Productivity of combinations of Spanish anatomical themes with symptom suffixes based on quantitative analysis

Carlos Herrero Zorita
Hiroto Ueda
Antonio Moreno Sandoval

Topic

The study of the frequency, productivity and compatibility of a series of Spanish medical *prefixal* themes related to parts of the human body and *suffixal* themes related to symptoms in a medical corpus

Topic

e.g.

Prefixal theme: *cardio*- (related to de heart)

Suffixal theme: *-algia* (pain)

Topic

Questions:

- What is their frequency of use?

- Can they all combine, for example, to form the word *cardialgia*?

- Is the combination of anatomical prefixes and symptoms suffixes random?

Data

- **Corpus:**

- Spanish subcorpus of the *Multimédica* corpus.

- Written medical corpus.

- Scientific and informative discourse.

- More than 4 million words.

Object of study

- 13 medical prefixes related to body:

cefal(o)-	Head
cerebr(o)-	Brain
dermat(o)-	Skin
arteri(o)-	Arteries
artr(o)-	Articulations
oste(o)-	Bones
neum(o)-	Lungs
tiroid(o)-	Thyroidal gland
miel(o)-	Marrow
hepat(o)-	Liver
hem(o)-	Blood
cardi(o)-	Heart
ocul(o)-	Eyes

Object of study

- 14 medical suffixes related to symptons:

-algia	Pain
-itis	Infection
-tomía	Cut, incision
-patía	Illness
-osis	Pathology
-génesis	Generation
-malacia	Softening
-blasto	Embryonic state of development
-oma	Tumour
-oide	Resemblance
-cito	Cell
-megalia	Irregular enlargement
-tóxico	Poison
-cele	Tumour

Tools

- *NUMEROS-web: Programs for quantitative data analysis*

- Developed by Hiroto Ueda

- Online tool.

- Statistical operations based on integration analyses.

Integration analyses

- Aim:

  - Distribution structure in matrixes

  - Regrouping rows and columns from numerical information

  - Distance and correlation in Euclidean space

Integration analyses

- Example:

	L-1	L-2	L-3	L-4	Value
d-1	v	v			1.581
d-2			v		1.000
d-3		v			2.160
d-4			v	v	3.000
d-5	v	v	v		3.536
Value	2.236	2.160	4.082	5.000	

→

	L-2	L-1	L-3	L-4	Value
d-3	v				1.000
d-1	v	v			1.581
d-5	v	v	v		2.160
d-2			v		3.000
d-4			v	v	3.536
Value	2.160	2.236	4.082	5.000	

Perea and Ueda (2012)

Integration analyses

- **Analyses:**

  - Distance to point zero analysis

  - Correspondence analysis

  - Cluster analysis

Integration analyses

- Distance to point zero analysis

  - Hiroto Ueda

  - Distance to point zero

  - Rows and columns

 - Diagonalised distribution

Integration analyses
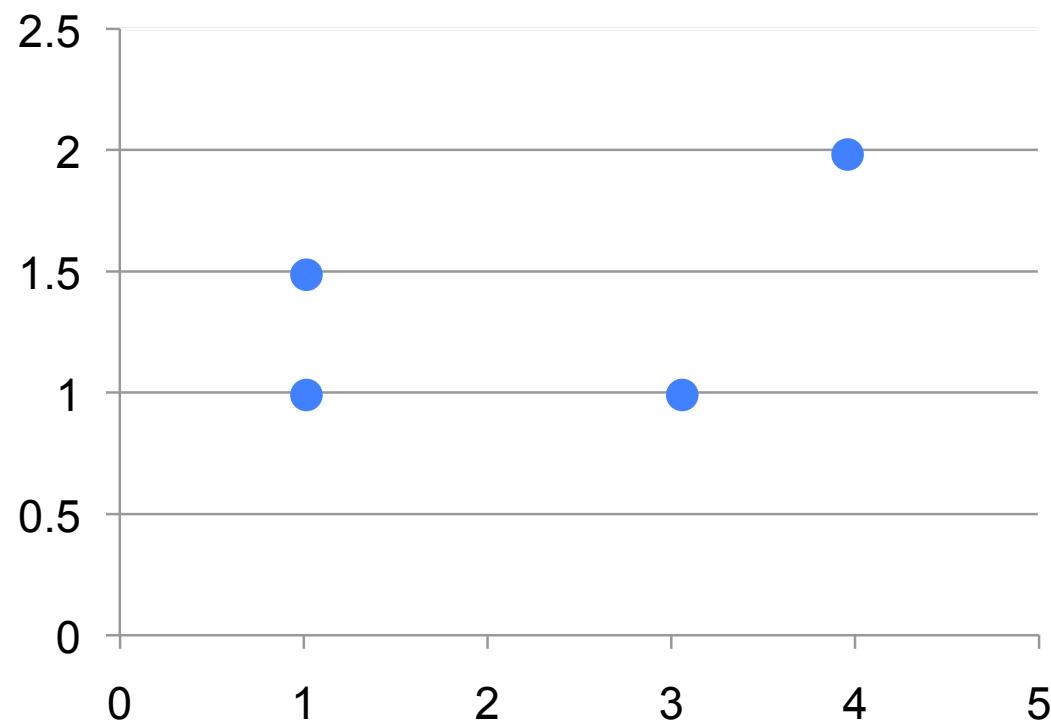
- Distance to point zero analysis

Integration analyses

- Distance to point zero analysis
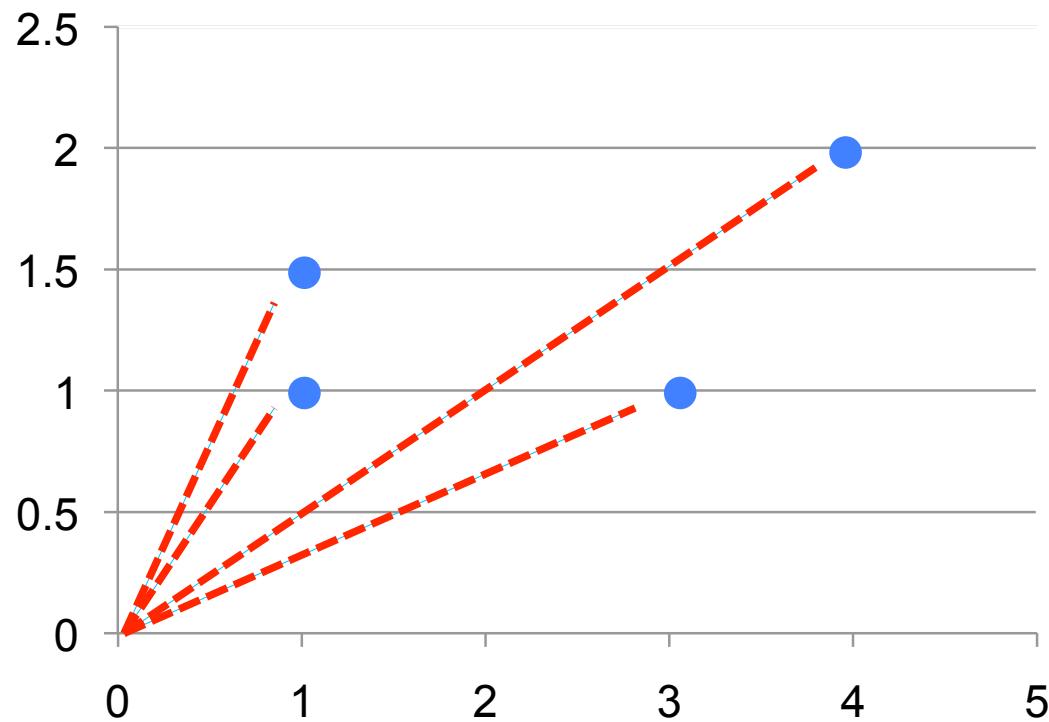
Integration analyses

- **Correspondence analysis**

  - Maximun level of correlation

  - Rows and columns
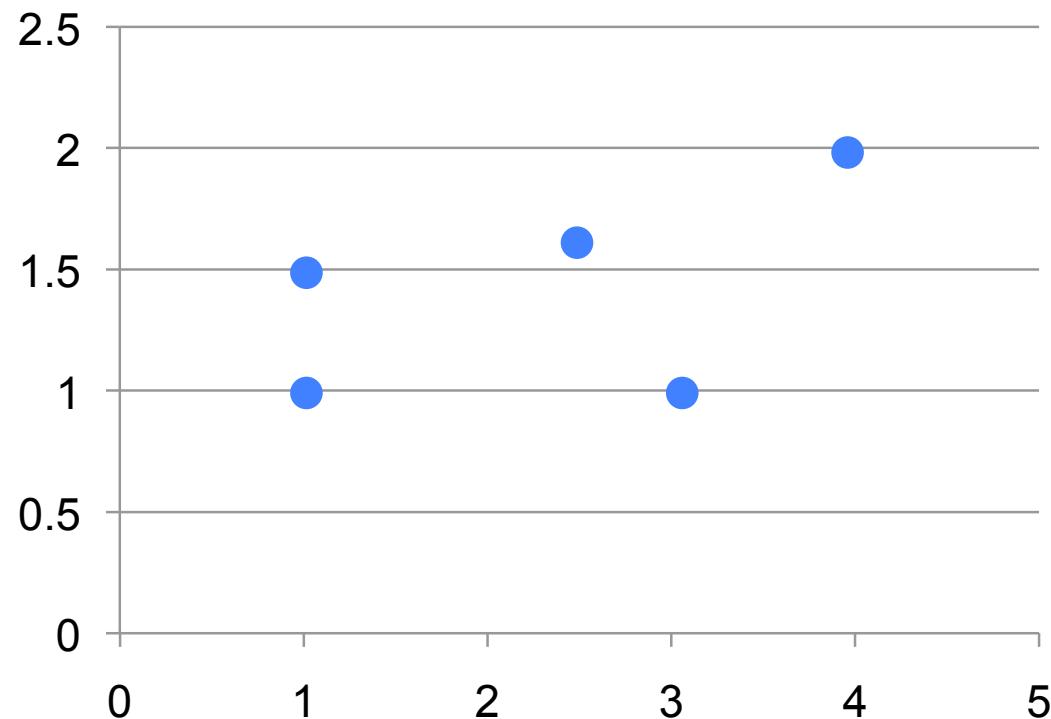
Integration analyses

- Correspondence analysis

Integration analyses

- Correspondence analysis
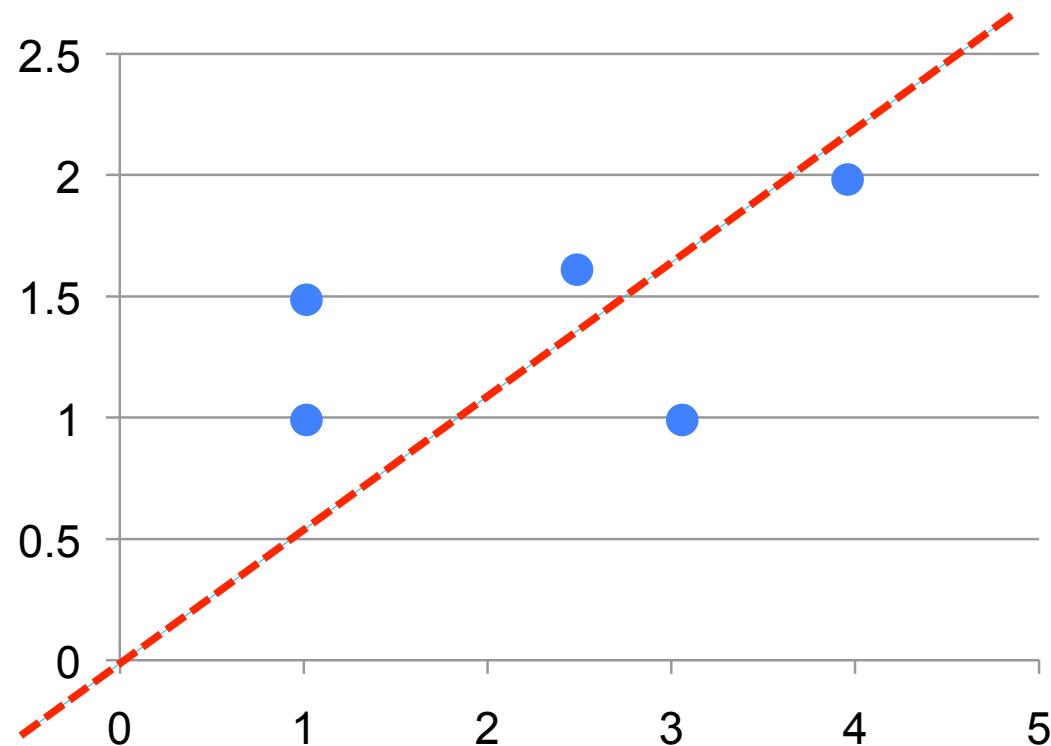
Integration analyses

- **Cluster analysis**

  - Distance among members

  - Group distribution

  - Rows and columns

Integration analyses
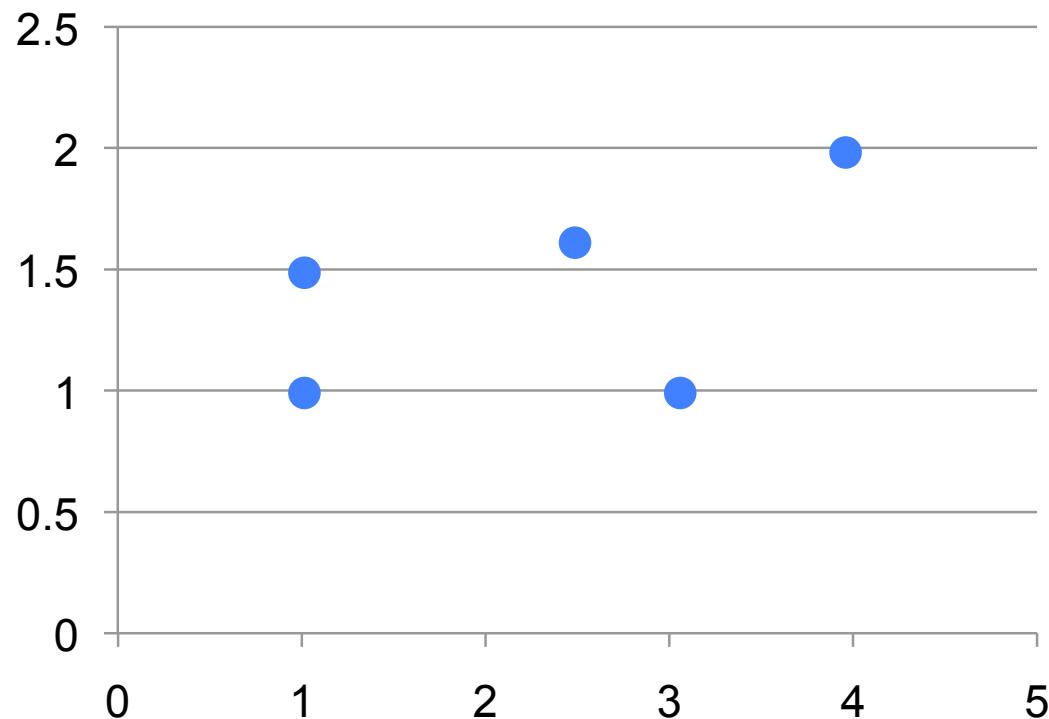
- Cluster analysis

Integration analyses

- Cluster analysis
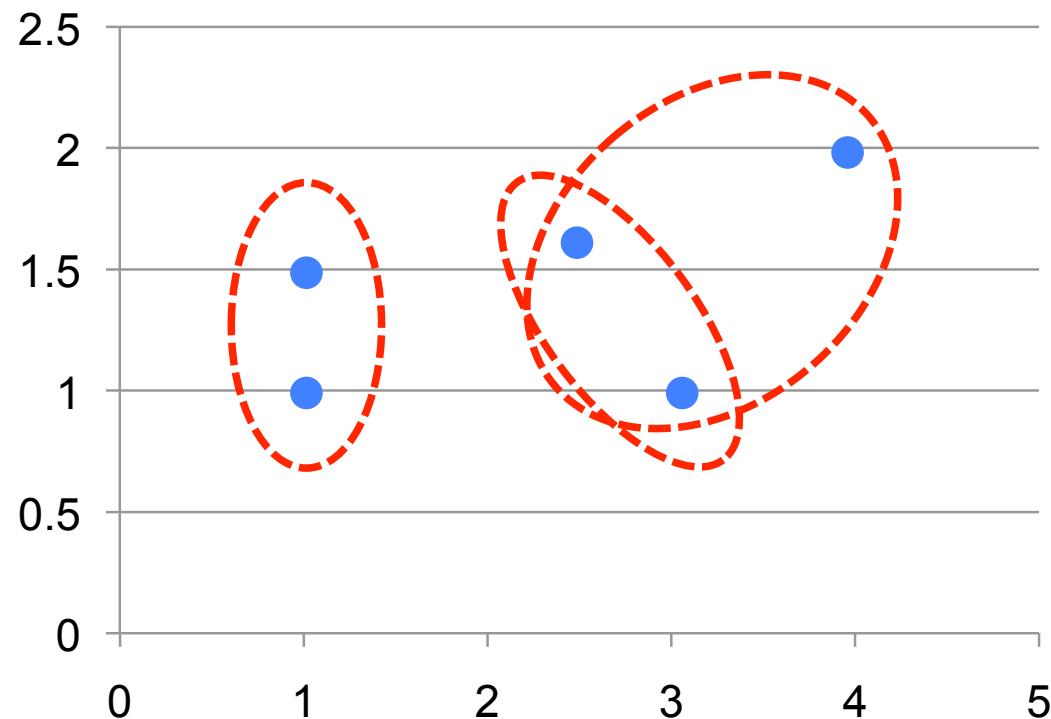
Objectives

- To observe the frequency of use of a series of medical prefixes and suffixes

- To study the compatibility between the prefixes and the suffixes

- To apply three statistical methods based on matrixes to test this compatibility

Results

Frequencies and productivity of prefixes and suffixes

Results – Frequency: Prefixes

PREFIX	ABSOLUTE	NORMALISED	TYPES
HEMO-	4091	1014,84	68
ARTERI-	2628	651,92	55
HEPAT-	2616	648,94	40
CARDIO-	2264	561,62	34
CEREBR-	2262	561,13	33
NEUMO-	1708	423,70	30
ARTR-	1202	298,18	24
OSTEO-	1167	289,49	23
CEFAL-	1158	287,26	23
MUSCUL-	1155	286,52	15
TIROID-	974	241,62	11
ABDOMIN-	905	224,50	10
MIELO-	758	188,03	10
OCUL-	705	174,89	5
DERMAT-	441	109,40	4

Results – Productivity: Prefixes

PREFIX	LEMMA	NORMALISED	RELATIVE
HEMO-	Hemorragia	419,23	41,31
HEPAT-	Hepatitis	392,19	60,16
CEREBR-	Cerebral	360,44	55,54
ARTERI-	Arterial	328,69	58,52
CARDIO-	Cardiovascular	315,79	56,28
ABDOMIN-	Abdominal	222,76	52,58
ARTR-	Artritis	214,33	71,88
NEUMO-	Neumonía	208,38	71,98
CEFAL-	Cefalea	134,45	46,80
TIROID-	Tiroideo	121,55	42,42
OCUL-	Ocular	103,44	42,81
DERMAT-	Dermatitis	54,57	24,31
OSTEO-	Osteoporosis	51,10	27,18
MIELO-	Mieloma	49,37	28,23
MUSCUL-	Musculatura	15,88	14,51

TERMS WITH ARTR-

Artritis

Artralgia

Artropatía

Artrópodo

Artrosis

Artroplastia

Artrítico

Artroscopia

Artroscópico

Artrotomía

Artroconidio

Artrografía

Artromialgia

Artrocalásico

Artritisartralgia

Artritógeno

Artrocentesis

...

Results – Productivity: Prefixes
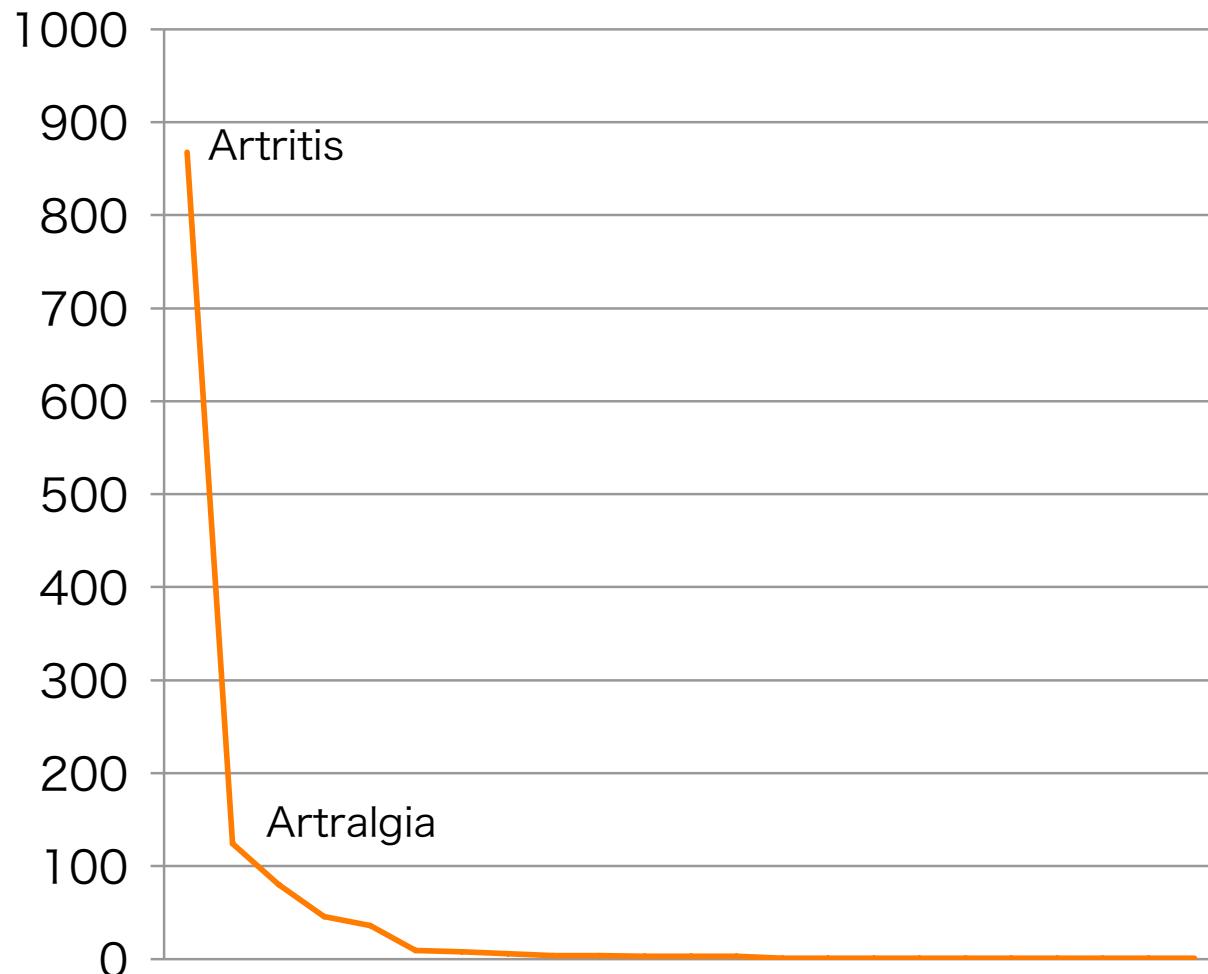
Results – Frequency: Suffixes

SUFFIX	ABSOLUTE	NORMALISED	TYPES
-ITIS	11683	2898,16	439
-OSIS	11642	2887,99	282
-PATÍA	5444	1350,48	195
-OMA	4646	1152,52	172
-TOMÍA	1382	342,83	124
-OIDE	1350	334,89	116
-ALGIA	808	200,44	55
-MEGALIA	451	111,88	42
-GÉNESIS	386	95,75	36
-CITO	288	71,44	23
-TÓXICO	225	55,82	22
-MALACIA	66	16,37	15
-BLASTO	42	10,42	11
-CELE	18	4,47	5

Results – Productivity: Suffixes

SUFFIX	LEMMA	NORMALISED	RELATIVE
-ITIS	Hepatitis	393,68	13,58
-OSIS	Estenosis	215,07	7,45
-PATÍA	Neuropatía	175,14	12,97
-OMA	Linfoma	132,22	11,47
-OIDE	Reumatoide	84,09	25,11
-ALGIA	Mialgia	49,86	24,88
-MEGALIA	Esplenomegalia	39,44	35,25
-TÓXICO	Citotóxico	30,51	54,67
-TOMÍA	Anatomía	22,57	6,58
-CITO	Linfocito	17,36	24,31
-GÉNESIS	Angiogénesis	16,62	17,36
-MALACIA	Osteomalacia	14,88	90,91
-BLASTO	Citoblasto	3,47	33,33
-CELE	Mielomeningocele	0,50	11,11

Results – Frequency: Prefixes

Chart showing frequency of prefixes and suffixes. Y-axis ranges from 0 to 14000 in increments of 2000.

Labels on the chart:
- -ITIS
- -OSIS
- -PATÍA
- -OMA
- HEMO-
- HEPAT-

Legend:
- PREFIX (green)
- SUFFIX (blue)
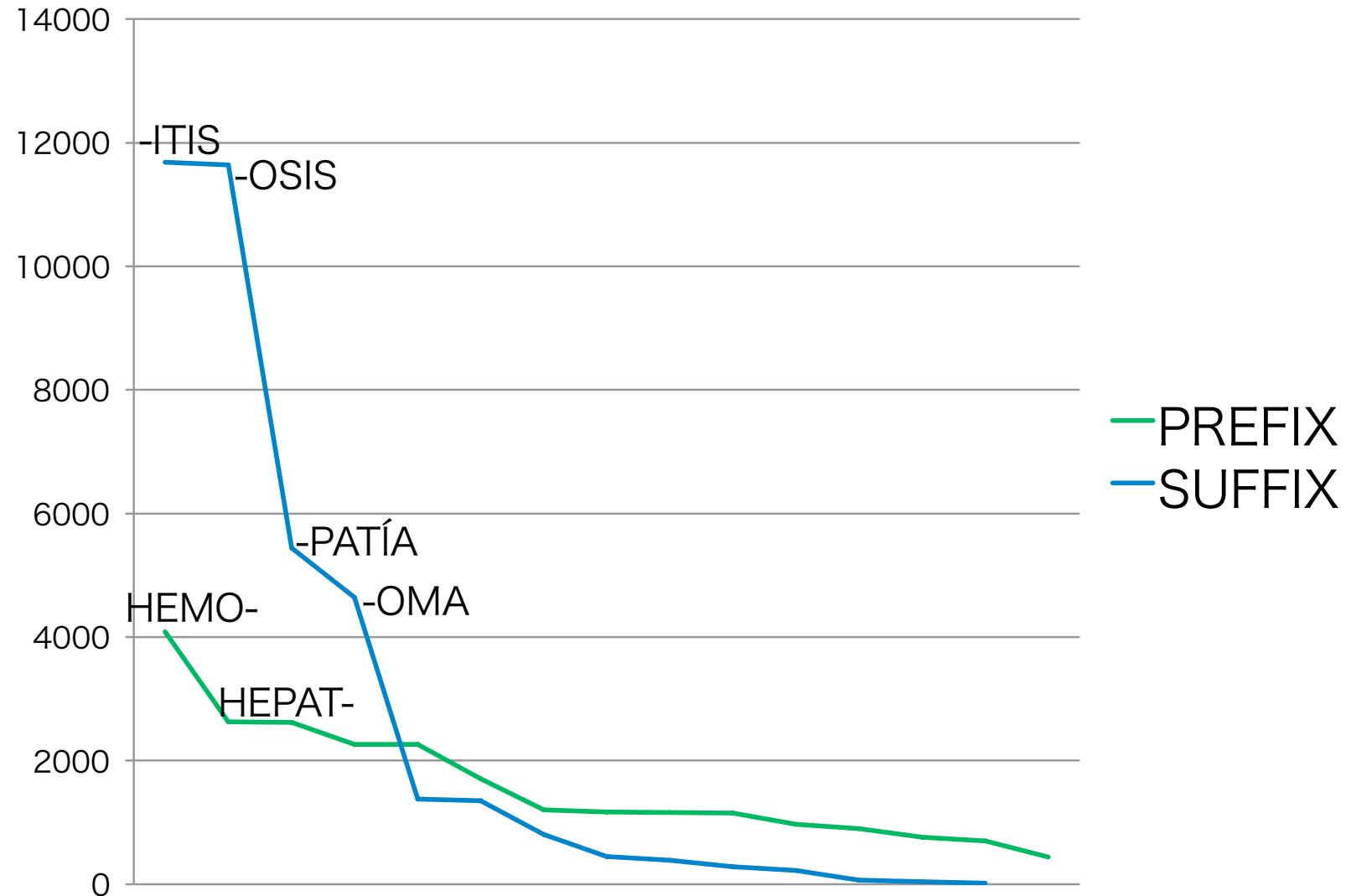
Results

Analyses regarding combination of prefixes and suffixes

Matrix array of combination results

Prefixes ↓   **Suffixes** ↓

	-algia	-blasto	-cele	-cito	-génesis	-itis	-malacia	-megalia	-oide	-oma	-osis	-patía	-tomía	-tóxico
arteri(o)-					1	2					4	1	1	
artr(o)-	3					3					2	2	1	
cardi(o)-								1			1	2		1
cefal(o)-	1													
cerebr(o)-						2					1			
dermat(o)-						4				2	7			
hem(o)-		2							1		2	3	1	
hepat(o)-		1		1		2		2		3	1	1	1	1
miel(o)-		1	1	1		1		1	1	2	2			1
neum(o)-				1		1				1	2	1	1	
ocul(o)-												1		
oste(o)-		2		1	2	3	1		1	5	8	2	1	
tiroid(o)-						1						1	1	

Matrix array of combination results

Prefixes → Suffixes →

	-algia	-blasto	-cele	-cito	-génesis	-itis	-malacia	-megalia	-oide	-oma	-osis	-patía	-tomía	-tóxico
arteri(o)-					1	2					4	1	1	
artr(o)-	3					3					2	2	1	
cardi(o)-								1			1	2		1
cefal(o)-	1													
cerebr(o)-						2					1			
dermat(o)-						4				2	7			
hem(o)-		2							1		2	3	1	
hepat(o)-		1		1		2		2		3	1	1	1	1
miel(o)-		1	1	1		1		1	1	2	2			1
neum(o)-				1		1				1	2	1	1	
ocul(o)-												1		
oste(o)-		2		1	2	3	1		1	5	8	2	1	
tiroid(o)-						1						1	1	

Matrix array of combination results

Prefixes Suffixes

	-algia	-blasto	-cele	-cito	-génesis	-itis	-malacia	-megalia	-oide	-oma	-osis	-patía	-tomía	-tóxico
arteri(o)-					1	2					4	1	1	
artr(o)-	3				3						2	2	1	
cardi(o)-							1				1	2		1

Cardiopatía (*cardiopathy*, "illness of the heart")

Cardiomiopatía (*cardiomyopathy*, "illness of the heart muscle")

Matrix array of combination results

	-algia	-blasto	-cele	-cito	-génesis	-itis	-malacia	-megalia	-oide	-oma	-osis	-patía	-tomía	-tóxico
arteri(o)-					1	2					4	1	1	
artr(o)-	3					3					2	2	1	
cardi(o)-								1			1	2		1
cefal(o)-	1													
cerebr(o)-						2					1			
dermat(o)-						4				2	7			
hem(o)-		2							1		2	3	1	
hepat(o)-		1		1		2		2		3	1	1	1	1
miel(o)-		1	1	1		1		1	1	2	2			1
neum(o)-				1		1				1	2	1	1	
ocul(o)-												1		
oste(o)-		2	1	2	3	1			1	5	8	2	1	
tiroid(o)-						1						1	1	

- ## Distance to zero point analysis

Dst.cct.	-algia	-génesis	-itis	-malacia	-osis	-oma	-tomía	-cito	-oide	-cele	-blasto	-patía	-megalia	-tóxico	Value
cefal(o)-	1														1,00
cerebr(o)-			2		1										3,79
dermat(o)-			4		7	2									4,67
arteri(o)-		1	2		4		1					1			5,92
artr(o)-	3		3		2		1					2			6,16
oste(o)-		2	3	1	8	5	1	1	1		2	2			6,67
neum(o)-			1		2	1	1	1				1			7,09
tiroid(o)-			1				1					1			8,21
miel(o)-			1		2	2		1	1	1	1		1	1	8,85
hepat(o)-			2		1	3	1	1			1	1	2	1	9,04
hem(o)-					2		1		1		2	3			9,74
cardi(o)-					1							2	1	1	11,64
ocul(o)-												1			12,00
Value	4,36	5,42	5,86	6,00	6,36	7,39	7,66	8,15	8,91	9,00	9,08	9,15	10,31	10,41	

- ## Distance to zero point analysis

Dst.cct.	-algia	-génesis	-itis	-malacia	-osis	-oma	-tomía	-cito	-oide	-cele	-blasto	-patía	-megalia	-tóxico	Value
cefal(o)-	1														1,00
cerebr(o)-			2		1										3,79
dermat(o)-			4		7	2									4,67
arteri(o)-		1	2		4		1					1			5,92
artr(o)-	3		3		2		1					2			6,16
oste(o)-		2	3	1	8	5	1	1	1		2	2			6,67
neum(o)-			1		2	1	1	1				1			7,09
tiroid(o)-			1				1					1			8,21
miel(o)-			1		2	2		1	1	1	1		1	1	8,85
hepat(o)-		2			1	3	1	1			1	1	2	1	9,04
hem(o)-					2		1			1	2	3			9,74
cardi(o)-					1							2	1	1	11,64
ocul(o)-												1			12,00
Value	4,36	5,42	5,86	6,00	6,36	7,39	7,66	8,15	8,91	9,00	9,08	9,15	10,31	10,41	

- # Correspondence analysis

Dst.cct.	-algia	-itis	-tomía	-patía	-osis	-génesis	-malacia	-blasto	-oma	-oide	-cito	-megalia	-tóxico	-cele	Value
cefal(o)-	1														- 7,01
artr(o)-	3	3	1	2	2										- 2,06
tiroid(o)-		1	1	1											- ,33
ocul(o)-				1											- ,29
cerebr(o)-		2			1										- ,19
arteri(o)-		2	1	1	4	1									- ,01
dermat(o)-		4			7				2						,12
neum(o)-		1	1	1	2						1	1			,18
hem(o)-			1	3	2			2		1					,21
oste(o)-		3	1	2	8	2	1	2	5	1	1				,36
cardi(o)-				2	1							1	1		,41
hepat(o)-		2	1	1	1			1	3		1	2	1		,53
miel(o)-		1			2			1	2	1	1	1	1	1	,79
Value	- 4,81	- ,26	- ,24	- ,20	,13	,34	,52	,60	,60	,66	,68	,83	,84	1,15	

- # Correspondence analysis

Dst.cct.	-algia	-itis	-tomía	-patía	-osis	-génesis	-malacia	-blasto	-oma	-oide	-cito	-megalia	-tóxico	-cele	Value
cefal(o)-	1														- 7,01
artr(o)-	3	3	1	2	2										- 2,06
tiroid(o)-		1	1	1											- ,33
ocul(o)-				1											- ,29
cerebr(o)-		2			1										- ,19
arteri(o)-		2	1	1	4	1									- ,01
dermat(o)-		4			7				2						,12
neum(o)-		1	1	1	2				1		1				,18
hem(o)-			1	3	2			2		1					,21
oste(o)-		3	1	2	8	2	1	2	5	1	1				,36
cardi(o)-				2	1							1	1		,41
hepat(o)-		2	1	1	1			1	3		1	2	1		,53
miel(o)-		1			2			1	2	1	1	1	1	1	,79
Value	- 4,81	- ,26	- ,24	- ,20	,13	,34	,52	,60	,60	,66	,68	,83	,84	1,15	

- ## Cluster analysis

Dst.cct.	-algia	-blasto	-oide	-cito	-oma	-génesis	-malacia	-itis	-osis	-patía	-tomía	-cele	-megalia	-tóxico	Value
arteri(o)-						1		2	4	1	1				
dermat(o)-					2			4	7						,06
oste(o)-		2	1	1	5	2	1	3	8	2	1				,14
neum(o)-			1	1				1	2	1	1				,26
artr(o)-	3							3	2	2	1				,49
cerebr(o)-								2	1						,67
hepat(o)-		1		1	3			2	1	1	1		2	1	1,02
miel(o)-		1	1	1	2			1	2			1	1	1	1,23
cardi(o)-									1	2			1	1	1,65
ocul(o)-										1					1,78
hem(o)-		2	1						2	3	1				1,97
tiroid(o)-								1		1	1				2,27
cefal(o)-	1														2,84
Value		,55	,61	,86	,97	1,21	1,26	1,58	1,71	2,11	2,31	2,81	3,09	3,13	

- Cluster analysis

Dst.cct.	-algia	-blasto	-oide	-cito	-oma	-génesis	-malacia	-itis	-osis	-patía	-tomía	-cele	-megalia	-tóxico	Value
arteri(o)-						1		2	4	1	1				
dermat(o)-					2			4	7						,06
oste(o)-		2	1	1	5	2	1	3	8	2	1				,14
neum(o)-				1	1			1	2	1	1				,26
artr(o)-	3							3	2	2	1				,49
cerebr(o)-								2	1						,67
hepat(o)-		1		1	3			2	1	1	1		2	1	1,02
miel(o)-		1	1	1	2			1	2			1	1	1	1,23
cardi(o)-									1	2			1	1	1,65
ocul(o)-										1					1,78
hem(o)-		2	1						2	3	1				1,97
tiroid(o)-								1		1	1				2,27
cefal(o)-	1														2,84
Value		,55	,61	,86	,97	1,21	1,26	1,58	1,71	2,11	2,31	2,81	3,09	3,13	

Conclusions

1. This work has provided an analysis of the frequency of use of our selected themes as well as a glossary of the medical terms they form.

- The frequency of suffixes –isis, -osis, -patía and –oma is higher than any other morpheme.

- The frequency of prefixes is lower but more equally distributed than the frequency of the suffixes.

Conclusions

2. This work has provided an opportunity for applying a series of integration analyses in the field of medical term morphology.

- These analyses allow to draw similarities in usage between medical anatomical and symptoms themes based only on quantitative linguistic information.

- The results show that the process of combination is not random and does not necessary belong only to semantic similarity.

Conclusions

- Distance to zero point analysis has shown pairs of anatomical prefixes with similar usage and semantic information:

  cefalo- / cerebr-: Head / Brain

  artro- / osteo-: Articulations / Bones

  neumo- / tiroid(o)- / miel(o)- / hepat(o)-: Lungs / Tyroidal gland / Liver

  hemo- / cardio-: Blood, heart

Conclusions

- Correspondence analysis and cluster analysis seem to draw similar conclusions regarding the suffixes, grouping them into three clear groups with apparently no semantic relationship:

  -itis / -tomía / -patía / -osis : Inflammation, incision, illness, patholgy.

  -blasto / -oma / -oide / -cito: Embryo, tumour, resemblance, cell.

  -megalia / -tóxico: irregular enlargement, poison

Thank you

Carlos Herrero Zorita
carlos.herrero@uam.es

Antonio Moreno Sandoval
antonio.msandoval@uam.es

Hiroto Ueda
hiroto.ueda.tokio@gmail.com

NUMEROS-web: Programs for quantitative data analysis
http://lecture.ecc.u-tokyo.ac.jp/~cueda/numeros/