tro {19} si no<n> si yo enbiasse por mas. Et otro si te<n>go por bie<n> & mando q<ue> q<ua>ndo yo enbiare por estos Caueros assi como sobredicho es. o el Conceio los enbiaredes a mj por p<ro> de u<uest>ro Conceio. q<ue> trayan cada Cauero tres tres {20} bestias & non mas. Et estas bestias q<ue> gelas apreçie<n> dos Jurados & dos Alcaldes q<ua>les el Conceio escogierre por esto. cada una q<ua>nto uale q<ua>ndo fazen la muebda del logar dont los enbia<n>. q<ue> si por aue<n>tura {21} alguna daq<ue>llas bestias muriere q<ue> sepades q<ue> auedes a dar el Conceio & el Pueblo por ella. & q<ue> dedes tanto por ella q<ua>nto fue apreçiada daq<ue>llos dos Jurados & dos Alcaldes assi como sobredicho es. Otro {22} si ma<n>do q<ue>los menestrales no<n> echen suerte en el Judgado por seer Juezes. ca el Juez deue tener la Senna & tengo q<ue> si affrue<n>ta uiniesse o alogar de periglo. & om<n>e uil o rafez touiesse la Senna. q<ue> podrie {23} caer el Conceio en gra<n>t onta. & en gra<n>t uerguença. & por end te<n>go por bie<n> q<ue> quila ouyere a tener q<ue> sea Cauero & om<n>e bueno & de uerguença. Et ot<ro> si se q<ue> en u<uest>ro Conceio q<ue> se fazen unas Confrad<ri>as {24} & unos ayuntamie<n>tos malos a me<n>gua de mio poder & de mio sennorio. & a danno de u<uest>ro Conceio & del pueblo. ho se fazen muchas malas encubiertas & malos paramie<n>tos. Et ma<n>do so pena de {25} los cuerpos & de q<ua>nto auedes. q<ue> estas co<n>fradrias q<ue>las desfagades. & q<ue> daq<ui> adelante no<n> las fagades fuera en tal manera pora soterrar muertos & pora luminarias. p<or>a dar a pobres. & pora co<n>fuerços. {26} mas q<ue> non pongades Alcaldes entre uos nj<n> coto malo. Et pues q<ue> uos do carrera poro fagades bie<n> & almosna & merced co<n> derecho. si uos a mas q<ui>siessedes passar. a otros cotos o aotros paramie<n>tos {27} o aponer Alcaldes. alos cuerpos & a q<ua>nto ouiessedes me tornaria por ello. Et ma<n>do q<ue> ni<n>guno no<n> sea osado de dar nj<n> de tomar calças por casar so parie<n>ta ca el q<ue>las tomasse pechar las hye dupladas al q<ue> {28} gelas diesse & pecharie Cinq<ue>nta m<o>r<abedis> en coto. los veynte & amj & los diez alos Jurados & los diez alos Alcaldes. & los ot<ro>s diez al q<ue>los descubriesse con uerdat. Et ma<n>do q<ue> todo om<n>e q<ue> casare co<n> {29} ma<n>çeba en cabello q<ue> nol de mas de Sessaenta m<o>r<abedis> pora pannos pora sus bodas. Et q<ui> casare con Bibda. nol de mas de q<ua>renta m<o>r<abedis> pora pannos pora sus bodas. Et q<ui> mas diesse desto q<ue> yo ma<n>do. pecha{30}rie Cinq<ue>nta mor<a>b<edi>s en coto. Los veynte a mj. & los diez Alos Jurados & los diez Alos Alcaldes. & los otros diez al q<ue>los mesturasse. Et ot<ro> si ma<n>do q<ue> no<n> coma<n> alas bodas mas de diez om<n>es. cinco dela {31} parte del Nouyo. & cinco dela parte dela Nouya q<ua>les el Nouyo & la Nouia q<ui>siere<n>. Et q<ua>ntos de mas hy comiessen pechar mie cada uno. Diez m<o>r<abedis> los siete amj & los tres aq<ui> los descubriesse. & esto {32} sea a buena fe sin escatima & sin cobdicia ni<n>guna. Et ma<n>do q<ue>las ot<ra>s cartas q<ue> yo dj tanbie<n> alos dela villa como delas Aldeas. q<ue>las aldeas fuessen Appartadas dela villa & la villa delas Aldeas. {33} q<ue> no<n> ualan. Et ma<n>do & deffiendo firme mie<n>tre q<ue> ni<n>guno no<n> sea osado de uenir cont<ra> esta mi carta nj<n> de quebrantar la nj<n> de me<n>guar la en ni<n>guna cosa ca el q<ue>lo fiziesse aurye la yra de dios & la {34} mia. & pechar mie en coto. mill. m<o>r<abedis>. [*[it:lat.:it]*: [!ffacta carta ap<u>d Sibilla Reg<e> exp<rimente>. xiij. die Ap<ri>lis.!] J<ohannes>. pet<ri> de Berla<n>ga fe<ci>t. [!ERA_M_CC_Lxxx_Nona!]]

#20:PRESENTACIÓN CRÍTICA

{h 1r} {1} Coñocida cosa sea a todos los que esta carta vieren cómo yo don Ferrando, por la gracia de Dios rey de Castiella, de Toledo, de León, de Gallizia, de Sevilla, de Córdova, de Murcia e de Jaén, embié {2} mis cartas a vós el concejo de Guadalfajara que embiássedes vuestros omnes buenos de vuestro concejo a mí, por cosas que avía de veer e de fablar convusco por buen paramiento de vuestra villa. E vós embiastes vuestros {3} omnes buenos ante mí, e yo fablé con ellos aquellas cosas que entendí que eran buen paramiento de la tierra. E ellos saliéronme bien e recudiéronme bien a todas las cosas que les yo dix, de guisa que les yo fui so pagado. {4} E esto passado rogáronme e pidiéronme mercet por su villa que les toviesse aquellos fueros e aquella vida e aquellos usos que ovieran en tiempo del rey don Alfonso mio avuelo e a su muerte, assí como gelos yo {5} prometí e gelos otorgué cuando fui rey de Castiella que gelo ternía e gelos guardaría ante mi madre, e ante mios ricos omnes, e ant'el arçobispo, e ante los obispos, e ante caveros de Castiella e {6} de Estremadura e ante toda mi corte. E bien coñosco e es verdat que cuando yo era más niño que aparté las aldeas de las villas en algunos logares. E a la

sazón que fiz esto érame más niño e non paré y {7} tanto mientes. E porque tenía que era cosa que devía a emendar, ove mio consejo con don Alfonso mio fijo, e con don Alfonso mio hermano, e con don Diago López, e con don Nuño Gonçálvez, e con don Rodrigo Alfonso, e {8} con el obispo de Palencia, e con el obispo de Segovia, e con el maestro de Calatrava, e con el maestro de Uclés, e con el maestro del Temple, e con el grant comendador del Hospital, e con otros ricos {9} omnes e caveros e omnes buenos de Castiella e de León, e tove por derecho e por razón de tornar las aldeas a las villas, assí como eran en días de mio avuelo e a su muerte, e que esse fuero e esse derecho {10} e essa vida oviessen los de las aldeas con los de las villas e los de las villas con los de las aldeas que ovieron en días de mio avuelo el rey don Alfonso e a su muerte. E pues que esto les fiz e este {11} amor, e tove por derecho de tornar las aldeas a las villas, mando otrossí a los de las villas e defiéndoles, so pena de mio amor e de mi gracia, e de los cuerpos e de cuanto que an, que ninguno, tan bien jurado {12} como alcalde como otro cavallero ninguno poderoso nin otro cualquiere, dé mala cuenta, nin dé mal despechamiento, nin mala premia, nin mala terrería, nin mal fuero fiziesse a los pueblos, tan bien de la villa como de las aldeas, {13} nin les tomasse conducho a tuerto nin a fuerça, que yo que me tornasse a ellos a fazerles justicia en los cuerpos e en los averes e en cuanto an, como omnes que tal yerro e tal tuerto e tal atrevimiento {14} fazen a señor. E maguer yo entiendo que todo esto devo a fazer e a vedar por mio debdo e por mio derecho como señor, plogo a ellos e otorgáronmelo, e tovieron que era derecho que yo que diesse aquella pena {15} que sobredicha es en los cuerpos e en los averes a aquellos que me errassen e tuerto me fiziessen a mios pueblos, assí como sobredicho es en esta carta. E mando e tengo por bien que cuando yo embiare por omnes {16} de vuestro concejo que vengan a mí por cosas que oviere de fablar con ellos, o cuando quisiéredes vós a mí embiar vuestros omnes buenos de pro de vuestro concejo, que vós catedes en vuestro concejo caveros atales cuales {17} toviéredes por guisados de embiar a mí, e aquellos caveros que en esta guisa tomáredes pora embiar a mí que les dedes despesa de concejo en esta guisa: que cuando vinieren fata Toledo, que dedes a cada {18} cavero medio morabedí cada día e non más; e de Toledo contra la frontera que dedes a cada cavero un morabedí cada día e non más. E mando e defiendo que estos que a mí embiáredes que non sean más de tres fata cuatro, {19} si non si yo embiasse por más. E otrossí tengo por bien e mando que cuando yo embiare por estos caveros, assí como sobredicho es, o el concejo los embiáredes a mí por pro de vuestro concejo, que trayan cada cavero tres tres {20} bestias e non más. E estas bestias que gelas aprecien dos jurados e dos alcaldes cuales el concejo escogiere por esto, cada una cuanto vale cuando fazen la muebda del logar dont los embían, que si por aventura {21} alguna d'aquellas bestias muriere, que sepades qué avedes a dar el concejo e el pueblo por ella, e que dedes tanto por ella cuanto fue apreciada d'aquellos dos jurados e dos alcaldes, assí como sobredicho es. Otrossí {22} mando que los menestrales non echen suerte en el judgado por seer juezes, ca el juez deve tener la seña, e tengo que si <a> afruenta viniesse o a logar de periglo e omne vil o rafez toviesse la seña que podrié {23} caer el concejo en grant onta e en grant vergüença. E por end tengo por bien que qui la oviere a tener que sea cavero e omne bueno e de vergüença. E otrossí sé que en vuestro concejo que se fazen unas confradrías {24} e unos ayuntamientos malos a mengua de mio poder e de mio señorío, e a daño de vuestro concejo e del pueblo, ó se fazen muchas malas encubiertas e malos paramientos. E mando, so pena de {25} los cuerpos e de cuanto avedes, que estas confradrías que las desfagades, e que d'aquí adelante non las fagades, fuera en tal manera pora soterrar muertos e pora luminarias, pora dar a pobres e pora confuerços. {26} Más que non pongades alcaldes entre vós nin coto malo. E pues que vos dó carrera por ó fagades bien, e almosna e merced con derecho, si vós a más quisiéssedes passar a otros cotos o a otros paramientos, {27} o a poner alcaldes, a los cuerpos e a cuanto oviéssedes me tornaría por ello. E mando que ninguno non sea osado de dar nin de tomar calças por casar so parienta, ca el que las tomasse pechar las ié dupladas al que {28} gelas diesse, e pecharié cincuenta morabedís en coto, los veinte a mí, e los diez a los jurados, e los diez a los alcaldes, e los otros diez al que los descubriesse con verdat. E mando que todo omne que casar con {29} manceba en cabello que no.l dé más de sessaenta morabedís pora paños pora sus bodas. E qui casare con bibda no.l dé más de cuarenta morabedís pora paños pora sus bodas. E qui más diesse d'esto que yo mando pecharié {30} cincuenta morabedís en coto, los veinte a mí, e los diez a los jurados, e los diez a los alcaldes e los otros diez al que los mesturasse. E otrossí mando que non coman a las bodas más de diez omnes, cinco de la {31} parte del novio e cinco de la parte de la novia, cuales el novio e la novia quisieren. E cuantos de más y comiessen pechar m'ié cada

uno diez morabedís, los siete a mí e los tres a qui los descubriesse; e esto {32} sea a buena fe sin escatima e sin cobdicia ninguna. E mando que las otras cartas que yo di tan bien a los de la villa como de las aldeas que las aldeas fuessen apartadas de la villa e la villa de las aldeas {33} que non valan. E mando e defiendo firmemientre que ninguno non sea osado de venir contra esta mi carta nin de quebrantarla nin de menguarla en ninguna cosa, ca el que lo fiziesse avrié la ira de Dios e la {34} mía, e pechar m'ié en coto mill morabedís. [!*[it:Facta carta apud Sibilla, rege exprimente, XIII die aprilis.!] Johannes Petri de Berlanga fecit. [!Era MCCLXXX nona.!]:it]*

C2　Word の Excel への読み込み

本研究では，Word で作成されたデータを Excel へ読み込んだ上で実験を行った。Word を Excel に読み込むプログラムのコードは，以下である：

```
Option Explicit
Dim Str1$, Str2$, Ptrn1$, Ptrn2$ '文字列型変数
Dim ObjDicRgn, ObjDicPrv, Ar1, Vr1, Ar2, key 'バリアント型変数
Dim ColNo%, OpOpt%, NumAtr%, Pos% '整数型変数
Const ColWid% = 80 '出力最大列幅
Sub ■Data_Import() 'データインポート

  Application.ScreenUpdating = False '画面表示更新ストップ
  Set ObjDic1 = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
  Set ObjDicRgn = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
  Set ObjDicPrv = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
  Set ObjRE = CreateObject("VBScript.RegExp") '正規表現オブジェクトを生成
    ObjRE.ignorecase = True '大小文字区別なし
    ObjRE.Global = True 'グローバル

  With New DataObject 'クリップボードデータ
    .GetFromClipboard：Str1$ = CStr(.GetText) '格納
  End With
  If Str1$ = "" Then MsgBox "Please select data!"：Exit Sub 'データが存在しないなら

  Str1$ = ObjReRep$(Str1$, "(¥r¥n)", "") '改行削除
  Str1$ = ObjReRep$(Str1$, "(¥{|#)", vbCrLf & "$1") '改行挿入
  Str1$ = ObjReRep$(Str1$, "(¥r¥n)+", vbCrLf) '改行連続は 1 つまで
  Str1$ = ObjReRep$(Str1$, "(^(¥r¥n)|(¥r¥n)$)", "") '先頭末尾の改行削除
  Str1$ = Trim(Replace(Str1$, vbTab, " ")) 'タブをスペース＆トリム

  ObjRE.Pattern = vbCrLf 'パターン
  Set Matches = ObjRE.Execute(Str1$) "'改行数をカウント
```

```
If Matches.Count = 0 Then MsgBox "Incorrect data!" : Exit Sub 'データが存在しないなら

Ar1 = Split(Str1$, vbCrLf) '改行記号でスプリット
NumAtr% = 0 '初期値
For Each Vr1 In Ar1 '各要素について
   Vr1 = Trim(Vr1) 'トリム
   If Left(Vr1, 1) = "#" Then '先頭が#なら
      Pos% = InStr(Vr1, ":") '位置
      If Trim(Mid(Vr1, 2, Pos% - 2)) > NumAtr% Then
         NumAtr% = Trim(Mid(Vr1, 2, Pos% - 2)) '属性の個数
      End If
   End If
Next

IP2 = Sheets("Prv>Prv").UsedRange 'Prv>Prv シート
For m% = 1 To UBound(IP2, 1) 'Provincia の数まで
   If Not ObjDicPrv.exists(IP2(m%, 1)) Then '新しい Provincia ならば
      ObjDicPrv.Add IP2(m%, 1), IP2(m%, 2) '連想配列に登録
   End If
Next
Erase IP2 '配列消去

IP2 = Sheets("Prv>Rgn").UsedRange 'Prv>Rgn シート
For m% = 1 To UBound(IP2, 1) 'Provincia の数まで
   If Not ObjDicRgn.exists(IP2(m%, 1)) Then '新しい Provincia ならば
      ObjDicRgn.Add IP2(m%, 1), IP2(m%, 2) '連想配列に登録
   End If
Next
Erase IP2 '配列消去

ReDim Ar2(1 To NumAtr%)   '配列サイズ再定義
For Each Vr1 In Ar1 '各要素について
   Vr1 = Trim(Vr1) 'トリム
   Pos% = InStr(Vr1, ":") '位置
   If Left(Vr1, 1) = "#" Then '先頭が#なら
      Ar2(Mid(Vr1, 2, Pos% - 2)) = Mid(Vr1, Pos% + 1)   '属性を格納

      If Left(Vr1, 3) = "#1:" Then '先頭が#1:ID なら
         If IsNumeric(Trim(Mid(Vr1, Pos% + 1))) Then 'ID が数値なら
```

```
            Str2$ = CInt(Trim(Mid(Vr1, Pos% + 1))) '格納
            Ar2(Mid(Vr1, 2, Pos% - 2)) = CInt(Mid(Vr1, Pos% + 1))   '属性を上書き
         Else 'ID が整数でないなら（文字列なら）
            Str2$ = Trim(Mid(Vr1, Pos% + 1)) '格納
         End If
      End If

      If Left(Vr1, 3) = "#5:" Then '先頭が#5:Fecha なら
         If Mid(Vr1, Pos% + 1) <> "Fecha" Then 'ラベルでなければ
            If IsNumeric(Left(Mid(Vr1, Pos% + 1), 4)) = True Then '年代ならば
               Ar2(Mid(Vr1, 2, Pos% - 2)) = Left(Mid(Vr1, Pos% + 1), 4) '年代を格納
            Else '年代が未知ならば
               Ar2(Mid(Vr1, 2, Pos% - 2)) = "s.a." '格納
            End If
         End If
      End If

      If Left(Vr1, 3) = "#7:" Then '先頭が#7:Provincia なら
         Ar2(7) = ObjDicPrv(Ar2(7)) '#7:Provincia から#7:Provincia へ変換
      End If

      If Left(Vr1, 3) = "#8:" Then '先頭が#8:Region なら
         Ar2(8) = ObjDicRgn(Ar2(7)) '#7:Provincia から#8:Region へ変換
      End If

      If Left(Vr1, 4) = "#" & NumAtr% & ":" Then '先頭が#20 なら
         If Not ObjDic1.exists(Str2$) Then '新しい ID ならば
            ObjDic1.Add Str2$, Ar2   '連想配列に登録
            Erase Ar2 :  ReDim Ar2(1 To NumAtr%)   '配列サイズ再定義
         End If
      End If
   End If
Next

'属性の修正----------------------------
IP = Sheets("Modificacion").UsedRange '修正箇所格納
For i& = 2 To UBound(IP, 1) '修正箇所まで繰り返し
   Str2$ = IP(i&, 1) '該当文書の ID 格納
      If ObjDic1.exists(Str2$) = True Then 'ID が存在すれば
```

```
            ReDim Ar2(1 To NumAtr%)
            For l% = 1 To NumAtr% '属性の数だけ
               Ar2(l%) = ObjDic1(Str2$)(l%)    '属性コピー
            Next

            If IP(i&, 2) = 5 Then '年代#5:なら
               Ar2(IP(i&, 2)) = Left(IP(i&, 3), 4)    '上書き
            Else 'それ以外なら
               Ar2(IP(i&, 2)) = Trim(IP(i&, 3))    '上書き
            End If
            Ar2(7) = ObjDicPrv(Ar2(7)) '#7:Provincia から#7:Provincia へ変換
            Ar2(8) = ObjDicRgn(Ar2(7)) '#7:Provincia から#8:Region へ変換

            ObjDic1.Remove (Str2$) '連想配列から削除
            ObjDic1.Add Str2$, Ar2    '連想配列に登録
            Erase Ar2 '配列消去
         End If
      Next
      Erase IP '配列消去
      '-------------------------------------------------

      ReDim OP(1 To Matches.Count, 1 To NumAtr%) '配列サイズ再定義
      i& = 2 : k& = 2 '初期値
      OpOpt% = 1920 '19:VP, 20:PC, 1920:VP&CP

      For Each Vr1 In Ar1 '各行について
         If Left(Vr1, 1) = "#" Then '先頭が#なら
            Pos% = InStr(Vr1, ":") '位置
            If Left(Vr1, 3) = "#1:" Then '先頭が#1:なら
               If IsNumeric(Trim(Mid(Vr1, Pos% + 1))) Then 'ID が数値なら
                  Str2$ = CInt(Trim(Mid(Vr1, Pos% + 1))) '格納
               Else 'ID が整数でないなら（文字列なら）
                  Str2$ = Trim(Mid(Vr1, Pos% + 1)) '格納
               End If
               If i& > k& Then k& = i& '行番号調整
               If i& < k& Then i& = k& '行番号調整
            End If
            ColNo% = Mid(Vr1, 2, Pos% - 2) '列番号取得
         Else '先頭が#でないなら
```

```
        If OpOpt% = ColNo% Then 'VPorPC を出力なら
          For l% = 1 To ColNo% - 1
            OP(i&, l%) = ObjDic1(Str2$)(l%) '属性格納
          Next
          OP(i&, ColNo%) = Vr1 'テキスト格納
          i& = i& + 1 'インクリメント
        ElseIf OpOpt% = 1920 Then 'VP&PC を出力なら
            If ColNo% = 19 Then 'VP なら
              For l% = 1 To ColNo% - 1
                OP(i&, l%) = ObjDic1(Str2$)(l%) '属性格納
              Next
              OP(i&, ColNo%) = Vr1 'テキスト格納
              i& = i& + 1 'インクリメント
            ElseIf ColNo% = 20 Then 'PC なら
              OP(k&, ColNo%) = Vr1 'テキスト格納
              If i& - 1 < k& Then
                For l% = 1 To ColNo% - 2
                  OP(k&, l%) = ObjDic1(Str2$)(l%) '属性格納
                Next
              End If
              k& = k& + 1 'インクリメント
            End If
        End If
    End If
Next

For l% = 1 To UBound(OP, 2) '出力列まで繰り返し
    OP(1, l%) = ObjDic1("ID")(l%) 'タイトル行
Next

'出力----------------------------------------------
Sheets.Add after:=Sheets(Sheets.Count) 'シート挿入
With ActiveSheet
    .Range([A1], Cells(UBound(OP, 1), UBound(OP, 2))) = OP 'ペースト
    .UsedRange.Select '全範囲選択
        With Selection:
          .Columns.AutoFit：.Rows.AutoFit '行列幅自動調整
          For l% = 1 To UBound(OP, 2)
            If .Columns(l%).ColumnWidth > ColWid% Then .Columns(l%).ColumnWidth = ColWid% '列幅調整
```

```
        Next
        .WrapText = True
    End With
  Range("B2").Select：ActiveWindow.FreezePanes = True   'ウィンドウ枠固定
End With

'属性表出力-------------------------------------------------------
Erase OP：ReDim OP(1 To ObjDic1.Count, 1 To NumAtr%) '配列サイズ再定義
i& = 1 '初期値
  For Each key In ObjDic1.Keys
    For j& = 1 To UBound(OP, 2)
      OP(i&, j&) = ObjDic1(key)(j&)
    Next
    i& = i& + 1 'インクリメント
  Next
  Sheets.Add after:=Sheets(Sheets.Count) 'シート挿入
  With ActiveSheet
    .Range([A1], Cells(UBound(OP, 1), UBound(OP, 2))) = OP 'ペースト
    .UsedRange.Select '全範囲選択
      With Selection:
        .Columns.AutoFit：.Rows.AutoFit '行列幅自動調整
        For l% = 1 To UBound(OP, 2)
          If .Columns(l%).ColumnWidth > ColWid% Then .Columns(l%).ColumnWidth = ColWid% '列幅調整
        Next
        .WrapText = True
      End With


    '並び替え----------------------------------
    With .Sort
      .SortFields.Clear
      .SortFields.Add key:=Range("A2"), _
          SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:=xlSortNormal
      .SetRange Range(Cells(2, 1), Cells(UBound(OP, 1), UBound(OP, 2))) '範囲設定
          .Header = xlNo：.Orientation = xlTopToBottom：.Apply
    End With
    '--------------------------
    Range("B2").Select：ActiveWindow.FreezePanes = True   'ウィンドウ枠固定
  End With
```

```
    '-------------------------------------------------------
    Set ObjDic1 = Nothing 'ディクショナリオブジェクトの解放
    Set ObjDicRgn = Nothing 'ディクショナリオブジェクトの解放
    Set ObjDicPrv = Nothing 'ディクショナリオブジェクトの解放
    Set ObjRE = Nothing '正規表現オブジェクトの解放
    Erase Ar1 : Erase Ar2 : Erase OP '配列消去
    Application.ScreenUpdating = True '画面表示を更新
End Sub

Function ObjReRep$(Str1$, Ptrn1$, Ptrn2$) '正規表現置換
    ObjRE.Pattern = Ptrn1$ '検索パターン
    Str1$ = ObjRE.Replace(Str1$, Ptrn2$) '置換
    ObjReRep = Str1$
End Function
```

C3　文字 *n*-gram 抽出

文字 *n*-gram 抽出のプログラムのコードは，以下である：

```
Option Explicit
'Dim ObjDic1 As Object, ObjRE As Object, objRE2 As Object, objRE3 As Object, Match As Object, Matches As Object,
Matches2 As Object
'Dim i&, j&, k&, h&, l%
Dim D(1 To 2500, 1 To 10) '配列
Public RE, Ltr1, Ltr2, Ltr3, Ltr4, NLtr, Query1 'バリアント型変数
Dim Str1$, Str2$, StrRE2$ '文字列型変数
Public ColTxt%, ColID%, ColFch%, ColTpl%, ColLgr%, ColPrv%, ColCA%, NumDoc%, Ngram% '整数型変数
Dim Sum1&, Vcb&    '長整数型変数

Sub ■Ngram() 'Ngram 抽出
    '準備-----------------------------------------------------------------------
    If ShtLstIdx% <= 0 Then MsgBox "シートを選択してください。" : Exit Sub 'リストボックスが選択されていなければ
    Application.ScreenUpdating = False '画面表示更新ストップ
    Set ObjDic1 = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
    Set ObjRE = CreateObject("VBScript.RegExp") '正規表現オブジェクトを生成
        ObjRE.ignorecase = Kwsk.chkNgramLcase '大小文字区別
        ObjRE.Global = True 'グローバル
    Set objRE2 = CreateObject("VBScript.RegExp") '正規表現オブジェクトを生成
    Set objRE3 = CreateObject("VBScript.RegExp") '正規表現オブジェクトを生成
        objRE3.ignorecase = Kwsk.chkNgramLcase '大小文字区別
```

```
    objRE3.Global = True 'グローバル

If SheetExists("RE") = True Then 'シート「RE」があれば選択して
    For h& = 2 To ActiveSheet.UsedRange.Rows.Count '列数まで
        If Sheets("RE").Cells(h&, 1) = "¥L" Then NLtr = Cells(h&, 2) '¥L を再定義
        If Sheets("RE").Cells(h&, 1) = "¥l1" Then Ltr1 = Cells(h&, 2) '¥l1 を再定義
        If Sheets("RE").Cells(h&, 1) = "¥l2" Then Ltr2 = Cells(h&, 2) '¥l2 を再定義
        If Sheets("RE").Cells(h&, 1) = "¥l3" Then Ltr3 = Cells(h&, 2)：Exit For '¥l3 を再定義
    Next
Else：MsgBox "「RE」シートがありません"：Exit Sub
End If

'前処理----------------------------------------------------------------------------
Sheets(ShtLstIdx%).Select 'シート選択
'Sheets("corpus_zikken").Select 'zikken
IP = ActiveSheet.UsedRange '使用範囲格納
For h& = 1 To UBound(IP, 2) '文書属性の列番号取得
    If IP(1, h&) = Lbl(2, 1) Then D(1, 1) = IP(1, h&)：ColID% = h& 'ID
    If IP(1, h&) = Lbl(11, 1) Then D(1, 2) = IP(1, h&)：ColTpl% = h& 'Tipologia
    If IP(1, h&) = Lbl(6, 1) Then D(1, 3) = IP(1, h&)：ColFch% = h& 'Fecha
    If IP(1, h&) = Lbl(7, 1) Then D(1, 4) = IP(1, h&)：ColLgr% = h& 'Lugar
    If IP(1, h&) = Lbl(8, 1) Then D(1, 5) = IP(1, h&)：ColPrv% = h& 'Provincia
    If IP(1, h&) = Lbl(9, 1) Then D(1, 6) = IP(1, h&)：ColCA% = h& 'CA

    If Kwsk.optNgramWord = True Then '単語 Ngram まらば
        If IP(1, h&) = Lbl(21, 1) Then ColTxt% = h& 'Presentacion Critica
    Else '文字 Ngram もしくは省略を含む単語の Ngram ならば
        If IP(1, h&) = Lbl(20, 1) Then ColTxt% = h& 'Transcripcion Paleografica
    End If
Next

If Sqr(ColTxt%) * Sqr(ColTpl%) * Sqr(ColFch%) * Sqr(ColLgr%) * Sqr(ColPrv%) * Sqr(ColCA%) = 0 Then '積がゼロなら
    MsgBox "該当列が存在しません。シートが正しいか確認してください。"：Exit Sub
End If

'文書を配列 D に格納
Str1$ = IP(2, ColTxt%) 'Texto 初期値
NumDoc% = 1 '文書数初期値
For i& = 2 To UBound(IP, 1) '行数まで繰り返し
```

```
    If i& <> UBound(IP, 1) Then '最終行でなければ
      If IP(i&, ColID%) = IP(i& + 1, ColID%) Then 'ID が同じなら
        Str1$ = Str1$ & IP(i& + 1, ColTxt%) '文書結合
      Else 'ID が異なるなら，D に登録
        D(NumDoc%, 1) = IP(i&, ColID%) 'ID
        D(NumDoc%, 2) = IP(i&, ColFch%) 'Fecha
        D(NumDoc%, 3) = IP(i&, ColPrv%) 'Provincia
        D(NumDoc%, 4) = IP(i&, ColCA%) 'CA
        D(NumDoc%, 5) = IP(i&, ColTpl%) 'Tipologia
        D(NumDoc%, 6) = Rep1(Str1$) 'Texto
        NumDoc% = NumDoc% + 1 : Str1$ = "" : Str1$ = IP(i& + 1, ColTxt%)
      End If
    ElseIf i& = UBound(IP, 1) Then '最終行なら
        D(NumDoc%, 1) = IP(i&, ColID%) 'ID
        D(NumDoc%, 2) = IP(i&, ColFch%) 'Fecha
        D(NumDoc%, 3) = IP(i&, ColPrv%) 'Provincia
        D(NumDoc%, 4) = IP(i&, ColCA%) 'CA
        D(NumDoc%, 5) = IP(i&, ColTpl%) 'Tipologia
        D(NumDoc%, 6) = Rep1(Str1$) 'Texto
    End If
Next

ReDim OP(1 To 100000, 1 To NumDoc% + 2) '配列サイズ再定義
OP(1, 1) = Lbl(2, 1) : OP(2, 1) = Lbl(6, 1) : OP(3, 1) = Lbl(8, 1):
OP(4, 1) = Lbl(9, 1) : OP(5, 1) = Lbl(11, 1) '列タイトル

'検索 ----------------------------------------------------------------------------
    Atr1% = 4 '文書属性数
    Ngram% = Kwsk.txtNgram 'N を指定
    Vcb& = 0 : Query1 = "" '初期化

    For l% = 1 To Ngram% 'N まで繰り返し
      If Kwsk.optNgramWord = True Then '単語 Ngram まらば
        Query1 = Query1 & Ltr1 & " " '単語 Ngram
      ElseIf Kwsk.optNgramChr = True Then '文字 Ngram ならば
        Query1 = Query1 & Ltr2 '文字 Ngram
      ElseIf Kwsk.optNgramAbbr = True Then '省略を含む単語の Ngram ならば
        Query1 = Query1 & Ltr3 & "<" & Ltr3 & ">" & Ltr3 & " "    '省略を含む単語 Ngram
      End If
```

```
Next：ObjRE.Pattern = Trim(Query1) '正規表現検索パターン

For h& = 1 To NumDoc%　'文書数 NumDoc%まで繰り返し
  Str2$ = "" '初期化
  If Kwsk.chkNgramLcase = True Then
    Str2$ = LCase(D(h&, 6)) '小文字に変換
  Else：Str2$ = D(h&, 6)：End If
  For i& = 1 To Atr1% + 1 '文書の属性数まで繰り返し
    OP(i&, h& + 1) = D(h&, i&) '文書の属性取得
  Next

  '語末の Ngram 検索漏れを防ぐ
  If Ngram% >= 2 Then '2gram 以上なら
    If Kwsk.optNgramWord = True Then '単語 Ngram ならば
      Str2$ = Str2$ & " ######"　　'文末記号を追加
    ElseIf Kwsk.optNgramChr = True Then '文字 Ngram ならば
      Str2$ = Str2$ & "##########"　'文末記号を追加
    End If
  End If

  For l% = 1 To Ngram% 'N まで繰り返し
    If Kwsk.optNgramWord = True Then '単語 Ngram ならば
      If Ngram% >= 2 Then '2gram 以上なら
        Str2$ = "# " & Str2$ & " #" '文頭記号と文末記号を追加
      End If
    ElseIf Kwsk.optNgramChr = True Then '文字 Ngram ならば
      If Ngram% >= 2 Then '2gram 以上なら
        Str2$ = "#" & Str2$ & "#" '文頭記号と文末記号を追加
      End If
    ElseIf Kwsk.optNgramAbbr = True Then '省略を含む単語の Ngram ならば
      If Ngram% >= 2 Then '2gram 以上なら
        Str2$ = "# " & Str2$ & " #" '文頭記号と文末記号を追加
      End If
    End If

    Set Matches = ObjRE.Execute(Str2$) 'マッチング

    If Matches.Count > 0 Then '一つ以上マッチすれば
      For Each RE In Matches '各マッチについて
```

```
            If Right(RE, 2) <> "##" And Right(RE, 3) <> "# #" Then '文末記号の連続でなければ
              If Kwsk.chkNgramLcase = True Then
                RE = LCase(RE) '小文字に変換
              Else：RE = RE：End If

              If Kwsk.optNgramWord = True Then '単語 Ngram ならば
                  objRE3.Pattern = "(^| )" & RE & "(?=$| )" 'パターン"
              ElseIf Kwsk.optNgramChr = True Then '文字 Ngram ならば
                  objRE3.Pattern = RE    'パターン"
              ElseIf Kwsk.optNgramAbbr = True Then '省略を含む単語の Ngram ならば
                  objRE3.Pattern = "(^| )" & RE & "(?=$| )" 'パターン"
              End If

              If Not ObjDic1.exists(RE) Then    '新規の単語ならば
                Vcb& = Vcb& + 1 '単語数をインクリメント
                ObjDic1.Add RE, Vcb& 'ディクショナリーに追加
                OP(Atr1% + Vcb& + 1, 1) = RE '単語代入
                Set Matches2 = objRE3.Execute(Str2$) 'マッチング
                OP(Atr1% + Vcb& + 1, h& + 1) = Matches2.Count '頻度代入
              Else '既存のキーなら
                If OP(Atr1% + ObjDic1(RE) + 1, h& + 1) = "" Then '該当文書の未検索語ならば
                  Set Matches2 = objRE3.Execute(Str2$) 'マッチング
                  OP(Atr1% + ObjDic1(RE) + 1, h& + 1) = Matches2.Count '頻度代入
                End If
              End If
            End If
          Next
        End If
      Next
    Next
    If Vcb& = 0 Then MsgBox "出力がありません。"：Exit Sub '出力がなければ

    'データ整形----------------------------------------------
    If Ngram% = 1 Then '1 グラムならば
      For i& = 2 + Atr1% To Vcb& + Atr1% + 1
        If OP(i&, 1) = " " Then OP(i&, 1) = "Space"：Exit For "' "を"Space"に
      Next
    End If
```

```
    If Kwsk.optNgramQuant Then '量的データなら
      For i& = 2 + Atr1% To Vcb& + Atr1% + 1 'グラム数まで繰り返し
        For j& = 2 To NumDoc% + 1 '文書数まで繰り返し
          If OP(i&, j&) = "" Then OP(i&, j&) = 0 '空白セルに 0 を記入
        Next
      Next
    ElseIf Kwsk.optNgramQual Then '質的データなら
      For i& = 2 + Atr1% To Vcb& + Atr1% + 1 'グラム数まで繰り返し
        For j& = 2 To NumDoc% + 1 '文書数まで繰り返し
          If OP(i&, j&) >= 1 Then '1 以上ならば
            OP(i&, j&) = 1 '1 に変換
          Else：OP(i&, j&) = 0 '0 を代入
          End If
        Next
      Next
    End If

'頻度の小計----------------------------------------------
If Kwsk.chkSubTotal Then
    Sum1& = 0 '初期化
    For i& = Atr1% + 2 To Vcb& + Atr1% + 1 'グラム数まで繰り返し
      For j& = 2 To NumDoc% + 1 '文書数まで繰り返し
        Sum1& = Sum1& + OP(i&, j&) '行和
      Next
      OP(i&, NumDoc% + 2) = Sum1&：Sum1& = 0 '初期化
    Next

    For j& = 2 To NumDoc% + 2 '文書数まで繰り返し
      For i& = Atr1% + 2 To Vcb& + Atr1% + 1 'グラム数まで繰り返し
        Sum1& = Sum1& + OP(i&, j&) '列和
      Next
      OP(Atr1% + Vcb& + 2, j&) = Sum1&：Sum1& = 0    '初期化
    Next
End If

'出力----------------------------------------------------------------
Sheets.Add after:=Sheets(Sheets.Count) '新シート挿入
Range([A1], Cells(Atr1% + Vcb& + 2, UBound(OP, 2))) = OP 'ペースト
Erase IP：Erase OP：Erase D '配列消去
```

```
    '書式設定
    ActiveSheet.UsedRange.Select '全範囲選択
    With Selection
        .Columns.AutoFit：.Rows.AutoFit '行列幅自動調整
    End With
    Cells(Atr1% + 2, 2).Select：ActiveWindow.FreezePanes = True 'ウィンドウ枠固定

    ■SheetsName_Load

    Set ObjDic1 = Nothing 'ディクショナリオブジェクトの解放
    Set ObjRE = Nothing '正規表現オブジェクトの解放
    Set objRE2 = Nothing '正規表現オブジェクトの解放
    Set objRE3 = Nothing '正規表現オブジェクトの解放
    Application.ScreenUpdating = True '画面表示を更新
End Sub
Public Function Rep1(StrRE2$)    '記号類を削除
    objRE2.ignorecase = Kwsk.chkNgramLcase '大小文字区別なし
    objRE2.Global = True 'グローバル

    If Kwsk.optNgramWord = False Then '単語 Ngram でないならば
        'アクセント記号付きの母音を置換
        StrRE2$ = Replace(StrRE2$, ChrW(&HE1), "a") 'a/ > a
        StrRE2$ = Replace(StrRE2$, ChrW(&HE0), "a") 'a¥ > a
        StrRE2$ = Replace(StrRE2$, ChrW(&HE9), "e") 'e/ > e
        StrRE2$ = Replace(StrRE2$, ChrW(&HED), "i") 'i/ > i
        StrRE2$ = Replace(StrRE2$, ChrW(&HF3), "o") 'o/ > o
        StrRE2$ = Replace(StrRE2$, ChrW(&HFA), "u") 'u/ > u
    End If

    'テキスト前処理
    objRE2.Pattern = "(¥[it:|:it¥])" 'パターン
    StrRE2$ = objRE2.Replace(StrRE2$, "") '[it:...:it]を削除

    objRE2.Pattern = "¥{.+?¥}" 'パターン
    StrRE2$ = objRE2.Replace(StrRE2$, "") '{...}を削除

    StrRE2$ = Replace(StrRE2$, ChrW(&H22), "") '置換
    StrRE2$ = Replace(StrRE2$, ChrW(&H27), "") '置換
```

```
StrRE2$ = Replace(StrRE2$, ChrW(&H2D), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H34), "") '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H39), "") '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H60), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HA1), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HA7), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HAD), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HB4), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HB6), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HB7), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HBA), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HBF), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HAB), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&HBB), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H2DD), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H2EE), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H2018), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H2019), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H201C), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(&H201D), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(8230), "")    '置換
StrRE2$ = Replace(StrRE2$, ChrW(9472), "")    '置換

StrRE2$ = Replace(StrRE2$, "[blanco]", "") '置換
StrRE2$ = Replace(StrRE2$, "[christus]", "") '置換
StrRE2$ = Replace(StrRE2$, "[crismon]", "") '置換
StrRE2$ = Replace(StrRE2$, "[cruz]", "") '置換
StrRE2$ = Replace(StrRE2$, "[doblez]", "") '置換
StrRE2$ = Replace(StrRE2$, "[firma]", "") '置換
StrRE2$ = Replace(StrRE2$, "[hueco]", "") '置換
StrRE2$ = Replace(StrRE2$, "[mancha]", "") '置換
StrRE2$ = Replace(StrRE2$, "[raya]", "") '置換
StrRE2$ = Replace(StrRE2$, "[roto]", "") '置換
StrRE2$ = Replace(StrRE2$, "[sello]", "") '置換
StrRE2$ = Replace(StrRE2$, "[sic]", "") '置換

StrRE2$ = Replace(StrRE2$, "[encabezamiento", "") '置換
StrRE2$ = Replace(StrRE2$, "[firma", "") '置換
StrRE2$ = Replace(StrRE2$, "[interlineado", "") '置換
```

```
StrRE2$ = Replace(StrRE2$, "[interneado", "") '置換
StrRE2$ = Replace(StrRE2$, "[lat.", "") '置換
StrRE2$ = Replace(StrRE2$, "[mano", "") '置換
StrRE2$ = Replace(StrRE2$, "[raspado", "") '置換
StrRE2$ = Replace(StrRE2$, "[rubrica", "") '置換
StrRE2$ = Replace(StrRE2$, "[signo", "") '置換
StrRE2$ = Replace(StrRE2$, "[tachada", "") '置換
StrRE2$ = Replace(StrRE2$, "[tachado", "") '置換

objRE2.Pattern = "¥[¥!.+?¥!¥]" '作成年代と場所
StrRE2$ = objRE2.Replace(StrRE2$, "") '置換

objRE2.Pattern = "(¥[mano ¥d+?|¥[margen mano ¥d+?|¥[interlineado mano ¥d+?)" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "") '置換

objRE2.Pattern = "(¥(.+?¥)|[=:;,&%!/¥-¥¥¥$¥^¥|¥.¥+¥?¥*¥(¥)¥{¥}¥[¥]])" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "") '置換

objRE2.Pattern = "¥d+?" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "") '置換

objRE2.Pattern = "¥<¥s+" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "<") '置換

objRE2.Pattern = "¥s+¥>" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, ">") '置換

objRE2.Pattern = "¥<¥>" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "") '置換

'カッコの向きを修正
objRE2.Pattern = "(¥<[^¥>]+?)(¥<)" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "$1>") '置換

objRE2.Pattern = "(¥>)([^¥<]+?¥>)" 'パターン
StrRE2$ = objRE2.Replace(StrRE2$, "<$2") '置換

objRE2.Pattern = "¥s+" '2 スペース以上を 1 つに
StrRE2$ = objRE2.Replace(StrRE2$, " ") '置換
```

```
  If Kwsk.chkNgramAbbr = True Then '省略箇所を展開しないなら
    objRE2.Pattern = "<" & "[^¥<¥>]+?" & ">"   'パターン
    StrRE2$ = objRE2.Replace(StrRE2$, "@") '置換
  End If

  If Kwsk.optNgramWord = True Then '単語 Ngram ならば
    StrRE2$ = Replace(StrRE2$, "<", "") '置換
    StrRE2$ = Replace(StrRE2$, ">", "") '置換
  End If

  Rep1 = Trim(StrRE2$) '返り値
End Function
```

C4　カーネル平滑化と文書分類

カーネル平滑化と文書分類のプログラムのコードは，以下である：

```
Option Explicit
Dim delta#, delta1#, delta2#, delta3#, Wi_1&, Wtype%, NgramType&, Cntnue%, n1&, n2&, n3&, n4&, NgrmOrd%, Pkn#
Dim OP2gram, Str1$, Str2$, Str3, ShtStr1$
Dim StrSpa$, ArSpa, ArTpl
Dim SumWght1#, WghtCnt%, LhogSum#, VarTmp#, SgmYear2#, VarSpa#, SgmPrv#, SgmRgn#, ThrTmp%, ThrSpa%
Dim SumWght2#, SumMAE1&, SumRSME1&, DocCnt1%
Dim MKN, kVal%, kFold%, Coef1#, NA1&, u%, TopK%, SumFch#, VarFch#, SumGeo#, VarGeo#
Dim ObjDicGeo, Place2LatLon, ObjDicDocCnt, ObjDicPriPrb, ObjDicGeoPrb, ObjDicErrFch, ObjDicErrGeo, Bln1
Dim ObjDicWi_1, ObjDicLambda_Wi_1, ObjDicContinue, ObjDicPkn_1 'Ngram の連想配列
Dim SmthPrm1#, VocSize%, Addtv# 'スムージングパラメータ
Dim FOPColCnt%, FOPRowCnt%, AbsErrCol%, FchCol2%, MaxErr1%, MinErr1%, CumDocCnt1% '整数型
Dim var1, var2 'ヴァリアント型
Dim Rng1 As Range, Rng2 As Range 'レンジオブジェクト
Dim SumNgrm#, Sum1#, P_x#, Q_x#, R_x#, SimMsr#, Ord1, Coef2# 'JSD
Dim SgmGeo#, RowGeo%, SumMAE2&, SumRSME2&, CumDocCnt2%, MaxErr2%, MinErr2%, AbsErrGeoCol%,
GeoAccCnt% '整数型
Dim ArSgmT, ArSgmS, ArAddSmth, vSgmT, vSgmS, vAdd 'バリアント型変数
Dim WghtLat#, WghtLon#, VarLat#, VarLon#

Sub ■Dating() '年代推定・場所推定
  With Kwsk
    '準備----------------------------------------------------------
```

```
    If ShtLstIdx% = -1 Then MsgBox "シートを選択してください。" : Exit Sub 'リストボックスが選択されていなければ Exit
    Application.ScreenUpdating = False '画面表示更新ストップ
    Set ObjDicGeo = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
    Set Place2LatLon = CreateObject("Scripting.Dictionary")  'ディクショナリオブジェクトを生成

    '二地点間の距離-------------------------
    If .optFchPrv Or .optFchCA Or .optPrv Or .optCA Then '場所推定するなら
      If .optFchPrv Then '年代推定・場所推定（Prv）なら
        IP = Sheets("DistancePrv2").UsedRange 'Distance シート
      ElseIf .optFchCA Then '年代推定・場所推定（CA）なら
        IP = Sheets("DistanceCA2").UsedRange 'Distance シート
      ElseIf .optPrv Then '場所推定（Prv）なら
        IP = Sheets("DistancePrv2").UsedRange 'Distance シート
      ElseIf .optCA Then '場所推定（CA）なら
        IP = Sheets("DistanceCA2").UsedRange 'Distance シート
      End If

      StrSpa$ = "" '初期化
      For i& = 2 To UBound(IP, 1) '場所数まで繰り返す
        StrSpa$ = StrSpa$ & IP(i&, 1) & "/" '全場所名を格納
        For j& = 2 To UBound(IP, 2) '場所数まで
          Str1$ = IP(i&, 1) & ":" & IP(1, j&) '格納
          If Not ObjDicGeo.exists(Str1$) Then '新しい地域のペアならば
            ObjDicGeo.Add Str1$, IP(i&, j&) '連想配列に登録
          End If
        Next
      Next
      StrSpa$ = Left(StrSpa$, Len(StrSpa$) - 1) '格納
      Str1$ = "" '初期化
      Erase IP '配列消去
    End If

    '地点の座標-------------------------
    If .optFchPrv Or .optFchCA Or .optPrv Or .optCA Then '場所推定するなら
      If .optFchPrv Then '年代推定・場所推定（Prv）なら
        IP = Sheets("CoordinatePrv").UsedRange 'Coordinate シート
      ElseIf .optFchCA Then '年代推定・場所推定（CA）なら
        IP = Sheets("CoordinateCA").UsedRange 'Coordinate シート
```

```
    ElseIf .optPrv Then '場所推定（Prv）なら
        IP = Sheets("CoordinatePrv").UsedRange 'Coordinate シート
    ElseIf .optCA Then '場所推定（CA）なら
        IP = Sheets("CoordinateCA").UsedRange 'Coordinate シート
    End If

    For i& = 2 To UBound(IP, 1) '場所数まで繰り返す
        If Not Place2LatLon.exists(IP(i&, 1)) Then '新しい地点ならば
            Place2LatLon.Add IP(i&, 1), Array(IP(i&, 2), IP(i&, 3)) '座標を連想配列に登録
        End If
    Next
    Erase IP '配列消去
End If

'変数値指定----------------------------------------
Atr1% = 4 '属性数
kFold% = .cmbKN_kfold '交差検定の回数
NgrmOrd% = .txtNgramOrd 'n-gram order
If .optDatNB_MV Then NgrmOrd% = 1 'ナイーブベイズならば
If .optPrv Or .optCA Then TopK% = 10 Else TopK% = 100 ' トップ K 個の文書数

'パラメータ空間------------------------------------
If .optDatLM Then '言語モデルなら
    ArAddSmth = Split("0.001", "/") 'ラプラススムージングパラメータ
ElseIf .optDatNB_MV Then 'ナイーブベイズ MV なら
    ArAddSmth = Split("0.001/0.01/0.1/1", "/") 'ラプラススムージングパラメータ
ElseIf .optDatJSD Then 'JSD なら
    ArAddSmth = Split("#") 'ラプラススムージングパラメータ
ElseIf .optDatCosine Then 'Cosine なら
    ArAddSmth = Split("#") 'ラプラススムージングパラメータ
End If

'平滑化パラメータ設定
If .chkKrnSmth Then '平滑化するなら
    If .optFch Then '年代推定だけなら
        ArSgmT = Split("3/5/10/20", "/") '時間カーネル平滑化パラメータ
        ArSgmS = Split("#") '空間カーネル平滑化パラメータ
    ElseIf .optFchCA Or .optFchPrv Then '年代推定・場所推定なら
        ArSgmT = Split("3/5/10", "/") '時間カーネル平滑化パラメータ
```

```
            ArSgmS = Split("25/50/75/100", "/") '空間カーネル平滑化パラメータ
        ElseIf .optPrv Or .optCA Then '場所推定だけなら
            ArSgmT = Split("#") '時間カーネル平滑化パラメータ
            ArSgmS = Split("75/100", "/") '空間カーネル平滑化パラメータ
        End If
    Else '平滑化しないなら
        ArSgmT = Split("#") '時間カーネル平滑化パラメータ
        ArSgmS = Split("#") '空間カーネル平滑化パラメータ
    End If

    '試行----------------------------------------------------------------------------------------
    For Each vSgmT In ArSgmT '時間カーネル平滑化パラメータ
        For Each vSgmS In ArSgmS '空間カーネル平滑化パラメータ
            For Each vAdd In ArAddSmth 'ラプラススムージングパラメータ
                '最終出力準備----------------------------------------------------------------
                If Not SheetExists(NgrmOrd% & "gramL") Then MsgBox "シート" & NgrmOrd% & "gramL" & "が存在しません。"：Exit Sub 'シートが存在しなければ Exit
                Sheets(NgrmOrd% & "gramL").Select 'シート選択
                IP = ActiveSheet.UsedRange '使用範囲格納
                ReDim FOP(1 To UBound(IP, 2), 1 To 20) '最終出力の配列サイズ再定義
                FOP(1, 4) = "Error(year)"：FOP(1, 5) = "Abs(Error(year))"
                FOP(1, 12) = "Top" & TopK% & ":*Fecha"：FOP(1, 13) = "Top" & TopK% & ":Abs(Err(year))"：FOP(1, 14) = "Top" & TopK% & ":S.D.(year)"
                FOP(1, 15) = "Error(km)"：FOP(1, 16) = "Top" & TopK% & ":Error(km)"：FOP(1, 17) = "Top" & TopK% & ":Horiz.S.D.(km)":：FOP(1, 18) = "Top" & TopK% & ":Vert.S.D.(km)" 'タイトル行

                Set ObjDicErrFch = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
                For Each var1 In Array(5, 10, 20, 30, 40, 50, 100, 1000)：ObjDicErrFch.Add var1, 0：Next '登録

                Set ObjDicErrGeo = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
                For Each var2 In Array(50, 100, 200, 300, 400, 500, 1000, 10000)：ObjDicErrGeo.Add var2, 0：Next '登録

                RowFch% = 0：RowPrv% = 0：RowCA% = 0：RowTpl% = 0 '初期化
                SumWght2# = 0：SumMAE1& = 0：SumRSME1& = 0：DocCnt1% = 0：CumDocCnt1% = 0 '初期化
                SumMAE2& = 0：SumRSME2& = 0：CumDocCnt2% = 0：GeoAccCnt% = 0 '初期化
                MaxErr1% = 0：MinErr1% = 1000：  MaxErr2% = 0：MinErr2% = 5000 '初期化
                u% = 2 'FOP 行番号の初期値

                For h& = 1 To UBound(IP, 1) '文書属性の行番号取得
```

```
                If IP(h&, 1) = Lbl(2, 1) Then FOP(1, 1) = IP(h&, 1)：RowID% = h& 'ID
                If IP(h&, 1) = Lbl(6, 1) Then FOP(1, 2) = IP(h&, 1)：FOP(1, 3) = "*" & FOP(1, 2)：RowFch% = h& 'Fecha
                If IP(h&, 1) = Lbl(8, 1) Then FOP(1, 6) = IP(h&, 1)：FOP(1, 7) = "*" & FOP(1, 6)：RowPrv% = h& 'Provincia
                If IP(h&, 1) = Lbl(9, 1) Then FOP(1, 8) = IP(h&, 1)：FOP(1, 9) = "*" & FOP(1, 8)：RowCA% = h& 'CA
                If IP(h&, 1) = Lbl(11, 1) Then FOP(1, 10) = IP(h&, 1)：FOP(1, 11) = "*" & FOP(1, 10)：RowTpl% = h&
'Tipologia
                If Sqr(RowFch%) * Sqr(RowPrv%) * Sqr(RowCA%) * Sqr(RowTpl%) > 0 Then Exit For    '積がゼロより大き
ければ Exit
            Next
            Erase IP '配列消去

            If Sqr(RowFch%) * Sqr(RowPrv%) * Sqr(RowCA%) * Sqr(RowTpl%) = 0 Then '積がゼロなら
               MsgBox "該当列が存在しません。シートが正しいか確認してください。"：Exit Sub
            End If

            If .optFchPrv Or .optPrv Then '場所推定が Provincia なら
               RowGeo% = RowPrv% '格納
            ElseIf .optFchCA Or .optCA Then '場所推定が CA なら
               RowGeo% = RowCA% '格納
            Else '場所推定ではないなら
               RowGeo% = RowPrv% '格納
            End If

            For kVal% = 0 To kFold% - 1 '交差検定数の回数まで繰り返し
               Set ObjDicDocCnt = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
               Set ObjDicPriPrb = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
               Sheets(NgrmOrd% & "gramL").Select '実験用シート選択
               IP2 = ActiveSheet.UsedRange 'IP2 に元データを格納
               Set ObjDic1 = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成

               ReDim OP(1 To UBound(IP2, 2), 1 To UBound(IP2, 1)) '配列サイズ再定義:文書×単語
               For i& = Atr1% + 2 To UBound(IP2, 1) '単語数まで繰り返し
                  OP(1, i& - Atr1%) = IP2(i&, 1) '単語名
               Next

               k& = 0 '初期化
               For j& = 2 To UBound(IP2, 2) '文書数まで繰り返し
                  If (IP2(RowID%, j&) - 1) Mod kFold% <> kVal% Then '推定を行う文書でないならば
                     If IsNumeric(IP2(RowFch%, j&)) = True Then '年代が既知ならば
```

```
                    If IP2(RowGeo%, j&) <> "s.l." And IP2(RowGeo%, j&) <> "SL" Then '場所が既知ならば
                      Str1$ = "" '初期化
                      If .optFch Then '年代推定のみなら
                        Str1$ = IP2(RowFch%, j&) '年代
                      ElseIf .optFchPrv Or .optFchCA Then '年代推定と場所推定なら
                        Str1$ = IP2(RowFch%, j&) & ":" & IP2(RowGeo%, j&)    '年代：場所
                      ElseIf .optPrv Or .optCA Then '場所推定のみなら
                        Str1$ = IP2(RowGeo%, j&) '場所
                      ElseIf .optDoc Then '文書間の比較なら
                        Str1$ = IP2(RowFch%, j&) & ":" & IP2(RowID%, j&) & ";" & IP2(RowPrv%, j&) 'ID：年代：場所
                      End If

                      If Str1$ <> "" Then '空文字列でないなら
                        If Not ObjDic1.exists(Str1$) Then    '新しいクラスならば
                          k& = k& + 1 'インクリメント
                          ObjDic1.Add Str1$, k& '連想配列に登録
                          OP(k& + 1, 1) = Str1$ '年代格納
                          For i& = Atr1% + 2 To UBound(IP2, 1) '単語数まで繰り返し
                            OP(k& + 1, i& - Atr1%) = IP2(i&, j&) '格納
                          Next
                          ObjDicDocCnt.Add Str1$, 1 '連想配列に登録
                          If Not .chkKrnSmth Then ObjDicPriPrb.Add Str1$, 1 '連想配列に登録
                        Else '既存のクラスなら
                          For i& = Atr1% + 2 To UBound(IP2, 1) '単語数まで繰り返し
                            If IP2(i&, j&) >= 1 Then '1 以上ならば
                              OP(ObjDic1(Str1$) + 1, i& - Atr1%) = OP(ObjDic1(Str1$) + 1, i& - Atr1%) + IP2(i&, j&)    '
加算
                            End If
                          Next
                          ObjDicDocCnt(Str1$) = ObjDicDocCnt(Str1$) + 1 'インクリメント
                          If Not .chkKrnSmth Then ObjDicPriPrb(Str1$) = ObjDicPriPrb(Str1$) + 1 'インクリメント
                        End If '新しいクラスならば
                      End If '空文字列でないなら
                    End If '場所が既知ならば
                  End If '年代が既知ならば
                End If '推定を行う文書でないならば
              Next '文書数まで繰り返し

              '出力
```

```
If Not SheetExists("w" & NgrmOrd% & "gram") Then 'シートが存在しなければ
   Sheets.Add after:=Sheets(Sheets.Count)：ActiveSheet.Name = "w" & NgrmOrd% & "gram" 'シート作成
End If
Sheets("w" & NgrmOrd% & "gram").Select 'シート選択
ActiveSheet.UsedRange.Clear '使用範囲クリア
Range([A1], Cells((ObjDic1.Count + 1), UBound(IP2, 1) - Atr1%)) = OP 'ペースト

'並び替え----------------------------------
With ActiveSheet.Sort
   .SortFields.Clear
   .SortFields.Add key:=Range("A2"), _
        SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:=xlSortNormal
   .SetRange Range(Cells(2, 1), Cells((ObjDic1.Count + 1), UBound(IP2, 1) - Atr1%)) '範囲設定
        .Header = xlNo：.Orientation = xlTopToBottom：.Apply
End With
ActiveSheet.UsedRange.Select '全範囲選択
With Selection：.Columns.AutoFit：.Rows.AutoFit：End With '行列幅自動調整
Range("B2").Select：ActiveWindow.FreezePanes = True   'ウィンドウ枠固定
Erase IP2：Erase OP '配列消去
Set ObjDic1 = Nothing 'ディクショナリオブジェクトの解放

'カーネル平滑化
If .chkKrnSmth = True Then '平滑化するならば
   ■Kernel_Smoothing ("w" & NgrmOrd% & "gram") 'カーネル平滑化
   Sheets("w" & NgrmOrd% & "gramKrnlSmth").Select 'シート選択
Else '平滑化しないならば
   Sheets("w" & NgrmOrd% & "gram").Select 'シート選択
End If

IP = ActiveSheet.UsedRange '使用範囲格納
ReDim OP(1 To UBound(IP, 1), 1 To UBound(IP, 2)) '配列サイズ再定義
For i& = 2 To UBound(IP, 1)：OP(i&, 1) = IP(i&, 1)：Next 'タイトル行
For j& = 2 To UBound(IP, 2)：OP(1, j&) = IP(1, j&)：Next 'タイトル列

If .optDatLM = True Then '言語モデルならば
   '■KneNey (2) 'KN スムージング
   ■AdditiveSmoothing (NgrmOrd%) '加算スムージング
   Call  ■OP_Sheet("w" & NgrmOrd% & "gramFreq", "0.000_ ") '出力
   Erase IP：Erase OP '配列消去
```

```
                ■Estimation_LM '推定
            ElseIf .optDatNB_MV Then 'ナイーブベイズ MV ならば
                ■AdditiveSmoothing_NB_MV (NgrmOrd%)   '加算スムージング
                Call  ■OP_Sheet("w" & NgrmOrd% & "gramFreq", "0.000_ ") '出力
                Erase IP：Erase OP '配列消去
                ■Estimation_LM '推定
            Else '類似度ならば
                If .optDatJSD Then 'JS 情報量ならば
                    ■JSD_PDF '確率分布を計算
                ElseIf .optDatCosine Then 'コサイン類似度ならば
                    ■Cosine_Normalize 'ベクトル正規化
                End If
                Call  ■OP_Sheet("w" & NgrmOrd% & "gramFreq", "0.000_ ") '出力
                Erase IP：Erase OP '配列消去
                ■Estimation_Similarity '推定
            End If
            Set ObjDicDocCnt = Nothing 'ディクショナリオブジェクトの解放
            Set ObjDicPriPrb = Nothing 'ディクショナリオブジェクトの解放
            'Set ObjDicGeoPrb = Nothing 'ディクショナリオブジェクトの解放
        Next '交差検定の回数まで繰り返し
        ■FinalOutPut '最終出力
    Next '時間カーネル平滑化パラメータ
  Next '空間カーネル平滑化パラメータ
Next 'ラプラススムージングパラメータ

Set ObjDicGeo = Nothing 'ディクショナリオブジェクトの解放
Set Place2LatLon = Nothing 'ディクショナリオブジェクトの解放
Application.ScreenUpdating = True '画面表示を更新
End With
End Sub

Sub ■AdditiveSmoothing_NB_MV(NgrmOrd%)   'Additive Smoothing for Naive Bayes Multivariate Bernoulli Model
'   VocSize% = 30 + 1 '語彙サイズ
Addtv# = 1 + vAdd 'スムージングパラメータ

For i& = 2 To UBound(IP, 1) 'クラス数まで繰り返し
    For j& = 2 To UBound(IP, 2) '単語数まで繰り返し
        OP(i&, j&) = (IP(i&, j&) + (Addtv# - 1)) / (ObjDicPriPrb(CStr(OP(i&, 1))) + 2 * (Addtv# - 1))     '確率
    Next
```

```
    Next
End Sub
Sub ■Error() '誤差集計
        OP(1, 1) = IP(RowID%, j&) & ":" & IP(RowFch%, j&) & ":" & IP(RowPrv%, j&) _
                        & ":" & IP(RowCA%, j&) & ":" & IP(RowTpl%, j&) '年代推定の文書の情報
        If Kwsk.optDatLM Or Kwsk.optDatNB_MV Then 'LM ならば
            OP(1, 2) = "LogL" 'ラベル
            Call ■OP_Sheet("w5", "0.0_ ") 'ペースト
            Ord1 = xlDescending '昇順
        ElseIf Kwsk.optDatJSD Then 'JSD ならば
            OP(1, 2) = "JSDiv." 'ラベル
            Call ■OP_Sheet("w5", "0.0000_ ") 'ペースト
            Ord1 = xlAscending '昇順
        ElseIf Kwsk.optDatCosine Then 'Cosine ならば
            OP(1, 2) = "Cosine" 'ラベル
            Call ■OP_Sheet("w5", "0.0000_ ") 'ペースト
            Ord1 = xlDescending '昇順
        End If

        '並び替え----------------------------------
            With ActiveSheet.Sort
                .SortFields.Clear
                .SortFields.Add key:=Range("B2"), _
                    SortOn:=xlSortOnValues, Order:=Ord1, DataOption:=xlSortNormal
                .SetRange Range(Cells(2, 1), Cells(UBound(OP, 1), UBound(OP, 2))) '範囲設定
                    .Header = xlNo：.Orientation = xlTopToBottom：.Apply
            End With
            ActiveSheet.UsedRange.Select '全範囲選択
            With Selection
                .Columns.AutoFit：.Rows.AutoFit '行列幅自動調整
            End With

            OP3 = ActiveSheet.UsedRange '格納

            FOP(u%, 1) = IP(RowID%, j&) 'ID
            FOP(u%, 2) = IP(RowFch%, j&) '実年代
            FOP(u%, 6) = IP(RowPrv%, j&) 'Provincia
            FOP(u%, 8) = IP(RowCA%, j&) 'CA
            FOP(u%, 10) = IP(RowTpl%, j&) 'Tipologia
```

```
        FOP(u%, 11) = "*" '*Tipologia

        If Kwsk.optFch Then '年代推定なら
          FOP(u%, 7) = "*" '*Provincia
          FOP(u%, 9) = "*" '*CA
        ElseIf Kwsk.optPrv Then 'Provincia 推定ならば
          FOP(u%, 7) = OP3(2, 1) '*Provincia
          FOP(u%, 9) = "*" '*CA
        ElseIf Kwsk.optCA Then 'CA 推定ならば
          FOP(u%, 7) = "*" '*Provincia
          FOP(u%, 9) = OP3(2, 1) '*CA
        ElseIf Kwsk.optFchPrv Then '年代推定・場所推定（Prv）なら
          FOP(u%, 7) = Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1) '*Provincia
          FOP(u%, 9) = "*" '*CA
        ElseIf Kwsk.optFchCA Then   '年代推定・場所推定（CA）なら
          FOP(u%, 7) = "*" '*Provincia
          FOP(u%, 9) = Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1)   '*CA
        ElseIf Kwsk.optDoc Then '文書間の比較なら
          FOP(u%, 7) = Mid(OP3(2, 1), InStr(OP3(2, 1), ";") + 1) '*Provincia
          FOP(u%, 9) = "*" '*CA
          End If

        If Kwsk.optPrv Or Kwsk.optCA Then '場所推定ならば-------------------------
          FOP(u%, 3) = "*" '推定年代
          FOP(u%, 4) = "*" '推定年代誤差
          FOP(u%, 5) = "*" '推定年代絶対値誤差
          FOP(u%, 12) = "*" 'TopK 個の推定年代の平均
          FOP(u%, 13) = "*" 'TopK 個の推定年代の絶対値誤差
          FOP(u%, 14) = "*" 'TopK 個の推定年代の分散

          'TopK 個の推定場所の平均・分散
          SumWght2# = 0：WghtLat# = 0：WghtLon# = 0：VarLat# = 0：VarLon# = 0：SumGeo# = 0：VarGeo# = 0 '初
期化
          For i& = 2 To TopK% + 1
            If Kwsk.optDatLM Or Kwsk.optDatNB_MV Then 'LM なら
              Coef2# = Exp(OP3(i&, 2) - OP3(2, 2)) '重み
            ElseIf Kwsk.optDatJSD Then 'JSD ならば
              Coef2# = 1 / OP3(i&, 2) '重み
            ElseIf Kwsk.optDatCosine Then 'Cosine ならば
```

```
            Coef2# = OP3(i&, 2) '重み
        End If
        SumWght2# = SumWght2# + Coef2# '重みの和
        VarLat# = VarLat# + Coef2# * (Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(0) ^ 2) 'TopK 個の重み付き経度の二乗和
        VarLon# = VarLon# + Coef2# * (Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(1) ^ 2) 'TopK 個の重み付き緯度の二乗和

        WghtLat# = WghtLat# + Coef2# * Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(0) 'TopK 個の重み付き経度の和
        WghtLon# = WghtLon# + Coef2# * Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(1) 'TopK 個の重み付き緯度の和
    Next

    VarLat# = VarLat# / SumWght2# 'TopK 個の重み付き経度の二乗平均
    VarLon# = VarLon# / SumWght2# 'TopK 個の重み付き緯度の二乗平均
    WghtLat# = WghtLat# / SumWght2# 'TopK 個の重み付き経度
    WghtLon# = WghtLon# / SumWght2# 'TopK 個の重み付き緯度

    FOP(u%, 17) = HubenyVarLat(Sqr(Round(VarLat# - (WghtLat# ^ 2), 10)), WghtLon#) 'TopK 個の推定場所の横方向の標準偏差：V[X]=E[X^2]-(E[x])^2
    FOP(u%, 18) = HubenyVarLon(Sqr(Round(VarLon# - (WghtLon# ^ 2), 10)), WghtLon#)    'TopK 個の推定場所の縦方向の標準偏差：V[X]=E[X^2]-(E[x])^2

    If IsNumeric(IP(RowFch%, j&)) = True Then '年代が既知なら
      If IP(RowGeo%, j&) <> "s.l." And IP(RowGeo%, j&) <> "SL" Then '場所が既知なら
        FOP(u%, 15) = HubenyDist(CDbl(Place2LatLon(Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1))(0)), CDbl(Place2LatLon(Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1))(1)), CDbl(Place2LatLon(IP(RowGeo%, j&))(0)), CDbl(Place2LatLon(IP(RowGeo%, j&))(1))) 'Top1 個の作成場所の距離誤差
        FOP(u%, 16) = HubenyDist(WghtLat#, WghtLon#, CDbl(Place2LatLon(IP(RowGeo%, j&))(0)), CDbl(Place2LatLon(IP(RowGeo%, j&))(1))) 'TopK 個の重み付き作成場所の距離誤差
        SumMAE2& = SumMAE2& + FOP(u%, 15) 'MAE
        SumRSME2& = SumRSME2& + FOP(u%, 15) ^ 2 'RMSE
        DocCnt1% = DocCnt1% + 1 '年代&場所が既知の文書数
        If FOP(u%, 15) = 0 Then GeoAccCnt% = GeoAccCnt% + 1 '場所推定が正しい文書数

        If FOP(u%, 15) > MaxErr2% Then MaxErr2% = FOP(u%, 15) '絶対値誤差の最大値
        If FOP(u%, 15) < MinErr2% Then MinErr2% = FOP(u%, 15) '絶対値誤差の最小値
```

```
            If FOP(u%, 15) <= 50 Then
               ObjDicErrGeo(50) = ObjDicErrGeo(50) + 1 'インクリメント
            ElseIf FOP(u%, 15) <= 100 Then
               ObjDicErrGeo(100) = ObjDicErrGeo(100) + 1 'インクリメント
            ElseIf FOP(u%, 15) <= 200 Then
               ObjDicErrGeo(200) = ObjDicErrGeo(200) + 1 'インクリメント
            ElseIf FOP(u%, 15) <= 300 Then
               ObjDicErrGeo(300) = ObjDicErrGeo(300) + 1 'インクリメント
            ElseIf FOP(u%, 15) <= 400 Then
               ObjDicErrGeo(400) = ObjDicErrGeo(400) + 1 'インクリメント
            ElseIf FOP(u%, 15) <= 500 Then
               ObjDicErrGeo(500) = ObjDicErrGeo(500) + 1 'インクリメント
            ElseIf FOP(u%, 15) <= 1000 Then
               ObjDicErrGeo(1000) = ObjDicErrGeo(1000) + 1 'インクリメント
            ElseIf FOP(u%, 15) > 1000 Then
               ObjDicErrGeo(10000) = ObjDicErrGeo(10000) + 1 'インクリメント
            End If
         Else '場所が不詳なら
            FOP(u%, 15) = "*" '場所推定誤差
            FOP(u%, 16) = "*" '場所推定誤差
         End If
      Else '年代が不詳なら
         FOP(u%, 15) = "*" '場所推定誤差
         FOP(u%, 16) = "*" '場所推定誤差
      End If
   ElseIf Kwsk.optFch Then '年代推定なら----------------------------------------------------------------------
      FOP(u%, 15) = "*" '場所推定の誤差
      FOP(u%, 16) = "*" 'TopK 個の場所推定の絶対値誤差
      FOP(u%, 17) = "*" 'TopK 個の推定場所の横方向の標準偏差：V[X]=E[X^2]-(E[x])^2
      FOP(u%, 18) = "*"   'TopK 個の推定場所の縦方向の標準偏差：V[X]=E[X^2]-(E[x])^2

      'TopK 個の推定年代と推定場所の平均・分散
      For i& = 2 To TopK% + 1
         If Kwsk.optDatLM Or Kwsk.optDatNB_MV Then 'n-gram モデルなら
            Coef2# = Exp(OP3(i&, 2) - OP3(2, 2))
         ElseIf Kwsk.optDatJSD Then 'kNN_JSD ならば
            Coef2# = 1 / OP3(i&, 2) '係数
         ElseIf Kwsk.optDatCosine Then 'kNN_Cosine ならば
            Coef2# = OP3(i&, 2) '係数
```

```
            End If

            SumWght2# = SumWght2# + Coef2# '事後確率の和
            SumFch# = SumFch# + Coef2# * Left(OP3(i&, 1), 4) 'TopK 個の文書年代の和
            VarFch# = VarFch# + Coef2# * (Left(OP3(i&, 1), 4) ^ 2) 'TopK 個の文書年代の重み付二乗和
        Next

        FOP(u%, 12) = SumFch# / SumWght2# 'TopK%個の推定年代の平均
        If (VarFch# / SumWght2#) - (FOP(u%, 12) ^ 2) < 0 Then 'ゼロ以下なら
            FOP(u%, 14) = 0
        Else '正なら
            FOP(u%, 14) = Sqr((VarFch# / SumWght2#) - (FOP(u%, 12) ^ 2)) 'TopK 個の文書年代の重み付き標準偏差：
V[X]=E[X^2]-(E[x])^2
        End If
        SumWght2# = 0：SumFch# = 0：VarFch# = 0：SumGeo# = 0：VarGeo# = 0 '初期化

        If IsNumeric(IP(RowFch%, j&)) = True Then '年代が既知なら
            If IP(RowGeo%, j&) <> "s.l." And IP(RowGeo%, j&) <> "SL" Then '場所が既知なら
                '年代誤差----------
                FOP(u%, 3) = Left(OP3(2, 1), 4) '推定年代
                FOP(u%, 4) = FOP(u%, 3) - FOP(u%, 2) '推定年代誤差
                FOP(u%, 5) = Abs(FOP(u%, 4)) '推定年代絶対値誤差
                FOP(u%, 13) = Abs(FOP(u%, 12) - IP(RowFch%, j&)) 'TopK 個の推定年代の絶対値誤差
                SumMAE1& = SumMAE1& + FOP(u%, 5) 'MAE
                SumRSME1& = SumRSME1& + FOP(u%, 5) ^ 2 'RMSE
                DocCnt1% = DocCnt1% + 1 '年代&場所が既知の文書数

                If FOP(u%, 5) > MaxErr1% Then MaxErr1% = FOP(u%, 5)    '絶対値誤差の最大値
                If FOP(u%, 5) < MinErr1% Then MinErr1% = FOP(u%, 5)    '絶対値誤差の最小値

                If FOP(u%, 5) <= 5 Then
                    ObjDicErrFch(5) = ObjDicErrFch(5) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 10 Then
                    ObjDicErrFch(10) = ObjDicErrFch(10) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 20 Then
                    ObjDicErrFch(20) = ObjDicErrFch(20) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 30 Then
                    ObjDicErrFch(30) = ObjDicErrFch(30) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 40 Then
```

```
                    ObjDicErrFch(40) = ObjDicErrFch(40) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 50 Then
                    ObjDicErrFch(50) = ObjDicErrFch(50) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 100 Then
                    ObjDicErrFch(100) = ObjDicErrFch(100) + 1 'インクリメント
                ElseIf FOP(u%, 5) > 100 Then
                    ObjDicErrFch(1000) = ObjDicErrFch(1000) + 1 'インクリメント
                End If
            Else '場所が不詳なら
                FOP(u%, 3) = Left(OP3(2, 1), 4) '推定年代
                FOP(u%, 4) = "*" '推定年代誤差
                FOP(u%, 5) = "*" '推定年代絶対値誤差
                FOP(u%, 13) = "*" '年代推定誤差
            End If
        Else '年代不詳ならば
            FOP(u%, 3) = Left(OP3(2, 1), 4) '推定年代
            FOP(u%, 4) = "*" '推定年代誤差
            FOP(u%, 5) = "*" '推定年代絶対値誤差
            FOP(u%, 13) = "*" '年代推定誤差
        End If
    Else '年代推定・場所推定ならば----------------------------------------------------------------
        'Top10 個の推定年代と推定場所の平均・分散
        SumWght2# = 0：SumFch# = 0：VarFch# = 0：WghtLat# = 0：WghtLon# = 0：VarLat# = 0：VarLon# = 0：
SumGeo# = 0：VarGeo# = 0 '初期化
        For i& = 2 To TopK% + 1
            If Kwsk.optDatLM Or Kwsk.optDatNB_MV Then 'n-gram モデルなら
                Coef2# = Exp(OP3(i&, 2) - OP3(2, 2))
            ElseIf Kwsk.optDatJSD Then 'kNN_JSD ならば
                Coef2# = 1 / OP3(i&, 2) '係数
            ElseIf Kwsk.optDatCosine Then 'kNN_Cosine ならば
                Coef2# = OP3(i&, 2) '係数
            End If

            SumWght2# = SumWght2# + Coef2# '重み和
            SumFch# = SumFch# + Coef2# * Left(OP3(i&, 1), 4) 'TopK 個の文書年代の重み付き年代和
            VarFch# = VarFch# + Coef2# * (Left(OP3(i&, 1), 4) ^ 2) 'TopK 個の文書年代の重み付き二乗和
            VarLat# = VarLat# + Coef2# * (Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(0) ^ 2) 'TopK 個の作
成場所の重み付き経度の二乗和
            VarLon# = VarLon# + Coef2# * (Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(1) ^ 2) 'TopK 個の作
```

```
成場所の重み付き緯度の二乗和
            WghtLat# = WghtLat# + Coef2# * Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(0) 'TopK 個の作成場所の重み付き経度の和
            WghtLon# = WghtLon# + Coef2# * Place2LatLon(Mid(OP3(i&, 1), InStr(OP3(i&, 1), ":") + 1))(1) 'TopK 個の作成場所の重み付き緯度の和
        Next

        FOP(u%, 12) = SumFch# / SumWght2# 'TopK%個の推定年代の平均
        If (VarFch# / SumWght2#) - (FOP(u%, 12) ^ 2) < 0 Then 'ゼロ以下なら
          FOP(u%, 14) = 0
        Else '正なら
          FOP(u%, 14) = Sqr((VarFch# / SumWght2#) - (FOP(u%, 12) ^ 2)) 'TopK 個の文書年代の重み付き標準偏差
        End If

        VarLat# = VarLat# / SumWght2# 'TopK 個の重み付き経度の二乗平均
        VarLon# = VarLon# / SumWght2# 'TopK 個の重み付き緯度の二乗平均
        WghtLat# = WghtLat# / SumWght2# 'TopK 個の重み付き経度
        WghtLon# = WghtLon# / SumWght2# 'TopK 個の重み付き緯度

        FOP(u%, 17) = HubenyVarLat(Sqr(Round(VarLat# - (WghtLat# ^ 2), 10)), WghtLon#) 'TopK 個の推定場所の横方向の標準偏差：V[X]=E[X^2]-(E[x])^2
        FOP(u%, 18) = HubenyVarLon(Sqr(Round(VarLon# - (WghtLon# ^ 2), 10)), WghtLon#)    'TopK 個の推定場所の縦方向の標準偏差：V[X]=E[X^2]-(E[x])^2

        If IsNumeric(IP(RowFch%, j&)) = True Then '年代が既知なら
          If IP(RowGeo%, j&) <> "s.l." And IP(RowGeo%, j&) <> "SL" Then '場所が既知なら
            '年代推定誤差----------
            FOP(u%, 3) = Left(OP3(2, 1), 4) '推定年代
            FOP(u%, 4) = FOP(u%, 3) - FOP(u%, 2) '推定年代誤差
            FOP(u%, 5) = Abs(FOP(u%, 4)) '推定年代絶対値誤差
            FOP(u%, 13) = Abs(FOP(u%, 12) - IP(RowFch%, j&)) 'TopK 個の推定年代の絶対値誤差
            SumMAE1& = SumMAE1& + FOP(u%, 5) 'MAE
            SumRSME1& = SumRSME1& + FOP(u%, 5) ^ 2 'RMSE
            DocCnt1% = DocCnt1% + 1 '年代&場所が既知の文書数

            If FOP(u%, 5) > MaxErr1% Then MaxErr1% = FOP(u%, 5) '絶対値誤差の最大値
            If FOP(u%, 5) < MinErr1% Then MinErr1% = FOP(u%, 5) '絶対値誤差の最小値

            If FOP(u%, 5) <= 5 Then
```

```
                    ObjDicErrFch(5) = ObjDicErrFch(5) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 10 Then
                    ObjDicErrFch(10) = ObjDicErrFch(10) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 20 Then
                    ObjDicErrFch(20) = ObjDicErrFch(20) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 30 Then
                    ObjDicErrFch(30) = ObjDicErrFch(30) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 40 Then
                    ObjDicErrFch(40) = ObjDicErrFch(40) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 50 Then
                    ObjDicErrFch(50) = ObjDicErrFch(50) + 1 'インクリメント
                ElseIf FOP(u%, 5) <= 100 Then
                    ObjDicErrFch(100) = ObjDicErrFch(100) + 1 'インクリメント
                ElseIf FOP(u%, 5) > 100 Then
                    ObjDicErrFch(1000) = ObjDicErrFch(1000) + 1 'インクリメント
                End If

                '場所推定誤差--------------------------------------------------------------------------
                FOP(u%, 15) = HubenyDist(CDbl(Place2LatLon(Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1))(0)),
CDbl(Place2LatLon(Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1))(1)), CDbl(Place2LatLon(IP(RowGeo%, j&))(0)),
CDbl(Place2LatLon(IP(RowGeo%, j&))(1))) 'ヒュベニの公式
                FOP(u%, 16) = HubenyDist(WghtLat#, WghtLon#, CDbl(Place2LatLon(IP(RowGeo%, j&))(0)),
CDbl(Place2LatLon(IP(RowGeo%, j&))(1))) 'TopK 個の重み付き作成場所の距離誤差

                SumMAE2& = SumMAE2& + FOP(u%, 15) 'MAE
                SumRSME2& = SumRSME2& + FOP(u%, 15) ^ 2 'RMSE
                If FOP(u%, 15) = 0 Then GeoAccCnt% = GeoAccCnt% + 1 '場所推定が正しい文書数

                If FOP(u%, 15) > MaxErr2% Then MaxErr2% = FOP(u%, 15) '絶対値誤差の最大値
                If FOP(u%, 15) < MinErr2% Then MinErr2% = FOP(u%, 15) '絶対値誤差の最小値

                If FOP(u%, 15) <= 50 Then
                    ObjDicErrGeo(50) = ObjDicErrGeo(50) + 1 'インクリメント
                ElseIf FOP(u%, 15) <= 100 Then
                    ObjDicErrGeo(100) = ObjDicErrGeo(100) + 1 'インクリメント
                ElseIf FOP(u%, 15) <= 200 Then
                    ObjDicErrGeo(200) = ObjDicErrGeo(200) + 1 'インクリメント
                ElseIf FOP(u%, 15) <= 300 Then
                    ObjDicErrGeo(300) = ObjDicErrGeo(300) + 1 'インクリメント
```

```
                ElseIf FOP(u%, 15) <= 400 Then
                    ObjDicErrGeo(400) = ObjDicErrGeo(400) + 1 'インクリメント
                ElseIf FOP(u%, 15) <= 500 Then
                    ObjDicErrGeo(500) = ObjDicErrGeo(500) + 1 'インクリメント
                ElseIf FOP(u%, 15) <= 1000 Then
                    ObjDicErrGeo(1000) = ObjDicErrGeo(1000) + 1 'インクリメント
                ElseIf FOP(u%, 15) > 1000 Then
                    ObjDicErrGeo(10000) = ObjDicErrGeo(10000) + 1 'インクリメント
                End If
            Else '場所が不詳なら
                FOP(u%, 3) = Left(OP3(2, 1), 4) '推定年代
                FOP(u%, 4) = "*" '推定年代誤差
                FOP(u%, 5) = "*" '推定年代絶対値誤差
                FOP(u%, 13) = "*" 'TopK 個の推定年代の絶対値誤差
                FOP(u%, 15) = "*" '場所推定誤差
                FOP(u%, 16) = "*" '場所推定誤差
            End If
        Else '年代不詳ならば
            FOP(u%, 3) = Left(OP3(2, 1), 4) '推定年代
            FOP(u%, 4) = "*" '推定年代誤差
            FOP(u%, 5) = "*" '推定年代絶対値誤差
            FOP(u%, 13) = "*" 'TopK 個の推定年代の絶対値誤差
            FOP(u%, 15) = "*" '場所推定誤差
            FOP(u%, 16) = "*" '場所推定誤差
        End If
    End If
    u% = u% + 1 'インクリメント
    Erase OP：Erase OP3 '配列消去
End Sub

Sub ■Estimation_LM() 'n-gram モデルによる推定
    Sheets(NgrmOrd% & "gramL").Select 'シート選択
    IP = ActiveSheet.UsedRange '格納

    Sheets("w" & NgrmOrd% & "gramFreq").Select 'シート選択
    IP2 = ActiveSheet.UsedRange '格納

    For j& = 2 To UBound(IP, 2) '文書数まで繰り返す
        If (IP(RowID%, j&) - 1) Mod kFold% = kVal% Then '推定を行う文書ならば
```

```
        ReDim OP(1 To UBound(IP2, 1), 1 To 5) '配列サイズ再定義
        LogSum# = 0 '初期値
        For k& = 2 To UBound(IP2, 1) 'クラス数まで繰り返し
          For i& = 2 + Atr1% To UBound(IP, 1) 'グラム数まで繰り返す
            If Kwsk.optDatLM = True Then '言語モデルならば
              If IP(i&, j&) >= 1 Then '頻度が 1 以上ならば
                If IP2(k&, i& - Atr1%) = 0 Then 'グラムの確率が 0 ならば
                  LogSum# = NA1& : MsgBox "Log(0)!" : Exit For '終了
                ElseIf IP2(k&, i& - Atr1%) > 0 Then 'グラムの確率が 0 より大きければ
                    LogSum# = LogSum# + IP(i&, j&) * Log(IP2(k&, i& - Atr1%)) '対数尤度
                End If
              End If
            ElseIf Kwsk.optDatNB_MV Then 'ナイーブベイズ MV ならば
              If IP(i&, j&) = 1 Then '頻度が 1 以上ならば
                LogSum# = LogSum# + 1 * Log(IP2(k&, i& - Atr1%)) '対数尤度
              ElseIf IP(i&, j&) = 0 Then '頻度が 0 ならば
                LogSum# = LogSum# + 1 * Log(1 - IP2(k&, i& - Atr1%)) '対数尤度
              End If
            End If
          Next

          '対数尤度＋対数事前確率
          If Kwsk.optFch Then '年代推定なら
            LogSum# = LogSum# '対数尤度
          ElseIf Kwsk.optPrv Or Kwsk.optCA Then '場所推定なら
            LogSum# = LogSum# '対数尤度
          ElseIf Kwsk.optFchPrv Or Kwsk.optFchCA Or Kwsk.optDoc Then '年代推定・場所推定・文書レベルなら
            If Kwsk.optNonPrior Then '年代も場所も未知とするなら
              LogSum# = LogSum# '対数尤度
            ElseIf Kwsk.optGeoKnown Then '場所を既知とするなら
              SgmGeo# = 0.0001 ^ 2 '標準偏差
              If IP(RowGeo%, j&) <> "s.l." And IP(RowGeo%, j&) <> "SL" Then '文書の作成場所が既知なら
SgmGeo#)) '重み係数
                LogSum# = LogSum# _
                          - Log(Sqr(SgmGeo#)) _
                          - ((ObjDicGeo(IP(RowGeo%, j&) & ":" & Mid(IP2(k&, 1), InStr(IP2(k&, 1), ":") + 1)) ^ 2) / (2 *
SgmGeo#)) '重み係数
              Else '文書の作成場所が不明なら
                LogSum# = LogSum# '対数尤度
```

```
                End If
            ElseIf Kwsk.optYearKnown Then '年代を既知とするなら
                If IsNumeric(IP(RowFch%, j&)) Then    '文書の作成年代が既知なら
                    SgmYear2# = 100 ^ 2 '標準偏差
                    LogSum# = LogSum# + Log(ObjDicPriPrb(IP2(k&, 1))) _
                                - Log(Sqr(SgmYear2#)) _
                                - ((IP(RowFch%, j&) - Left(IP2(k&, 1), 4)) ^ 2) / (2 * SgmYear2#) '重み係数
                Else '文書の作成年代が不明なら
                    LogSum# = LogSum# '対数尤度
                End If
            End If
            End If
            OP(k&, 1) = IP2(k&, 1) 'クラス名
            OP(k&, 2) = LogSum# '対数尤度
            LogSum# = 0 '初期化
        Next 'クラス数まで繰り返し
        ■Error '誤差集計
    End If '推定を行う文書ならば
Next '文書数まで繰り返す
Erase IP : Erase IP2 '配列消去
End Sub

Sub ■Estimation_Similarity() '類似度による推定
    Sheets(NgrmOrd% & "gramL").Select 'シート選択
    IP = ActiveSheet.UsedRange '格納

    Sheets("w" & NgrmOrd% & "gramFreq").Select 'シート選択
    IP2 = ActiveSheet.UsedRange '格納

    For j& = 2 To UBound(IP, 2) '文書数まで繰り返し
        If (IP(RowID%, j&) - 1) Mod kFold% = kVal% Then '年代推定を行う文書ならば
            ReDim OP(1 To UBound(IP2, 1), 1 To 5) '配列サイズ再定義
            SimMsr# = 0 : Sum1# = 0 '初期化

            If Kwsk.optDatJSD Then 'JSD ならば
                For i& = 2 + Atr1% To UBound(IP, 1) 'グラム数まで繰り返す
                    Sum1# = Sum1# + IP(i&, j&)    'グラムの総数
                Next
            ElseIf Kwsk.optDatCosine Then    'コサイン類似度ならば
```

```
    For i& = 2 + Atr1% To UBound(IP, 1) 'グラム数まで繰り返す
      If IP(i&, j&) > 0 Then '頻度が 0 より大きいなら
        If Kwsk.chkkNN_Binary Then '二値ベクトルなら
            Sum1# = Sum1# + 1 ^ 2 '二乗和
        Else '頻度ベクトルなら
          Sum1# = Sum1# + IP(i&, j&) ^ 2 '二乗和
        End If
      End If
    Next
    Sum1# = Sqr(Sum1#) 'L2 ノルム
  End If

  For k& = 2 To UBound(IP2, 1) 'クラス数まで繰り返す
    If Kwsk.optDatJSD Then 'JSD ならば
      For i& = 2 + Atr1% To UBound(IP, 1) 'グラム数まで繰り返す
        P_x# = IP(i&, j&) / Sum1# '確率
        Q_x# = IP2(k&, i& - Atr1%) '確率
        R_x# = (P_x# + Q_x#) / 2 '確率

        If P_x# > 0 Then '確率が 0 より大きければ
          SimMsr# = SimMsr# + P_x# * Log(P_x# / R_x#) 'JS 情報量
        End If

        If Q_x# > 0 Then '確率が 0 より大きければ
          SimMsr# = SimMsr# + Q_x# * Log(Q_x# / R_x#) 'JS 情報量
        End If
      Next
      SimMsr# = SimMsr# / 2 ' 1/2 ( Σ_xP(x)logP(x)/R(x) + Σ_xQ(x)logQ(x)/R(x) )
    ElseIf Kwsk.optDatCosine Then 'コサイン類似度ならば
      'ベクトル正規化
      For i& = 2 + Atr1% To UBound(IP, 1) 'グラム数まで繰り返す
        If IP(i&, j&) > 0 Then '頻度が 0 より大きいなら
          If Kwsk.chkkNN_Binary Then '二値ベクトルなら
            SimMsr# = SimMsr# + (1 / Sum1#) * IP2(k&, i& - Atr1%) 'コサイン類似度
          Else '頻度ベクトルなら
            SimMsr# = SimMsr# + (IP(i&, j&) / Sum1#) * IP2(k&, i& - Atr1%) 'コサイン類似度
          End If
        End If
      Next
```

```
            End If

            '類似度＋対数事前確率--------------------------------
            If Kwsk.optFch Then '年代推定なら
               SimMsr# = SimMsr#    '類似度
            ElseIf Kwsk.optPrv Or Kwsk.optCA Then '場所推定ならば
               SimMsr# = SimMsr#    '類似度
            ElseIf Kwsk.optFchPrv Or Kwsk.optFchCA Or Kwsk.optDoc Then '年代推定・場所推定もしくは文書レベルなら
               If Kwsk.optNonPrior Then '年代も場所も未知とするなら
                  SimMsr# = SimMsr# '類似度
               ElseIf Kwsk.optGeoKnown Then '場所を既知とするなら
                  If IP(RowGeo%, j&) <> "s.l." And IP(RowGeo%, j&) <> "SL" Then '文書の作成年代が既知なら
                     SgmGeo# = 1000 ^ 2 '標準偏差
                     SimMsr# = SimMsr# * (1 / Exp(-((ObjDicGeo(IP(RowGeo%, j&) & ":" & Mid(IP2(k&, 1), InStr(IP2(k&, 1),
":") + 1)) ^ 2) / (2 * SgmGeo#)))) '重み係数
                  '                  Log (Exp(-(ObjDicGeo(IP(RowPrv%, j&) & ":" & Mid(OP3(2, 1), InStr(OP3(2, 1), ":") + 1)) ^ 2) /
(2 * VarSpa#))) '重み係数
                  Else '文書の作成場所が未知なら
                     SimMsr# = SimMsr# '類似度
                  End If
               ElseIf Kwsk.optYearKnown Then '年代を既知とするなら
                  If IsNumeric(IP(RowFch%, j&)) Then '文書の作成年代が既知なら
                     SgmYear2# = 100 ^ 2    '標準偏差
                     SimMsr# = SimMsr# * (1 / Exp(-((IP(RowFch%, j&) - Left(IP2(k&, 1), 4)) ^ 2) / (2 * SgmYear2#))) '重み係数
                  Else '文書の作成年代が未知なら
                    SimMsr# = SimMsr# '類似度
                  End If
               End If
            End If
            OP(k&, 1) = IP2(k&, 1) 'クラス名
            OP(k&, 2) = SimMsr# '類似度
            SimMsr# = 0 '初期化
         Next 'クラス数まで繰り返す
         ■Error '誤差集計
      End If '年代推定を行う文書ならば
   Next '文書数まで繰り返し
   Erase IP：Erase IP2 '配列消去
End Sub
```

```
Sub ■Kernel_Smoothing(Str1$) 'カーネル平滑化
  FchInc% = 1100：FchFin% = 1700 '年代の範囲
  ThrTmp% = 20：If IsNumeric(vSgmT) Then VarTmp# = vSgmT ^ 2 '年代の平滑化パラメータ
  ThrSpa% = 200：If IsNumeric(vSgmS) Then VarSpa# = vSgmS ^ 2 '場所の平滑化パラメータ
  ArSpa = Split(StrSpa$, "/")：ArTpl = Split("*", "/") '配列
  Wght1# = 0：SumWght1# = 0：WghtCnt% = 0 '初期値
  h& = 2 '初期値

  Sheets(Str1$).Select 'シート選択
  IP2 = ActiveSheet.UsedRange '使用範囲格納

  If Kwsk.optFch Then '時間カーネル平滑化なら
    ReDim OP(1 To (FchFin% - FchInc% + 1) + 1, 1 To UBound(IP2, 2)) '配列サイズ再定義
  ElseIf Kwsk.optPrv Or Kwsk.optCA Then '空間カーネル平滑化なら
    ReDim OP(1 To (UBound(ArSpa) + 1) + 1, 1 To UBound(IP2, 2))  '配列サイズ再定義
  Else '時空間カーネル平滑化なら
    ReDim OP(1 To (FchFin% - FchInc% + 1) * (UBound(ArSpa) + 1) + 1, 1 To UBound(IP2, 2)) '配列サイズ再定義
  End If
  For j& = 2 To UBound(IP2, 2)：OP(1, j&) = IP2(1, j&)：Next '単語名コピー

  If Kwsk.optFch Then '時間カーネル平滑化なら
    For i& = 2 To (FchFin% - FchInc% + 1) + 1 '年代数まで繰り返し
      OP(h&, 1) = FchInc% + i& - 2 '年代
      ObjDicPriPrb.Add OP(h&, 1), 0 '連想配列にキーのみ登録
      h& = h& + 1 'インクリメント
    Next
  ElseIf Kwsk.optPrv Or Kwsk.optCA Then '空間カーネル平滑化なら
    For Each Str3 In ArSpa '場所数まで繰り返し
      OP(h&, 1) = Str3 '場所
      ObjDicPriPrb.Add OP(h&, 1), 0 '連想配列にキーのみ登録
      h& = h& + 1 'インクリメント
    Next
  Else '時空間カーネル平滑化なら
    For Each Str3 In ArSpa '場所数まで繰り返し
      For i& = 2 To FchFin% - FchInc% + 2 '年代数まで繰り返し
        OP(h&, 1) = FchInc% + i& - 2 & ":" & Str3 '年代：場所
        ObjDicPriPrb.Add OP(h&, 1), 0 '連想配列にキーのみ登録
        h& = h& + 1 'インクリメント
      Next
```

```
    Next
  End If

  For i& = 2 To UBound(OP, 1) '全クラス数まで繰り返し
    For k& = 2 To UBound(IP2, 1) '観測クラス数まで繰り返し
      If Kwsk.optFch Then '時間カーネル平滑化なら
        If Abs(IP2(k&, 1) - OP(i&, 1)) <= ThrTmp% Then '年代差が閾値以下なら
          Wght1# = Exp(-((IP2(k&, 1) - OP(i&, 1)) ^ 2) / (2 * VarTmp#)) '重み係数
        End If
      ElseIf Kwsk.optPrv Or Kwsk.optCA Then '空間カーネル平滑化なら
        If ObjDicGeo(IP2(k&, 1) & ":" & OP(i&, 1)) <= ThrSpa% Then '二地点の距離が閾値以下ならば
          Wght1# = Exp(-(ObjDicGeo(IP2(k&, 1) & ":" & OP(i&, 1)) ^ 2) / (2 * VarSpa#)) '重み係数
        End If
      Else '時空間カーネル平滑化なら
        If Abs(Left(IP2(k&, 1), 4) - Left(OP(i&, 1), 4)) <= ThrTmp% Then '年代差が閾値以下なら
          If ObjDicGeo(Mid(IP2(k&, 1), InStr(IP2(k&, 1), ":") + 1) & ":" & Mid(OP(i&, 1), InStr(OP(i&, 1), ":") + 1)) <=
ThrSpa% Then '距離の差が閾値以下ならば
            Wght1# = Exp(-((Left(IP2(k&, 1), 4) - Left(OP(i&, 1), 4)) ^ 2) / (2 * VarTmp#)) * _
                Exp(-(ObjDicGeo(Mid(IP2(k&, 1), InStr(IP2(k&, 1), ":") + 1) & ":" & Mid(OP(i&, 1), InStr(OP(i&, 1),
":") + 1)) ^ 2) / (2 * VarSpa#))   '重み係数
          End If
        End If
      End If

      'ナダラヤワトソンの分母の重み和はクラス共通だから無視して，分子の重み付き頻度和だけ計算
      If Wght1# > 0 Then '重みが 0 より大きければ
        For j& = 2 To UBound(IP2, 2)   '単語数まで繰り返し
          If IP2(k&, j&) > 0 Then   '頻度がゼロより大きければ
            OP(i&, j&) = OP(i&, j&) + Wght1# * IP2(k&, j&) '単語の重み付き頻度和
          End If
        Next

        If Kwsk.optDatNB_MV Then 'ナイーブベイズならば
          ObjDicPriPrb(CStr(OP(i&, 1))) = ObjDicPriPrb(CStr(OP(i&, 1))) + Wght1# * ObjDicDocCnt(CStr(IP2(k&, 1))) '文
書数のカーネル平滑化頻度
        End If
        Wght1# = 0 '初期化
      End If
```

```
        If Not Kwsk.optPrv And Not Kwsk.optCA Then '場所推定でなければ
          If k& <> UBound(IP2, 1) Then '最終行でなければ
            If Left(IP2(k& + 1, 1), 4) - Left(OP(i&, 1), 4) > ThrTmp% Then Exit For '観測年代が閾値以上離れているなら
          End If
        End If
    Next '観測クラス数まで繰り返し

Next '全クラス数まで繰り返し

'出力----------------------------------------------------------------------------
ReDim OP2(1 To UBound(OP, 1), 1 To UBound(OP, 2)) '配列サイズ再定義
For j& = 2 To UBound(OP, 2)：OP2(1, j&) = OP(1, j&)：Next 'タイトル列
h& = 2 '初期値
For i& = 2 To UBound(OP, 1) 'クラス数まで繰り返し
    For j& = 2 To UBound(OP, 2) '単語数まで繰り返し
        If OP(i&, j&) > 10 ^ (-5) Then '頻度が 10 ^ (-5)より大きいものがあれば
            For k& = 2 To UBound(OP, 2) '単語数まで繰り返し
                If OP(i&, k&) > 0 Then '頻度が 0 より大きければ
                    OP2(h&, k&) = OP(i&, k&) '格納
                ElseIf OP(i&, k&) = "" Then '頻度が 0 ならば
                    OP2(h&, k&) = 0 '格納
                End If
            Next
            OP2(h&, 1) = OP(i&, 1) 'クラス名
            h& = h& + 1 'インクリメント
            Exit For
        End If
    Next
Next

'出力
If Not SheetExists(Str1$ & "KrnlSmth") Then 'シートが存在しなければ
    Sheets.Add after:=Sheets(Sheets.Count)：ActiveSheet.Name = Str1$ & "KrnlSmth" 'シート作成
End If
Sheets(Str1$ & "KrnlSmth").Select 'シート選択
With ActiveSheet
    .UsedRange.Clear 'クリア
    .Range([A1], Cells(h& - 1, UBound(OP2, 2))) = OP2 'ペースト
    .Range([B2], Cells(h& - 1, UBound(OP2, 2))).NumberFormatLocal = "0.0_ "
```

```
    .UsedRange.Select '全範囲選択
    With Selection
      .Columns.AutoFit：.Rows.AutoFit '行列幅自動調整
    End With
    Range("B2").Select：ActiveWindow.FreezePanes = True　'ウィンドウ枠固定
  End With
  Erase IP2：Erase OP：Erase OP2 '配列消去
End Sub
Sub ■AdditiveSmoothing(NgrmOrd%) 'Additive Smoothing
  VocSize% = 30 + 1 '語彙サイズ
  Addtv# = vAdd 'スムージングパラメータ

  For i& = 2 To UBound(IP, 1) 'クラス数まで繰り返し
    '履歴（Wi_1）と頻度を連想配列に格納
    Set ObjDicWi_1 = CreateObject("Scripting.Dictionary") 'ディクショナリオブジェクトを生成
    For j& = 2 To UBound(IP, 2) '単語数まで繰り返し
      If Not ObjDicWi_1.exists(Left(IP(1, j&), NgrmOrd% - 1)) Then '新しいパターンなら
        If IP(i&, j&) > 0 Then '頻度が 0 より大きければ
          ObjDicWi_1.Add Left(IP(1, j&), NgrmOrd% - 1), IP(i&, j&) '履歴（Wi_1）を登録
        End If
      Else '既存のパターンなら
        If IP(i&, j&) > 0 Then '頻度が 0 より大きければ
          ObjDicWi_1(Left(IP(1, j&), NgrmOrd% - 1)) = ObjDicWi_1(Left(IP(1, j&), NgrmOrd% - 1)) + IP(i&, j&) 'インクリメント
        End If
      End If
    Next

    For j& = 2 To UBound(IP, 2) '単語数まで繰り返し
      OP(i&, j&) = (IP(i&, j&) + Addtv#) / (ObjDicWi_1(Left(IP(1, j&), NgrmOrd% - 1)) + Addtv# * VocSize%) '確率
    Next
    Set ObjDicWi_1 = Nothing 'ディクショナリオブジェクトの解放
  Next
End Sub

Sub ■JSD_PDF() '確率分布を計算
  SumNgrm# = 0 '初期化
  For i& = 2 To UBound(IP, 1) 'クラス数まで繰り返し
    For j& = 2 To UBound(IP, 2) 'Ngram のタイプ数まで繰り返し
```

```
        SumNgrm# = SumNgrm# + IP(i&, j&)    'Ngram の総数
      Next

      For j& = 2 To UBound(IP, 2)
        OP(i&, j&) = IP(i&, j&) / SumNgrm# '確率
      Next
      SumNgrm# = 0 '初期化
    Next
End Sub

Sub ■FinalOutPut() '最終出力
    Sheets.Add after:=Sheets(Sheets.Count) 'シート挿入
    With ActiveSheet
      .Range([A1], Cells(UBound(FOP, 1), UBound(FOP, 2))) = FOP 'ペースト
      .UsedRange.Select '全範囲選択
      With Selection：.Columns.AutoFit：.Rows.AutoFit：End With '行列幅自動調整
      Range("B2").Select：ActiveWindow.FreezePanes = True    'ウィンドウ枠固定
    End With
    Range(Cells(2, 12), Cells(UBound(FOP, 1), 12 + 6)).NumberFormat = "0_ " '表示桁数

    '並び替え----------
    With ActiveSheet.Sort
      .SortFields.Clear
      .SortFields.Add key:=Range("A2"), _
          SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:=xlSortNormal
      .SetRange Range(Cells(2, 1), Cells(UBound(FOP, 1), UBound(FOP, 2))) '範囲設定
          .Header = xlNo：.Orientation = xlTopToBottom：.Apply
    End With
    Erase FOP '配列消去

  '記述統計--------------------------------------------------
    ReDim OP(0 To 29, 1 To 3) '配列サイズ
    FOPColCnt% = ActiveSheet.UsedRange.Columns.Count '列数
    FOPRowCnt% = ActiveSheet.UsedRange.Rows.Count '行数

    For h& = 1 To FOPColCnt% '列番号取得
      If Cells(1, h&) = Lbl(6, 1) Then FchCol2% = h& 'Fecha の列番号取得
      If Cells(1, h&) = "Abs(Error(year))" Then AbsErrCol% = h& 'Abs(Error(year))の列番号取得
      If Cells(1, h&) = "Error(km)" Then AbsErrGeoCol% = h&：Exit For 'Error(km)の列番号取得
```

```
Next

If Sqr(FchCol2%) * Sqr(AbsErrCol%) * Sqr(AbsErrGeoCol% + 1) = 0 Then '積がゼロなら
   MsgBox "該当列が存在しません。シートが正しいか確認してください。"：Exit Sub
End If

Set Rng1 = Range(Cells(2, AbsErrCol%), Cells(FOPRowCnt%, AbsErrCol%)) 'レンジ格納
Set Rng2 = Range(Cells(2, AbsErrGeoCol%), Cells(FOPRowCnt%, AbsErrGeoCol%)) 'レンジ格納

With Application.WorksheetFunction
   'タイトル列
   OP(0, 1) = "Error"：OP(1, 1) = "MAE(year)"：OP(2, 1) = "RMSE(year)"：OP(3, 1) = "MedAE(year)"
   OP(12, 1) = "Max.Error(year)"：OP(13, 1) = "Min.Error(year)"
   OP(16, 1) = "MAE(km)"：OP(17, 1) = "RMSE(km)"：OP(18, 1) = "MedAE(km)"
   OP(27, 1) = "Max.Error(km)"：OP(28, 1) = "Min.Error(km)"：OP(29, 1) = "Geo.Acc."

   If Kwsk.optPrv Or Kwsk.optCA Then '場所推定のみなら
      For h& = 1 To 13：OP(h&, 2) = "*"：Next
      h& = 0 '初期値
      For Each var1 In Array(5, 10, 20, 30, 40, 50, 100, 1000) '±誤差基準を配列に格納
         OP(4 + h&, 1) = "Error：±" & var1
         OP(4 + h&, 2) = "*" '絶対度数
         OP(4 + h&, 3) = "*" '相対度数
         h& = h& + 1 'インクリメント
      Next

      OP(16, 2) = SumMAE2& / DocCnt1% 'MAE
      OP(17, 2) = Sqr(SumRSME2& / DocCnt1%) 'RMSE
      OP(18, 2) = .Median(Rng2) 'MedAE

      h& = 0 '初期値
      For Each var2 In Array(50, 100, 200, 300, 400, 500, 1000, 10000) '±誤差基準を配列に格納
         OP(19 + h&, 1) = "Error：±" & var2 & "km"
         OP(19 + h&, 2) = "'" & CumDocCnt2% + ObjDicErrGeo(var2) & "/" & DocCnt1% '絶対度数
         OP(19 + h&, 3) = 100 * (CumDocCnt2% + ObjDicErrGeo(var2)) / DocCnt1% & "%" '相対度数
         CumDocCnt2% = CumDocCnt2% + ObjDicErrGeo(var2) '閾値以下の累積文書数
         h& = h& + 1 'インクリメント
      Next
      OP(27, 2) = MaxErr2% '場所推定誤差の最大値
```

```
    OP(28, 2) = MinErr2% '場所推定誤差の最大値
    OP(29, 2) = "'" & GeoAccCnt% & "/" & DocCnt1% '場所推定の精度
    OP(29, 3) = 100 * (GeoAccCnt% / DocCnt1%) & "%" '場所推定の精度
  Else '年代推定ならば
    OP(1, 2) = SumMAE1& / DocCnt1% 'MAE
    OP(2, 2) = Sqr(SumRSME1& / DocCnt1%) 'RMSE
    OP(3, 2) = .Median(Rng1) 'MedAE
    OP(12, 2) = MaxErr1% '絶対値誤差の最大値
    OP(13, 2) = MinErr1% '絶対値誤差の最大値

    h& = 0 '初期値
    For Each var1 In Array(5, 10, 20, 30, 40, 50, 100, 1000) '±誤差基準を配列に格納
      OP(4 + h&, 1) = "Error：±" & var1
      OP(4 + h&, 2) = "'" & CumDocCnt1% + ObjDicErrFch(var1) & "/" & DocCnt1% '絶対度数
      OP(4 + h&, 3) = 100 * (CumDocCnt1% + ObjDicErrFch(var1)) / DocCnt1% & "%" '相対度数
      CumDocCnt1% = CumDocCnt1% + ObjDicErrFch(var1) '閾値以下の累積文書数
      h& = h& + 1 'インクリメント
    Next

    If SumMAE2& > 0 Then '場所推定もするなら
      OP(16, 2) = SumMAE2& / DocCnt1% 'MAE
      OP(17, 2) = Sqr(SumRSME2& / DocCnt1%) 'RMSE
      OP(18, 2) = .Median(Rng2) 'MedAE
      OP(27, 2) = MaxErr2% '絶対値誤差の最大値
      OP(28, 2) = MinErr2% '絶対値誤差の最大値
      OP(29, 2) = "'" & GeoAccCnt% & "/" & DocCnt1% '場所推定の精度
      OP(29, 3) = 100 * (GeoAccCnt% / DocCnt1%) & "%" '場所推定の精度
    Else '年代推定のみなら
      OP(16, 2) = "*" 'MAE
      OP(17, 2) = "*" 'RMSE
      OP(18, 2) = "*" 'MedAE
      OP(27, 2) = "*" '絶対値誤差の最大値
      OP(28, 2) = "*" '絶対値誤差の最大値
      OP(29, 2) = "*" '場所推定の精度
      OP(29, 3) = "*" '場所推定の精度
    End If

    h& = 0 '初期値
    For Each var2 In Array(50, 100, 200, 300, 400, 500, 1000, 10000) '±誤差基準を配列に格納
```

```
            OP(19 + h&, 1) = "Error：±" & var2
            If Kwsk.optFchPrv Or Kwsk.optFchCA Then '年代推定・場所推定なら
               OP(19 + h&, 2) = "'" & CumDocCnt2% + ObjDicErrGeo(var2) & "/" & DocCnt1% '絶対度数
               OP(19 + h&, 3) = 100 * (CumDocCnt2% + ObjDicErrGeo(var2)) / DocCnt1% & "%" '相対度数
               CumDocCnt2% = CumDocCnt2% + ObjDicErrGeo(var2) '閾値以下の累積文書数
            Else '年代のみ推定なら
               OP(19 + h&, 2) = "*" '絶対度数
               OP(19 + h&, 3) = "*" '相対度数
            End If
            h& = h& + 1 'インクリメント
         Next
      End If
      Set ObjDicErrFch = Nothing 'ディクショナリオブジェクトの解放
      Set ObjDicErrGeo = Nothing 'ディクショナリオブジェクトの解放
   End With

   '出力設定
   Range(Cells(1, FOPColCnt% + 3), Cells(UBound(OP, 1) + 1, FOPColCnt% + 3 + UBound(OP, 2) - 1)) = OP 'OP をペースト
   Range(Cells(1, FOPColCnt% + 3), Cells(UBound(OP, 1) + 1, FOPColCnt% + 3 + UBound(OP, 2) - 1)).Select '範囲選択
   With Selection '外枠罫線
      .Borders(xlEdgeLeft).LineStyle = xlContinuous：.Borders(xlEdgeTop).LineStyle = xlContinuous
      .Borders(xlEdgeBottom).LineStyle = xlContinuous：.Borders(xlEdgeRight).LineStyle = xlContinuous
      .Columns.AutoFit '列幅調整
   End With

   Range(Cells(2, FOPColCnt% + 3 + 1), Cells(UBound(OP, 1) + 1, FOPColCnt% + 3 + 2)).HorizontalAlignment = xlRight '右揃
え
   Range(Cells(2, FOPColCnt% + 3 + 1), Cells(UBound(OP, 1) + 1, FOPColCnt% + 3 + 2)).NumberFormat = "0.00_ " '表示桁数
   Range(Cells(2, FOPColCnt% + 3 + 2), Cells(UBound(OP, 1) + 1, FOPColCnt% + 3 + 2)).Style = "Percent" 'パーセント表示
   Erase OP '配列消去

   'シート名
   If Kwsk.optDatLM Then '言語モデルなら
      ShtStr1$ = "LM" '言語モデル
   ElseIf Kwsk.optDatJSD Then 'JSD なら
      ShtStr1$ = "JSD" 'JS 情報量
   ElseIf Kwsk.optDatNB_MV Then 'ナイーブベイズならば
      ShtStr1$ = "NB" 'Naive Bayes
   End If
```

```
  If Kwsk.optFch Then '年代推定・場所推定なら
    ActiveSheet.Name = ShtStr1$ & ",n" & NgrmOrd% & ",F_#" & ",T" & vSgmT & ",S" & vSgmS & ",Lp" & vAdd 'シート名
  ElseIf Kwsk.optFchCA Then '年代推定・場所推定（自治州）なら
    ActiveSheet.Name = ShtStr1$ & ",n" & NgrmOrd% & ",F_CA" & ",T" & vSgmT & ",S" & vSgmS & ",Lp" & vAdd 'シート名
  ElseIf Kwsk.optFchPrv Then '年代推定・場所推定（県）なら
    ActiveSheet.Name = ShtStr1$ & ",n" & NgrmOrd% & ",F_Prv" & ",T" & vSgmT & ",S" & vSgmS & ",Lp" & vAdd 'シート名
  ElseIf Kwsk.optCA Then '場所推定（自治州）なら
    ActiveSheet.Name = ShtStr1$ & ",n" & NgrmOrd% & ",#_CA" & ",T" & vSgmT & ",S" & vSgmS & ",Lp" & vAdd 'シート名
  ElseIf Kwsk.optPrv Then '場所推定（県）なら
    ActiveSheet.Name = ShtStr1$ & ",n" & NgrmOrd% & ",#_Prv" & ",T" & vSgmT & ",S" & vSgmS & ",Lp" & vAdd 'シート名
  End If
End Sub

Sub ■OP_Sheet(Str1$, Str2$) 'OP を指定したシート Str1$に出力する
  If Not SheetExists(Str1$) Then 'シートが存在しなければ
    Sheets.Add after:=Sheets(Sheets.Count)：ActiveSheet.Name = Str1$ 'シート作成
  End If

  Sheets(Str1$).Select 'シート選択
  With ActiveSheet
    .UsedRange.Clear 'クリア
    .Range([A1], Cells(UBound(OP, 1), UBound(OP, 2))) = OP 'ペースト
    .Range([B2], Cells(UBound(OP, 1), UBound(OP, 2))).NumberFormatLocal = Str2$ '表示桁数
    .UsedRange.Select '全範囲選択
    With Selection
      .Columns.AutoFit：.Rows.AutoFit '行列幅自動調整
    End With
    Range("B2").Select：ActiveWindow.FreezePanes = True　'ウィンドウ枠固定
  End With
End Sub
```