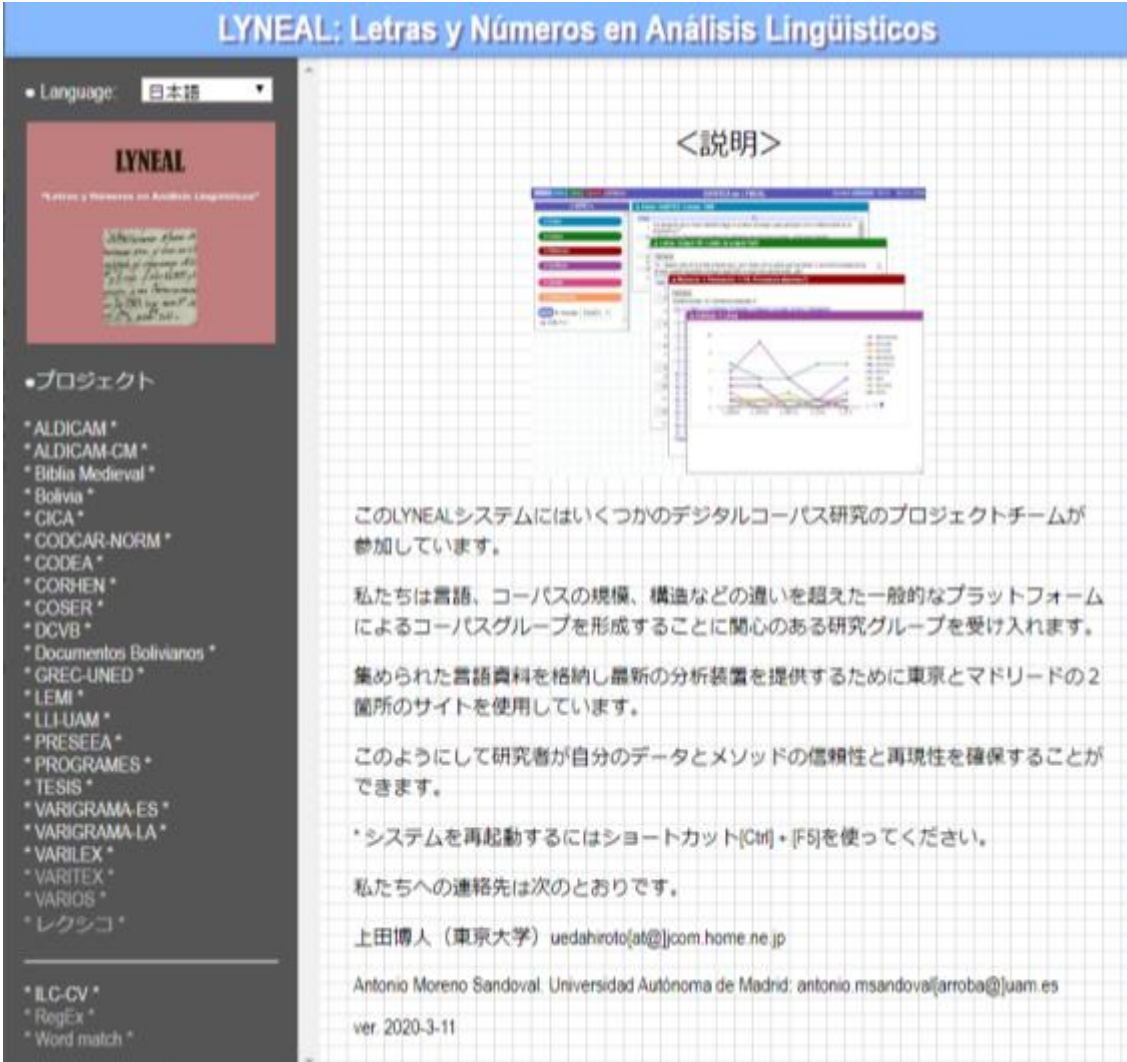


LYNEAL の使い方

上田博人 ver. 2020-2-5



The screenshot displays the LYNEAL software interface. On the left is a sidebar with a 'Language' dropdown set to '日本語' and a list of projects including ALDICAM, Biblia Medieval, CODCAR-NORM, CODEA, CORHEN, COSER, DCVB, Documentos Bolivianos, GREC-UNED, LEMI, LLI-UAM, PRESEEA, PROGRAMES, TESIS, VARIGRAMA-ES, VARIGRAMA-LA, VARILEX, VARTEX, VARIOS, and レクシコ. The main window features a title bar 'LYNEAL: Letras y Números en Análisis Lingüísticos', a '<説明>' (Description) section with Japanese text, and a line graph showing data trends. The text in the description section reads: 'このLYNEALシステムにはいくつかのデジタルコーパス研究のプロジェクトチームが参加しています。私たちは言語、コーパスの規模、構造などの違いを超えた一般的なプラットフォームによるコーパスグループを形成することに関心のある研究グループを受け入れます。集められた言語資料を格納し最新の分析装置を提供するために東京とマドリードの2箇所のサイトを使用しています。このようにして研究者が自分のデータとメソッドの信頼性と再現性を確保することができます。*システムを再起動するにはショートカット[Ctrl]+[F5]を使ってください。私たちへの連絡先は次のとおりです。上田博人（東京大学）uedahiroto[at]jcom.home.ne.jp Antonio Moreno Sandoval Universidad Autónoma de Madrid: antonio.msandoval@aroba[uam].es ver. 2020-3-11'

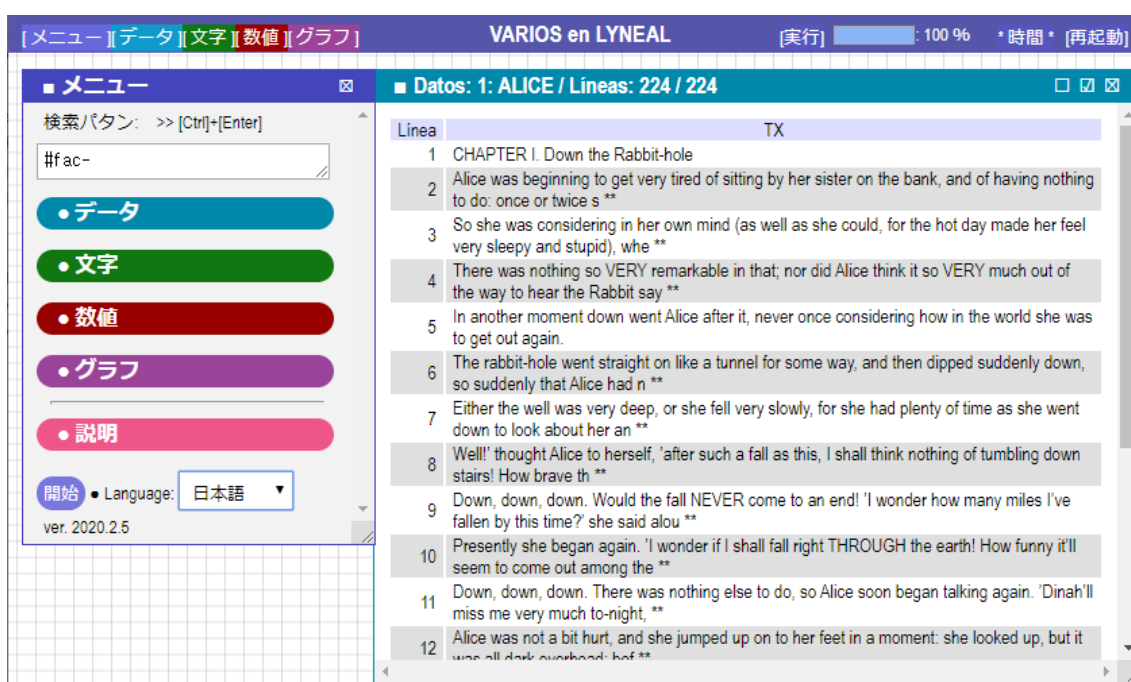
東京大学: <http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/>

マドリード自治大学: <http://shimoda.lllf.uam.es/ueda/lyneal/>

上のサイトにアクセスし、左上の●Languageで「日本語」を選択すると前頁(表紙)の画面になります。

LYNEALはスペイン語 *Letras y Números en Análisis Lingüísticos* 「言語分析の文字と数字」の頭文字です。それぞれの機能は次のように順番で並んで配置されています。ユーザーがボタンをクリックするとその機能がすぐに確認できるインタラクティブなシステムです。操作は直感的で簡単です。

左の VARIOS をクリックし、左下の●Languageするで「日本語」を選択すると次の画面になります。



ページのヘッダー（最上部）の左上の部分に[メニュー], [データ], [文字], [数値], [グラフ]のボタンがあります。中央にページのタイトル, そして右に[実行]ボタン, プログレスバーと%と m.s.(ミリセカンド)と[再起動]ボタンが見えます。「メニュー」をクリックすると左欄が消え, 再度クリックすると左欄が現れます。大きな画面で出力結果を見たいときに利用してください。「実行」ボタンを押したとき, または[Ctrl]+[Enter]キーを押したときにプログラムが実行されます。「●数値」と「●グラフ」はその設定を変更したときにプログラムが自動的に実行されます。「●文字」の設定を変えたときには「実行ボタン」をクリックするか, [Ctrl]+[Enter]キーを押してください。左側のメニューと右側の区画はタイトルバーをドラッグアンドドロップして移動することができます。

1. データ

左の区画の「■メニュー」の「●データ」をクリックするとデータのセクションが展開します。

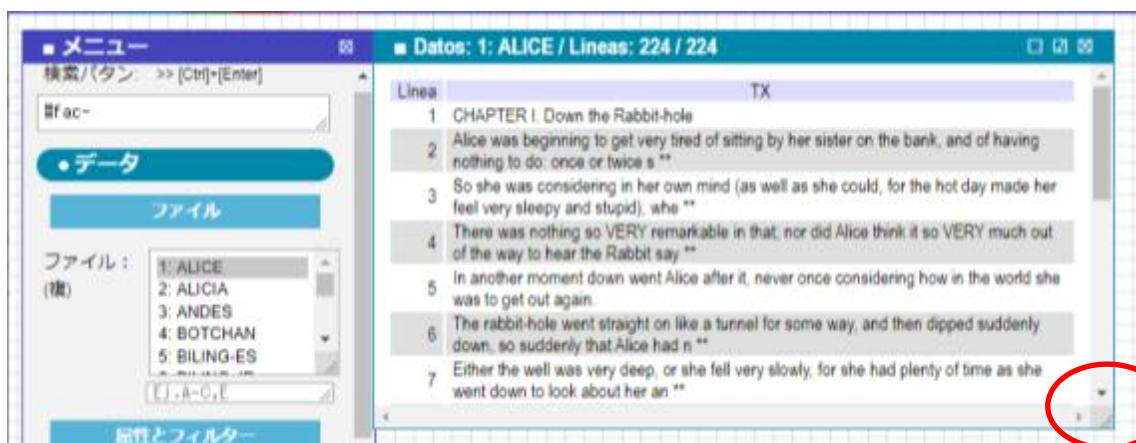


1.1. ファイル

左の区画の「●データ」の「ファイル」をクリックします。



最初のファイル 1:ALICE は、ルイス・キャロル Lewis Carroll 原作の「不思議の国のアリス」Alice's Adventures in Wonderland の各章の抜粋です。



上の赤の○印に囲まれたハンドルを下にドラッグすると出力範囲を拡

大することができます。この範囲拡大機能はシステム全体に共通です。左欄や出力部の右下の隅に同じハンドルがあるときは、それを下にドラッグすると範囲を拡大することができます。

1.2. 属性とフィルター

「●データ」→「属性とフィルター」→「属性(複):」で属性を選択すると、その下に「フィルタ(複):」が現れます。下のようにフィルタを選択すると、出力の■データはフィルタによって選択されたデータだけが現れます。選択されたデータが分析の対象になります。

■データ : 1: VARITEX / 行数 : 218 / 1088

行	TX	ID	TEMA	CIUDAD
1	Un profesor de Tokio llega a la recepción de un hotel de la ciudad de México. Va a participar como conferencista en un s **	3	A_Hotel	3_México
2	- Buenos días... Desearía una habitación individual para pasar tres noches. ¿Qué precio tiene?	3	A_Hotel	3_México
3	- ¡Buenos días, señor! ¡Bienvenido! El hotel tiene habitaciones individuales con televisión. Algunas de ellas tienen vis **	3	A_Hotel	3_México
4	- ¿A los tépel qué?	3	A_Hotel	3_México
5	- A los volcanes, que se llaman Popocatépetl o "Montaña que humea"; e Iztaccihuatl, o "Mujer dormida", en náhuatl.	3	A_Hotel	3_México
6	- ¡Ah, vaya! ¡Ojalá ¿Y la habitación tiene baño y teléfono?	3	A_Hotel	3_México
7	- Sí, claro. No se lo dije porque eso es lo normal en los hoteles... El precio de la habitación es de 100 dólares, unos **	3	A_Hotel	3_México
8	- No, no, muy bien. La que me ha descrito es perfecta para mí.	3	A_Hotel	3_México

1.3. ユーザーデータ

これまでではサーバーに置かれたデータを使いましたが、ユーザーが自分のデータをテキストボックスにコピー&ペーストしたり、自分のファイルをシステムにロードして分析することができます。これらのデータはユーザーの環境だけで使用され、サーバーにセーブしたり、一般に公開されたりはしません。

ユーザーデータ

テキストボックス

ユーザーファイル

1.3.1. テキストボックス

左欄の「データ」の「ファイル」で「テキストボックス」を選択すると、次のように右に空欄が現れます。[Ctrl]+C でコピーしたテキストをこの中をクリックして[Ctrl]+V でペーストしてください。ペースト

した後に、テキストボックスの下の「*クリック*」を押してください。



テキストファイルは次のようにタブコードで分離されていれば、最初の行が列と属性のタイトルになります。Excel シートからコピーすることも可能です。テキスト部を「TX」とし、属性には最初に「##」をつけてください。

TX	##OBRA	##Folio	##Línea
Delos sos oios tan fuerte mientras lorando	13a.Cid	1r	1
Tornaua la cabeça & estaua los catando	13a.Cid	1r	2
Vio puertas abiertas & vços sin ca<n>nados	13a.Cid	1r	3
Alcandaras uazias sin pieles & sin mantos	13a.Cid	1r	4
E sin falcones & sin adtores mudados	13a.Cid	1r	5

データは次の「##OBRA::13a.Cid」のように、属性が変更されたところだけにそのことをタイトル（##OBRA）とコロン2つ「::」と属性名を書くこともできます。属性が変わる位置で、たとえば「##OBRA::13b.Fazienda」のように属性名だけを変えて指定してください。タイトル名を変えれば複数のタイトルと属性が指定できます。

TX
##OBRA::13a.Cid
Delos sos oios tan fuerte mientras lorando
Tornaua la cabeça & estaua los catando
Vio puertas abiertas & vços sin ca<n>nados
##OBRA::13b.Fazienda
emont por la gracia de dios. arço=
~bispo de Toledo. a don almeric. arçi=
~diano de antiochia con grant a=

次の入力データ形式も可能です。

TX	##PAPEL	##SEXO	##EDAD	##VARIEDAD
	##LENGUA			
\$SE	Entrevistador(a)	*	*	*

\$\$I Informante Mujer 35 * Castellano

¿Qué opinas de la violencia familiar, y porque las crees que las mujeres no reaccionen ante una agresión de sus parejas? E

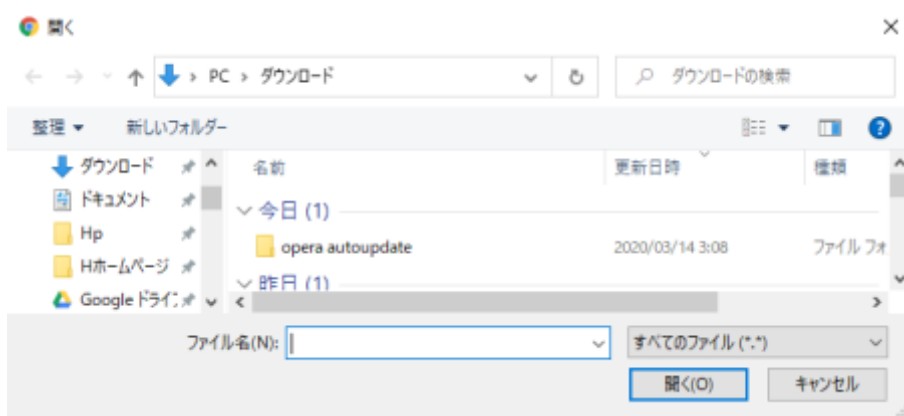
¡Por qué si te das de cuenta!, en todos los episodios si te vas a toda la historia de la humanidad ya sea he...bíblicamente, históricamente siempre la mujer ha estado metida, o tratando de tener poder atreves del hombre.

I

上の形式では、最初の行でそれぞれのカラムのタイトルを設定し、次の行の\$\$EはEがEntrevistador(a) * * * * *で決定されます。データ行のEは、\$\$Eで決定されたカラムのデータを展開します。このようにすれば、繰り返される同一のカラム群が1回の規定(\$\$E)だけで済ますことができます。

1.3.2. ユーザーファイル

左欄の「データ」の「ファイル」で「ユーザーファイル」を選択すると次のようなユーザーのディレクトリーの画面が現れます。



ファイルの構造はテキストボックスの場合と同じです。ファイルは拡張子が".txt"であるテキストファイルを使用してください。

1.3.3. 数値行列データ

データはテキストだけでなく数値行列であってもかまいません。次は、1:0データの例です。

N.D1.0:1	A	B	C	D
i.1	1	1	0	0
i.2	0	0	1	0
i.3	0	1	0	0
i.4	0	0	1	1

i.5	1	1	1	0
-----	---	---	---	---

次は頻度データの例です。

N.D2.fre.	A	B	C	D	E
i.1	10	19	14	7	12
i.2	11	7	10	0	1
i.3	0	0	1	12	1
i.4	0	1	2	3	3

次は小数点のあるデータ例です。

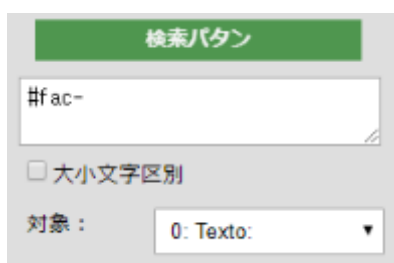
N.D4.ss-ff	-ss-	#ss-	-ff-	#ff-		
R: León	4.8	5.8	1.4	6.1		
R: C. la Vieja	9.7	9.6	1.5	10.1		
R: C. la Nueva	1.8	2.4	0.5	1.9		
R: Aragón	11	0.9	2.1	3		
A: 1200	16.8	3.1	1.2	6.2		
A: 1225	5.7	0.2	0.5	0		
A: 1250	12.2	3.1	1.5	6.8		
A: 1275	13.1	13.8	1.5	12.4		
A: 1300	13.4	14.8	4	22.1		
A: 1325	14.3	20.5	3.6	23.8		
A: 1350	8.9	10.6	3.1	9.4		
A: 1375	3.7	3.7	1.9	3.3		
A: 1400	6.6	1	0.9	1.2		
A: 1425	3.9	0.5	0.7	1.1		
A: 1450	7.8	0.5	1.1	2.3		
A: 1475	7.3	0	0.9	0.8		
L: cortesana	0.6	0.5	0.7	0.8		
L: albalaes	19.2	19.4	3	12.9		
L: privilegios	7.9	1.6	3.1	7.9		
L: gótica	2.8	0.6	0.6	2.2		
L: gótica cursiva		10.7	15.8	2.8	19.6	
L: gótica cursiva [albalaes]			9.2	19.2	1.8	20.3
L: gótica cursiva [precortesana]				0.2	2	0.3 1.7
L: gótica libraria		10.3	9	3.5	11	
L: gótica redonda		12.8	1.1	0.7	1.6	
L: humanística	3.3	0	3.3	0		
L: humanística redonda			0.5	3.7	0	1
L: precortesana		3.9	3.9	0.7	2.7	

(1) データのそれぞれの値はタブで区切られていること、(2) 第1行に属性のタイトルがあること、(3) 第1列に個体のタイトルがあること、そして(4)左上の最初のセルが「N.」（大文字エヌとピリオド）で始まることが必須の条件です。この条件を満たせば LYNEAL はこのデータをテキストではなく、数値行列として扱い、文字の処理をしないで数値分析を行います。

2. 文字



2.1. 検索パターン



2.1.1. シャープ・ハイフン・イコール：#, -, =

言語形式を検索するために特殊な記号で表現する「パターン」を使います。たとえば、シャープ「#」とハイフン「-」はパターンの中では次を意味します。

- # 語の境界であること
- 文字が1個以上あること

語の境界(#)は、たとえば A, a で始まる語を探す次のパターンなどで使います。

#a-

このパターンで Alice, and, a, as などのように A, a で始まる語を検索することができます。

イコール(=)の記号はハイフン(-)と同じように a だけでなく、a に続く文字があるかぎり、それが途切れるまで検索します。そしてイコール(=)は「文字がない」ということも示すので、「#a%」は a だけでもマッチします。ここで「文字」は単語の中で使用するアルファベットや各国語の文字（ひらがな、カタカナ、漢字など）を意味します。「%」や「?」などの記号やアラビア数字(1, 2, ...)は含みません。

同様にして、パターン「-d#」は tired, and, had などのように語尾が d である語を検索します。

パターンは次のように、[Enter]キーで改行していくつでも書くことができます。

-d#
-r#

大文字と小文字を区別する必要があるときには、「大小文字区別」のチェックボタンを押してください。

「#」の代わりにブランク（空白）を使うこともできます。パターンの始点と終点のブランク（空白）は「#」と見なします。

ハイフン「-」は 1 個以上の文字連続を示すので、スペースなどで囲まれた 1 単語を検索します。したがって 2 語の連続は「- -」で検索します。

2.1.2. アットマーク : @

アットマーク「@」は 1 文字を検索します。たとえば「#d@s-」は *das-*, *des-*, *dis-*, などで始まる語を検索します。

2.1.3. ピリオド : .

ピリオド(.)は文字、数字、記号、スペースに対応します。

2.1.4. バックスラッシュ+d: ¥d

数字を検索するときにはパターン「\d」を使います。

2.1.5. 角括弧: [...]

個別の文字をグループにするときには角括弧[...]を使います。たとえば「#amig[oa]#」は amigo と amiga を検索します。母音で終わる語を検索するときには「&[aeiou]」を使います。

角括弧の中でハイフン(-)を使って、一連の文字を文字コードの小さな文字と大きな文字でならべます。たとえば「#[a-k]&」は小文字の a, b, c, ..., k で始まる語を検索します。曲アクセント記号(^)を[の直後に置くと、続く文字以外の文字を指定します。たとえば「#[^a-k]&」は小文字の a, b, c, ..., k で始まる語を検索します。このとき[...]の中に空白を含めます。

2.1.6. 丸括弧と縦棒: (...|...)

複数のパターンを丸かっこの中に縦棒で区切って並べてまとめることができます。たとえば「(king|queen)」は king と queen を同時に検索します。また、「#(have|has|had)#」と「#ha(ve|s|d)#」はどちらも have, has, had を検索します。「(king|queen)」は丸かっこを外して「king|queen」としても同じ結果になります。

2.1.7. バックスラッシュ: ¥

パターンで使用する記号そのものを検索するときにはバックスラッシュ「\」（または半角の¥記号）をその記号の前につけます。たとえば「\-」というパターンはハイフン(-)を検索します。

2.1.8. 量の指定: ?, *, +, {***}

次の記号によって文字やパターンが連続する数を指定することができます。

記号	意味	例
?	先行する文字やパタ	books?で book と books を検索します

	ンがあってもなくてもよい	
*	先行する文字やパターンがゼロ個以上ある	[aeo]*h で h, ah, aah, aaah, ..., eh, aoooh などを検索します。
+	先行する文字やパターンがゼロ個以上ある	[aeo]+h で ah, aah, aaah, ..., eh, aoooh などを検索しますが、h は検索しません。
{x}	先行する文字やパターンが x 個ある	\d{3} で 124, 366, 889 などの数値を検索します。
{x,}	先行する文字やパターンが x 個以上ある	\d{3,} で 124, 3667, 8899 などの 3 桁の数値を検索します。
{x,y}	先行する文字やパターンが x 個以上 y 個以下ある	\d{3,4} で 124, 3662, 8891 などの 3-4 桁の数値を検索します。

2.1.9. 文脈 : {***}

たとえば単語 + 前置詞 (of, in, on, from) を検索するには「-(of|in|on|from)」というパターンを使います。

Núm.	Contexto anterior (A).	Forma	Contexto posterior (P)
1	Alice was beginning to get very	tired of	sitting by her sister on the bank, and of having *
2	*was beginning to get very tired of sitting by her	sister on	the bank, and of having nothing to do: once or tw*
3	* very tired of sitting by her sister on the bank,	and of	having nothing to do: once or twice she had peepe*

ここで前置詞をフォームに含めずに、文脈として後文に位置させるにはパターンを「- {{(of|in|on|from)}}」とします。

Núm.	Contexto anterior (A).	Forma	Contexto posterior (P)
1	Alice was beginning to get very	tired	of sitting by her sister on the bank, and of havin*
2	*was beginning to get very tired of sitting by her	sister	on the bank, and of having nothing to do: once or *
3	* very tired of sitting by her sister on the bank,	and	of having nothing to do: once or twice she had pee*

前文脈でも同様です。たとえば「{{(of|in|on|from)}} -」は次を検索します。

Núm.	Contexto anterior (A).	Forma	Contexto posterior (P)
1	Alice was beginning to get very tired of	sitting	by her sister on the bank, and of having nothing *
2	*ning to get very tired of sitting by her sister on	the	bank, and of having nothing to do: once or twice *
3	*tired of sitting by her sister on the bank, and of	having	nothing to do: once or twice she had peeped into *

* 以上の場合には文脈またはフォームに対応するパターンに空白（スペース）を含めなければなりません。

2.1.10. 繰り返し : (***) ¥1

丸括弧(...)と「\1」は丸括弧にあるパターンが繰り返されることを示します。たとえば「(-)\1」は次の検索します。

行	前文	パターン	フォーム	後文	Chap*
10	~the top of the house!' (Which was very likely true.) /	(#&#)\1	Down, down,	down. Would the fall NEVER come to an end! 'I wonder h~	1
12	~o ask: perhaps I shall see it written up somewhere.' /	(#&#)\1	Down, down,	down. There was nothing else to do, so Alice soon bega~	1
143	~/ 'Perhaps it hasn't one,' Alice ventured to remark. /	(#&#)\1	Tut, tut,	child!' said the Duchess. 'Everything's got a moral, i~	9

丸括弧(...)が 2 つあるときは\1 と\2 を使って順に対応するパターンを繰り返します。

2.1.11. 定義 : ***==***#a%

次のように複雑なパターンを簡単な文字（列）に==を使って定義し、繰り返し使用することができます。このとき、たとえば

```
c==[bcdfghjklmnpqrstvwxyz]
```

のように、パターンの中にある文字列を使ってはいけません。

```
C==[bcdfghjklmnpqrstvwxyz]
V==[aeiou]
{{V}}C{{V}}
{{V}}CC{{V}}
{{V}}CCC{{V}}
```

Line	Anterior text	Pattern	Form	Posterior text	Chap*
2	~*** / CHAPTER I. Down the Rabbit-ho	{{V}}C{{V}}		e / undefined~	1
2	~*** / CHA	{{V}}CC{{V}}	PT	ER I. Down the Rabbit-hole / undefined~	1
2	~*** / CHAPTER I. Down the Ra	{{V}}CC{{V}}	bb	it-hole / undefined~	1
3	~CHAPTER I. Down the Rabbit-hole / A	{{V}}C{{V}}		ice was beginning to get very tired of sitting by her s~	1

■ Numbers: 1. Score. 1. Absolute frequency														
→	Pattern	1	2	3	4	5	6	7	8	9	10	11	12	*Reset*
1	{{V}}C{{V}}	214	157	237	193	184	170	222	267	237	136	214	137	{{V}}C{{V}}
2	{{V}}CC{{V}}	142	93	142	109	138	107	127	114	131	73	111	74	{{V}}CC{{V}}
3	{{V}}CCC{{V}}	11	22	18	21	15	14	18	17	17	31	14	11	{{V}}CCC{{V}}
Reset		1	2	3	4	5	6	7	8	9	10	11	12	

上の検索式を単純に

```
C
CC
CCC
```

とすると、Cの中にCCのケースが含まれてしまいます。

2.1.12. 置換 : ***=>***

パターンは検索だけでなく置換にも使われます。たとえば「a=>A」はテキストの中のすべての小文字の a を大文字の A に変えます。「a=>」はすべての小文字の a を消去します。中世スペイン語の省略形は ynf<n>t のように「<...>」で囲って転記されるので、完全形にして検索するときには次のようにして、「<」と「>」を置換で消去してから検索します。「&t#」は part, ynfant などを検索します。

```
[<>]=>
&t#
```

省略された文字列を消去するときは「<&=>」という置換パターンによって、たとえば gra<n>t は grat に変わります。

2.1.13. 除外

パターンの後に{-}をつけると、検索された形式がそのパターンに一致するとき除外されて検索されません。たとえば次の%d は語末が d の

{nuestro}	(nuestr[oa]s?)
{vuestro}	(vuestr[oa]s?)

パターン	正規表現
{art}	({el} {un})
{clit}	(me te nos os se l[oa]s?)
{conj}	(como cuando cuanto donde o pero mas ni aunque conque mientras porque pues que sino si)
{dem}	({este} {ese} {aquel})
{indef}	(algo alguien nada nadie algún algun[oa]s? ningún ningun[oa] mism[oa]s? otr[oa]s?)
{inf}	&[aeií]r
{ger}	&([á]ndo i[eé]ndo)
{interr}	(adónde cómo cuál cuándo cuánto dónde qué quién)
{num}	(un un[oa]s? dos tres cuatro cinco seis siete ocho nueve diez once doce trece catorce quince veinte treinta cuarenta cincuenta sesenta setenta ochenta noventa cien ciento (dos tres cuatro seis sete ocho nove)cient[oa]s quinient[oa]s mil millón)
{ord}	%((primer segund tercer cuart quint sext séptim octav noven décim)[oa]s? primer tercer)
{pos}	({mi} {tu} {su} {nuestro} {vuestro})
{pp}	&([aií]d abiert cubiert escrit muert puest resuelt rot vist vuel dich hech)[oa]s?
{prep}	(en por a con p[oa]ra sobre so ante bajo contra sin entre [fh]asta desde de tras cabe? faza hacia)
{pron}	(yo tú [nv]osotr[oa]s ell[oa]s ustedes usted)
{rel}	(que donde quien cuy[oa]s? cual(es)?)

LYNEAL は次の変換規則を備えているので、たとえば"llam{ar}"や"com{er}"というパターンで、その規則変化形はすべて検索します。この検索は、形式を検索するので、意味や機能は考慮されません。たとえば com{er}は *como* を検索しますが、これは接続詞も同時に検索します。

パターン	正規表現
{ar}	((IPa} ar{FC} {Pa} {Sa} {SPa} {Ia} ar ad[oa]s? [á]ndo ad ares)
{er}	((Ipe} er{FC} {Pe} {Se} {SPE} {Ie} er id[oa]s? i[eé]ndo ed eres)
{ir}	((Ipe} ir{FC} {Pe} {Se} {SPE} {Ii} ir id[oa]s? i[eé]ndo id ires)

パターン	正規表現
{Ia}	(o amos áis a[sn]? ades)

{Ie}	(o emos éis e[sn]? [ée]des)
{Ii}	(o imos ís e[sn]? [íi]des)
{Ie2}	(emos éis e[sn]? [ée]des)
{Ii2}	(imos ís e[sn]? [íi]des)
{IPa}	(abais ábamos aba[sn]?)
{IPe}	(íamos íais ía[ns]?)
{Pa}	(é asteis ó amos aste aron)
{Pa2}	(asteis ó amos aste aron)
{Pe}	(í isteis ió imos iste ieron)
{RP}	(e isteis o imos iste ieron)
{F}	(é á[sn]? emos éis)
{FC}	(é á[sn]? emos éis íamos íais ía[sn]?)
{Sa}	(emos éis e[sn]? [ée]des)
{Se}	(amos áis a[sn]? ades)
{SPa}	(a(ra se)(is [sn]?) á(ra se)mos)
{SPe}	(ie(ra se)(is [sn]?) ié(ra se)mos)
{SP2}	(e(ra se)(is [sn]?) é(ra se)mos)

2.1.15. 大小文字の区別

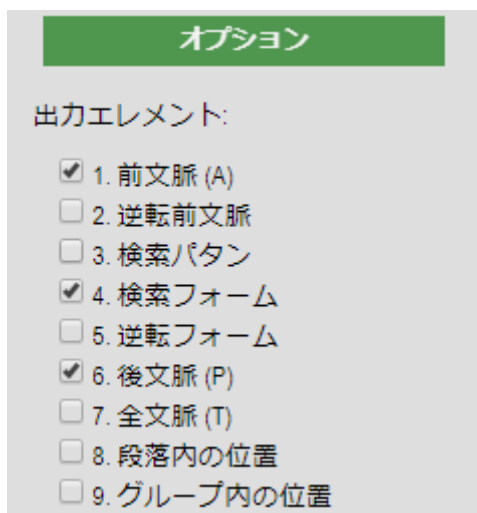
ふつう検索は大文字と小文字の区別なく行いますが、「大小文字区別」のボタンをチェックすると大文字と小文字を区別した検索になります。たとえば **Qué** は大文字の **Q** がある文頭の疑問詞を検索します。

2.1.16. 対象

テキスト以外の属性を対象に選択することができます。

2.2. オプション

2.2.1. 出力エレメント



オプション

出力エレメント:

- 1. 前文脈 (A)
- 2. 逆転前文脈
- 3. 検索パターン
- 4. 検索フォーム
- 5. 逆転フォーム
- 6. 後文脈 (P)
- 7. 全文脈 (T)
- 8. 段落内の位置
- 9. グループ内の位置

標準の出力リストは前文脈，検索フォーム，後文脈です。他に上の図にあるさまざまなオプションがあります。

それぞれの項目は次を意味します：

出力エレメント	説明
1. 前文脈	検索文字列の前文（前にある文の一部）
2. 逆転前文脈	検索文字列の前文脈の逆配列
3. パターン	ユーザーが入力したパターン
4. フォーム	パターンに一致した文字列
5. 逆転フォーム	フォームの逆配列
6. 後文	検索文字列の後文（後にある文の一部）
7. 全文脈	文の中にフォームがある形式
8. 段落内の位置	
9. グループ内の位置	

2.2.2. 出力属性・プロセス・出力限度

出力属性 (複) (No)
 1: Chapter.

プロセス:

1. 漸増検索
 2. 文書数
 3. フォームで並べ替え

出力限度: 1 000 ▲:▼

漸増検索では、多くの検索は連続して行います。たとえば「@@」で2文字を検索すると次の結果になります。ここでは CHAPTER の CH, AP, TE が検索されています。

■ 文字: 出力:100 / 出力限度:100

行	前文	ボタン	フォーム	後文
2	~undefined /	@@	CH	APTER I. Down the Rabbit-hole / undefined~
2	~undefined / CH	@@	AP	TER I. Down the Rabbit-hole / undefined~
2	~undefined / CHAP	@@	TE	R I. Down the Rabbit-hole / undefined~
2	~undefined / CHAPTER I.	@@	Do	wn the Rabbit-hole / undefined~
2	~undefined / CHAPTER I. Do	@@	wn	the Rabbit-hole / undefined~

このような連続検索では一度検索された部分は次の検索の範囲に含まれません。ここで「●文字」→「オプション」→「プロセス」の「漸増検索」をチェックすると次のように、CH, HA, AP, PT, TE, ER, ... のように次々と1文字ずつ移動しながら検索を続けるので、すべての可能な2文字連続が見つかります。

■ 文字: 出力:100 / 出力限度:100

行	前文	ボタン	フォーム	後文
2	~*** /	@@	CH	APTER I. Down the Rabbit-hole / undefined~
2	~*** / C	@@	HA	PTER I. Down the Rabbit-hole / undefined~
2	~*** / CH	@@	AP	TER I. Down the Rabbit-hole / undefined~
2	~*** / CHA	@@	PT	ER I. Down the Rabbit-hole / undefined~
2	~*** / CHAP	@@	TE	R I. Down the Rabbit-hole / undefined~
2	~*** / CHAPT	@@	ER	I. Down the Rabbit-hole / undefined~
2	~*** / CHAPTER I.	@@	Do	wn the Rabbit-hole / undefined~

2語の連続は「#& &」を使います。「& &」とすると語の途中から検索するので、完全な語の連続にはなりません。なお、この処理は時間がかかります。

行	前文	パタン	フォーム	後文
2	~*** /	#& &	CHAPTER I	. Down the Rabbit-hole / undefined~
2	~*** / CHAPTER I.	#& &	Down the	Rabbit-hole / undefined~
2	~*** / CHAPTER I. Down	#& &	the Rabbit	-hole / undefined~
3	~CHAPTER I. Down the Rabbit-hole /	#& &	Alice was	beginning to get very tired of sitting by her sister o~
3	~CHAPTER I. Down the Rabbit-hole /	#& &	was beginning	to get very tired of sitting by her sister o~

2.3. 文脈

文脈

前文脈パタン (A)

後文脈パタン (P)

全文脈パタン (T)

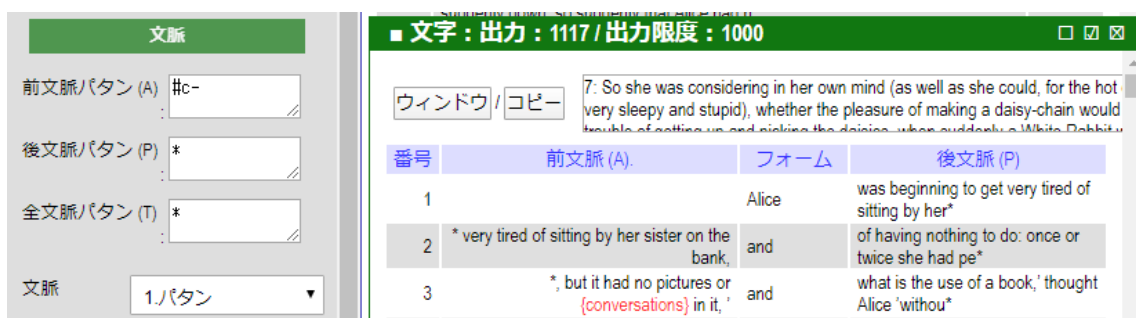
文脈

単位:

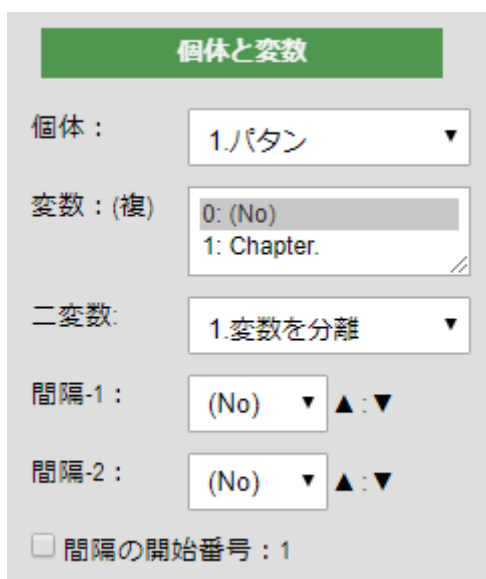
前文脈文字数: ▲:▼

後文文字数: ▲:▼

前文脈，後文脈，全文脈にパタンを設定すると，パタンに一致した文字列が赤字でハイライトされます。



2.4. 個体と変数



たとえば章番号や地点や年代などの変数がある言語データを扱うとき、それらの変数と検索されたケース（語や文字列など）の関係を探ることがあります。次は変数として 1: Chapter を選択したところです。次の図は検索パタンの Alice, Rabbit, King, Queen とこの変数と coche, auto, carro という個体との関係を示します。

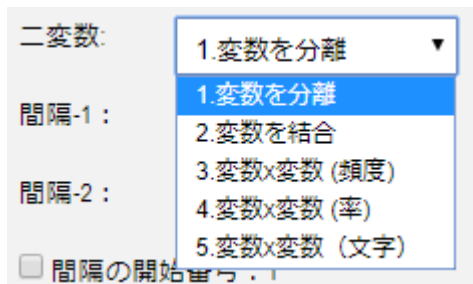


間隔を 2 とし、間隔の開始番号:1 にして実行すると次の図が得られます。たとえば Alice は Chapter 1, 2 で 19 回現れています。



2.5. 二個の変数

「●文字」で変数を2個以上選択すると「変数を分離」「変数を結合」「変数 x 変数」のいずれかを選択することができます。ここではVARITEXのCIUDADとTEMAを選択します。



2.5.1. 変数を分離

下図のように2個の変数を分離して出力します。



2.5.2. 変数を結合

下図のように2個の変数を結合して出力します。



2.5.3. 変数 x 変数 (頻度)

下図のように 2 個の変数を行列にして出力します。パターン: #fa-



パターン: #sa-



2.5.4. 変数 x 変数 (率)

パターン: #fa- [改行] #sa-

■ 数値: 1. AF: 絶対頻度 (*) - 1. 得点分析

行数: 10 / 列数: 5

ウインドウ / コピー: 書式あり / コピー: 書式なし

→	絶対頻度	1_Madrid	2_Sevilla	3_México	4_Lima	5_B.A.
1	A_Hotel				0.333	
2	B_Camino	0.143	0.400	0.143	0.125	0.167
3	C_Funcionarios	0.500		0.250		0.500
4	D_Estudiantes	0.667	1.000	0.667	0.400	0.500
5	E_Capadocia	0.400	0.250	0.200		0.375

2.5.5. 変数 x 変数 (文字)

パターン: #fa- [改行] #sa-

■ 数値: 1. AF: 絶対頻度 (*) - 1. 得点分析

行数: 10 / 列数: 5

ウインドウ / コピー: 書式あり / コピー: 書式なし

*	1_Madrid	2_Sevilla	3_México	4_Lima	5_B.A.
A_Hotel	#sa-3;	#sa-1;	#sa-3;	#fa-1; #sa-2;	#sa-2;
B_Camino	#fa-1; #sa-6;	#fa-2; #sa-3;	#fa-1; #sa-6;	#fa-1; #sa-7;	#fa-1;
C_Funcionarios	#fa-3; #sa-3;	#sa-3;	#fa-1; #sa-3;	#sa-2;	#fa-1;
					#sa-1;

3. 数値

「数値」のセクションにあるさまざまな機能について説明します。

- 数値
 - 分析方法
 - オプション

*この説明書では LYNEAL の操作法だけを説明します。数値処理、確率、統計については次の『限度データ分析法』→2.3『言語研究のための数値データ分析法』を参照してください。

● 説明

- * 使い方 (スペイン語) [pdf]
- * 使い方 (日本語) [pdf]
- * 引用の仕方(スペイン語) [pdf]
- * 引用の仕方(日本語) [pdf]
- * Análisis de datos lingüísticos [pdf]
- * 『言語データ分析法』 [pdf]

- 『言語研究のための数値データ分析法』
 - → 第1章 グラフ[PDF]
 - → 第2章 行列[PDF]
 - → 第3章 確率[PDF]
 - → 第4章 統計[PDF]
 - → 第5章 得点[PDF]
 - → 第6章 関係[PDF]
 - → 第7章 分析[PDF]
 - → 第8章 推測[PDF]
 - → 第9章 検定[PDF]

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

3.1. 分析方法

分析方法

行列：

分析：

個体 属性 両方

相関：

グループ：

無集中化

個体と変数の集中化：

個体の集中化

変数の集中化

3.1.1. 行列

出力の「数値」には次の数値行列が出力されます。左欄の「数値」セクションの「行列」から数値行列の種類を選択します。



絶対頻度

次の「絶対頻度」は各個体が各属性に出現した回数を示します。

→	A.F.	0	2	4	6	8	10	12
1	#Alice#	13	17	24	31	37	18	12
2	#King#	0	0	0	0	4	12	8
3	#Queen#	0	0	0	2	23	4	1

相対頻度

(3) 横パーセント・縦パーセント

■ 数値: 1. 得点. 3. 横パーセント (%)								
→	ボタン	2	4	6	8	10	12	*リセット*
1	#Alice#	12.5	13.8	14.5	26.3	19.7	13.2	#Alice#
2	#Queen#	.0	.0	6.7	66.7	10.0	16.7	#Queen#
3	#King#	.0	.0	.0	16.7	.0	83.3	#King#
	リセット	2	4	6	8	10	12	

■ 数値: 1. 得点. 4. 縦パーセント (%)								
→	パターン	2	4	6	8	10	12	*リセット*
1	#Alice#	100.0	100.0	91.7	62.5	90.9	44.4	#Alice#
2	#Queen#	.0	.0	8.3	31.3	9.1	11.1	#Queen#
3	#King#	.0	.0	.0	6.3	.0	44.4	#King#
	リセット	2	4	6	8	10	12	

3.1.2. 正規化頻度

絶対頻度はそれぞれの属性の中での出現数なので、属性間で比較するためには使えません。同様に横パーセントも属性間で比較するためには使えません。それぞれの属性の規模が異なるからです。ここでの属性は各章(Chap.)ですから、それぞれの章の大きさが異なることは明らかです。

そこで先の(2)で見た文字数と語数を使って絶対頻度を正規化します。次がそれぞれの属性(章)に文字数を 1000 に揃えて正規化した数値(千字率)を示します。

■ 数値: 1. 得点. 5. 千字率 (%)								
→	パターン	2	4	6	8	10	12	*リセット*
1	#Alice#	2.5	2.7	3.4	4.7	4.4	3.3	#Alice#
2	#Queen#	.0	.0	.3	2.4	.4	.8	#Queen#
3	#King#	.0	.0	.0	.5	.0	3.3	#King#
	リセット	2	4	6	8	10	12	

たとえば Alice: Chap.2 の頻度が 19 であり、Chap.2 の文字数が 7624 なので、 $19 : x = 7624 : 1000$, $119 * 1000 / 7624 = 2.5$ が Alice: Chap.2 の 1000 字あたりの頻度になります。ここでは Alice, King, Queen という単語の頻度を調べているので、正規化頻度はむしろ 1000 語を基準にすべきです。次が千字率を示しています。

■ 数値: 1. 得点. 6. 千字率 (%)								
→	パターン	2	4	6	8	10	12	*リセット*
1	#Alice#	9.6	10.6	13.5	19.0	17.3	13.2	#Alice#
2	#Queen#	.0	.0	1.2	9.5	1.7	3.3	#Queen#
3	#King#	.0	.0	.0	1.9	.0	13.2	#King#
	リセット	2	4	6	8	10	12	

たとえば Alice: Chap.2 の頻度が 19 であり、Chap.2 の語数が 1982 なので、 $19 * 1000 / 1982 = 9.6$ が Alice: Chap.2 の千語率です。

* 日本語は単語が空白で分離されていないので、テキストのままでは語数が計算できません。そこで疑似的な文節として「漢字またはカタカナの連続」またはそれにひらがなの連続、またはひらがなの連続を設定してこれを語数の近似値の計算に使用します。

日本語の疑似的な文節： $(\backslash z|\backslash k)+\backslash h*\backslash h+$

ここで $\backslash h$ はひらがな ([あ-けー])、 $\backslash k$ はカタカナ ([ア-ロ])、 $\backslash z$ は漢字 ([一-籲々ヾ]) を示します。

3.1.3. 調整相対頻度

相対頻度、パーセント、パーミル、そして正規化頻度(パーミليون)は、どれも実測値(絶対頻度)の規模(総和または平均値)と異なります。ここで、相対頻度の規模を実測値の規模と等しくした値を「調整相対頻度」(Adjusted Relative Frequency: ARF)と呼ぶことにします。

たとえば、次のデータは 5 種のテキストの中で現れた特定の要素を示します。

OF	1 Madrid	2 Sevilla	3 México	4 Lima	5 B.A.	O.sum
#fa-	24	24	17	15	19	99

それぞれのテキストの総語数は次のように異なります。

Total	1 Madrid	2 Sevilla	3 México	4 Lima	5 B.A.	T.sum
Palabra	5505	5230	5898	4680	5271	26584

それぞれのテキストの総語数が異なるので、上の実測値(OF)はそのままでは比較できません。そこで、実測値を総語数で割って、次の相対頻度(RF)を求めます。

RF	1 Madri	2 Sevill	3 Méxic	4 Lima	5 B.A.	R.sum
#fa-	0.00436	0.00458	0.002882	0.00320	0.00360	0.01864

相対頻度はとても小さな数値になり、扱いにくいので、その和(R.sum)が実測値の和(O.sum)と等しくなるように、全体に $O.sum/R.sum$ を掛けて、次の調整相対頻度(ARF)を求めます。

ARF	1 Madrid	2 Sevilla	3 México	4 Lima	5 B.A.	A.sum
-----	----------	-----------	----------	--------	--------	-------

#fa-	23.15408	24.37155	15.30798	17.02233	19.14407	99
------	----------	----------	----------	----------	----------	----

上の調整相対頻度の和(A.sum)は実測値の和(O.sum)と等しくなります。実際は ARF の小数点以下 1 桁または 0 桁を使用します。実測値が 3 桁以上であれば、小数点以下は不要です。そのとき、調整相対頻度の和(A.sum)と実測値の和(O.sum)が完全に一致しないこともあります。この非常に小さな誤差は無視して構わないでしょう。

同じように、次のような行列でも、その調整相対頻度が計算できます。

OF	1 Madri	2 Sevill	3 Méxic	4 Lim	5 B.A.	sum.O
A Hotel	0	0	0	1	0	1
B Camino	1	2	1	1	1	6
C Funcionario	2	0	0	0	0	2
D Estudiante	2	2	1	1	2	8
E Consultorio	4	2	2	0	3	11
F_Fiesta	1	3	3	1	0	8
G Política	1	3	1	0	1	6
H_Compras	8	7	5	6	6	32
I Fútbol	0	1	0	1	3	5
J Despedida	5	4	4	4	3	20
sum.O	24	24	17	15	19	99

Total	1 Madri	2 Sevill	3 Méxic	4 Lim	5 B.A.	sum.T
A Hotel	411	256	495	418	444	2024
B Camino	473	456	516	417	525	2387
C Funcionario	527	407	533	408	440	2315
D_Estudiante	471	474	522	407	488	2362
E Consultorio	668	696	697	583	589	3233
F Fiesta	504	493	537	370	437	2341
G Política	729	663	748	516	701	3357
H Compras	464	484	491	336	449	2224
I Fútbol	384	427	386	395	397	1989
J Despedida	874	874	973	830	801	4352

sum.T	5505	5230	5898	4680	5271	26584
-------	------	------	------	------	------	--------------

RF	1 Madri	2 Sevill	3 Méxic	4 Lim	5 B.A.	sum.R
A Hotel	0.00000	0.00000	0.00000	0.0023	0.0000	0.0023
B Camino	0.00211	0.00439	0.00194	0.0024	0.0019	0.0127
C Funcionari	0.00380	0.00000	0.00000	0.0000	0.0000	0.0038
D_Estudiante	0.00425	0.00422	0.00192	0.0024	0.0041	0.0169
E Consultorio	0.00599	0.00287	0.00287	0.0000	0.0050	0.0168
F Fiesta	0.00198	0.00609	0.00559	0.0027	0.0000	0.0163
G Política	0.00137	0.00452	0.00134	0.0000	0.0014	0.0086
H Compras	0.01724	0.01446	0.01018	0.0178	0.0133	0.0731
I Fútbol	0.00000	0.00234	0.00000	0.0025	0.0075	0.0124
J Despedida	0.00572	0.00458	0.00411	0.0048	0.0037	0.0229
sum.R	0.04246	0.04347	0.02795	0.0351	0.0371	0.1862

ARF	1 Madri	2 Sevill	3 Méxic	4 Lim	5 B.A.	sum.A
A Hotel	0.0	0.0	0.0	1.3	0.0	1.3
B Camino	1.1	2.3	1.0	1.3	1.0	6.7
C Funcionario	2.0	0.0	0.0	0.0	0.0	2.0
D Estudiante	2.3	2.2	1.0	1.3	2.2	9.0
E Consultorio	3.2	1.5	1.5	0.0	2.7	8.9
F Fiesta	1.1	3.2	3.0	1.4	0.0	8.7
G Política	0.7	2.4	0.7	0.0	0.8	4.6
H Compras	9.2	7.7	5.4	9.5	7.1	38.9
I_Fútbol	0.0	1.2	0.0	1.3	4.0	6.5
J Despedida	3.0	2.4	2.2	2.6	2.0	12.2
sum.A	22.6	22.9	14.8	18.7	19.8	98.8

このように計算された調整相対頻度は実際の観測値の中で相対化された数値なので、現実的な規模で数値を比較することができます。よって調整相対頻度は、全体の傾向を実際の規模の中で把握するときに便利です。

3.1.4. パーブロック

検索された形式が見つかるブロック数を示します。ブロック数は全体を 10, 100, 1000, ... のブロックに分割して計算します。全体のブロック数(10, 100, 1000, ...)は「乗数」のボックスで指定します。

→	100 bloques	1200	1250	1300	1350	1400	1450	*Resetear*
1	1.#[^[^]%	39	68	60	53	55	57	1.#[^[^]%
2	2.#ff%	12	63	50	45	35	17	2.#ff%
3	3.#s[^s]%	39	70	62	53	55	58	3.#s[^s]%
4	4.#ss%	5	51	44	39	24	5	4.#ss%
Resetear		1200	1250	1300	1350	1400	1450	

3.1.5. 標準得点

標準得点は (絶対頻度 - 平均) / 標準偏差で求めます。標準得点の平均はゼロ、標準偏差は 1 になります。横行と縦列でそれぞれの標準得点を求めます。次は横標準得点の出力です。

■ 数値: 1. 得点. 13. 横標準得点								
→	パタン	2	4	6	8	10	12	*リセット*
1	#Alice#	-.847	-.580	-.446	1.962	.624	-.713	#Alice#
2	#Queen#	-.722	-.722	-.433	2.165	-.289	.000	#Queen#
3	#King#	-.548	-.548	-.548	.000	-.548	2.191	#King#
リセット		2	4	6	8	10	12	

3.1.6. 順位得点

行と列の数値の順 (昇順・降順) を出力します。次は横行の昇順の順位得点です。

■ 数値: 1. 得点. 15. 横昇順位得点								
→	パタン	2	4	6	8	10	12	*リセット*
1	#Alice#	1	3	4	6	5	2	#Alice#
2	#Queen#	1	1	3	6	4	5	#Queen#
3	#King#	1	1	1	5	1	6	#King#
リセット		2	4	6	8	10	12	

3.1.7. 0:1 データ

ゼロを超える値をすべて 1 として出力します。データの有無だけを考慮する得点です。

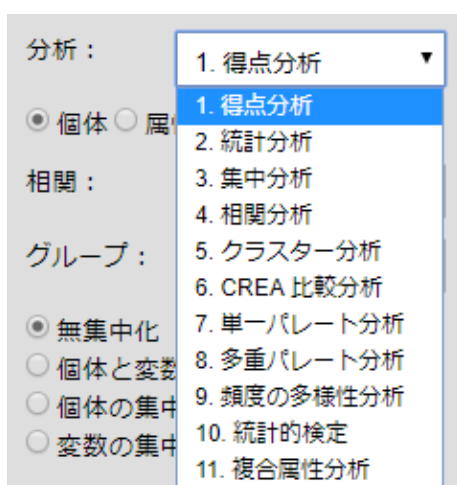
■ 数値: 1. 得点: 19. 0:1データ								
→	パターン	2	4	6	8	10	12	*リセット*
1	#Alice#	1	1	1	1	1	1	#Alice#
2	#Queen#	0	0	1	1	1	1	#Queen#
3	#King#	0	0	0	1	0	1	#King#
	リセット	2	4	6	8	10	12	

3.1.8. 対数得点

極端な大きな数値があるときには、データ全体を対数に変換することがあります。対数の底のオプションとして 2 と 10 を用意しました。

■ 数値: 1. 得点: 20. 対数変換(底: 2)								
→	パターン	2	4	6	8	10	12	*リセット*
1	#Alice#	4.248	4.392	4.459	5.322	4.907	4.322	#Alice#
2	#Queen#	.000	.000	1.000	4.322	1.585	2.322	#Queen#
3	#King#	.000	.000	.000	2.000	.000	4.322	#King#
	リセット	2	4	6	8	10	12	

3.2. 分析



3.2.1. 得点分析

最初の「得点」の内容は先に見た「行列」になります。以下では「統計量」と「関係」のセクションを説明します。それぞれのセクションで使用する行列の例として「絶対頻度」を使いますが、これはパーセントなどの相対頻度や正規化頻度でもそれぞれの分析が可能です。目的に応じて選択してください。

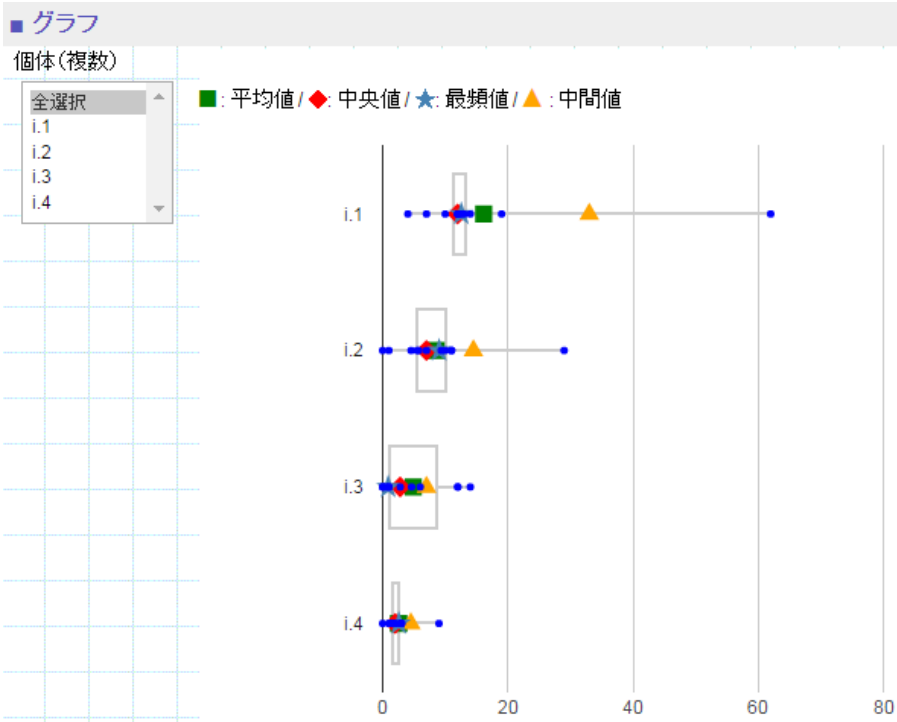
3.2.2. 統計分析

個体の和、平均、最小値、最大値などの統計量が出力されます。属性の統計量を調べるときには、「入力行列の転置」をチェックしてください。

■ データファイル: D2 / 行数: 4						
行	N.D2.fre.	A	B	C	D	E
1	i.1	10	19	14	7	12
2	i.2	11	7	10	0	1
3	i.3	0	0	1	12	1
4	i.4	0	1	2	3	3

■ 数値. 1. 絶対頻度. 2. 統計量													
→	N.D2.fre.	和	平均値	最小値	最大値	範囲	中間値	第1四分位	中央値	第3四分位	最頻値	標準偏差	リセット*
1	i.1	62	12.4	7	19	12	13	10	12	14	12	4.03	i.1
2	i.2	29	5.8	0	11	11	5.5	1	7	10	9.333	4.534	i.2
3	i.3	14	2.8	0	12	12	6	0	1	1	0.5	4.622	i.3
4	i.4	9	1.8	0	3	3	1.5	1	2	3	2.667	1.166	i.4
リセット*		和	平均値	最小値	最大値	範囲	中間値	第1四分位	中央値	第3四分位	最頻値	標準偏差	

統計量の一部である、平均値、中央値、最頻値、中間値をボックスチャートで視覚化します。ボックスの左端と右端は第1四分位と第3四分位を示します。



* 最頻値は階級化されていないデータには適さないなので代わりに大数最頻値を使います。

3.2.3. 集中分析

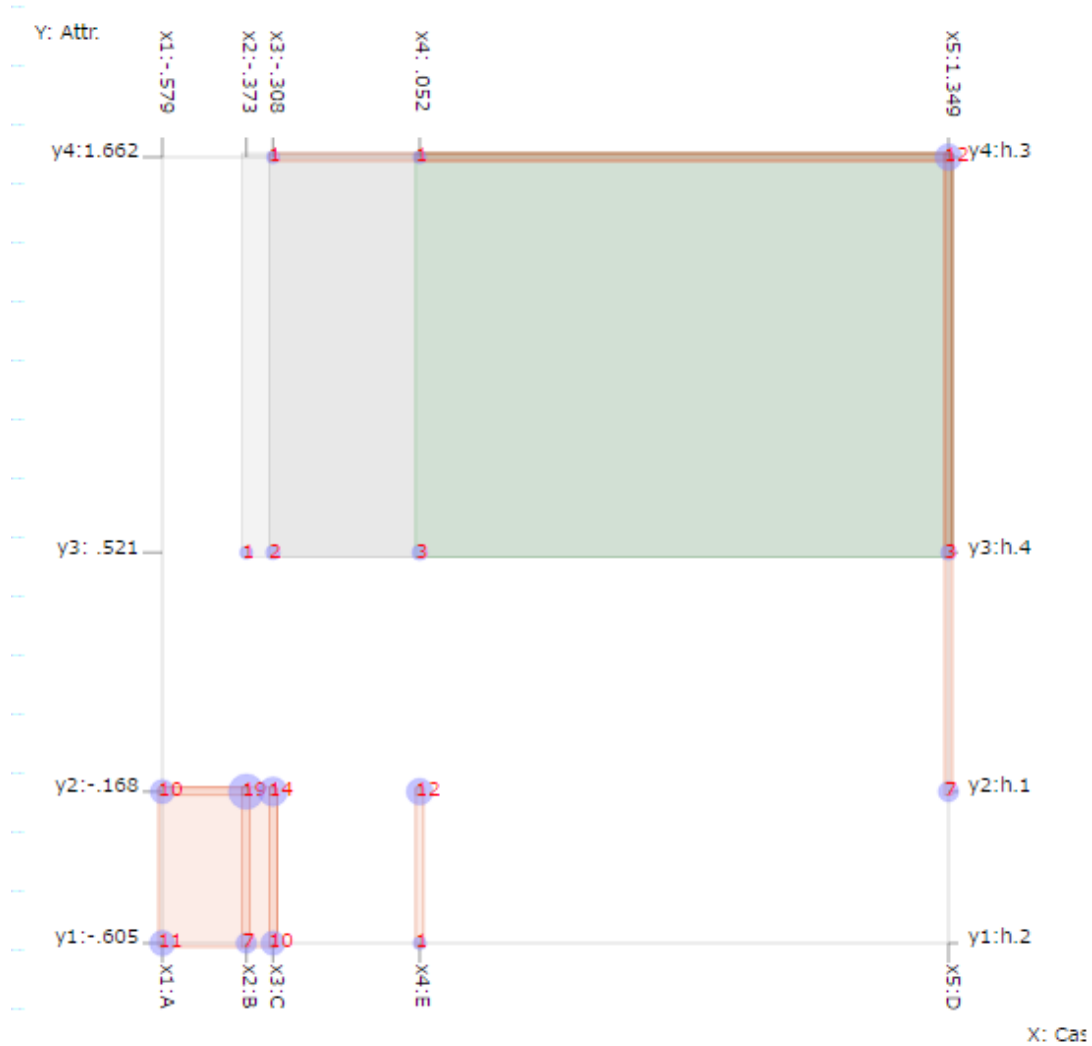
集中分析によって下左の入力行列の行と列の順を変えて下右のように数値の分布が対角線に集中した行列に変えます。

n.	Cnp	A	B	C	D	E	値
1	h.4	0	1	2	3	3	4.000
2	h.3	0	0	1	12	1	3.000
3	h.2	11	7	10	0	1	2.000
4	h.1	10	19	14	7	12	1.000
5	値	1.000	2.000	3.000	4.000	5.000	

n.	Cnp	A	B	C	E	D	値
1	h.3	0	0	1	1	12	1.662
2	h.4	0	1	2	3	3	.521
3	h.1	10	19	14	12	7	-.168
4	h.2	11	7	10	1	0	-.605
5	値	-.579	-.373	-.308	.052	1.349	

集中化の方法としては、対応分析、原点距離分析、クラスター分析が選択できます。行と列の順を変えることの他に、行または列の順だけを変えることもできます。

グラフでは行と列の値に従って、それぞれの間隔を決め、分布の数値の大きさにいって、色または青円の大きさを決めます。数値の集中度が高い領域を灰→青→緑→赤の色で塗ります。



3.2.4. 相関分析

属性間の相関関係をさまざまな指標によって調べます。連関係数は 0:1 データに適用し、グッドマンとクラスカルの連関係数は順位得点に適用します。

■ データ: ファイル: D4 / 行数: 28

行	N.D4.ss-ff	-ss-	#ss-	-ff-	#ff-
1	R: León	4.8	5.8	1.4	6.1
2	R: C. la Vieja	9.7	9.6	1.5	10.1
3	R: C. la Nueva	1.8	2.4	0.5	1.9
4	R: Aragón	11	0.9	2.1	3
5	A: 1200	16.8	3.1	1.2	6.2
6	A: 1225	5.7	0.2	0.5	0

(データの一部)

行列:	1. 絶対頻度	6	A: 1225	5.7	0.2	0.5	0
分析:	3. 相関	■ 数値. 3. 相関. 1. 絶対頻度. 1. ピアソン相関係数 R.					
Select:	1. ピアソン相関係数						
グラフ:	1. ピアソン相関係数 R.						
個性グル	2. C. 共起: a						
属性グル	3. J. $a/(a+b+c)$						
	4. D. $2a/(2a+b+c)$						
	5. U. $(2a-b-c)/(2a+b+c)$						
	6. O. $a/\sqrt{(a+b)(a+c)}$						
	7. G.K. 順序連関係数						

→	N.D4.ss-ff	-ss-	#ss-	-ff-	##f-	*リセット*
1	-ss-	1.000	.594	.542	.622	-ss-
2	#ss-	.594	1.000	.602	.929	#ss-
3	-ff-	.542	.602	1.000	.698	-ff-
4	##f-	.622	.929	.698	1.000	##f-
リセット	-ss-	#ss-	-ff-	##f-		

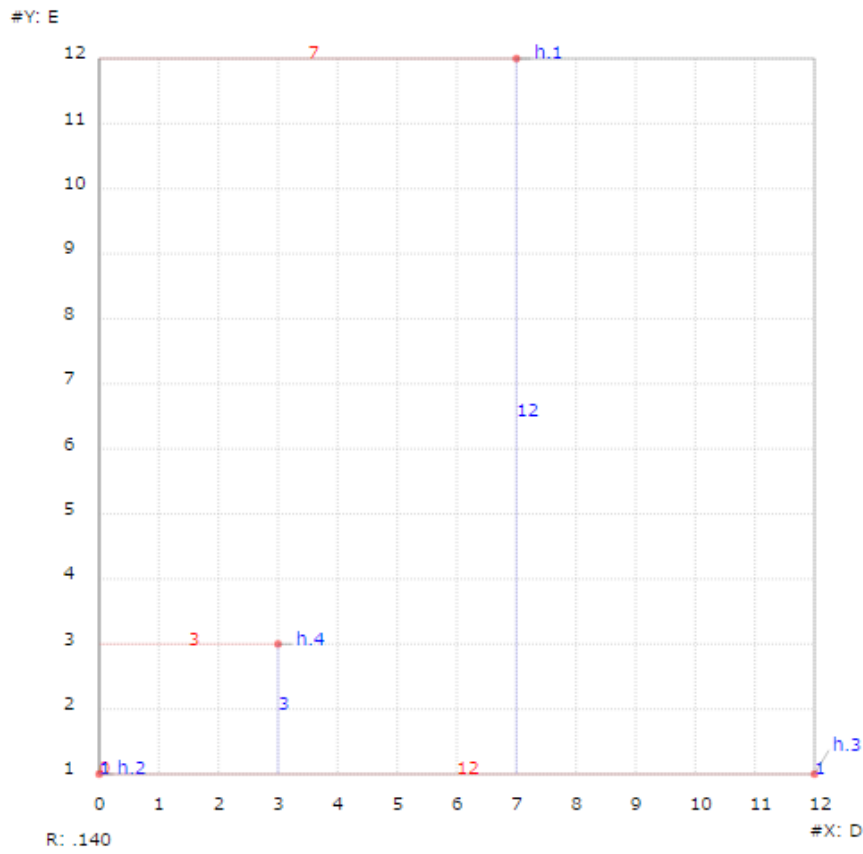
グラフではカラーテーブル、サイズテーブルで全体の相関を示し、散佈図で2つの属性間の相関を示します。

n.	Rpp	A	B	C	D	E
1	A	1.000	.787	.944	-.480	.436
2	B	.787	1.000	.945	-.092	.896
3	C	.944	.945	1.000	-.331	.709
4	D	-.480	-.092	-.331	1.000	.140
5	E	.436	.896	.709	.140	1.000

カラーテーブル

n.	Rpp	A	B	C	D	E
1	A	1.000	.787	.944	-.480	.436
2	B	.787	1.000	.945	-.092	.896
3	C	.944	.945	1.000	-.331	.709
4	D	-.480	-.092	-.331	1.000	.140
5	E	.436	.896	.709	.140	1.000

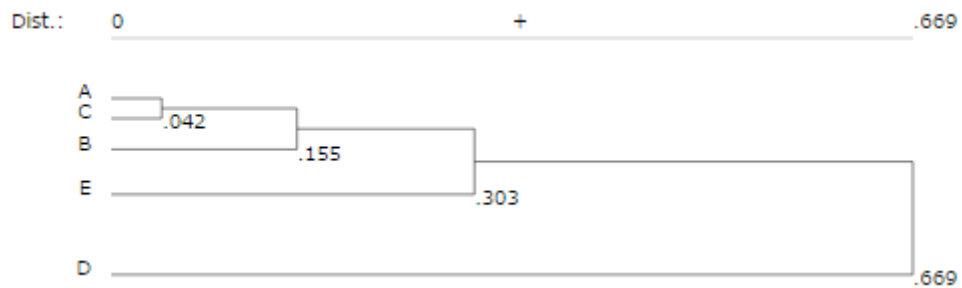
サイズテーブル



散布図

3.2.5. クラスタ分析

クラスタ分析（平均法）の結果を次のようなデンドログラムで出力します。



3.2.6. CREA 比較分析

■ 数値: 1. AF: 絶対頻度 (*) - 6. CREA 比較分析

ウィンドウ / コピー : 書式あり / コピー : 書式なし

→	Comp.CREA	全体	計 N.	CREA.N	N-C	(N-C)/(N+C)
1	favor	978.0	978.0	143.0	835.0	0.745
2	faldas	75.2	75.2	7.6	67.7	0.817
3	fachada	37.6	37.6	17.1	20.6	0.376
4	falta	714.7	714.7	243.8	470.9	0.491
5	fatal	263.3	263.3	14.3	249.0	0.897
6	fácilmente	112.8	112.8	33.8	79.1	0.539
7	facu	150.5	150.5		150.5	1.000
8	fácil	225.7	225.7	112.8	112.9	0.334
9	facultad	188.1	188.1	38.9	149.2	0.657
10	fastidies	37.6	37.6		37.6	1.000
11	familia	639.5	639.5	245.0	394.5	0.446
12	fabricarnos	37.6	37.6		37.6	1.000
13	falda	150.5	150.5	13.9	136.5	0.830
14	factura	75.2	75.2	12.3	62.9	0.719
15	favorito	75.2	75.2	12.0	63.2	0.725
16	fanático	37.6	37.6	3.8	33.8	0.818
17	facilidad	37.6	37.6	30.8	6.8	0.100
18	facilísima	37.6	37.6		37.6	1.000
19	faltar	112.8	112.8	7.8	105.1	0.871
20	fátima	263.3	263.3	4.1	259.2	0.969
21	faltan	37.6	37.6	19.5	18.1	0.316
22	facilísimo	37.6	37.6		37.6	1.000

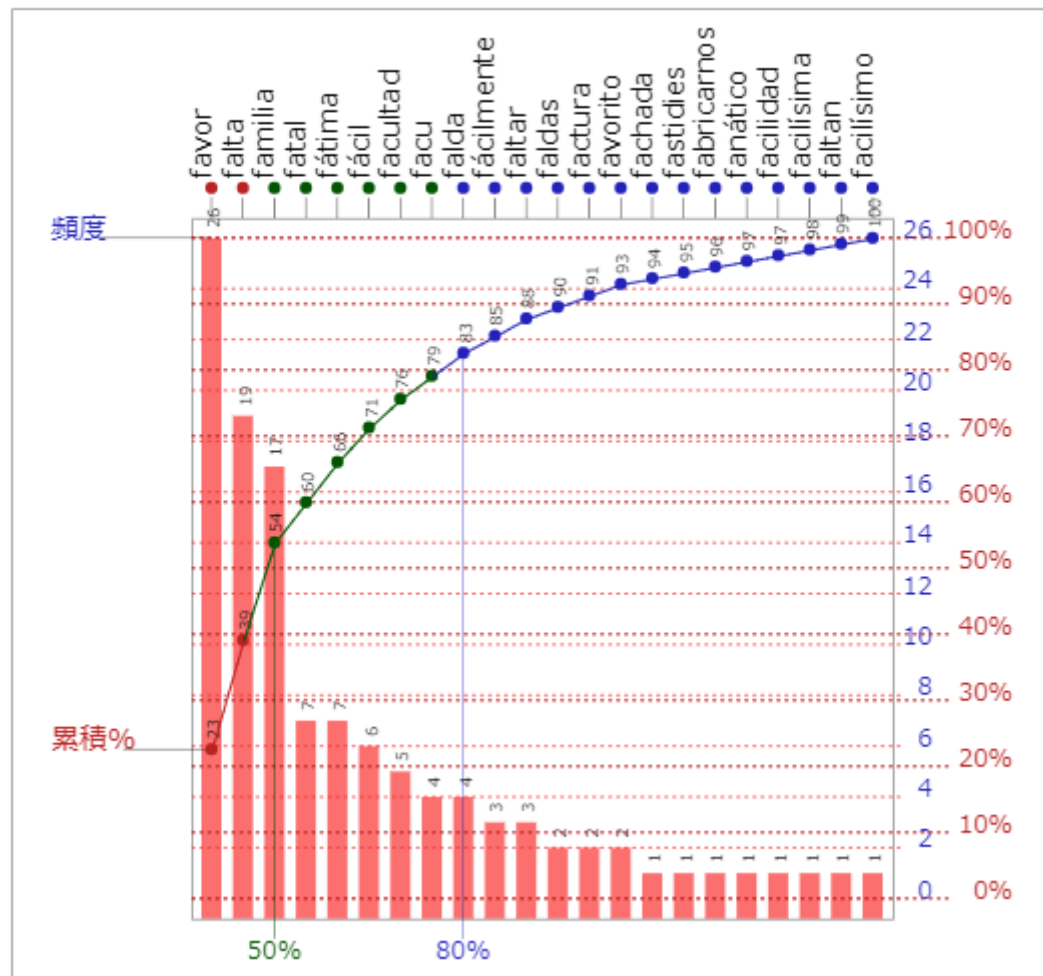
パターン : #fa-のフォーム

3.2.7. 単一パレート分析

Pareto	計
パレート指数 (Pi)	.593
H: 半数ポイント (50%)	3
N: データ数	22
H/N (%)	14 %
A: 半数累積度数	62
T: 総度数	115
A/T (%)	54 %
ポイント 50%	3
ポイント 80%	9
ポイント 95%	17

Pareto	頻度	%	累積頻度	累積%
favor	26	22.6	26	22.6
faldas	2	1.7	28	24.3
fachada	1	.9	29	25.2
falta	19	16.5	48	41.7
fatal	7	6.1	55	47.8
fácilmente	3	2.6	58	50.4
facu	4	3.5	62	53.9
fácil	6	5.2	68	59.1
facultad	5	4.3	73	63.5
fastidies	1	.9	74	64.3
familia	17	14.8	91	79.1
fabricarnos	1	.9	92	80.0
falda	4	3.5	96	83.5
factura	2	1.7	98	85.2
favorito	2	1.7	100	87.0
fanático	1	.9	101	87.8
facilidad	1	.9	102	88.7
facilísima	1	.9	103	89.6
faltar	3	2.6	106	92.2
fátima	7	6.1	113	98.3
faltan	1	.9	114	99.1
facilísimo	1	.9	115	100.0

■ グラフ: 7. 単一パレート分析

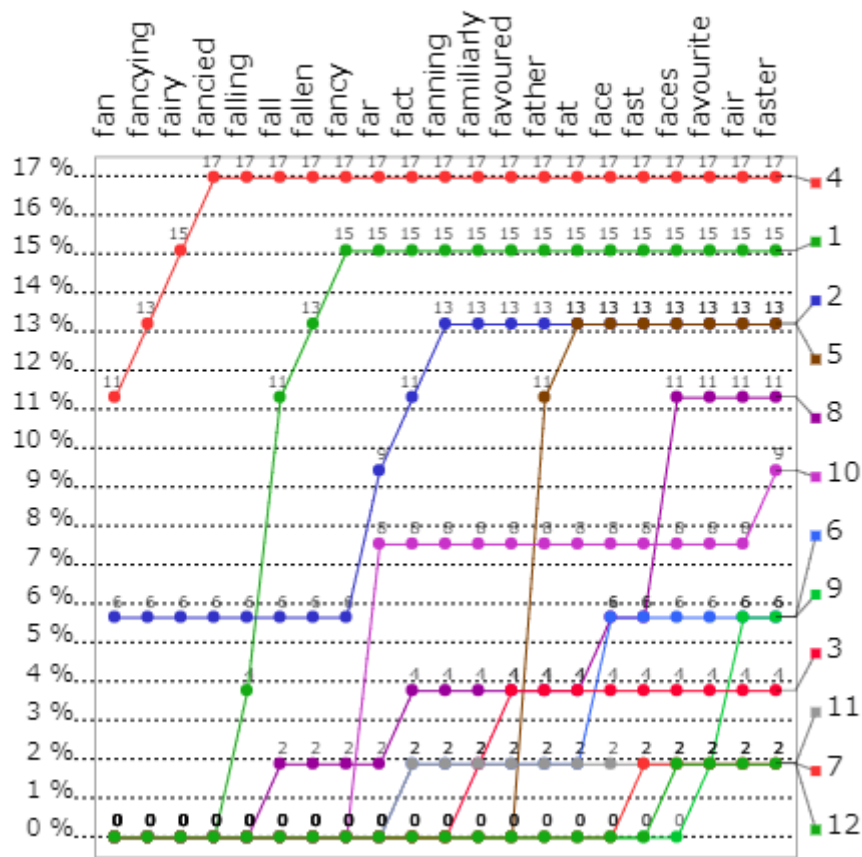


3. 2. 8. 多重パレート分析

Pareto	1	8	2	10	6	11	4	3	5	7	12	9
パレート指数 (Pi)	.953	.946	.946	.990	.983	1.000	.964	.974	.993	1.000	1.000	.983
H: 半数ポイント (50%)	2	2	2	1	1	1	1	2	1	1	1	1
N: データ数	4	4	4	2	2	1	4	2	2	1	1	2
H/N (%)	50 %	50 %	50 %	50 %	50 %	100 %	25 %	100 %	50 %	100 %	100 %	50 %
A: 半数累積度数	6	4	5	4	2	1	6	2	6	1	1	2
T: 総度数	8	6	7	5	3	1	9	2	7	1	1	3
A/T (%)	75 %	67 %	71 %	80 %	67 %	100 %	67 %	100 %	86 %	100 %	100 %	67 %
ポイント 50%	2	2	2	1	1	1	1	2	1	1	1	1
ポイント 80%	3	3	3	2	2	1	3	2	1	1	1	2
ポイント 95%	4	4	4	2	2	1	4	2	2	1	1	2

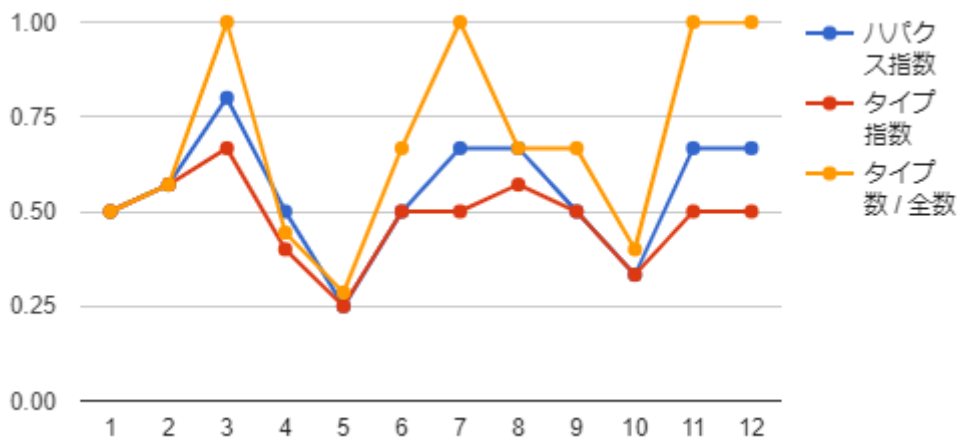
順位	1	8	2	10	6	11	4	3	5	7	12	9
1	fall: 4 (50 %)	faces: 3 (50 %)	fan: 3 (43 %)	far: 4 (80 %)	face: 2 (67 %)	fact: 1 (100 %)	fan: 6 (67 %)	familiarly: 1 (50 %)	father: 6 (86 %)	fast: 1 (100 %)	faces: 1 (100 %)	fair: 2 (67 %)
2	falling: 2 (25 %)	fall: 1 (17 %)	far: 2 (29 %)	faster: 1 (20 %)	fact: 1 (33 %)	-	fancying: 1 (11 %)	favoured: 1 (50 %)	fat: 1 (14 %)	-	-	favourite: 1 (33 %)
3	fallen: 1 (13 %)	fact: 1 (17 %)	fact: 1 (14 %)	-	-	-	fairy: 1 (11 %)	-	-	-	-	-
4	fancy: 1 (13 %)	face: 1 (17 %)	fanning: 1 (14 %)	-	-	-	fancied: 1 (11 %)	-	-	-	-	-

■ グラフ: 8. 多重パレート分析



3.2.9. 頻度の多様性分析

→ 頻度の多様性	1	2	3	4	5	6	7	8	9	10	11	12
1 全数	8	7	2	9	7	3	1	6	3	5	1	1
2 タイプ数	4	4	2	4	2	2	1	4	2	2	1	1
3 最大値	4	3	1	6	6	2	1	3	2	4	1	1
4 ハバクス	2	2	2	3	1	1	1	3	1	1	1	1
5 ハバクス指数	0.500	0.571	0.800	0.500	0.250	0.500	0.667	0.667	0.500	0.333	0.667	0.667
6 タイプ指数	0.500	0.571	0.667	0.400	0.250	0.500	0.500	0.571	0.500	0.333	0.500	0.500
7 タイプ数/全数	0.500	0.571	1.000	0.444	0.286	0.667	1.000	0.667	0.667	0.400	1.000	1.000



3.2.10. 統計的検定

→ 頻度の検定	1	2	3	4	5	6	7	8	9	10
1 a	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0
2 E	0.000055	0.000055	0.000055	0.000055	0.000055	0.000055	0.000055	0.000055	0.000055	0.000055
3 N	1104	1032	879	726	765	956	967	1015	861	1015
4 M	0.1	0.1				0.1	0.1	0.1		
5 X	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0
6 左側/右側	左側	左側	/右側	/右側	/右側	左側	左側	左側	/右側	左側
7 P値	0.941	0.945				0.949	0.948	0.946		0.945
8 有意性	NS	NS	S	S	S	NS	NS	NS	S	NS

3.3. グループ

個体と属性の数が多いときは分布に近い個体を個体グループにまとめ、分布に近い属性を属性グループにまとめると、それぞれのグループの特徴が明らかになり、また出力データの全体が様子が変わりやすくなります。

次のデータ D3 は個体が 47 個、属性が 15 個あります。

■ Data. File: D3 / Lines:47

Line	N.D3.farmer	MX	CU	RD	PR	C5	PN	VE	CO	EC	PE	BO	CH	PA	UR	AR
1	01 cacahuero	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
2	02 cafetalista	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
3	03 camilucho	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
4	04 campero	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
5	05 camperuso	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
6	06 campirano	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
7	07 campiruso	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
8	08 campista	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0

かなり大きなデータ行列なので、10 個の個体グループと 5 個の属性グループにまとめました。

→	Gnp	[A.1]	[A.2]	[A.3]	[A.4]	[A.5]
1	C.01	2	7	25	3	0
2	C.02	1	2	0	0	0
3	C.03	1	0	0	6	34
4	C.04	14	1	1	0	0
5	C.05	2	13	0	0	0
6	C.06	0	1	2	10	3
7	C.07	1	0	0	2	1
8	C.08	1	0	0	0	0
9	C.09	2	2	0	0	0
10	C.10	0	0	0	1	0

個体グループ: ▲ ▼

属性グループ: ▲ ▼

個体グループ	個体
C.01	01 cacahuero; 05 camperuso; 06 campirano; 13 conuquero; 15 chagrero; 25 huasicama; 31 lanudo; 32 llanero; 36 montubio; 37 paisano; 38 pajuerano; 40 payazo; 44 sabanero;
C.02	02 cafetalista;
C.03	03 camilucho; 04 campero; 09 campusano; 12 comparsa; 16 changador; 18 chuncano; 20 estanciero; 21 gaucho; 28 invernador; 39 partidario; 41 piona; 46 viñatero;
C.04	07 campiruso; 08 campista; 10 campuso; 17 chilero; 23 guanaco; 27 hulero;
C.05	11 colono; 19 enmaniguado; 22 guajiro; 29 jibaro; 35 montero; 42 ranchero;
C.06	14 coquero; 24 guaso; 30 lampero; 47 yanacón;
C.07	26 huertero;
C.08	33 macanero;
C.09	34 manuto; 45 veguero;
C.10	43 rondín;

属性グループ

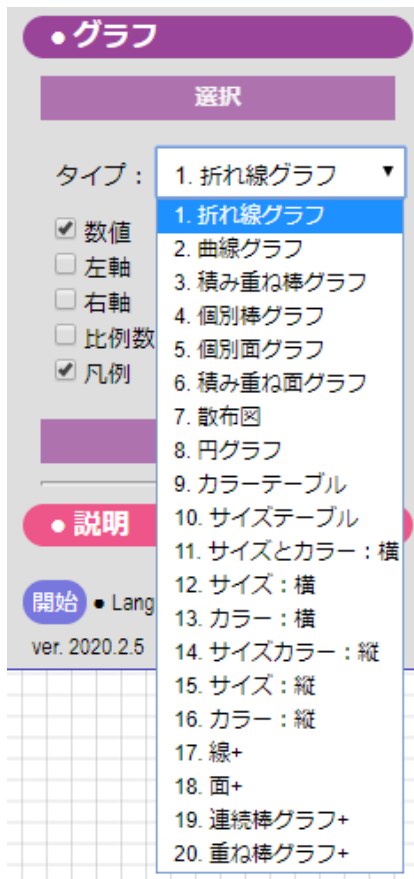
グループ	属性
[A.1]	MX; C5; PN;
[A.2]	CU; RD; PR;
[A.3]	VE; CO; EC;
[A.4]	PE; BO; CH;
[A.5]	PA; UR; AR;

*このようなまとめ方にはいろいろな方法がありますが、ここでは規定近接係数によるクラスター分析の平均法を使用しました。

4. グラフ



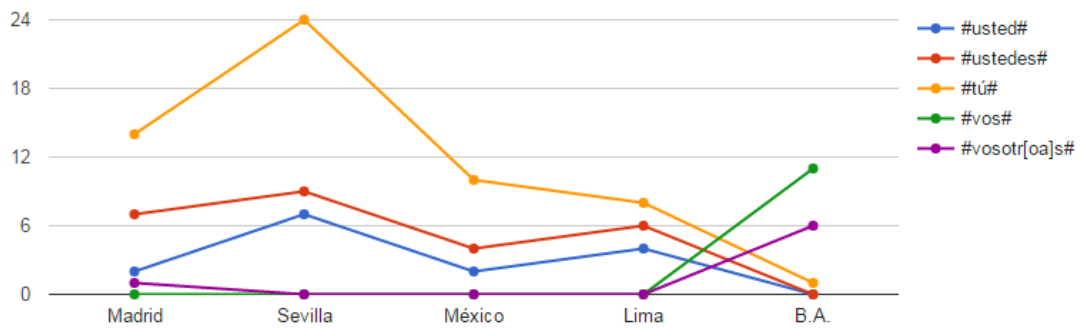
各種の得点の全体の傾向や特徴を見るためにグラフが役立ちます。



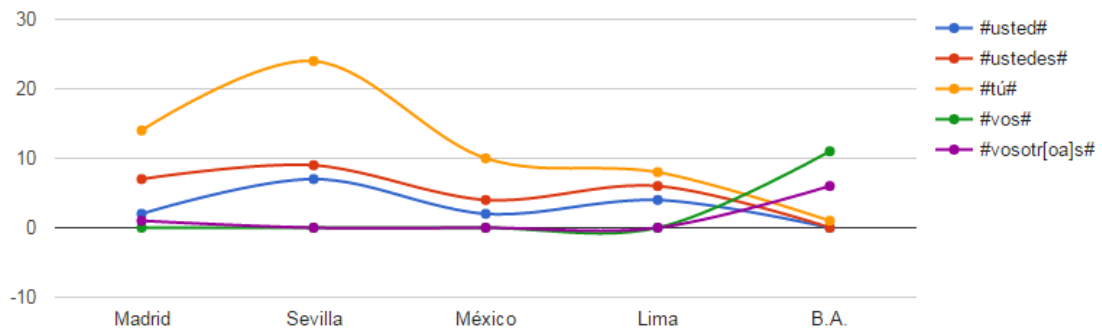
次の絶対頻度得点を例にします。

Núm.	Patrón	Madrid	Sevilla	México	Lima	B.A.
3	#ú#	14	24	10	8	1
1	#usted#	2	7	2	4	0
2	#ustedes#	7	9	4	6	0
4	#vos#	0	0	0	0	11
5	#vosotr[oa]s#	1	0	0	0	6

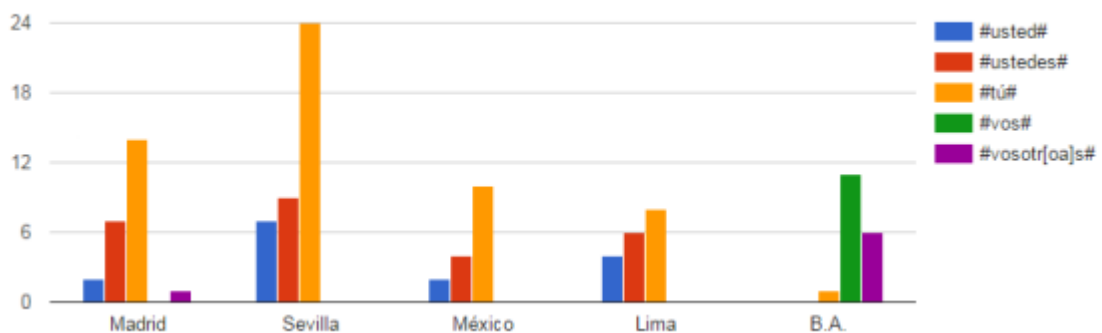
折れ線グラフ



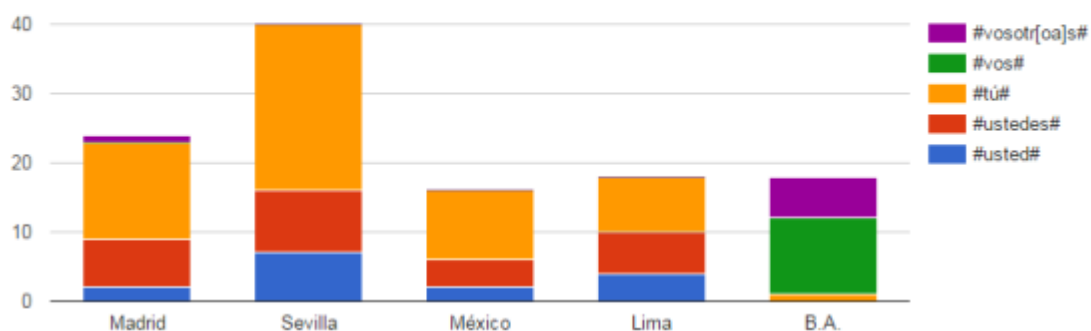
曲線グラフ



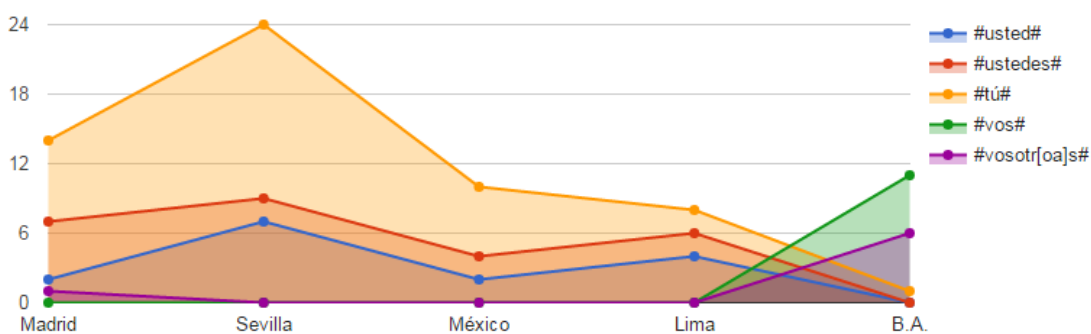
個別棒グラフ



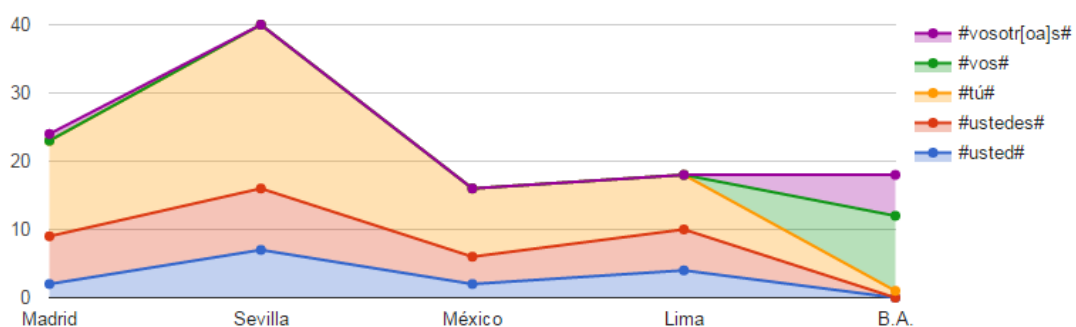
積み重ね棒グラフ



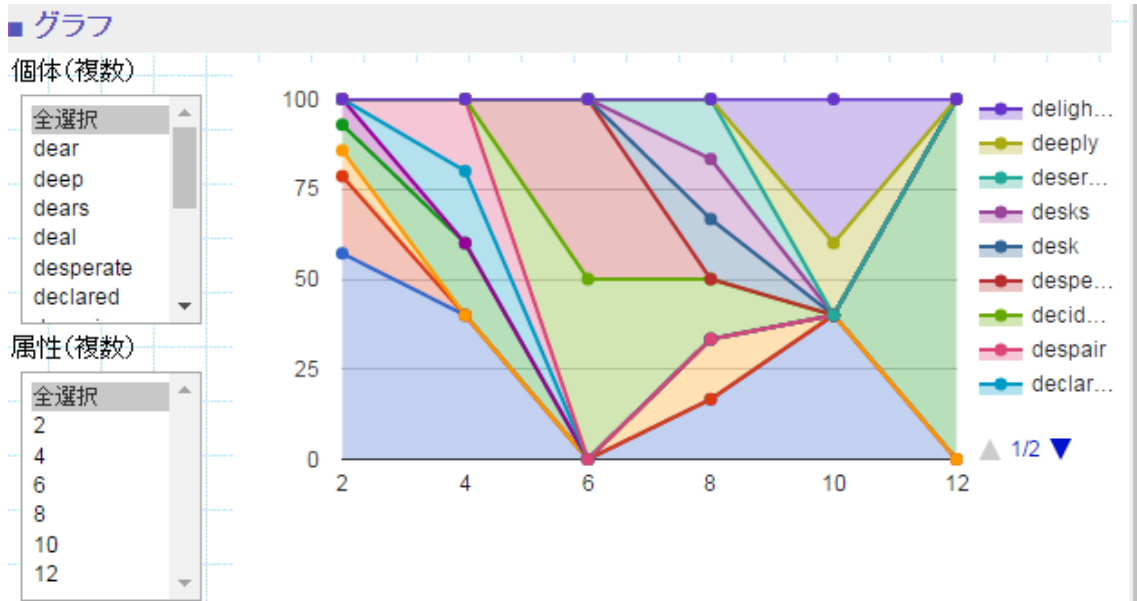
個別面グラフ



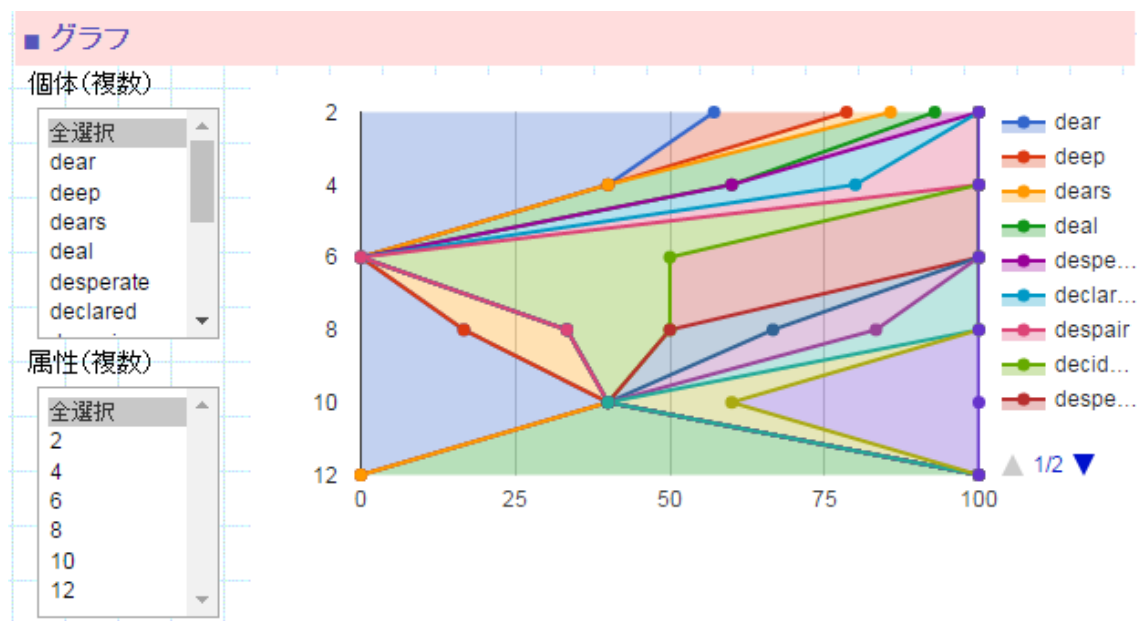
積み重ね面グラフ



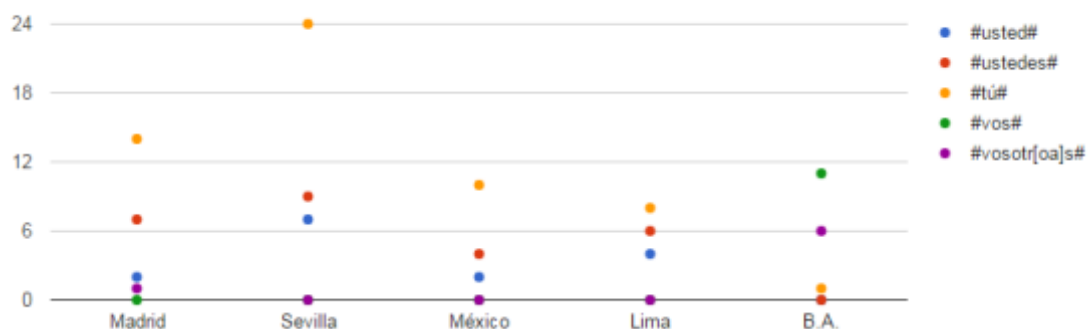
左欄の「行列」で「縦パーセント」を選択すると、次のように 100% 全体の中で占める割合がグラフ化されます。



左欄の「縦向きグラフ」をチェックするとグラフは次のように 90 度回転します。



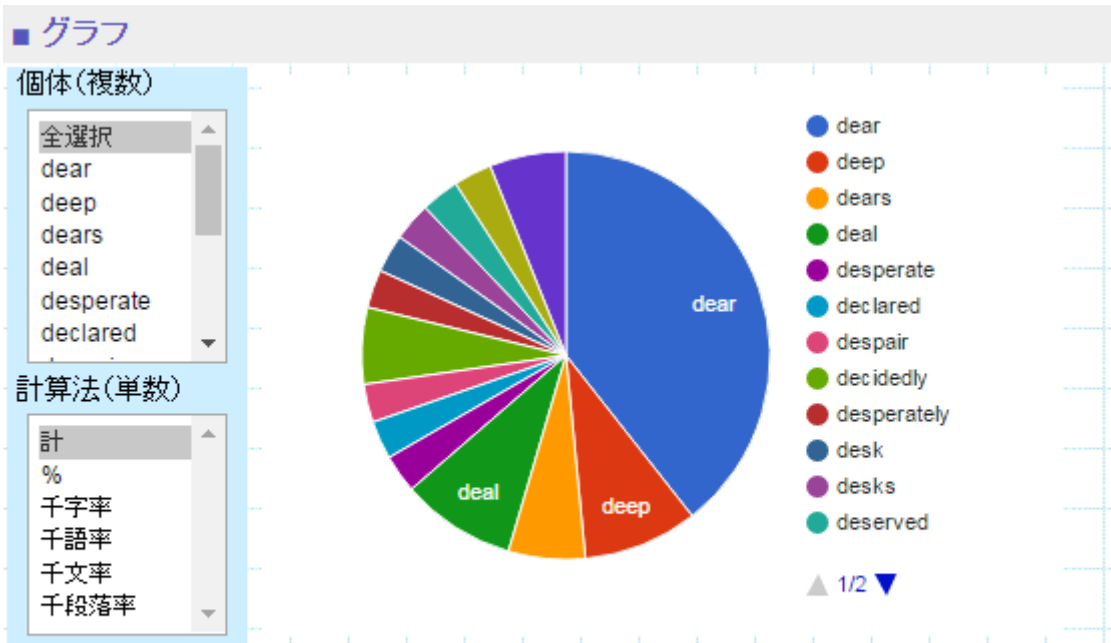
散布図



円グラフ

属性が1つだけの数値データでは横並びのグラフを出力せず、次のような円グラフを出力します。また、数表も次のように計、%、千字率、千語率、千文率、千段落率を出力します。

■ 数値: 1. 得点. 属性なし									
→	フォーム	計	%	千字率	千語率	千文率	千段落率	*リセット*	
1	dear	13.0	39.4	.3	1.2	15.1	61.3	dear	
2	deep	3.0	9.1	.1	.3	3.5	14.2	deep	
3	dears	2.0	6.1	.0	.2	2.3	9.4	dears	
4	deal	3.0	9.1	.1	.3	3.5	14.2	deal	
5	desperate	1.0	3.0	.0	.1	1.2	4.7	desperate	



以上のグラフの出力には次のグーグルチャート Google Chart を使用しました。<https://developers.google.com/chart/>.

その他に次のカラーテーブルとサイズテーブルをプログラムしました。

n.	ボタン	2	4	6	8	10	12
1	#Alice#	19	21	22	40	30	20
2	#Queen#	0	0	2	20	3	5
3	#King#	0	0	0	4	0	20

n.	ボタン	2	4	6	8	10	12
1	#Alice#	19	21	22	40	30	20
2	#Queen#	-	-	2	20	3	5
3	#King#	-	-	-	4	-	20

上のカラーテーブルでは第 1 四分位以下を青、第 3 四分位以上を赤、その間を緑にしました。下のサイズテーブルでは数値の大きさと四角の枠の大きさが比例します。

[FIN]