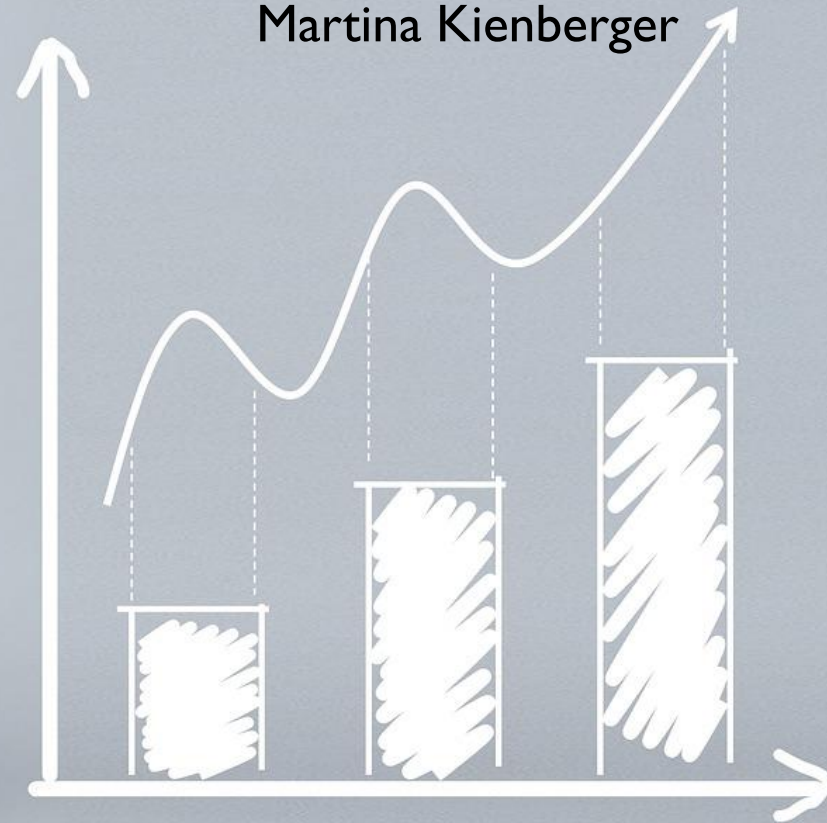


Análisis de encuestas con Python

Martina Kienberger



UNIVERSIDAD
DE GRANADA



Contenidos

1. Crear encuestas
2. Mi proyecto de investigación
3. Pasos en el análisis
4. Herramientas



1. Crear encuestas

Are you certainly likely, very likely, likely, somewhat likely, unlikely, very unlikely or certainly not likely to take an overly specific online survey?



Algunos puntos a considerar

- Administración del cuestionario
- Proceso de análisis
- Tipos de preguntas/ítems adecuados
- Pool de ítems por tema
- Definir preguntas y opciones de respuesta con ayuda de literatura, expertos y/o miembros de la población de interés
- Diseño, introducción, instrucciones, secuencia de preguntas
- Evaluar, hacer un estudio piloto



Tipos de preguntas

Preguntas cerradas	Preguntas abiertas
Respuestas a elegir	Respuestas libres (texto libre)



Preguntas cerradas

Dicotómicas

p. ej. ¿Te gustan los animales? ☐ Sí. ☐ No.

Nominal-politómicas

p. ej. ¿Qué animal te gusta más? ☐ Gato ☐ Perro ☐ Tortuga

Ordinal-politómicas

p. ej. ¿Cuánto te gustan los gatos? ☐ Nada. ☐ Un poco. ☐ Bastante. ☐ Mucho.


Continuas (intervalos exactos)

p. ej. ¿Cuántas mascotas tienes? [valores numéricos]

De elección múltiple

p. ej. ¿Qué animales te gustan? ☐ Gato ☐ Perro ☐ Tortuga

2. Mi proyecto de investigación

A young woman with brown hair tied back, wearing black-rimmed glasses and a dark grey top, is looking down at an open book. The book is open to a page with yellowed text. A thought bubble is positioned above her head on the right side of the image.

Haus =
house
(casa)?

Diseño de investigación

Desarrollo de las herramientas de investigación

Encuesta *online* destinada a estudiantes (FES-U)

Encuesta *online* destinada a docentes (FES-UL)



Estudio piloto (1er cuatrimestre 2017/18)

Estudiantes de la Universidad de Salamanca (N = 68)
Análisis cualitativo y cuantitativo

Docentes de la Universidad de Salamanca (N = 22)
Análisis cualitativo (y cuantitativo)



Estudio principal (2018-2019)

Recogida de datos a nivel nacional, online,
versión adicional de la encuesta en papel (N = 333, total 401)

Recogida de datos a nivel nacional, online
(N = 43, ges. 53)



Análisis de datos

Estadística descriptiva, análisis factorial, análisis de clústeres,
MANOVA, *Random Forest*/Árboles de decisión

Estadística descriptiva

Estrategias para comprender palabras desconocidas en alemán

Aquí tiene un ejemplo de un texto escrito en alemán:

(Desplácese hacia abajo para ver las preguntas correspondientes.)

Nach Adelboden zur digitalen Entgiftung

BERN. Flugmodus für die Seele: Adelboden Tourismus eröffnet das erste Digital Detox Camp der Schweiz. Vier Tage lang werden die Besucher von Pushs und Whatsapp-Nachrichten abgeschirmt. Anstatt Surfen im Internet stehen Yoga und Entspannungsübungen auf dem Programm. Selfiesticks haben dabei nichts verloren, dafür werden in der Berg- hütte Freilax in Tronegg Bleistifte und Notizbücher verteilt. «Bei uns auf der Alp kann man Antworten auf die tägliche Hektik finden», sagt Tourismus- direktor Urs Pfenninger. Die Teilnehmer sollen in erster Linie

wieder lernen, offline zu sein. Für diesen «kalten Entzug» stehen den Gästen Coaches zur Seite, die Teilnehmer auf eine Sinnesreise mitnehmen.

Doch das Camp vom 7. bis 10. September ist erst ein Experiment. Deshalb sucht Adelboden Tourismus derzeit nach fünf Freiwilligen, die sich dieser Challenge stellen. **MIW**



Yoga statt Whatsapp in den Bergen, DIGITALDETOX

Fuente: periódico suizo "20 Minuten" (edición regional de Berna, p. 5), 2 de agosto de 2017

Probablemente, en este texto encuentre palabras o grupos de palabras que no conoce aún. Sin embargo, seguro que logra entender algunas de estas palabras nuevas sin ayuda, bien enseguida o después de pensar un poco.

Lea el texto y piense qué hace para descubrir el significado o la función de las palabras que desconoce. También puede pensar en otras situaciones en las que se haya encontrado con vocabulario nuevo.

¿Qué estrategias o técnicas usa para averiguar el significado de palabras desconocidas?

¿Qué características de las palabras, frases o textos le ayudan?

Apunte todas sus ideas aquí:

Siga con la encuesta pasados 10 minutos como máximo.

Ejemplo y pregunta abierta para empezar
(activación metacognitiva)

Diseño de la encuesta (parte principal):

- Preguntas cerradas, información sobre estrategias en escala de Likert (frecuencia)

5 tipos de estrategias cognitivas

Comparo palabras desconocidas en alemán con palabras inglesas que se escriben de forma igual o similar. p. ej. al. <i>Licht</i> - ing. <i>light</i> (esp. <i>luz</i>).	1	2	3	4	5
--	---	---	---	---	---

+ estrategias metacognitivas

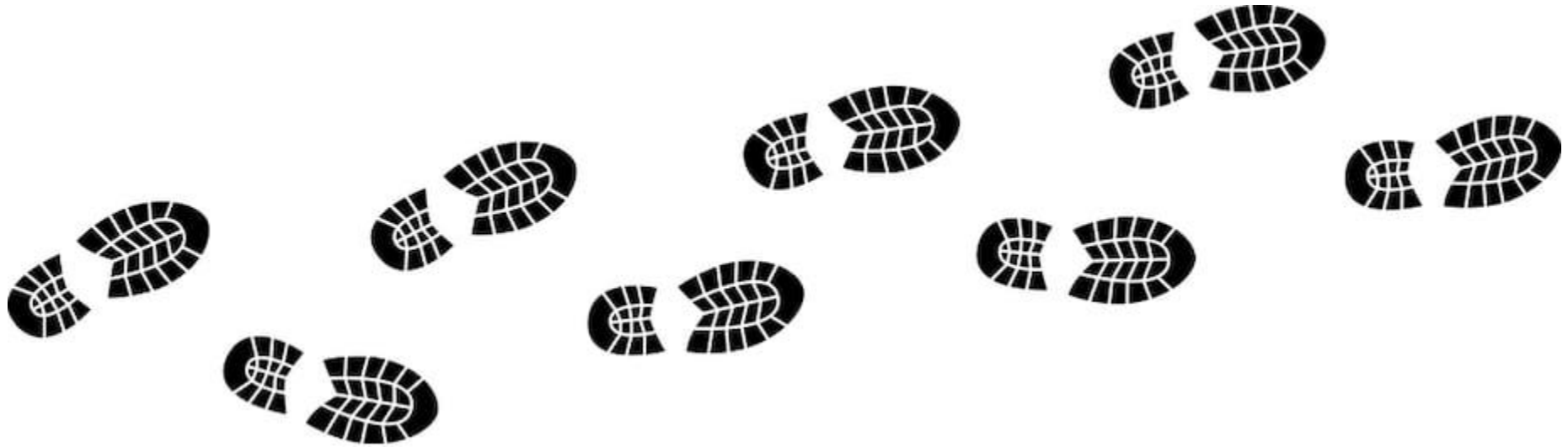
Combino varias estrategias para asegurarme de encontrar el significado correcto de las palabras desconocidas.	1	2	3	4	5
---	---	---	---	---	---

+ opinión sobre el uso de estrategias (grado de aprobación)

- Recogida de información sobre los encuestados (→ variables):

- | | | |
|---------------|---|----------|
| - Universidad | - Tipo y nivel del curso de alemán | - Edad |
| - Estudios | - Experiencia en aprendizaje de idiomas | - Género |
| - Curso | - Conocimientos de idiomas | |

3. Pasos en el análisis



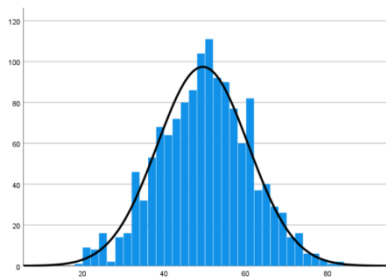
El proceso de análisis

Limpieza de
datos

Evaluación
psicométrica

Estadística
descriptiva

Estadística
inferencial





4. Herramientas

Limpieza de datos

Estadística descriptiva



Definición de variables

Cronbach's Alpha

Correlación

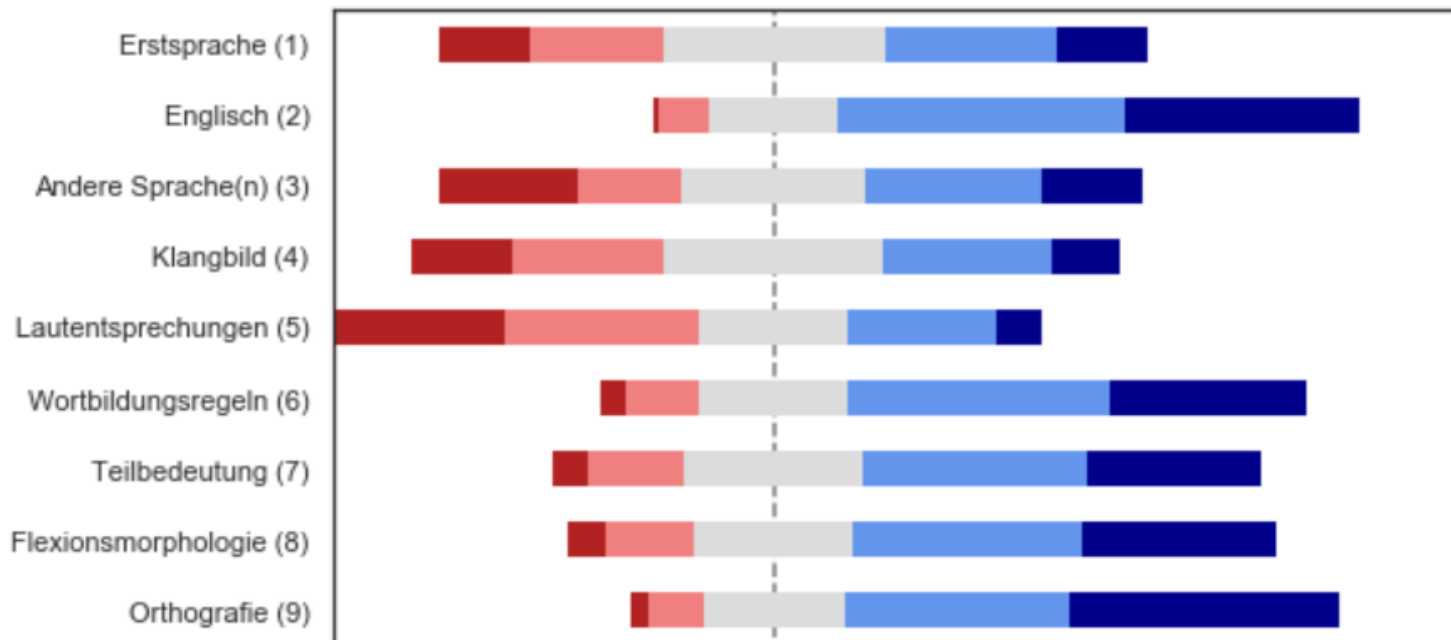
```
In [80]: Daten_A3[Liste_A_Q3].apply(pd.value_counts).fillna(0).astype(int).T
```

```
Out[80]:
```

	0	1	2	3	4	5
A-Q3_1	0	51	74	124	96	51
A-Q3_2	1	4	28	72	160	131
A-Q3_3	2	78	58	103	99	56
A-Q3_4	0	57	85	122	95	37
A-Q3_5	0	96	109	83	83	25
A-Q3_6	0	15	41	83	147	110
A-Q3_7	0	20	54	100	125	97
A-Q3_8	0	22	49	89	128	108
A-Q3_9	0	9	31	79	126	151
A-Q3_28	0	11	33	100	135	117
A-Q3_10	0	30	93	108	110	55

In [162...

```
graf_likert(Daten3_N,16)
```



In [78]:

```
df_describe(Daten_A3[Daten_A3[Liste_A_Q3]!=0],Liste_A_Q3)
```

Out[78]:

	count	mean	std	min	25%	50%	75%	max	mode	median	kurtosis	skew
A-Q3_1	396.0	3.055556	1.208386	1.0	2.00	3.0	4.0	5.0	3.0	3.0	-0.843564	-0.106998
A-Q3_2	395.0	3.977215	0.943580	1.0	3.00	4.0	5.0	5.0	4.0	4.0	0.069314	-0.756139
A-Q3_3	394.0	2.992386	1.327893	1.0	2.00	3.0	4.0	5.0	3.0	3.0	-1.114192	-0.130166
A-Q3_4	396.0	2.924242	1.183994	1.0	2.00	3.0	4.0	5.0	3.0	3.0	-0.854739	-0.036763
A-Q3_5	396.0	2.575758	1.237318	1.0	2.00	2.0	4.0	5.0	2.0	2.0	-1.036283	0.276562
A-Q3_6	396.0	3.747475	1.087032	1.0	3.00	4.0	5.0	5.0	4.0	4.0	-0.198458	-0.685463
A-Q3_7	396.0	3.568182	1.146441	1.0	3.00	4.0	4.0	5.0	4.0	4.0	-0.613826	-0.456901
A-Q3_8	396.0	3.633838	1.166957	1.0	3.00	4.0	5.0	5.0	4.0	4.0	-0.526970	-0.571100

Estadística inferencial

Correlación



SciPy

Diferencias entre grupos

Análisis factorial

Aprendizaje automático



Clasificación

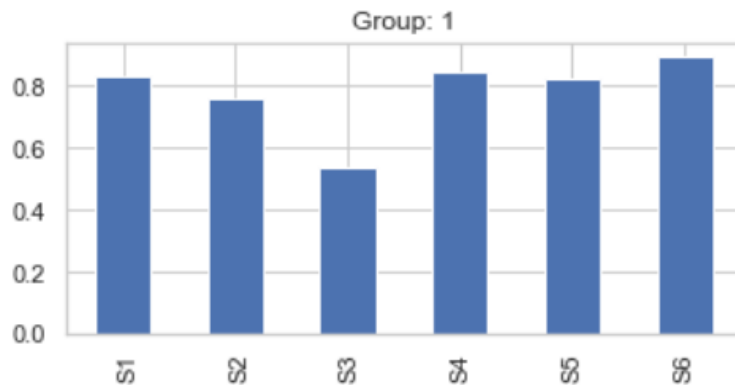
Predicción

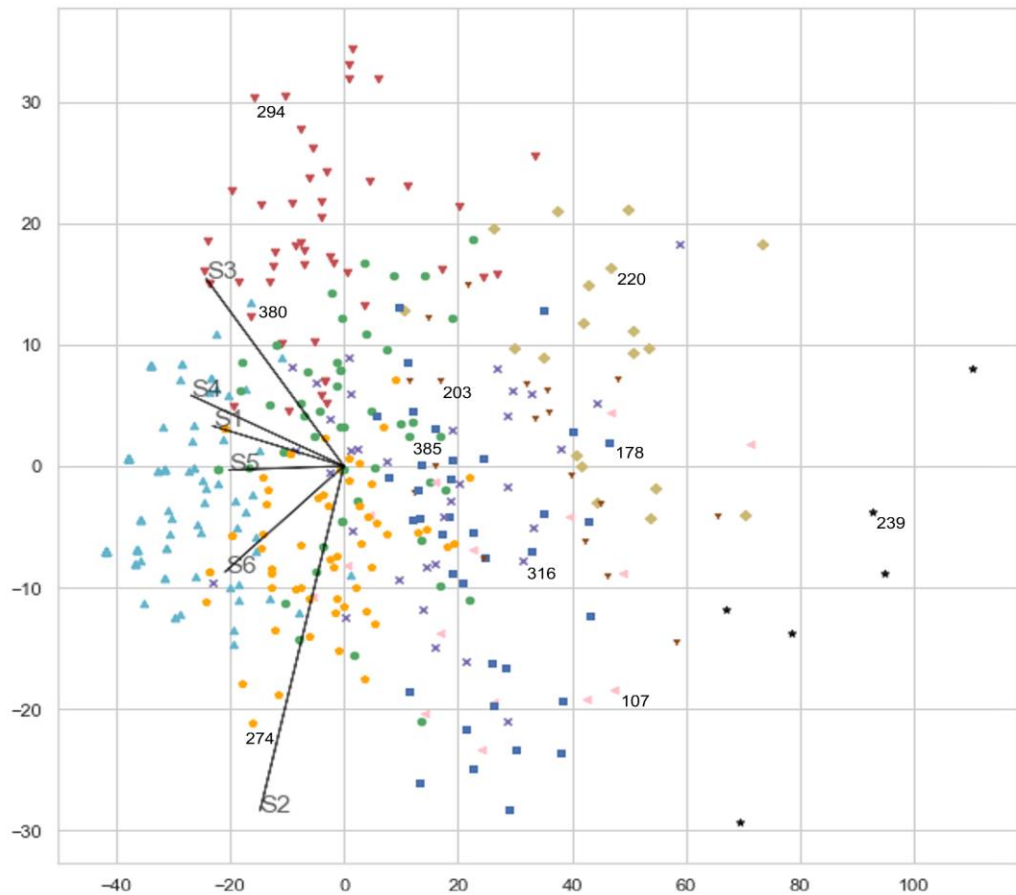
Análisis de clústeres

In [132...

```
for i in range(1,11):  
    print('Cluster '+str(i)+': '+str(Daten[Liste_A_Q3].loc[CL.labels_ == i].shape[0]))
```

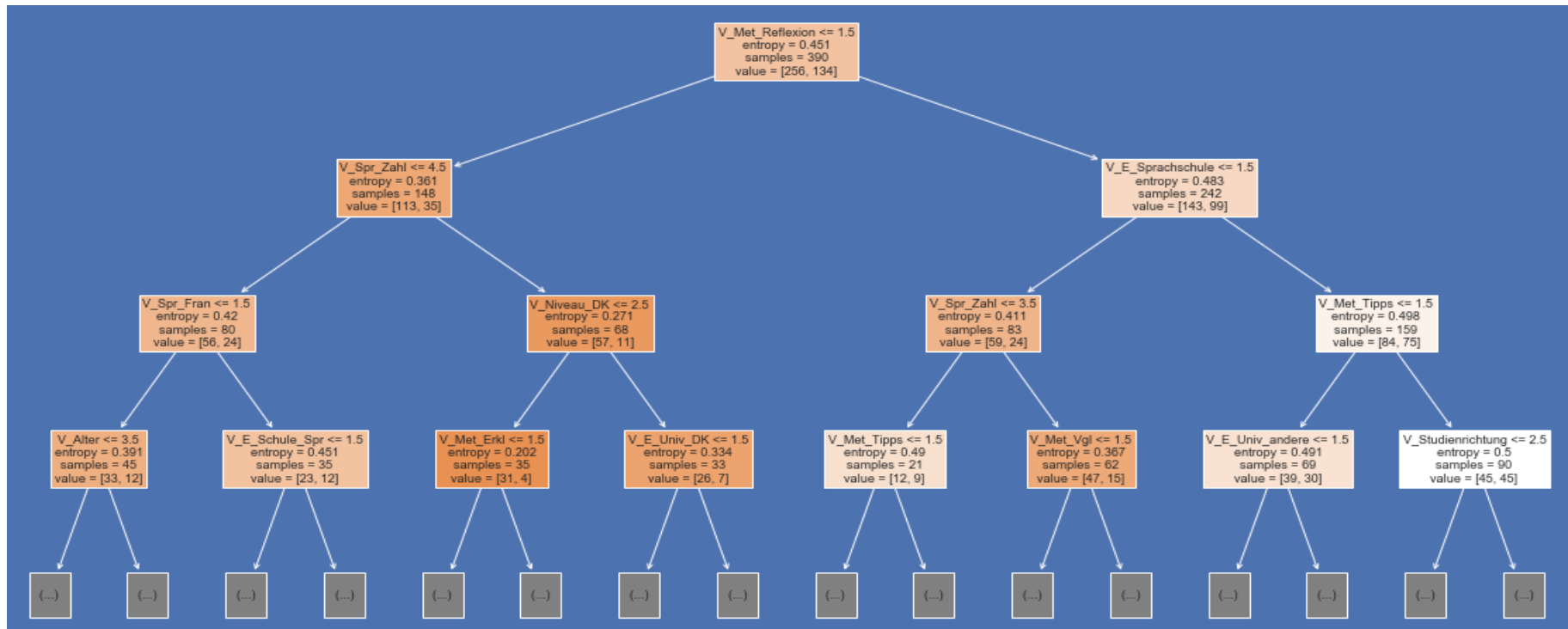
Cluster 1: 61
Cluster 2: 37
Cluster 3: 50
Cluster 4: 47
Cluster 5: 34
Cluster 6: 109
Cluster 7: 19
Cluster 8: 6
Cluster 9: 18
Cluster 10: 15





Biplot con resultados
integrados del análisis de
clústeres

(Pandas, Scikit-learn,
SciPy, BIOFES)



Ejemplo de un árbol de decisión (uso de un amplio rango de estrategias)

Análisis de texto libre

Tokenizar

spaCy

Lematizar

Etiquetado gramatical

Eliminar palabras vacías



Word cloud con los lemas más usados

(SpaCy, Scikit-learn,
Word_cloud)

Topics with most important 15 words

```
In [13]: print_top_words(nmfFR, features , 15)
```

Topic #0: contexto conozco compuesta otras sacar fijo verbo compuestas frase conozca oracion deducir entiendo raiz otra
Topic #1: ingles idiomas español conozco otros otras parecido aleman similitud lenguas similares parecidas parecen frances deducir
Topic #2: diccionario idioma otro traductor aleman online comprendo resulta similitud intuir otras fijo completa pons leo
Topic #3: texto frase leo entiendo leer idea tema verbos sentido deducir caso compuestas resto comprension tipo

```
In [28]: examples
```

```
Out[28]:
```

	EL_CON	SIM_ID	DICC	CON_AM	txt
4	0.000000	100.000000	0.000000	0.000000	Relacionarlas con palabras conocidas de mi lengua materna o de las lenguas que conozco y aprendo. La ortografía de las palabras o su sonido puede recordarme a otras palabras similares en otros idiomas.
75	34.145618	0.000000	0.000000	65.854382	Me fijo en el contexto, trato de averiguar el significado a partir de la etimología o de las palabras que forman los compuestos. Trato de entender el texto globalmente para comprender el tema del que se habla antes de imaginar lo que puede significar cada término.
78	13.453096	60.981811	0.000000	25.565093	Primero busco palabras que se parezcan a los idiomas que ya domino, como el español o el inglés y después intento entender qué significan las otras palabras gracias al contexto o a información que se me de acerca del texto.

In [14]:

```
res
```

Out[14]:

	EL_CON	SIM_ID	DICC	CON_AM	n
Cluster 1	14.76	69.93	0.47	14.85	155
Cluster 2	45.72	20.35	18.55	15.38	144
Cluster 3	19.04	1.01	76.24	3.72	14
Cluster 4	17.91	3.99	8.91	69.19	74
Total	25.29	34.03	13.97	26.71	387

Más información en: https://github.com/martinakienberger/FTSA_2020

Visualización

matplotlib

 **seaborn**

Y además...



pinguin

The logo for statsmodels, featuring a stylized blue 'S' made of dots and a line.

statsmodels

Vielen Dank für Ihre Aufmerksamkeit!

Contacto: mkienberger@ugr.es

<https://github.com/martinakienberger>

Bibliografía

- Dörnyei, Z. (2010). *Questionnaires in second language research: construction, administration, and processing*, 2nd ed. Routledge, Taylor & Francis Group, New York.
- Foster I. et al. (2016). *Big data and social science: A practical guide to methods and tools*. Chapman and Hall/CRC.
- Kienberger, M. (2020). *Das Potenzial des potenziellen Wortschatzes nutzen. Erschließungsstrategien für unbekannten Wortschatz unter DaF-Lernenden an spanischen Universitäten*. Tesis doctoral. Universität Wien. <https://utheses.univie.ac.at/detail/55663>
- Kienberger, M., & Vicente, V. (2022). Descubrir el significado de palabras desconocidas en alemán: estrategias empleadas por aprendientes universitarios. *Didáctica. Lengua y literatura*, (34), 95-108. <https://revistas.ucm.es/index.php/DIDA/article/view/81351>
- Petscher, Y., Schatschneider, C. & Compton, D. (eds.) (2013). *Applied quantitative analysis in education and the social sciences*. Routledge.

Librerías citadas

<https://eli5.readthedocs.io>

<https://matplotlib.org>

<https://numpy.org>

<https://pandas.pydata.org>

<https://pypi.org/project/biofes/>

<https://scikit-learn.org/>

<https://scipy.org>

<https://seaborn.pydata.org>

<https://spacy.io>