

Excel VBA による 言語データ分析

ver. 2015.4.8

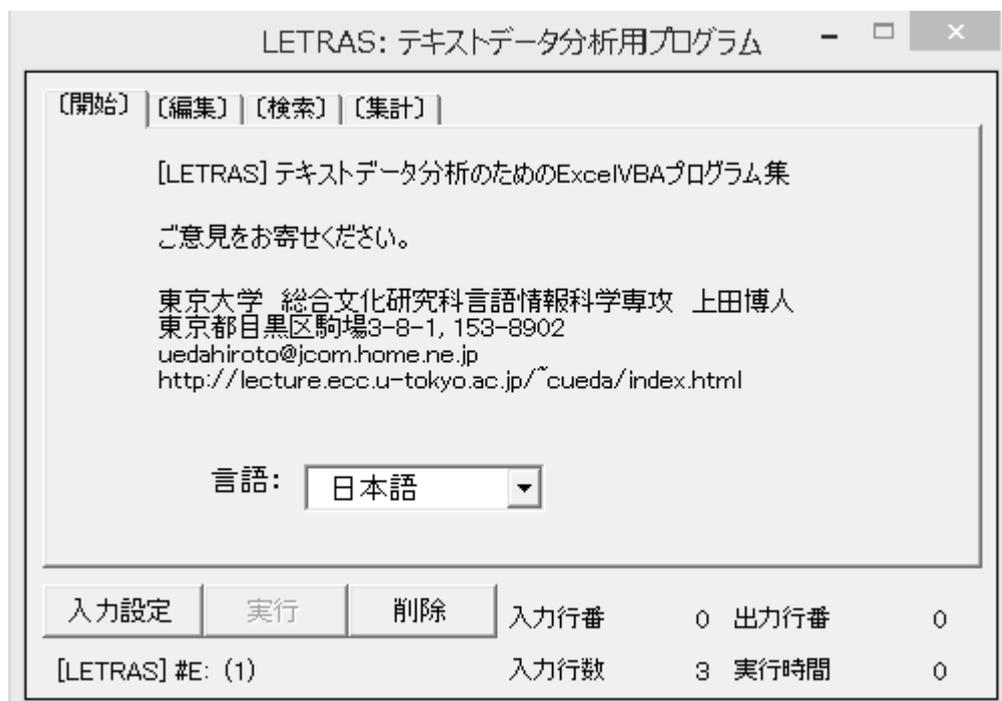
この文書は LETRAS.xlsm (以下では LETRAS とします) を簡単に解説したものです。LETRAS は随時改訂していますので、この文書も予告なしに改訂していきます。常に最終バージョンを次のサイトにアップロードします。ご使用になられた方はぜひご意見をお寄せください。私のメールアドレスは LETRAS の開始ページをご覧ください。参考にさせていただき、よりよいものを目指したいと思います。よろしくお願いいたします。

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/>

(東京大学・上田博人)

1. 開始

LETRAS のファイルを開き、マクロを有効にした後、ショートカット [Ctrl] + [Shift] + L(エル)を押すと LETRAS のマクロが起動します。次は「説明」のタブを開いたところでは



このフォームのタイトルバーに、プログラム名 (LETRAS ...)、最小化ボタン (—)、最大化ボタン (□)、終了ボタン (×) があります。分析中にこのフォームが邪魔になったときは最小化ボタンを押してください。再び最大化ボタンを押すと元の大きさに戻ります。「終了」ボタンを押すとフォームが消えます。再度立ち上げるときはショートカット [Ctrl]+[Shift]+L で起動してください。

【表紙】の下に作成者のメールアドレスが載せられています。プログラムの不具合や改善点などのご意見をお寄せください。なるべく多くの人に回答を差し上げられるようにいたします。

LETRAS を立ち上げていれば、他の Book も分析できます。Book 内で LETRAS のユーザーフォームを表示してください。

シート [L] の列 [A] のデータを次のように [A6]=3, [A8]=PT, [A10]=CM に変

更し、[A12]にテーマの色を使って塗りつぶしてください。

| English | Español | 日本語 |
|---|---|---|
| LETRAS for textual data analysis ver. 2013.10.5 | LETRAS para análisis de datos textuales « | LETRAS: テキストデータ分析用プログラム集 « |
| Select language in the cell [A6]: English=1; Spanish=2; Japanese=3, and restart LETRAS. | Seleccione el idioma en la celda [A6]: inglés = 1; español = 2; japonés = 3, y reinicie LETRAS. | 言語を選択してください。英語=1; スペイン語=2; 日本語=3 をセル [A6]に書き込み再度LETRASを起動してください。 |
| 3 | « | « |
| Select decimal separator in the cell [A8]: PT (point) or CM (comma), and restart LETRAS. | Seleccione el separador decimal en la celda [A8]: PT (punto) o CM (coma), y reinicie LETRAS. | 小数点を選択してください。(点)=PTまたはCM(コンマ)をセル[A8]に書き込み、再度LETRASを起動してください。 |
| PT | « | « |
| Select thousands separator in the cell [A10]: PT (point), CM (comma) or BL (blank), and restart LETRAS. | Seleccione el separador de miles en la celda [A10]: PT (punto), CM (coma) o BL (blanco), y reinicie LETRAS. | 千位点を選択してください。PT (点)、CM(コンマ)またはBL(ブランク)をセル[A10]に書き込み、LETRASを再起動してください。 |
| CM | « | « |
| Select background color in the cell [A12]. | Seleccione el color de fondo en la celda [A12]. | 背景色を[A12]に指定してください。 |
| Background color Color de fondo 背景色 | « | « |

[Ctl]+[Shift]+L で日本語バージョンの LETRAS を起動します。

このフォームには次のような共通のベースの上に、さまざまなタブのついたページが載せられています。

| | |
|------|--|
| 入力設定 | 現在選択されているシートと列を入力データに設定します。 |
| 実行 | 処理を実行します。 |
| 削除 | 選択されているシートを削除します。複数選択することもできます。開始時のシートを削除しようとする時確認を求められます。 |
| 入力行数 | 入力データの全行数が示されます。 |
| 入力行番 | 実行中に入力データの行番が順次表示されます。 |
| 出力行番 | 実行中に出力データの行番が順次表示されます。 |
| 実行時間 | 実行時間がミリ秒単位で表示されます。 |

次は LETRAS の Excel シートに載せたサンプルデータ Sample です。

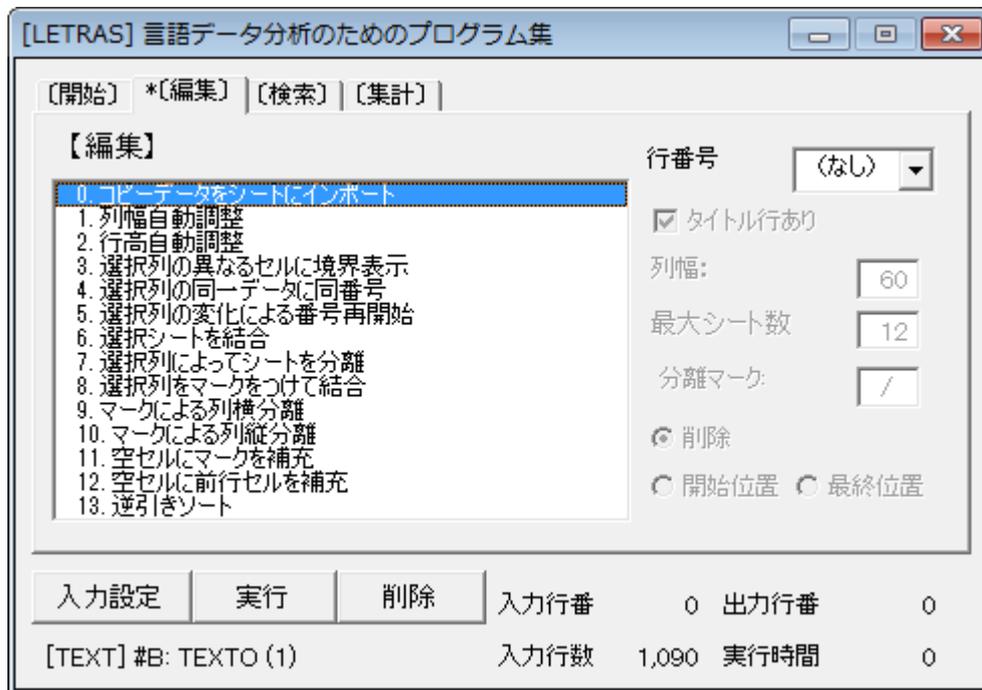
| テキスト | 見出し:1 | 見出し:2 | 行 |
|--|-----------|------------|---|
| A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región. | [A] Hotel | (a) Madrid | 1 |
| - ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene? | [A] Hotel | (a) Madrid | 2 |

サンプルはスペイン語圏各地の会話例です。データは、このように、初めの1行をタイトル行とします。A列がテキスト、B列以降はそれぞれの行についての付加情報です。

「表紙」以外のタブで実行が可能です。時間がかかる処理を中止するときは、[Esc]キーを押してください。

2. 編集

「テキスト編集」のタブに移動してください。ここで言語資料を作成したり整理したりします。



2.1. コピーデータをシートにインポート

はじめに言語資料（コーパス）を用意します。自分が持っている言語データや自分が作成した言語データを Excel のシート上にコピーして言語資料を作成しましょう。たとえば次のような形式のファイルを Excel に読み込むことができます。たとえば、#Tema:は出力シートの B 列の 1 行のセル[B1]に「Tema」という文字が入り、同じ列の次の行に Hotel が入ることを意味します。この Hotel は次に#Tema:が指定されるまで繰り返します。#Ciudad:も同様です。テキストには何も指定しません。

#Tema:[A] Hotel

#Ciudad:(a) Madrid

A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre

posibles visitas turísticas por la región.

– ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?

– ¡Buenos días, señor, bienvenido! El hotel cuenta con habitaciones individuales con baño, televisión y teléfono y algunas de ellas con vistas al Monasterio. El precio de la habitación por día, desayuno incluido, es de 12.000 pesetas. ¿Está bien así o prefiere usted otra cosa?

次が実行結果です。

| TX | Tema | Ciudad |
|--|-----------|------------|
| A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región. | [A] Hotel | (a) Madrid |
| – ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene? | [A] Hotel | (a) Madrid |
| – ¡Buenos días, señor, bienvenido! El hotel cuenta con habitaciones individuales con baño, televisión y teléfono y algunas de ellas con vistas al Monasterio. El precio de la habitación por día, desayuno incluido, es de 12.000 pesetas. ¿Está bien así o prefiere usted otra cosa? | [A] Hotel | (a) Madrid |

2.2. 列自動調節・行高自動調整

セル内のデータの量にしたがって列幅と行高を適切な規模に調整します。

| TX | Tema | Ciudad |
|--|-----------|------------|
| A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región. | [A] Hotel | (a) Madrid |
| – ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene? | [A] Hotel | (a) Madrid |
| – ¡Buenos días, señor, bienvenido! El hotel cuenta con habitaciones individuales con baño, televisión y teléfono y algunas de ellas con vistas al Monasterio. El precio de la habitación por día, desayuno incluido, es de 12.000 pesetas. ¿Está bien así o prefiere usted otra cosa? | [A] Hotel | (a) Madrid |

2.3. 異なるセル

2.3.1. 境界表示

選択した列の、直前のセルと異なるセルが出現したときに境界線を引きます。該当列をキーにして並べ替えをしておくといでしょう。次は、C列を選択した場合です。

| | | | | |
|---|-----------|---|----|------|
| – Muchas gracias. Y una última pregunta: yo soy socio de ***. ¿Hacen ustedes algún tipo de descuento en la estancia? Éste es mi carnet. | [A] Hotel | 1 | 12 | Vrtx |
| – Sí, señor. Gracias. Los socios tienen un 10% en todas nuestras tarifas. | [A] Hotel | 1 | 13 | Vrtx |
| – Bien, pues gracias otra vez. Me daré una ducha y después saldré a recorrer la ciudad que parece tan bonita. Hasta luego. | [A] Hotel | 1 | 14 | Vrtx |
| – Adiós, señor. ¡Y feliz estancia entre nosotros! | [A] Hotel | 1 | 15 | Vrtx |
| En la recepción de un hotel llega un viajero que viene atraído por el turismo para conocer la ciudad de Sevilla, debido a que tuvo la oportunidad de conocer la promoción de la misma en su país por medio de la televisión y la prensa. El turista dialoga con el consejero solicitándole información sobre comidas, costumbres y visitas. | [A] Hotel | 2 | 16 | Vrtx |

2.3.2. 同一セルに同番号

選択した列で同じデータが続く行に同一の番号をつけます。次はC列を選択した場合です。E列が3,4と変化します。

| | | | | |
|--------|--|-----------|-----|---|
| B.A. | – Llegó la hora de separarnos y de comenzar la tarea diaria. Nos veremos a la hora de almorzar, ¿querés? Yo estaré en el comedor a la una menos cuarto. | [C] Funci | 238 | 3 |
| B.A. | – Bueno; de acuerdo, Javi. Comeremos juntos. Nos vemos... | [C] Funci | 239 | 3 |
| B.A. | – Chau, Cholo. | [C] Funci | 240 | 3 |
| Madrid | Dos estudiantes universitarias, que estudian la carrera de Económicas en la Universidad de Madrid, van a coger juntas el metro hasta Ciudad Universitaria, en donde se encuentra la universidad. Susana López y Carmen García estudiaron en el mismo instituto de bachillerato; ahora también van juntas a la universidad y, de vez en cuando, quedan para estudiar. | [D] Estud | 241 | 4 |
| Madrid | – ¡Hola, Carmen! ¿Qué tal? ¿Coges el metro para la Uni? | [D] Estud | 242 | 4 |

2.3.3. 異なるセルで番号再開始

選択した列で、異なるセルになったとき、番号を1から再開始します。以下はID列を選択した場合です。

| | | | | |
|---|-----------|---|----|----|
| - Bien, pues gracias otra vez. Me daré una ducha y de recorrer la ciudad que parece tan bonita. Hasta luego. | [A] Hotel | 1 | 14 | 14 |
| - Adiós, señor. ¡Y feliz estancia entre nosotros! | [A] Hotel | 1 | 15 | 15 |
| En la recepción de un hotel llega un viajero que viene turismo para conocer la ciudad de Sevilla, debido a que oportunidad de conocer la promoción de la misma en su de la televisión y la prensa. El turista dialoga con el con drole información sobre comidas, costumbres y visitas. | [A] Hotel | 2 | 16 | 1 |
| - ¡Buenos días! ¿Tienen una habitación con baño para tres noches? | [A] Hotel | 2 | 17 | 2 |

2.4. シートの結合・分離

2.4.1. 複数のシートを結合

「Unite selected sheets (選択したシートを結合)」で複数のシートを連続させて1つのシートに結合させます。リストボックス「入力シート」で複数のシートを選択してください。

Sample-2

| Row | Text | Title-1 | Title-2 | Par |
|-----|---|--------------|---------|-----|
| 981 | ●10 Reunión de amigos. Despedida | 10 Despedida | | 1 |
| 982 | ■1 Madrid | 10 Despedida | Madrid | 1 |
| 983 | En casa de la familia García Gutiérrez tiene lugar una reunión entre amigos para Selene, una amiga venida de Salta, Argentina, que ha obtenido una beca para Salamanca. Aprovechando su estancia en España, Selene ha viajado hasta Madrid amigo a quien había conocido el año anterior en un congreso en México y para día siguiente, los García Gutiérrez organizan una reunión de amigos para despedir aquella se desarrolla en un ambiente muy cordial. | 10 Despedida | Madrid | 2 |
| 984 | - ¡Hola, Gonzalo! ¡Gracias por venir! Te presento a Selene, nuestra amiga que viene de Argentina. | 10 Despedida | Madrid | 3 |

出力:

| Row | Text | Title-1 | Title-2 | Par |
|-----|--|----------|----------|-----|
| 1 | ●1 En el hotel | 01 Hotel | - | 1 |
| 2 | ■1 Madrid | 01 Hotel | 1 Madrid | 1 |
| 3 | A la recepción de un hotel madrileño llega un profesor extranjero conferenciante en un seminario sobre Nutrición organizado por una universidad verano con sede en El Escorial. El profesor hablará con el conserje sobre los servicios del hotel, así como sobre posibles visitas turísticas. | 01 Hotel | 1 Madrid | 2 |

2.4.2. 選択列によってシートを分離

「Separate sheet by selected column (sorted)(ソートされた選択列によってシートを分離)」によって選択された列(シートの列を選択してください)にあるデータを基準にして異なるシートに分離します。選択されたデータの種類が多いと、その数だけシートが作成されてしまうので、最大シート数

の初期値を 10 としてあります。

入力：2.3.1 の出力シート（C 列を選択します。）

出力：2.3.1 の入力シート

2.5. マークによる列の結合・分離

2.5.1. 選択列をマークをつけて結合

隣接した複数の列を選択し、それらをマークをつけて結合します。初期設定では出力で結合する対象列を削除します。

入力：CD 列を選択

| | A | B | C | D | E |
|---|-----|---|--------------|----------|------|
| 1 | Row | Text | Title-1 | Title-2 | Par. |
| 2 | 981 | ●10 Reunión de amigos. Despedida | 10 Despedida | - | 1 |
| 3 | 982 | ■1 Madrid | 10 Despedida | 1 Madrid | 1 |
| 4 | 983 | En casa de la familia García Gutiérrez tiene lugar una reunión entre amigos para darle la bienvenida a Selene, una amiga venida de Salta, Argentina, que ha obtenido una beca para la Universidad de Salamanca. Aprovechando su estancia en España, Selene ha viajado hasta Madrid para visitar a su amigo a quien había conocido el año anterior en un congreso en México y para conocer a su familia. Al día siguiente, los García Gutiérrez organizan una reunión de amigos para presentarles a su invitada, y aquella se desarrolla en un ambiente muy cordial. | 10 Despedida | 1 Madrid | 2 |

出力：

| Row | Text | Title-1/Title-2 | Par. |
|-----|---|-----------------------|------|
| 981 | ●10 Reunión de amigos. Despedida | 10 Despedida/- | 1 |
| 982 | ■1 Madrid | 10 Despedida/1 Madrid | 1 |
| 983 | En casa de la familia García Gutiérrez tiene lugar una reunión entre amigos para darle la bienvenida a Selene, una amiga venida de Salta, Argentina, que ha obtenido una beca para la Universidad de Salamanca. Aprovechando su estancia en España, Selene ha viajado hasta Madrid para visitar a su amigo a quien había conocido el año anterior en un congreso en México y para conocer a su familia. Al día siguiente, los García Gutiérrez organizan una reunión de | 10 Despedida/1 Madrid | 2 |

2.5.2. マークによる列横分離

1 列の中にあるデータをマークによって要素の数だけ複数の列に分離します。

入力：2.3.4 の出力シート

出力：2.3.4 の入力シート

2.5.3. マークによる列縦分離

マークによって 1 列を複数の列に分離し、要素の数だけ縦に並べて出力します。次は分離するマークとしてコンマ(,)を指定しました。位置は「Last

position (最終位置)」です。

| Row | Text | Title-1 | Title-2 | Par. |
|-----|--|--------------|----------|------|
| 981 | ●10 Reunión de amigos. Despedida | 10 Despedida | - | 1 |
| 982 | ■1 Madrid | 10 Despedida | 1 Madrid | 1 |
| 983 | En casa de la familia García Gutiérrez tiene lugar una reunión entre amigos para darle la bienvenida a Selene, | 10 Despedida | 1 Madrid | 2 |
| 983 | una amiga venida de Salta, | 10 Despedida | 1 Madrid | 2 |
| 983 | Argentina, | 10 Despedida | 1 Madrid | 2 |

2.6. 逆引き配列による並べ替え(r)

対象の文字列の逆引き配列による整列をします。語尾による語形変化などの分析に役立ちます。A列を逆引きにするときはArのように指定します。

出力：

| テキスト | 見出し-1 | 見出し-2 | 段落 | 文 |
|--|-----------------|----------|----|---|
| ¡No sabes cómo gritaba la gente toda prendida ! | 09 Fútbol | 3 México | 13 | 2 |
| ¡Y todavía quieres que tome un taxi y me gaste un montón de lana ! | 02 Camino | 3 México | 7 | 3 |
| ¡Y me queda muy padre ! | 08 Compras | 3 México | 15 | 3 |
| ¡Qué es de tu vida, chaval ! | 09 Fútbol | 1 Madrid | 3 | 2 |
| ¡Qué bien, macho ! | 03 Funcionarios | 1 Madrid | 4 | 6 |

(...)

3. 検索

「検索」のコンボボックスには「アイテム」「文内アイテム」「文外アイテム」「文脈内アイテム」「語列内アイテム」「アイテム存在文」「アイテム不在文」「アイテム置換」の機能があります。それぞれ「正規表現」と「単語目録」のオプションによる検索式が使えます。「正規表現」のコンボボックスにはシート RE に用意したさまざまな検索式が表示されるので、その中から選択したり、また独自に検索式を設定することができます。「単語目録」のコンボボックスも同様です。



上下に並んだテキストボックスの上のテキストボックスで指定した検索式で入力シートの選択列のデータを置換・検索します。下のテキストボックスは、その下の小さなテキストボックスで指定したユニコードの文字や、その右の「ラテン補助」があるコンボボックスで指定した文字グループが表示されます。これを上の検索式テキストボックスにコピーして使用することができます。左下にあるスピンドタンによって、2つのテキストボックスの領域を変えることができます。その右のスピンドタンは上のテキストボックスの文字を拡大・縮小します。「チェック」ボタンを押すと検索式が一般の正規表現に変換されます。「保存」ボタンを押すと正規表現や田んぼ目録を、それぞれシート RE, WL に保存し、左のコンボボックス

に数字をつけて追加されます。「クリア」ボタンを押すとテキストボックスが空白になります。ふつうは「大小文字区別」をしませんが、区別するときにはこれをチェックしてください。「文内・アイテム」で「背景色」をチェックすると該当するセルに色が塗られます。「全出力」をチェックするとすべての置換、検索の結果が出力され、チェックを外すと、置換、検索された行だけが出力されます。

「検索」のオプション

| | |
|----------------|--|
| アイテム | 検索されたアイテムを文の標識（テキスト名や文番号などを一義的に示すもので、以下では ID と呼びます）の集合と共に表示します。 |
| 文内アイテム | 検索されたアイテムをテキストの中でマークします。 |
| 文外アイテム | 検索されたアイテムを独立した列に取り出して、元の文と一緒に表示します。 |
| 文脈内アイテム | 検索されたアイテムを中心に置き、その前後の文脈を左右に配置します。 |
| 語列内アイテム | 検索されたアイテムを中心に置き、その前後の語列を左右に配置します。 |
| アイテム存在文 | 検索されたアイテムを含むデータだけのシートを作成します。 |
| アイテム不在文 | 検索されたアイテムを含まないデータだけのシートを作成します。 |
| アイテム置換 | 検索されたアイテムを置換します。 |

検索式のオプション

| | |
|-------------|--|
| 正規表現 | 一般的な正規表現と LETRAS 拡張正規表現（後述）が使えます。 |
| 単語目録 | 単語使用文字の連続を単位とした置換・検索をします。多くの単語を置換・検索するときは、単純一致や正規表現よりも処理が高速です。 |

テキストボックス

| | |
|------------|-----------------------|
| 検索式 | 上のテキストボックスに検索式を入力します。 |
|------------|-----------------------|

| | |
|------------|----------------------------|
| 再定義 | 検索式に入力された文字列が再定義されて表示されます。 |
|------------|----------------------------|

チェックボックス

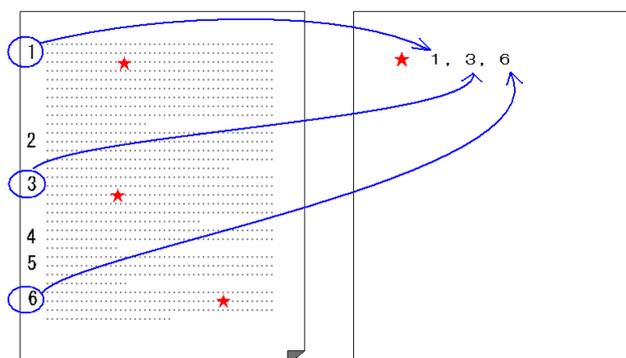
| | |
|---------------|---|
| 大小文字区別 | ふつうは大小文字を区別しませんが、区別するときのはチェックしてください。 |
| 制限字数 | 入力データのクラス内の字数を制限します。制限字数は下のテキストボックスで指定してください。 |

その他

| | |
|------------------|---|
| クリア | 検索式のテキストボックスをクリアします。 |
| スピントタン(1) | 上下のテキストボックスの領域を変えます。 |
| スピントタン(2) | 2つのテキストボックスのフォントサイズを調整します。 |
| 入力シート | 最終列にシート名と行番（行番号のみ）がある行を選択し、このボタンを押すと、その入力シートの該当行を表示します。入力面（シート）から出力面（シート）に移動するときもこのボタンを押してください。 |

3.1. アイテム

検索されたアイテムを文 ID¹の集合と共に表示します。「並べ替え」は「文字順」と「出現順」が選択できます。



「索引」のイメージ

¹ 文 ID については「1.3 準備」を参照してください。

3.1.1. 並べ替え：文字順

はじめに文アイテムによる整列をして出力させてみましょう。拡張正規表現の&は¥1+、つまり、単語を指定します。

#p% (正規表現)

| 検索式 | 焦点 | Título:1 | Título:2 | Fila |
|-----|-------------------|----------------|-------------|------|
| %d# | antigüedad | [C] Funcionar | (a) Madrid | 177 |
| %d# | barbaridad | [B] Camino | (b) Sevilla | 122 |
| %d# | barbaridad | [B] Camino | (e) B.A. | 164 |
| %d# | barbaridad | [E] Consultori | (a) Madrid | 327 |
| %d# | bondad | [H] Compras | (a) Madrid | 697 |

3.1.2. 並べ替え：出現順

| シート | 行 | 鍵語 | 見出し-1 | 見出し-2 | 段落 | 文 |
|-----|---|--------------------|----------|----------|----|---|
| Vt | 4 | +profesor | 01 Hotel | 1 Madrid | 2 | 1 |
| Vt | 4 | +para | 01 Hotel | 1 Madrid | 2 | 1 |
| Vt | 4 | +participar | 01 Hotel | 1 Madrid | 2 | 1 |
| Vt | 4 | +por | 01 Hotel | 1 Madrid | 2 | 1 |
| Vt | 4 | no | 01 Hotel | 1 Madrid | 2 | 1 |
| Vt | 5 | +profesor | 01 Hotel | 1 Madrid | 2 | 2 |

3.1.3. 連続検索

たとえば正規表現「@@@」（3文字）を使うと、recepciónの中でそれにマッチした字列を次のように次々に出力します。

| | |
|----------|------------|
| Sch.Flm. | @@@ |
| @@@ | rec |
| @@@ | epc |
| @@@ | ión |

このとき、正規表現「@@@」にマッチした字列は、次の検索の対象になりません。そこで、recepciónはrec, epc,iónのように区切られて出力されます。

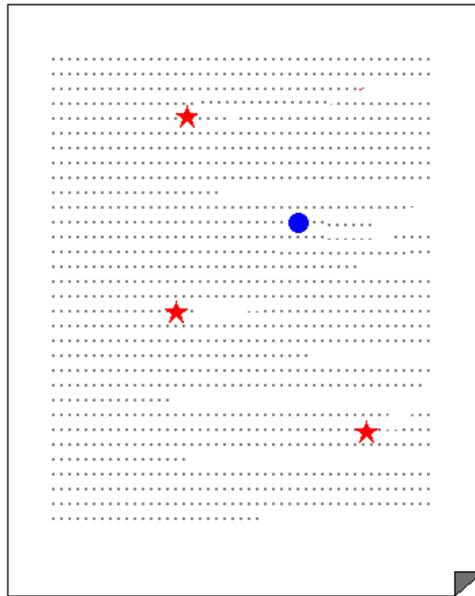
一方、3@のように、最初に数字をつけて検索すると、次のような出力になります。

| | |
|----------|-----|
| Sch.Flm. | 3@ |
| 3@ | rec |
| 3@ | ece |
| 3@ | cep |
| 3@ | epc |
| 3@ | pci |
| 3@ | ció |
| 3@ | ión |

このときは、recepción は rec, ece, cep, epc, pci, ció, ión のように、あらゆる 3 文字連続が出力されます。このような出力を「連続検索」とよぶことにします。連続検索の検索式は、字連続の場合は「3@」のように、数字+@とし、語連続の場合は「3&」のように、数字+&とします。連続検索は、アイテムの他に、以下の文外アイテム、文脈内アイテム、語列内アイテム、および「集計」で使うことができます。

3.2. 文内アイテム

言語資料を分析するとき、関心のある検索されたアイテム（一定の語、語の連続、語の一部）に注目して、テキストの中でそれらの出現を確認することがよくあります。LETRAS では単純一致、正規表現、単語リストを使って、Excel の列の中に見つかる検索されたアイテムを記号でマークすることができます。以下では、比較的複雑な正規表現を練習するために、とくに指定しないときは、すべて正規表現を使って検索します。



検索のイメージ

¥c¥c (子音 + 子音)

テキスト

Unidad 1 En el hotel

1.1. Ma{*dr*}id

A la rece{*pc*}ión de un hotel ma{*dr*}ileño {*ll*}ega un {*pr*}ofesor e{*xt*}ra{*nj*}ero para pa{*rt*}icipar como co{*nf*}ere{*nc*}ia{*nt*}e en un seminario so{*br*}e Nu{*tr*}ición o{*rg*}anizado por una unive{*rs*}idad de verano con sede en El E{*sc*}orial.

「検索マーク」をチェックすると、該当するセルに色が塗られます。

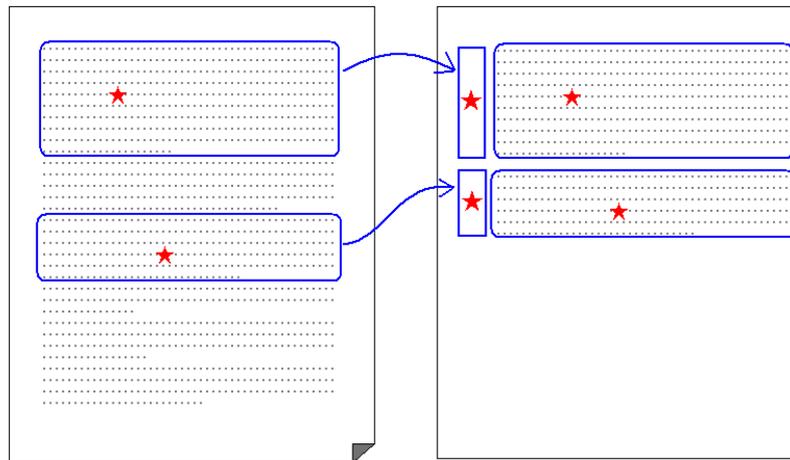
#m% (m で始まる単語) 「検索マーク」をチェック

| Row | Text | Title-1 | Title-2 | Par. |
|-----|---|-----------|----------|------|
| 1 | ● 1 En el hotel | 01 Hotel- | | 1 |
| 2 | ■ 1 {*Madrid*} | 01 Hotel | 1 Madrid | 1 |
| 3 | A la recepción de un hotel {*madrileño*} llega un profesor e participar co{*mo*} conferenciante en un se{*minario*} sob organizado por una universidad de verano con sede en El Esc hablará con el conserje, pidiéndole infor{*mación*} sobre los así co{*mo*} sobre posibles visitas turísticas por la región. | 01 Hotel | 1 Madrid | 2 |

3.3. 文外アイテム

検索されたアイテムを独立した A 列に取り出して、元の文と一緒に表示し

ます。「並べ替え」は「文字順」と「出現順」が選択できます。



「外置」のイメージ

%ndo% (ndo を含む単語)

鍵語 テキスト

afinando y vete {*afinando*} la voz para cantarle el "Cumpleaños feliz", ¿eh?

afinando y vete {*afinando*} la voz para cantarle "Las mañanitas", ¿eh?

aguantando – Aquí estamos, hijo, {*aguantando*} para no caer...

出力付加段落数

出力する段落数を増やして、検証する文脈の範囲を拡げることができます。
コンボボックスの数字を変えて試してください。

同：付加段落 1

鍵語 テキスト

afinando ¡Ah! y vete {*afinando*} la voz para cantarle el "Cumpleaños feliz",
¿eh? 6 Fiesta

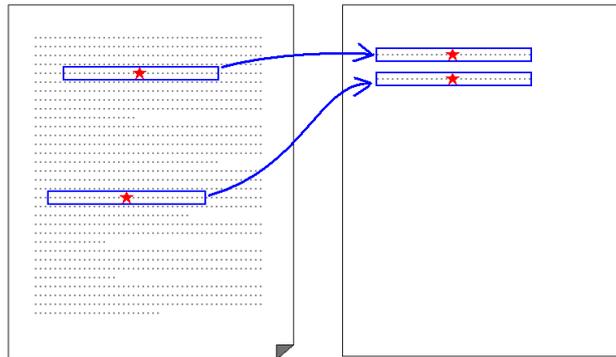
afinando ¡Ah! y vete {*afinando*} la voz para cantarle "Las mañanitas", ¿eh? 6
Fiesta

aguantando ¡Qué es de tu vida, chaval ! – Aquí estamos, hijo, {*aguantando*} para
no caer... 9 Fútbol

3.4. 文脈内アイテム

検索されたアイテムを中心に置き、その前後の文脈を左右に配置します。
検索されたアイテムを中心の 1 列に置くことで、検索されたアイテムの配

列が追跡しやすくなります。また、その前後の文脈の特徴も目立ちます。
「並べ替え」は「文字順」と「出現順」が選択できます。



内置のイメージ

()(%[ai]d[oa]s?)# (ado, ados, ada, adas, ...など)

| シート | 行 | 前文脈 | 鍵語 | 後文脈 |
|-----|------|--|-------------------|---|
| Vt | 1028 | Sus clases no son | aburridas | . |
| Vt | 245 | Que vas | acelerado | , para un poco, oye, ¿tú sabes dónde está ****? |
| Vt | 2995 | Lo sirves rodeado de una hoja de lechuga y | acompañado | con un trozo de camote y otro de choclo . |
| Vt | 1034 | Tiene un programa más | adecuado | a nuestros gustos, ¿no te parece? |
| Vt | 172 | - Si señor, estamos | afiliados | a varios programas de las líneas aéreas. |
| Vt | 810 | Yo también estoy | agobiada | con esa asignatura. |

類義語

ラテン語訳『創世記』には「(彼は)言った」という意味で主に **ait**, **dixit**, **inquit** という 3 つの変化形が使われています。ちょっと数が多いのですが、それぞれの検索されたアイテムの出現形をラテン語訳の原典であるヘブライ語テキストの中で探してみましょう。たとえば、**ait** の(1:11), (1:26), ...、**dixit** の(1:6)…という具合です。そうすると、これら全部は基本的に一つのヘブライ語動詞に対応していることがわかります。

それでは、ラテン語の 3 つの動詞はまったくの同義語でバリエーションに過ぎなかったのでしょうか。翻訳したヒエロニムスは 3 つの動詞を気ままに使っていたのでしょうか。

ここで、ふたたび「内置」で同じ検索されたアイテムを検索してみると、文末では **ait** だけが使われていて (**ait** は文末だけに限りませんが)、「～と言った」というような付け足しのような感じです。

| | | |
|--|-----|---|
| Egressus ad eos Loth post tergum accludens ostium | ait | . |
| Jacob vero convocata omni domo sua | ait | . Abjicite deos alienos qui in medio vestri sunt: et mundamini ac mu |
| ad vesperam Jacob de agro egressa est in occursum ejus Lia: et | ait | . Ad me inquit intrabis. Quia mercede conduxi te pro mandragoris f |
| uae. Pariebant omnes oves varios foetus. Quando vero e contrario | ait | . Alba quaeque accipies pro mercede. Omnes greges alba peperer |
| itis Esau sermonibus patris: irruit clamore magno. Et consternatus | ait | . Benedic etiam et mihi pater mi. |
| benedixitque ei et | ait | . Benedictus Abram Deo excelso: qui creavit caelum et terram: |
| Quos cum vidisset | ait | . Castra Dei sunt haec. Et appellavit nomen loci illius Manaim id est |

一方、**inquit** は文中に挿入され、日本語にすれば「～と彼は言った～」
 というような感じです。

| | | |
|--|---------------|---|
| At ille. Servus | inquit | Abraam sum. |
| Ad quos ille clementer. Accedite | inquit | ad me. Et cum accessissent prope. Ego sum ait Joseph frater veste |
| At ille. Septem | inquit | agnas accipies de manu mea: ut sint mihi in testimonium: quoniam e |
| Cui Judas. Quid respondebinus | inquit | domino meo? Vel quid loquemur aut juste poterimus obtendere? E |
| espondit. Filii mei sunt quos donavit mihi Deus in hoc loco. Adduc | inquit | eos ad me: ut benedicam illis. |
| unt pauulum. Tunc Joseph accersito dispensatore domus ait. Surge | inquit | et persequere viros. Et apprehensis dicito. Quare reddidistis mahum |

この2つの動詞に対して、**dixit** は文頭で用いられ（逆に文頭だからと言って
dixit になる、ということではありませんが）、「（彼が）言ったことは～」と
 いうようなしっかりとした意味合いが感じられます。

| | | |
|---|--------------|--|
| | Dixit | que Esau. Quenam sunt istae turmae quas obviam habui? Respon |
| | dixit | que fratribus suis. Afferte lapides. Qui congregantes fecerunt tumulu |
| | Dixit | que illi Isaac. Quis enim es tu? Qui respondit. Ego sum filius tuus pr |
| Bene | dixit | que illis Deus: et ait. Crescite et multiplicamini et replete terram et si |
| ergo semper Esau Jacob pro benedictione qua benedixerat ei pater. | Dixit | que in corde suo. Venient dies luctus patris mei: et occidam Jacob f |

スペイン語（だけではありませんが）はこれらの類義語（けっして「同義語」
 ではありません）の中から **dixit** だけを継承しました。**dixit** のはっきりとした意
 味と形が好まれたのだと思います。

「とは思わない」

次は私たちが編集したスペイン語・日本語辞典の用例の日本語の部分「思
 いません」を「内置」で検索した結果です。

| 1 | A | B | C | D |
|-----|-------|-----------------------------|------------|--|
| ID | | 前文脈 | 後文脈 | 対応文 |
| 191 | 8742 | それは移植に敏感な木なので根づくとは【 | 思】いません | Es un árbol muy sensible al trasplante y no creo que \$arraigue. |
| 192 | 10864 | どうぞいます。でも私は今その通りにすることがよいとは【 | 思】いません | Gracias por su consejo pero no estimo conveniente seguirlo por. |
| 193 | 17774 | 私たちはフランシスコにこの仕事ができるとは【 | 思】いません | No consideramos que Francisco \$sea capaz de realizar este trabajo. |
| 194 | 17783 | 私はこのスーツケースが 20キロ以上あるとは【 | 思】いません | No creo que esta maleta \$tenga más de veinte kilos. |
| 195 | 18024 | 私は彼女がその機会を無駄にするほど愚かだとは【 | 思】いません | No la creo tan torita como para desaprovechar esa ocasión. |
| 196 | 8640 | ホセのように誠実な人にそんな嘘が言えるなんて【 | 思】いもよらなかった | Es inconcebible que una persona tan honrada como José \$pueda decir una mentira así. |

検索されたアイテムの前の位置を見ると、ほとんどが「とは」で終わっていて、なぜか「と思いません」の例が見つかりませんでした。

次は夏目漱石『坊っちゃん』の全文を調べた結果です。「思う」の前は「と」に限られ、「思わない」の前ではほとんどが「とは」が現れています。このような分布は統計的な有意差を調べるまでもなく、見ただけで単に偶然で起きるはずがないことがわかります。

| 引用部 | 「思う」 | 「思わない」 |
|------|-------|--------|
| 「と」 | 183 例 | なし |
| 「とは」 | 1 例 | 11 例 |

「と思わない」と言ってもちゃんと日本語として成り立つのに、実際にほとんど使われない理由は、「思わない」というのが単に「思う」の否定形というのではなくて、ちょうど「…なんて信じられない」というような話者の何らかの評価を示しているからではないかと思われまます。その評価の対象が「…とは」で表示されているのでしょう。

スペイン語の「(私は)思う」*creo*の目的節には **que+直説法** が使われ、「思わない」*no creo* では **接続法** が使われます。そして学生の手紙や日本の文法書を見ると、そのほとんどが「…とは思いません」と訳しています。スペイン語文法では「否定」という意味要素が接続法の条件になる、と論じられることが多いのですが、*no creo* を単に動詞の否定形とするよりも、やはり「信じられない」という「評価」という話者の態度が関わっているのでしょう。この点で日本語とスペイン語の文法の間には偶然ではない、人間の普遍的な認知に関わる意味の平行関係があるように思われます。

私たちは理論的に導かれた文法の規則を実際の言語使用にあてはめることが多いのですが、実際の **言語使用の分析** が逆に理論的な文法の問題を再考するきっかけになることがあります。帰納法と演繹法を両立させることができればよいと思います。

3.5. 語列内アイテム

「3.5. 語列内アイテム」では、キーワードと一緒に現れる語をその位置にそって集計したり、集めたり、連続して扱ったりして、その関係を探ります。ここでは単語を単位として、その前後の数語との連続関係を分析します。出力の形式には「語形」「集計」「合同」「結合」がありますが、この中で「語形」だけが文の横のつながりを保持します。一方、「集計」「合

同「結合」では、横のつながりを切って分析します。「並べ替え」は「文字順」と「出現順」が選択できます。

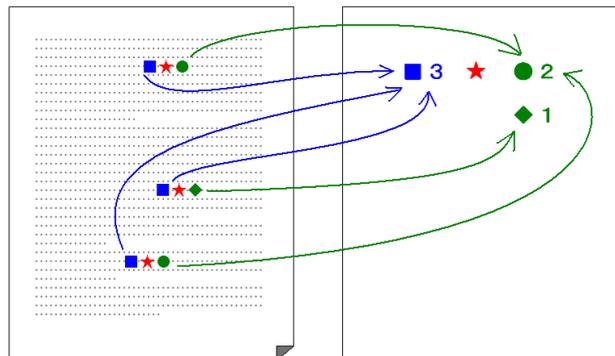
単語を単位としていることを除いて、「文脈内アイテム」とほぼ同じ機能を持ちます。たとえば、上の設定で実行すると次のように出力されます。

Output [#de#]

| Sum | Ant. 3 words | Sum | Key word | Sum | Pos. 3 words | Sum | Total 6 words |
|-----|--------------|------|----------|-----|--------------|-----|---------------|
| 174 | | 1012 | de | 207 | | 381 | |
| 142 | , | | | 192 | , | 334 | , |
| 141 | la | | | 182 | . | 278 | la |
| 116 | el | | | 137 | la | 197 | y |
| 83 | y | | | 114 | y | 182 | . |
| 72 | en | | | 60 | que | 129 | el |
| 60 | a | | | 59 | de | 118 | de |

■ 集計

上の「語形」の前語(W-1, W-2, ...)、鍵語、後語(W+1, W+2, ...)のそれぞれの列の単語をまとめて集計します。まとめているので、横の関係は切れています。キーワードのそばにある語の頻度を縦の列だけを区別して調べたいときに役に立ちます。出力は「文字順」、「出現順」、「頻度順」が選択できます。



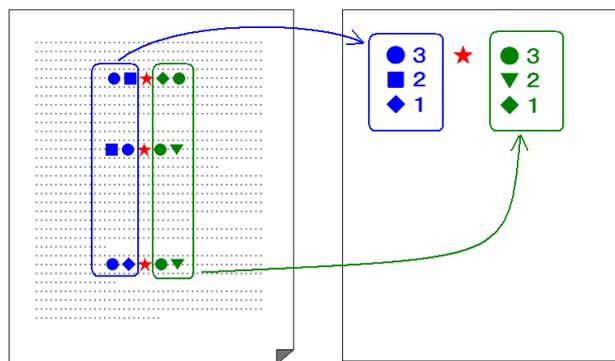
連続：集計のイメージ

por (単語目録) 並べ替え：頻度順

| Sum | W-3 | Sum | W-2 | Sum | W-1 | Sum | Key word | Sum | W+1 | Sum | W+2 | Sum | W+3 |
|-----|-----|-----|-----|-----|--------|------|----------|-----|---------|-----|-----|-----|-----|
| 120 | | 128 | la | 36 | , | 1012 | de | 110 | la | 137 | , | 203 | |
| 60 | en | 77 | el | 21 | y | | | 29 | las | 125 | . | 57 | . |
| 45 | de | 65 | , | 20 | es | | | 23 | los | 75 | y | 55 | , |
| 41 | , | 47 | | 13 | centro | | | 22 | todo | 30 | en | 45 | de |
| 41 | y | 42 | un | 13 | dentro | | | 19 | un | 26 | ... | 39 | y |
| 30 | el | 31 | a | 13 | lo | | | 17 | acuerdo | 24 | ? | 24 | que |
| 29 | a | 22 | una | 12 | sala | | | 17 | que | 19 | ! | 21 | se |

■ 集合

前語の列(W-1, W-2, ...)、後語の列(W+1, W+2, ...)、両者(W-1, W-2, ..., W+1, W+2, ...)をまとめてそれぞれを合計列に出力し、その中の単語を合同して、その頻度を集計します。キーワードのそばで連続する語の集合を見るときに使います。



連続：集合のイメージ

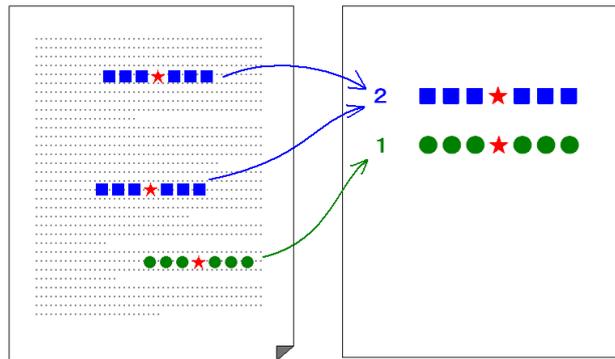
por (単語目録) 並べ替え：頻度順

| Sum | Ant. 3 words | Sum | Key words | Sum | Pos. 3 words | Sum | Total 6 words |
|-----|--------------|------|-----------|-----|--------------|-----|---------------|
| 174 | | 1012 | de | 207 | | 381 | |
| 142 | , | | | 192 | , | 334 | , |
| 141 | la | | | 182 | . | 278 | la |
| 116 | el | | | 137 | la | 197 | y |
| 83 | y | | | 114 | y | 182 | . |
| 72 | en | | | 60 | que | 129 | el |
| 60 | a | | | 59 | de | 118 | de |

■ 結合

前の数語、または後の数語の連続を切らずに、その連続の頻度を集計しま

す。キーワードに隣接するまとまった語数の特徴を調べることができます。



連続：連続のイメージ

por (単語目録) 並べ替え：頻度順

| Sum | Ant. 3 words | Sum | Key word | Sum | Pos.3 words |
|-----|------------------|------|----------|-----|---------------|
| 9 | en_el_centro | 1012 | de | 9 | tema_.¿ |
| 9 | en_la_sala | | | 9 | vez_en_cuando |
| 8 | _(dentro | | | 9 | espera_)_ |
| 8 | _me_alegro | | | 6 | acuerdo_. _ |
| 7 | _ _ | | | 6 | todo_. _ |
| 7 | ¿_qué_es | | | 6 | tu_vida_? |
| 7 | oye_. _cambiando | | | 5 | la_ciudad_y |

後期ラテン語から初期スペイン語を想像する

大学で習うラテン語はカエサルやキケロなどの紀元前後に書かれた文章を対象とする**古典ラテン語**です。一方、ここでテキスト例として見ている『創世記』のラテン語は紀元5世紀の「**後期ラテン語**」Late Latin とよばれるものです。

ラテン語は名詞が格変化し、それによって主語や目的語の関係がわかるので、とくに動詞の目的語の位置が定まっているわけではありませんが、ふつうは動詞の前におきます。そして動詞はふつう文末に置かれます。一方、現代スペイン語などラテン語から派生した言語では目的語は動詞の後に置くのがふつうです。

さて、次は名詞の対格の例として *aquam, arcam, terram* を選び、その直前の語を頻度順に並べたときの出力です。

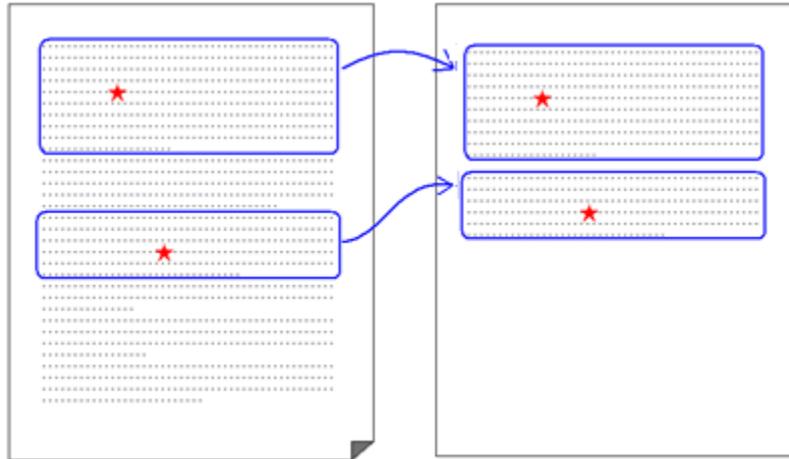
| SUM | W-1 | SUM | aquam arcam terram |
|-----|-------------|-----|--------------------------|
| 42 | super | 121 | terram |
| 38 | in | 12 | aquam |
| 8 | omnem | 11 | arcam |
| 7 | et | | |
| 6 | ad | | |
| 4 | universam | | |
| 3 | tibi | | |
| 3 | dabo | | |
| 3 | hauriendam | | |
| 2 | replete | | |
| 2 | operaretur | | |
| 1 | aridam | | |
| 1 | illuminent | | |
| 1 | deus | | |
| 1 | ingredieris | | |
| 1 | elevaverunt | | |
| 1 | aquae | | |
| 1 | dissipans | | |
| 1 | inter | | |
| 1 | exercere | | |
| 1 | abram | | |
| 1 | perambula | | |
| 1 | te | | |
| 1 | hydriam | | |
| 1 | hauriam | | |
| 1 | hauriret | | |
| 1 | hausit | | |
| 1 | reppererunt | | |
| 1 | invenimus | | |
| 1 | possideas | | |
| 1 | isaac. | | |
| 1 | habebat | | |
| 1 | sunt | | |
| 1 | attulit | | |
| 1 | eis | | |
| 1 | fames | | |
| 1 | praeter | | |

ここで興味深いのは、『創世記』ラテン語の対格（目的語）の位置が、まるで現代スペイン語のように、ほとんど例外なく動詞の直後になっていることです。このように当時の話し言葉が反映していると思われる後期ラテン語の様子から文献によって記録されていない**原始スペイン語 Proto-Spanish**のシンタックスを想像することができます。

3.6. 存在文と不在文

3.6.1. 存在文を抽出

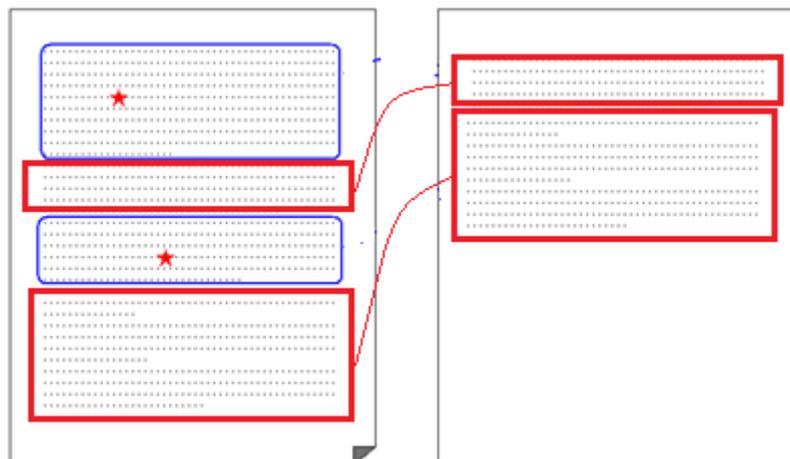
検索されたアイテムに一致したデータがあるセルを取り出して、新しいシートにコピーします。



「抽出」のイメージ

3.6.2. 不在文を抽出

検索されたアイテムに一致したデータがないセルを取り出して、新しいシートにコピーします。



「除外」のイメージ

3.7. 置換

Excel の置換の機能は限られています。LATRAS では正規表現を含めたさまざまな検索式を使って、言語テキストを分析に適した形に変換することができます。検索式は A=>B のように「=>」を使います。これの左辺が右辺に置換されます。

以下では次のデータをサンプルとします。

Text

- 1 En el hotel

- 1 Madrid

A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial.

正規表現

「Search fml.(検索式)」で「Regular expression(正規表現)」を選択します。たとえば次の正規表現の検索式を使うと r に続く英数字の連続が<R>に置換されます。

r¥w*=><R>

- 1 En el hotel

- 1 Mad<R>

A la <R>ón de un hotel mad<R>ño llega un p<R> ext<R> pa<R> pa<R> como confe<R> en un semina<R> sob<R> Nut<R>ón o<R> po<R> una unive<R> de ve<R> con sede en El Esco<R>.

スペイン語の ó などは単語文字として認識されていません。そこで、英語以外の言語では一般の正規表現を拡張して再定義した LETRAS 拡張正規表現を使用します。たとえば¥l(エル)は独・仏・西語などの西欧語の単語に使用される文字として再定義されます。

r¥l*=><R>

A la <R> de un hotel mad<R> llega un p<R> ext<R> pa<R> pa<R> como confe<R> en un semina<R> sob<R> Nut<R> o<R> po<R> una unive<R> de ve<R> con sede en El Esco<R>.

先の¥w では madrileño の ñ を含めませんが、¥l はこれを含めます。さらに%

が¥I*を再解釈します。したがって、r¥I*は r%とすることができます。

(...)による後方照応を使うと、正規表現に一致した文アイテムを再生することができます。

(r%)=><\$1>

A la <recepción> de un hotel mad<rileño> llega un p<rofesor> ext<ranjero> pa<ra> pa<rticipar> como confe<renciante> en un semina<rio> sob<re> Nut<rición> o<rganizado> po<r> una unive<rsidad> de ve<rano> con sede en El Esco<rial>.

上では r に始まる文字列を<...>で囲みました。r で始まる単語をマークするには次のような指定をします。

(¥L)(r%)=>\$1<\$2>

A la <recepción> de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial.

¥L は西欧語単語使用文字以外の文字（空白や記号類）を示します。これに一致した文字列が、=>の右辺の\$1 で再生され、さらに r%に一致した文字列が\$2 で再生されます。さらにデータの最初の語も含めるならば、(^¥L)としなければなりません。また、一般に語末まで含めた単語以外の文字列は(^|\$¥L)とします。これを「#」を使って再解釈します。したがって、(^|\$¥L)(r%)=>\$1<\$2>は、#(r%)=>\$1<\$2>のように書くことができます。

ñ や ó などの特殊文字は n@, o/のように書くと、それぞれ ñ と ó に再解釈されます。

(%n@%)=><\$1>

(%o/%)=><\$1>

A la <recepción> de un hotel <madrileño> llega un profesor extranjero para participar como conferenciante en un seminario sobre <Nutrición> organizado por una universidad de verano con sede en El Escorial.

上の連立検索式は(%(n@|o/)%)=><\$1>のように、(...|...)で括って示すこともできます。

3.8. 大小文字区別

「大小文字区別」をチェックして次の置換式を使うと小文字ではじまる語だけにマッチします。

(%e%)==><\$1> (大小区別なし)

A la <recepción> <de> un <hotel> <madrileño> <llega> un <profesor> <extranjero> para participar como <conferenciante> <en> un <seminario> <sobre> Nutrición organizado por una <universidad> <de> <verano> con <sede> <en> <El> <Escorial>.

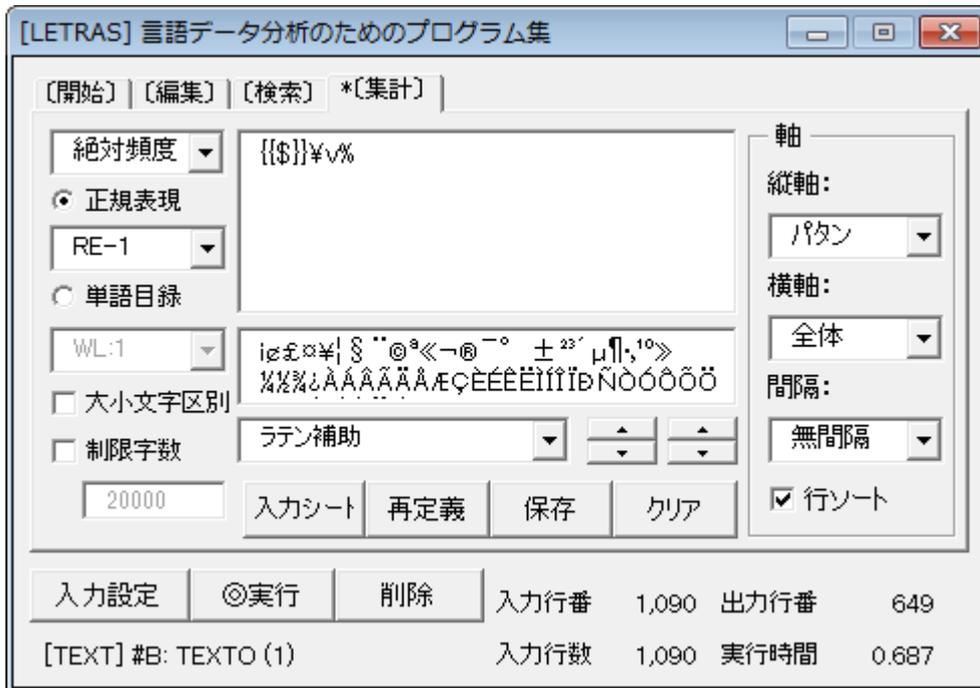
(%e%)==><\$1> (大小区別あり)

A la <recepción> <de> un <hotel> <madrileño> <llega> un <profesor> <extranjero> para participar como <conferenciante> <en> un <seminario> <sobre> Nutrición organizado por una <universidad> <de> <verano> con <sede> <en> El Escorial.

(%E%)==><\$1> (大小区別あり)

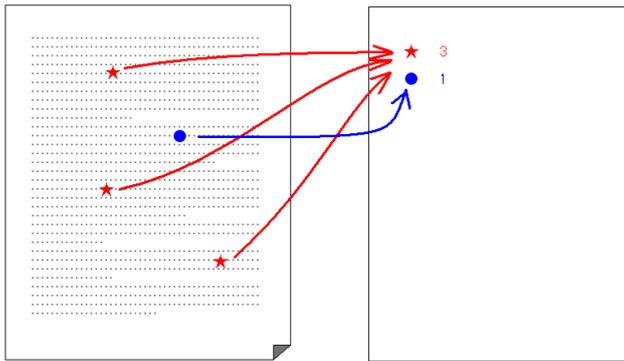
A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en <El> <Escorial>.

4. 集計



「集計」では検索されたアイテムの出現頻度を計算し、表にして出力します。「絶対頻度」「縦相対頻度(%)」「横相対頻度(%)」「千語率」「千字率」「アイテムリスト」が選択できます。

頻度の計算は「フォーム」(F)で「合計」(すべての検索式に一致した検索されたアイテムの頻度)、「検索式」(それぞれの検索式に一致した検索されたアイテムの頻度)、「アイテム」(それぞれの検索されたアイテムの頻度)が選択できます。分類の基準として「クラス」(C)を選択することができます。「クラス」のコンボボックスの最初は「全体」です。これを選択すると、シートごとに検索されたアイテムの分布を見ることができます。列文字(A, B, C...)を選択すると、その列にあるセルの内容ごとに検索されたアイテムの分布を見ることができます。



集計のイメージ

Output-1 [検索式: &#amp;#iend;縦列: F.合計; 横行: : 全体]

| | |
|--------------|-----|
| &#amp;#iend; | 全体 |
| 合計 | 153 |

Output-2 [検索式: &#amp;#iend; 縦列:F.検索式; 横行: : 全体]

| | |
|--------------|-----|
| &#amp;#iend; | 全体 |
| &#amp;#iend; | 129 |
| &#amp;#iend; | 24 |
| 計 | 153 |

Output-3 [検索式: &(a|ie)ndo#; 縦列:F.検索式; 横行: ID]

| &#amp;#iend; | B.A. | Lima | Madrid | México | Sevilla | 計 |
|--------------|------|------|--------|--------|---------|-----|
| &#amp;#iend; | 17 | 19 | 29 | 34 | 30 | 129 |
| &#amp;#iend; | 5 | 2 | 7 | 8 | 2 | 24 |
| 計 | 22 | 21 | 36 | 42 | 32 | 153 |

Output-4 [検索式: &(a|ie)ndo#; 縦列:F.アイテム; 横行: ID]

| &ando# &iendo# | B.A. | Lima | Madrid | México | Sevilla | 計 |
|-------------------|------|------|--------|--------|---------|----|
| afinando | | | 1 | 1 | | 2 |
| aguantando | | | 1 | | | 1 |
| añadiendo | | | 1 | | | 1 |
| aprovechando | 1 | | 2 | 1 | 1 | 5 |
| ayudando | | | 1 | 1 | | 2 |
| cambiando | 2 | 2 | 3 | 3 | | 10 |

Output-5 [検索式: &(a|ie)ndo#; 縦列:C.Tema; 横行: C.ID]

| &ando# &iendo# | B.A. | Lima | Madrid | México | Sevilla | 計 |
|-------------------|------|------|--------|--------|---------|-----|
| [A] Hotel | | | | 1 | | 1 |
| [B] Camino | 3 | 2 | 1 | | 4 | 10 |
| [C] Funcionarios | 4 | 5 | 4 | 3 | 5 | 21 |
| [D] Estudiantes | 4 | 3 | 6 | 5 | 3 | 21 |
| [E] Consultorio | 3 | 6 | 5 | 7 | 4 | 25 |
| [F] Fiesta | | 3 | 3 | 3 | 2 | 11 |
| [G] Política | 2 | | 7 | 15 | 5 | 29 |
| [H] Compras | | | | 3 | 1 | 4 |
| [I] Fútbol | 3 | 2 | 5 | 3 | 4 | 17 |
| [J] Despedida | 3 | | 5 | 2 | 4 | 14 |
| 計 | 22 | 21 | 36 | 42 | 32 | 153 |

Output-7 [検索式: &(a|ie)ndo#; 縦列:F 検索式; 横行:C.Línea Int: 100]]

| &ando# &iendo# | 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 | 計 |
|-------------------|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| &ando# | 1 | 15 | 25 | 17 | 12 | 13 | 12 | 4 | 14 | 11 | 5 | 129 |
| &iendo# | 1 | 3 | 4 | 3 | 2 | 5 | 5 | | | 1 | | 24 |
| 計 | 2 | 18 | 29 | 20 | 14 | 18 | 17 | 4 | 14 | 12 | 5 | 153 |

Output-8 [同; 縦相対頻度 (%)]

| &ando# &iendo# | 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|-------------------|-------|-------|-------|-------|-------|-------|-------|--------|--------|-------|--------|
| &ando# | 50.0% | 83.3% | 86.2% | 85.0% | 85.7% | 72.2% | 70.6% | 100.0% | 100.0% | 91.7% | 100.0% |
| &iendo# | 50.0% | 16.7% | 13.8% | 15.0% | 14.3% | 27.8% | 29.4% | 0.0% | 0.0% | 8.3% | 0.0% |

Output-9 [同; 横相対頻度 (%)]

| &ando#&iendo# | 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---------------|------|-------|-------|-------|------|-------|-------|------|-------|------|------|
| &ando#&iendo# | 0.8% | 11.6% | 19.4% | 13.2% | 9.3% | 10.1% | 9.3% | 3.1% | 10.9% | 8.5% | 3.9% |
| &ando#&iendo# | 4.2% | 12.5% | 16.7% | 12.5% | 8.3% | 20.8% | 20.8% | 0.0% | 0.0% | 4.2% | 0.0% |

Output-10 [同; 千語率]

| &ando#&iendo# | 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| &ando#&iendo# | 0.464 | 4.496 | 8.091 | 6.751 | 5.242 | 4.724 | 5.141 | 3.042 | 7.592 | 4.287 | 2.098 |
| &ando#&iendo# | 0.464 | 0.899 | 1.294 | 1.191 | 0.874 | 1.817 | 2.142 | 0.000 | 0.000 | 0.390 | 0.000 |

Output-11 [同; 千字率]

| &ando#&iendo# | 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| &ando#&iendo# | 0.102 | 1.084 | 1.965 | 1.605 | 1.236 | 1.117 | 1.205 | 0.729 | 1.819 | 0.994 | 0.485 |
| &ando#&iendo# | 0.102 | 0.217 | 0.314 | 0.283 | 0.206 | 0.429 | 0.502 | 0.000 | 0.000 | 0.090 | 0.000 |

Output-12 [同; アイテムリスト]

| &ando#&iendo# | 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---------------|-------------|-----------------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|----------------|------------------|-----------------|
| &ando#&iendo# | saliendo-50 | cogiendo-116 | viendo-203 | sufriendo-316 | corriendo-432 | viendo-529 | haciendo-603 | | | añadiendo-948 | |
| &ando#&iendo# | | haciendo-157 | compartiendo-22 | doliendo-337 | viviendo-451 | haciendo-556 | manteniendo-617 | | | | |
| &ando#&iendo# | | viendo-173 | insistiendo-244 | haciendo-377 | | viendo-561 | subiendo-624 | | | | |
| &ando#&iendo# | | | insistiendo-274 | | | manteniendo-570 | haciendo-627 | | | | |
| &ando#&iendo# | | | | | | haciendo-580 | viendo-669 | | | | |
| &ando#&iendo# | Fernando-99 | cuando-113 | cambiando-208 | Cuando-305 | cuando-407 | pensando-503 | paseando-602 | señalando-731 | cuando-835 | tomando-911 | Cuando-1002 |
| &ando#&iendo# | | fijando-119 | cuando-208 | Cuando-314 | cambiando-411 | afinando-509 | aprovechando-60 | cuando-741 | aguantando-837 | tirando-913 | paseando-1022 |
| &ando#&iendo# | | cuando-123 | Fernando-213 | llegando-315 | peleando-412 | cuando-510 | ayudando-603 | molestando-752 | gritando-846 | Cuando-922 | Aprovechando-10 |
| &ando#&iendo# | | conversando-143 | llegando-213 | cuando-320 | cuando-424 | pensando-521 | mamando-609 | cuando-752 | cantando-846 | Aprovechando-9 | Cuando-1068 |
| &ando#&iendo# | | mirando-147 | hablando-220 | cambiando-324 | cuando-425 | necesitando-522 | pasando-614 | | Cuando-846 | Cuando-938 | cuando-1079 |
| &ando#&iendo# | | fijando-155 | Cambiando-222 | protestando-325 | dando-436 | dando-555 | echando-616 | | cuando-852 | Cuando-948 | |
| &ando#&iendo# | | Cuando-161 | mejorando-223 | cuando-337 | cambiando-441 | aprovechando-55 | entrando-618 | | cuando-860 | cuando-949 | |
| &ando#&iendo# | | Cuando-168 | llegando-228 | Fernando-362 | pensando-467 | ayudando-556 | paseando-626 | | jugando-860 | Aprovechando-960 | |
| &ando#&iendo# | | cambiando-178 | Cambiando-237 | Fernando-364 | afinando-473 | pasando-567 | trabajando-627 | | dominando-861 | Cuando-980 | |
| &ando#&iendo# | | charlando-181 | cuando-237 | cuando-367 | cuando-474 | dando-579 | chupando-633 | | cuando-876 | cuando-981 | |
| &ando#&iendo# | | cuando-183 | cuando-242 | Fernando-368 | pensando-485 | pasando-588 | hablando-640 | | cuando-883 | cuando-982 | |
| &ando#&iendo# | | tirando-187 | Cuando-245 | cuando-378 | preparando-491 | pasando-589 | cuando-659 | | Cuando-887 | | |
| &ando#&iendo# | | cuando-190 | cambiando-246 | cambiando-382 | | cuando-590 | | | cuando-893 | | |
| &ando#&iendo# | | Cuando-191 | pasando-246 | protestando-383 | | | | | tomando-893 | | |
| &ando#&iendo# | | charlando-194 | cuando-254 | cuando-384 | | | | | | | |
| &ando#&iendo# | | | cuando-257 | cuando-395 | | | | | | | |
| &ando#&iendo# | | | pasando-261 | cuando-396 | | | | | | | |
| &ando#&iendo# | | | cuando-269 | | | | | | | | |
| &ando#&iendo# | | | cuando-272 | | | | | | | | |
| &ando#&iendo# | | | Cuando-275 | | | | | | | | |
| &ando#&iendo# | | | cambiando-276 | | | | | | | | |
| &ando#&iendo# | | | cuando-284 | | | | | | | | |
| &ando#&iendo# | | | cuando-287 | | | | | | | | |
| &ando#&iendo# | | | pasando-291 | | | | | | | | |

検索アイテムがあるセルを選択して「入力シート」ボタンをクリックすると、入力したテキストにジャンプします。

| Word list | B.A. | Lima | Madrid | México | Sevilla |
|--------------|----------------|----------------|-----------------|-------------------|----------------|
| #el % a# | El turista-82 | El turista-57 | el mapa-12 | el recepcionista- | El turista-17 |
| #el % a# | el mapa-92 | el día-71 | el tema-245 | el sistema-48 | el lila-188 |
| #el % a# | el día-442 | el Alianza-902 | el programa-250 | el mapa-52 | el tema-577 |
| #el % a# | el día-442 | | el tema-552 | El programa-280 | el sistema-593 |
| #el % a# | el tema-646 | | el agua-940 | el problema-398 | |
| #el % a# | el sistema-662 | | | el tema-599 | |
| #el % a# | | | | el tema-624 | |
| #el % a# | | | | el América-885 | |
| #el % a# | | | | El día-980 | |
| #el % a# | | | | el día-1040 | |
| #{ir}{{ a }} | iré-95 | vaya-79 | vas-105 | Va-38 | ir-111 |
| #{ir}{{ a }} | vas-155 | ir-139 | va-107 | ir-125 | vas-117 |
| #{ir}{{ a }} | voy-159 | vas-141 | va-109 | vas-127 | vas-117 |

Output-13:語の長さ

#@ { 1 } #

#@ { 2 } #

#@ { 3 } #

#@ { 4 } #

#@ { 5 } #

#@ { 6 } #

#@ { 7 } #

#@ { 8 } #

#@ { 9 } #

| * | 全体 |
|------------|--------|
| #@ { 1 } # | 1,668 |
| #@ { 2 } # | 6,751 |
| #@ { 3 } # | 4,081 |
| #@ { 4 } # | 3,105 |
| #@ { 5 } # | 3,586 |
| #@ { 6 } # | 2,631 |
| #@ { 7 } # | 2,002 |
| #@ { 8 } # | 1,145 |
| #@ { 9 } # | 743 |
| 計 | 25,712 |

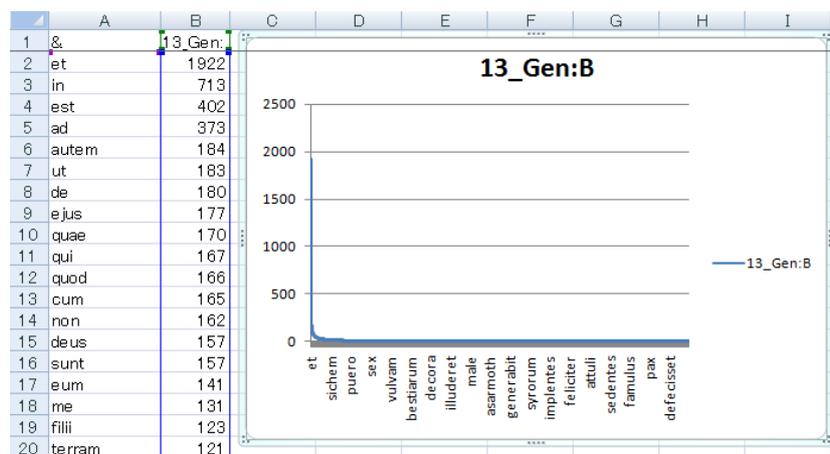
Output-14:文の長さ

##¥P{1,10}##
 ##¥P{11,20}##
 ##¥P{21,30}##
 ##¥P{31,40}##
 ##¥P{41,50}##

| * | Madrid | Sevilla | México | Lima | B.A. | 計 |
|---------------|--------|---------|--------|------|------|-------|
| ##¥P{1,10}## | 99 | 67 | 171 | 79 | 72 | 488 |
| ##¥P{11,20}## | 137 | 113 | 179 | 119 | 127 | 675 |
| ##¥P{21,30}## | 98 | 87 | 122 | 87 | 118 | 512 |
| ##¥P{31,40}## | 67 | 68 | 92 | 74 | 74 | 375 |
| ##¥P{41,50}## | 56 | 50 | 72 | 50 | 55 | 283 |
| 計 | 457 | 385 | 636 | 409 | 446 | 2,333 |

L 字形分布

ラテン語訳『創世記』の全出現語の度数を「分布」を使って計算し、そのグラフを描いてみると次のような形になります。



一番頻度の高い et (=英 and)は 1922, 次が in(=英 in)が 713, est(=英 is)が 402, ...と続きます。このように語の頻度を降順で辿っていくとその減少が急であることがわかります。そして et, in, est などの高頻度語の数が非常に少なく、逆に低頻度語は非常に数多くあります。『創世記』で頻度が 1 の語は

3480 もありました。その様子を示しているのが、上の図です。このような分布は「L字形分布」と呼ばれ、どのようなテキストでも比較的大きなものであれば、このような分布を示します。ここでは名詞や動詞などの変化形などを代表形にまとめこんだ計算をしていませんが、代表形にまとめてもその分布は同様の傾向を示します。

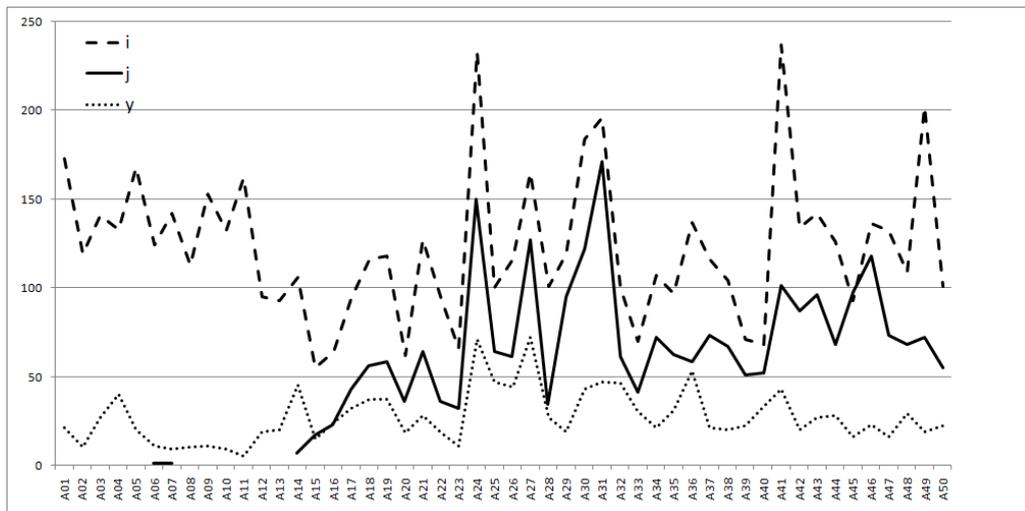
言語は、種類は少ないけれど非常によく使われる接続詞、前置詞、代名詞などの「機能語」(function word)と、種類はとても多いけれどあまり使われない名詞や動詞などの「内容語」(content word)から成り立っています。このように頻度が偏っていることはバランスがとれていないように見えますが、むしろ言語のあり方として合理的だと思われます。かりにすべての単語が同じ頻度で使われるとしたら、機能語の数はまったく不足してしまい、同じ内容語が何度も繰り返されることになります。これでは言語としての働きをなさなくなるでしょう。

ところで、『創世記』でもっとも頻度が高い内容語はやはり **Deus** (=英 **God**) でした。度数は 157 で、代名詞などよりも頻度が高く、全体で 15 位の位置を占めています。このような語はテキストの中で「特徴語」と呼ばれ、とくに注意が必要です。

対照頻度表

複数のシートを使った「分布」の結果は**対照頻度表**になり、各種の統計分析にかけることができます。また、対照頻度表そのものの観察も興味深いものがあります。

次の図は **LETRAS** の「分布」を使って中世スペイン語に翻訳された『創世記』50章の中にある i, j, y の文字の出現頻度を調べたものです²。



実は、これらの3つの文字は同じ語の中でも、mi, mj, my; amigo, amjgo のように交替して使われていました。このような分布を**自由変異 free variation**と呼びます。当時これらは同じ文字(**文字素 grapheme**)のバリエント(異文字: **allograph**)でした。

ところが、実際にその分布を調べてみるとグラフの実線が示すように、jの文字が途中まで(14章の途中まで)ほとんど使われていないのです。自由変異のバリエントならばどこにでも出現するはずですが。翻訳本の製本の状態を見ると14章の途中で帖(quire)が分かれています。書体の違いからここで写字生が交替したことがわかります。これらのバリエントの使い方には個人差があ

² Schonfield, Jeremy (ed.). 1992. *Companion volume to the facsimile edition. La Biblia de Alba. An illustrated manuscript Bible in Castilian*, by Rabbi Moses Arragel. Madrid. Fundación Amigos de Sefarad.

Ueda, Hiroto. 2009. "Palatal graphemes in a medieval Spanish biblical text: a corpus analysis of «i, j, y» in *Genesis, Biblia de Alba*", *Corpus analysis and variation in linguistics*, edited by Yuji Kawaguchi, Makoto Minegishi and Jacques Durand, John Benjamins Publishig Company, pp. 239-257.

ったようです。

5.【補説】正規表現と単語目録

5.1. 一般の正規表現

正規表現は複雑な文字列処理に適しています。正規表現の規則は非常に単純ですが、使い方については練習が必要です。何度でも実験して確認してください。

5.1.1. 特殊文字

| | |
|-----------|----------------------|
| ¥t | 水平タブに一致します。 |
| ¥b | 任意の英単語の境界に一致します。 |
| ¥B | 任意の英単語境界以外の位置に一致します。 |
| ¥n | 改行に一致します。 |

入力文：

The Universal Declaration of Human Rights Article 1. All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

正規表現:¥bin¥b:単語境界に挟まれた in

The Universal Declaration of Human Rights Article 1. All human beings are born free and equal {*in*} dignity and rights. They are endowed with reason and conscience and should act towards one another {*in*} a spirit of brotherhood.

beings の中の in は、単語境界に挟まれていないので一致しません。

5.1.2. 文字クラス

| | |
|----------------|---|
| [xyz] | 文字セットに含まれている任意の 1 文字に一致します。[...] の中では., ?, *などをエスケープする必要はありません。 |
| [^xyz] | 文字セットに含まれていない任意の 1 文字に一致します。 |
| .(ピリオド) | 改行(¥n)以外の任意の文字に一致します。 |
| ¥w | 英単語に使用される任意の文字(アルファベット、数字、アンダースコア)[a-zA-Z0-9_]に一致します。 |
| ¥W | 英単語に使用される文字以外の任意の文字に一致します。 |

| | |
|-----------|-------------------------------------|
| | [^a-zA-Z0-9[a-zA-Z_0-9]と同じです。 |
| ¥d | 任意の数字に一致します。[0-9]と同じです。 |
| ¥D | 任意の数字以外の文字に一致します。[^0-9]と同じです。 |
| ¥s | 任意のスペース文字に一致します。[¥t¥r¥n¥f]と同じです。 |
| ¥S | 任意の非スペース文字に一致します。[^ ¥t¥r¥n¥f]と同じです。 |

入力文：

All human beings are born free and equal in dignity and rights.

正規表現検索:[e-h] (e, f, g, h, i に一致)

All { *h* }uman b{ *e* }in{ *g* }s ar{ *e* } born { *f* }r{ *e* }{ *e* } and { *e* }qual in di{ *g* }nity and ri{ *g* }{ *h* }ts.

正規表現検索:[^a-v] (a-v 以外に一致)

All{ * *}human{ * *}beings{ * *}are{ * *}born{ * *}free{ * *}and{ * *}equal{ * }in{ * *}dignit{ *y* }{ * *}and{ * *}rights{ *.* }

正規表現検索:[c-i] (c, d, e, f, g, h, i に一致)

All human beings are born free and equal in dignity and rights.

結果 正規表現検索:[^c-i](c, d, e, f, g, h, i 以外に一致)

All human beings are born free and equal in dignity and rights.

5.1.3. 選択、グループ化、繰り返し

| | |
|-------|--|
| | 複数の句を1つの正規表現にまとめ、これらのうちの任意の句に一致します。たとえば、d(os a)は dos または da に一致します。 ¥ba¥b ¥bthe¥b のように(...)でも使うことができます。 |
| (...) | 複数の句をグループ化して1つの句を作成します。(ab)*c は abc または c に一致します。 |
| + | 1個以上の直前の文字に一致します。{1,}と同じです。e+で e, ee, eee, ...に一致します。 |
| * | ゼロ個以上の直前の文字またはぐるに一致します。{0,}と同じです。ah*で a, ah, ahh, ...に一致します。 |
| ? | ゼロ個または1個の直前の文字に一致します。{0,1}と同じです。 |

| | |
|-------|--|
| | books?で book と books に一致します。 |
| {a} | 先行する正規表現 a 個に一致します。[aeoiu]{2}で 2 母音の連続(ei, ee, ua など)に一致します。 |
| {a,} | 先行する正規表現 a 個以上の直前の文字に一致します。[aeoiu]{3,}で 3 母音の連続(aei, uai, auuu など)に一致します。 |
| {a,b} | 先行する正規表現 a 個以上、b 個以下に一致します。[aeoiu]{2,4}で 2-4 母音の連続(ei, aei, uai, auuu など)に一致します。 |

正規表現検索 (free|equal) (free と equal に一致)

All human beings are born {free*} and {equal*} in dignity and rights.

正規表現検索 [e-h]+ ([e-h]の連続に一致)

All {h*}uman b{e*}in{g*}s ar{e*} born {f*}r{ee*} and {e*}qual in di{g*}nity and ri{gh*}ts.

正規表現検索 [aeoiu]{2} (2 母音の連続)

All human b{ei*}ngs are born fr{ee*} and eq{ua*}l in dignity and rights.

結果 3 正規表現検索 [^aeoiu]{2,} (母音以外の文字 2 個以上の連続に一致)

A{*ll h*}uma{*n b*}ei{*ngs *}are{* b*}o{*rn fr*}ee a{*nd *}equa{*l *}i{*nd *}i{*gn*}i{*ty *}a{*nd r*}i{*ghts.*}

5.1.4. エスケープ文字

特殊文字の検索 (,), [,], {, }, ?, !, .(ピリオド), +, *, |, ¥を探るときは、その前に¥をつけてエスケープします。たとえば¥?でクエスチョンマークを検索します。

入力文：

¿Cómo está usted?

正規表現 ¥? (クエスチョンマーク)

¿Cómo está usted?

5.1.5. 置換文字

正規表現の後方参照を使うと、検索式の一部を参照することができます。句を括弧で囲み、\$の後に1つの数字を続けることによってその句を指定します。

| | |
|------|---------------------------|
| \$n | 検索パタンの n 番目の(...)に一致した文字列 |
| \$\$ | \$という文字 |

入力文：

Rumi: Hola, profesor.Prof. Rubio: Buenos di/as.Rumi: Buenos di/as. Nos encontramos ahora en la Universidad [[Complutense]] de Madrid. ¿Dónde nos vamos ahora?
Prof.: Bueno, vamos a iniciar hoy el [[Camino]] del [[Cid]], la primera parte.

正規表現：HTML コードを作成します。

```
a/=>&aacute;  
e/=>&eacute;  
i/=>&iacute;  
ó=>&oacute;  
ú=>&uacute;
```

正規表現：([aeiou])/=>&\$1acute;: 上の連立式を折りたたみます。

Rumi: Hola, profesor.Prof. Rubio: Buenos días.Rumi: Buenos días. Nos encontramos ahora en la Universidad Complutense de Madrid. ¿Dónde nos vamos ahora? Prof.: Bueno, vamos a iniciar hoy el Camino del Cid, la primera parte.

正規表現：#(c%)=>[[\$1]](c で始まる語を[[...]]でマークします。)

Rumi: Hola, profesor.Prof. Rubio: Buenos días.Rumi: Buenos días. Nos encontramos ahora en la Universidad [[Complutense]] de Madrid. ¿Dónde nos vamos ahora?
Prof.: Bueno, vamos a iniciar hoy el [[Camino]] del [[Cid]], la primera parte.

次は中世スペイン語の資料を文字化した資料です。

Otro(22)ssí mando que los menestrales non echen suerte en el jugado por seer

juezes, ca el juez deve tener la seña, e tengo que si <a> afruenta viniessse o a logar de periglo e omne vil o rafez toviessse la seña que podrié (23) caer el concejo en grant onta e en grant vergüença.

(22)は語の途中で改行され ssí 以下が 22 行目になることを示しています。文法研究のためには、これを Otrrossí (22)にする必要があります。これは次の置換式によって実現できます。

正規表現 : (¥(¥d+¥))(&)=>\$2 \$1

Otrrossí (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 viniessse o a logar de periglo e omne vil o rafez toviessse la seña que podrié (23) caer el concejo en grant onta e en grant vergüença.

5.1.6. 後方参照

後方参照を使うことで式の内容を記憶させ、それを後から参照させることができます。

| | |
|--------------------|---|
| (...)¥n | 検索文字列の(...)の式に一致した文字列が記憶され、それを n 回繰り返して参照します。 |
| (...)...(...)=>\$n | 検索文字列の(...)の式に一致した文字列が記憶され、置換文字列でそれを参照して再生します。n は(...)の順番に対応する番号です。 |

入力文 :

どんだんテーマが広がって、ますます興味がわいてきた。

正規表現 : (..)¥1:2 文字が 2 回繰り返す文字列

どんだんテーマが広がって、ますます興味がわいてきた。

参照する文字(列)がわかっているときは、検索式をたとえば「(どん){2}」のようにすることができますが、ここでは他にも「ますます」「ぐんぐん」のように、さまざまに変化する場合を想定しています。¥1 が先行する(..)を後方から参照しています。

5.2. 拡張正規表現

特殊文字を再定義

LETRAS.xlsm では一般の正規表現を拡張して次の検索字を再定義します。

| | |
|----|--|
| # | 単語の境界：#b%はbで始まる単語を検索します。 |
| ## | 文の境界：¥aを区切りとします。 |
| & | 単語文字1個以上 ¥l+ |
| % | 単語文字0個以上 ¥l* |
| ¥l | 西欧語単語文字 [A-Za-zÀ-ó] |
| ¥L | 西欧語単語文字以外 [^A-Za-zÀ-ó] |
| ¥i | キリル文字 |
| ¥I | キリル文字以外 |
| ¥g | ギリシャ文字 |
| ¥G | ギリシャ文字以外 |
| ¥e | ハングル |
| ¥E | ハングル以外 |
| ¥y | CJK互換漢字、統合漢字、漢字拡張文字 |
| ¥Y | CJK互換漢字、統合漢字、漢字拡張文字以外 |
| ¥v | 母音文字 [aeiouÀ-Æà-æÈ-Ëè-ëÏ-ïÒ-Öò-öÛ-Üù-ü] |
| ¥V | 母音文字以外 [^aeiouÀ-Æà-æÈ-Ëè-ëÏ-ïÒ-Öò-öÛ-Üù-ü] |
| ¥c | 子音文字 [bcdfghj-np-tv-zÇçÑñß] |
| ¥C | 子音文字以外 [^bcdfghj-np-tv-zÇçÑñß] |
| ¥a | 文頭句読点：[¿] |
| ¥p | 句読点：[¿!:.?!。 . ? ! : --] |
| ¥P | 句読点以外 |
| n@ | n字連続 |
| n& | n語連続 |

単語の境界

拡張正規表現：#(m%)=>[\$2]（語頭の単語境界#を使った置換式の置換文字列のトークンの数字は+1としてください。）

Otro(22)ssí [mando] que los [menstrales] non echen suerte en el jugado por seer juezes, ca el juez deve tener la seña, e tengo que si <a> afruenta viniesses 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.

拡張正規表現 : %[rs]

Otro(22)ssí [mando] que { *los* } [{ *menestrales* }] non echen suerte en el juzgado { *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.

日本語文字の再定義

日本語文字を検索するときは、¥h (ひらがな), ¥k (カタカナ), ¥z (漢字), ¥j (日本語文字) を使用します。

| | |
|----|--------------------------|
| ¥h | ひらがな [あ-んー] |
| ¥H | ひらがな以外 [^あ-んー] |
| ¥k | カタカナ [ア-ンー] |
| ¥K | カタカナ以外 [^ア-ンー] |
| ¥z | 漢字 [一-顛々ヅ] |
| ¥Z | 漢字以外 [^ 一-顛々ヅ] |
| ¥j | 日本語文字 [あ-んア-ンー一-顛々ヅ] |
| ¥J | 日本語文字以外 [^あ-んア-ンー一-顛々ヅ] |

入力文 : 『坊っちゃん』夏目漱石

親譲りの無鉄砲で小供の時から損ばかりしている。

拡張正規表現 ¥z{2} (漢字 2 文字の連続)

{ *親譲* } りの { *無鉄砲* } 砲で { *小供* } の時から損ばかりしている。

拡張正規表現 : ¥z+ (1 個以上の漢字)

{ *親譲* } りの { *無鉄砲* } で { *小供* } の { *時* } から { *損* } ばかりしている。

拡張正規表現 ¥z*¥h+ (漢字とひらがな : 土漢字+ひらがな)

『 { *坊っちゃん* } 』夏目漱石

{ *親譲りの* } { *無鉄砲で* } { *小供の* } { *時から* } { *損ばかりしている* } 。

外国語文字の再定義

「再定義」というシートには次のような設定をしています。これは自由に変

えることができます。変更したときは「更新」ボタンを押してください。
ここで使用している「/,",~,`,^」という特殊記号を検索するときには前に¥をつけてエスケープしてください。

| | |
|--------|--------|
| ¥/ | x128\$ |
| a/ | á |
| e/ | é |
| i/ | í |
| ó | ó |
| ú | ú |
| A/ | Á |
| E/ | É |
| I/ | Í |
| Ó | Ó |
| Ú | Ú |
| ¥" | x128\$ |
| a" | ä |
| e" | ë |
| i" | ï |
| o" | ö |
| u" | ü |
| A" | Ä |
| E" | Ë |
| I" | Ï |
| O" | Ö |
| U" | Ü |
| x128\$ | " |
| ¥~ | x128\$ |
| a~ | ã |
| e~ | e |
| i~ | i |
| o~ | õ |
| u~ | u |
| A~ | Ã |
| E~ | E |
| I~ | I |

| | |
|--------|--------|
| O~ | Õ |
| U~ | U |
| x128\$ | ~ |
| ¥` | x128\$ |
| a` | à |
| e` | è |
| i` | ì |
| o` | ò |
| u` | ù |
| A` | À |
| E` | È |
| I` | Ì |
| O` | Ò |
| U` | Ù |
| x128\$ | ` |
| ¥^ | x128\$ |
| a^ | â |
| e^ | ê |
| i^ | î |
| o^ | ô |
| u^ | û |
| A^ | Â |
| E^ | Ê |
| I^ | Î |
| O^ | Ô |
| U^ | Û |
| x128\$ | ^ |

前後の条件

{{正規表現}} 検索文字列の前後に付加して検索の条件とします。
 たとえば、**{{te }}**va%は te と空白に続くデータを検索します。

5.3. 単語目録

「単語目録」を選択すると単語を単位にして置換します。次の例では、(1:1)

の In は P に置換されますが、(1:2)の inanis の in は置換されません。大量の単語を置換したり検索するときは単純一致や正規表現と比べて処理が高速になります。

a=>A

de=>DE

en=>EN

A la recepción DE un hotel madrileño llega un profesor extranjero para participar como conferenciante EN un seminario sobre Nutrición organizado por una universidad DE verano con sede EN El Escorial.

この検索式は正規表現ではないので、(a|de|en)のようにまとめることができません。正規表現では、#(a|de|en)#=>\$1 <\$2> \$3 とします。