Linguistic determinants of English for Academic Purposes

Christoph Haase
Chemnitz University of Technology

Outline

1) Introduction
2) Hypotheses
3) Hedge expressions
4) Lexical gradience
5) Source/target domains
6) Conclusion

1. Introduction

1.1 The Corpus of Scientific and Popular Academic English (SPACE): Compilation and structure

binary structure: 1) academic texts,
2) popular-academic texts

academic texts: from 3 preprint servers for academic publications.
arXiv.org
Proceedings of the National Academy of Sciences
Public Library of Science - Medicine (PLoS)

popular-academic texts: *New Scientist*

1.2 What characterizes academic texts?

0104PN Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants

Phylogenetic relationships within **Plantaginaceae** were determined from a 4,730-nt data set consisting of portions of four chloroplast regions (ndhF, rbcL, and **intergenic spacers atpB rbcL** and trnLtrnF). Relationships within Plantago subgenus Plantago were analyzed from a **9,845-nt data set** containing two additional chloroplast regions (intergenic spacers psaAtrnS and trnCtrnD). **Maximum likelihood (ML) trees** were constructed with PAUP* by using the general time-reversible model, a gamma distribution with four rate categories, and an estimate of the proportion of invariant sites. The rate matrix, base frequencies, shape of the **gamma distribution**, and proportion of invariant sites were estimated before the MLanalysis from a neighbor-joining tree constructed from the data. Divergence times outside Plantaginaceae were taken from ref. 27. Those within the family were calculated by using a penalized likelihood approach (28) as implemented in the R8S program (29) and a time constraint of 48 million years (27) for the Antirrhinum Plantago split. The ML tree was used as the starting tree for the divergence time analysis. The starting tree was constructed by first constraining the taxa in the 4,730-nt data set to incorporate the alternative relationships within subgenus Plantago and then estimating branch lengths for this topology in PAUP*. A smoothing factor of three was determined by using the R8S cross-validation procedure.

0104NS Plant DNA shows speedy changes

The mitochondria of a group of nondescript flowering plants contain the fastest-evolving DNA yet known.

Until now, the mitochondrial genomes of plants were thought to evolve slowly. But when Jeffrey Palmer and colleagues at Indiana University in Bloomington compared mitochondrial DNA from nine species of plantain (members of the genus Plantago) and 41 other plants, they found that some Plantago sequences changed

1.3 Academic vs. popular-academic strategies

Popular versions use

1. syntactic compression and semantic simplification,

cf. *Mitochondrial substitution rates are extraordinarily elevated* vs.
Plant DNA shows speedy changes

2. "lack of lexical differentiation" as observed by Lorenz in second language learners (Lorenz 1998:58)

3. overuse of stylistic devices like amplifiers (*completely*, *absolutely*) and boosters (*very*, *highly*, *immensely*)

- similarity between authors of popular academic texts and second language learners

- lack of lexical differentiation mirrored by lack of scientific differentiation

1.4 The Academic Discourse situation

author/speaker & reader/listener: contract of diffusing knowledge
A B from A to B

Contract: A and B both share the knowledge that:

A knows that B does not know everything that A knows (about X)

A uses conventionalised strategies to express him/herself comprehensibly
→ lexical, semantic and pragmatic means to modify **propensity** of a statement

Propensity: degree of probability of a statement to hold true

1.5 Diffusion of knowledge in academic texts

Linguistic determinants of academic texts negotiate 1-3:

1. the author A of an academic text knows that reality has more than "ideal" cases of e.g. Newtonian mechanics but is blurred at microlevels
→ hedge an utterance by saying "Ideally,…" etc.

caveat: intended listeners share this knowledge and don't need the explicit hedge marking

2. the author B of a popular academic text (who at the same time is an informed reader of 1.) knows that his/her readers **do not** share the knowledge of 1.
→ the hedgy precision of 1.) has a different function
→ hedging, lexical downgrading used as simplification

3. the reader C of 2.) has a contract that contents will be processed and force-fed via hedges and metaphors

C knows that C does not know what A or B know so C expects lexicalization that coincides with C's knowledge by

transformation of specific knowledge (source domain) to
generic knowledge (target domains)

→ cognitive core of metaphorization

1.6 Linguistic determinants of academic speech (selection)

level	**lexicon**	**syntax**	**semantics**	**pragmatics**
N	**ontological depth**	count/mass	**metaphors** **source/target domains**	hedges commitment deixis
V	causatives resultatives	aux V causation	**$V_{perception}$** motion events	**hedges** **commitment**
Conj	cause clause effect clause			linkers
Det/Q			definiteness scope	
P			spatial domains	
Adv			modal adv	
clause	linking	biclausal causativity		
text				abstract conclusion

2. Hypotheses and data discussion

- observed and intuitive differences between academic and popular academic texts can be quantified

1) in terms of the propensity of their vague and hedgy expressions
2) in terms of the semantic depth of their content words

3) Academic texts are less subjective and show ontologically "deeper" content words
4) Popular-academic texts are more subjective and use more general content words
5) Academic text processing of non-native speakers mirrors the difference between Academic and popular-academic texts

3. Hedge expressions

3.1 Functions

Hedges: a metaphorical device (coined by Lakoff)
properties: to delimit the scope of an utterance i.e.

they…
a.) distance the speaker from the utterance

b.) blur quantities, attributes, specifications given in the utterance

c.) relativize notions of truth

Canonical examples: *sort of, kind of*

3.2 Usage

Hedges can be used to estimate therefore

a.) the commitment of a speaker/producer of a text to his/her utterance

b.) the amounts, causes, applications in question and

c.) the distance of the listener to fully commit to the semantic content or truth value of his/her utterance

hedging: enables therefore both to cross borders which are primarily borders of knowledge

3.3 Forms

Hedges follow pragmatic lexicalization patterns
cut across syntactic classes

→ there is no definite, taggable class of a hedge, only:

1. reporting verbs (*thinks*, *believes*, *claims*, *says*, etc.);
2. verbs of outcome and resultatives (*succeeds*, *finishes*, etc.)
3. prepositional phrases of mediation such as *by means of*, *on behalf of*, *etc.*
4. modal verbs in their deontic and epistemic meaning
5. modal adverbs (probably, likely, possibly, certainly…)
6. quantifiers (some, most, few…)

3.4 Post-processing

Lexical analysis:

- POS-tagged using Treetagger and Penn Treebank tagset

Pragmatic analysis:

- manually annotated for author commitment
- hedge expressions like *probably*, *normally*, *suggests that*, *some evidence for* etc. assigned a value between 1 and 10

1 (extremely low probability/propensity)
10 (certainty, extremely high probability/propensity)

e.g. **partially_AV0_M_6**

POS tag:	AV0 (adverb)
positional marker:	M for medial
propensity score:	6

3.5 Propensity tags

0090PN Topical DNA oligonucleotide therapy reduces UV-induced mutations and photocarcinogenesis in hairless mice

UV-induced DNA damage gives rise to mutations and skin cancer. We show that topical pTT pretreatment enhances the rate of DNA photoproduct removal, decreases UV-induced mutations, and reduces photocarcinogenesis in UV-irradiated hairless WT repair-proficient and Xpc heterozygous **partially_AV0_M_6** repair-deficient mice, both transgenic for the lacZpUR288 mutation-indicator gene. These data support the existence of inducible mammalian DNA damage responses that increase DNA repair capacity after DNA damage and hence reduce the impact of future exposures to environmental carcinogens. The ability of topically applied pTT to induce protective physiologic responses that **normally_AV0_M_8** result from DNA damage **suggests_VV_M_8** a previously undescribed means of reducing skin cancer in high-risk individuals. Skin cancer accounts for **at least_AJS_M_9** 40% of all human malignancies, 1,000,000 cases annually in the U.S. (1, 2). Incidence is **clearly_AV0_M_9** linked to UV exposure and increases exponentially with age (1, 3). Skin cancer risk is **greatly_AV0_M_8** increased in the rare disease xeroderma pigmentosum (XP), because of mutation in one of several DNA repair […]

0090NS Suntan lotion primes the skin's defences

It **might be_VM_M_4** possible to develop suntan lotions that kick-start the skin's protective mechanisms against cancer before you hit the beach. The key ingredient **could be_VM_M_6** a fragment of DNA just two bases long, called a TT dimer, that mimics one of the signs of DNA damage from ultraviolet light. Barbara Gilchrest's team from Boston University and colleagues in the Netherlands exposed hairless mice to a mild ultraviolet radiation, the equivalent of half an hour of afternoon sun. They found that genes involved in DNA repair were **more_AV0_M_8** active in mice that had the TT dimmer rubbed on their skin before exposure. And only 22 per cent of the treated mice developed skin cancers after 24 weeks compared with 88 percent of untreated mice. (Proceedings of the National Academy of Sciences, DOI:10.1073/pnas.0306389101). People who want a tan may **not even need to_VM_M_5** go out in the sun. Mouse skin does not produce melanin but earlier tests on guinea pigs **suggest that_VV_M_7** the TT dimer also triggers the tanning response. The team has not yet begun testing it on people.

3.6 The test

Elicitation of hedge expressions in gap filling in Academic texts

Test A with free gaps
Test B word class or list given
Test C word class or list given
Test D category of medium propensity elicited

Informants:

- 32 German students of English
  (1st year, Chemnitz University of Technology)
- 70 Czech students of English
  (3rd year, Masaryk University, Brno)

3.7 Academic text sample for testing

Test A	Test B
TEST A: Please complete the text in the most natural way by inserting lexical items	**TEST B: Please complete the text in the most natural way by either inserting items and/or by choosing from the lists**
Overcoming an evolutionary conflict: Removal of a reproductive organ _____ increases locomotor performance	**Suntan lotion primes the skin's defences**
One potential consequence of sexual size dimorphism is conflict among characters._____, a structure evolved for reproduction _____ impair performance during other activities (e.g., locomotion). Here we provide quantitative evidence for an animal overcoming an evolutionary conflict generated by differential scaling and sexual size dimorphism by obligatorily removing an undamaged reproductive organ, and thus _____ enhancing its locomotor performance. The spider genus Tidarren (Araneae, Theridiidae) is interesting because, within several species presenting extreme sexual size dimorphism (males representing #1% of the total mass of the female), males voluntarily remove one of their two disproportionately large pedipalps (modified copulatory organs; a single one	It MODAL: _____ possible to develop suntan lotions that kick-start the skin's protective mechanisms against cancer before you hit the beach. The key ingredient MODAL: _____be a fragment of DNA just two bases long, called a TT dimer, that mimics one of the signs of DNA damage from ultraviolet light. Barbara Gilchrest's team from Boston University and colleagues in the Netherlands exposed hairless mice to a mild ultraviolet radiation, the equivalent of half an hour of afternoon sun. They found that genes involved in DNA repair were extraordinarily extremely fully

POS tag	absolute #	relative #	POS tag	absolute #	relative #
#	117	8%	CJC	137	10%
# + VB	16	1%	CJS	12	1%
# + VV	2	0%	CJS + #	1	0%
/	221	16%	CJS + VV	1	0%
AJ0	38	3%	CJT	32	2%
AJ0 + NN + AT0	1	0%	DT0	24	2%
AJC	74	5%	DT0 + VB	2	0%
AJC + #	1	0%	NN	34	2%
AJS	9	1%	NN + #	3	0%
AJS + #	5	0%	PNQ + VB + AJC	1	0%
AT0	7	1%	VB	33	2%
AT0 + AJ0	1	0%	VB + #	4	0%
AV	156	11%	VM	240	17%
AV + #	1	0%	VM + XX0	2	0%
AV0	114	8%	VV	99	7%
AV0 + AV0	2	0%	VV + AT0	1	0%
AV0 + VB	1	0%	XX0	3	0%
AVQ + #	1	0%	XX0 + CJS + #	1	0%

Absolute and relative distribution of parts of speech (POS) for Czech results in test A

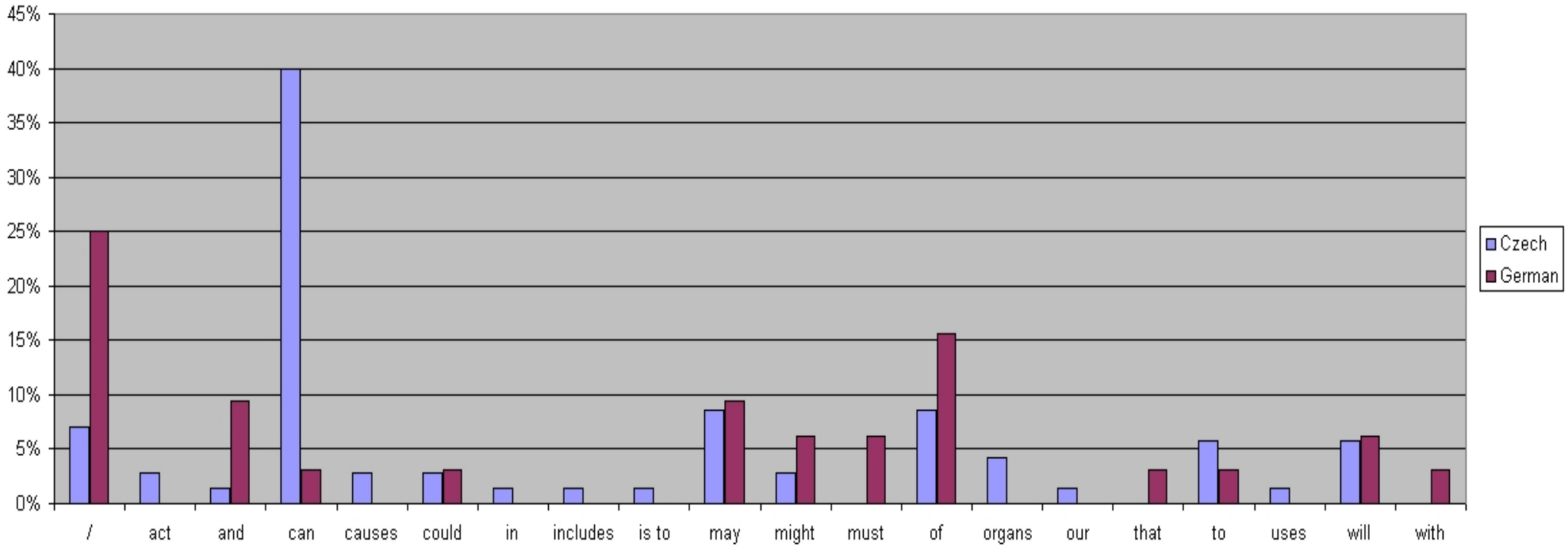
3.8 Distribution of lexical items by Czechs and Germans in test A, item 3

gap item supplied	Czech, *n* = 70	German, *n* = 32
/	7%	25%
act	3%	0
and	1%	9%
can	40%	3%
causes	3%	0%
could	3%	3%
in	1%	0%
may	9%	9%
might	3%	6%
must	0	6%
of	9%	16%
organs	4%	0%
that	0	3%
to	6%	3%
uses	1%	0%
will	6%	6%

3.9 Test A, gap 3

Overcoming an evolutionary conflict: Removal of a reproductive organ _____ increases locomotor performance

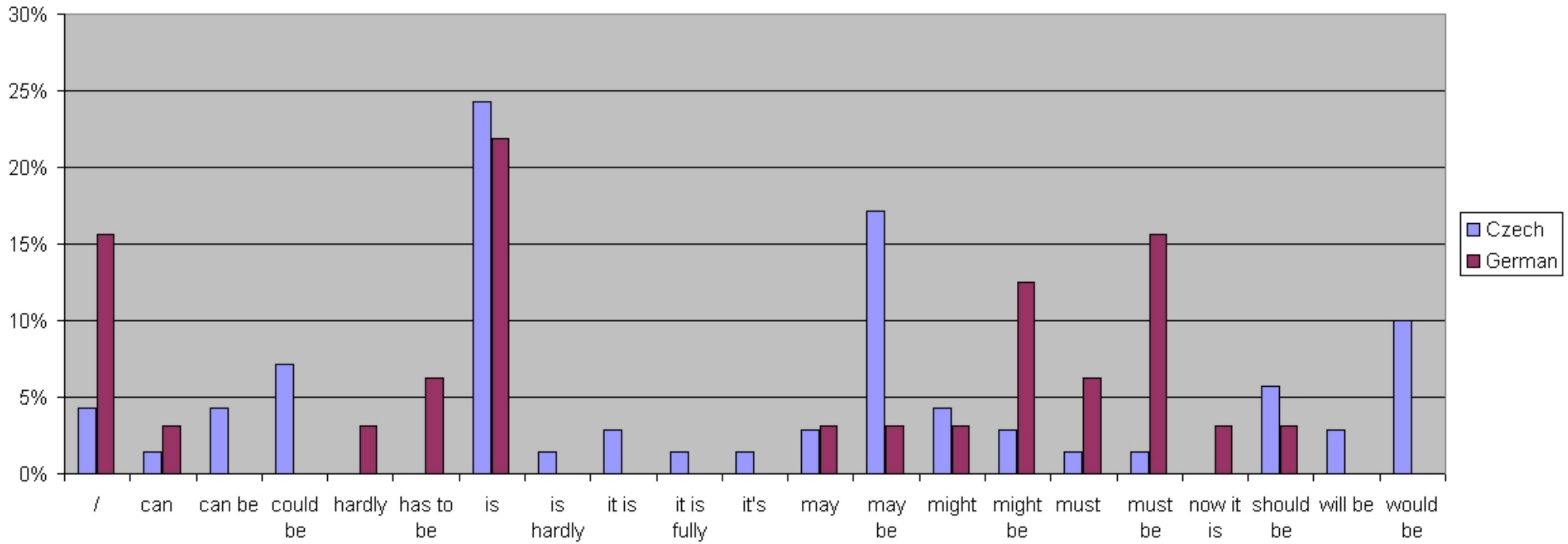
….dimorphism by obligatorily removing an undamaged reproductive organ,
and thus _____ enhancing its locomotor performance



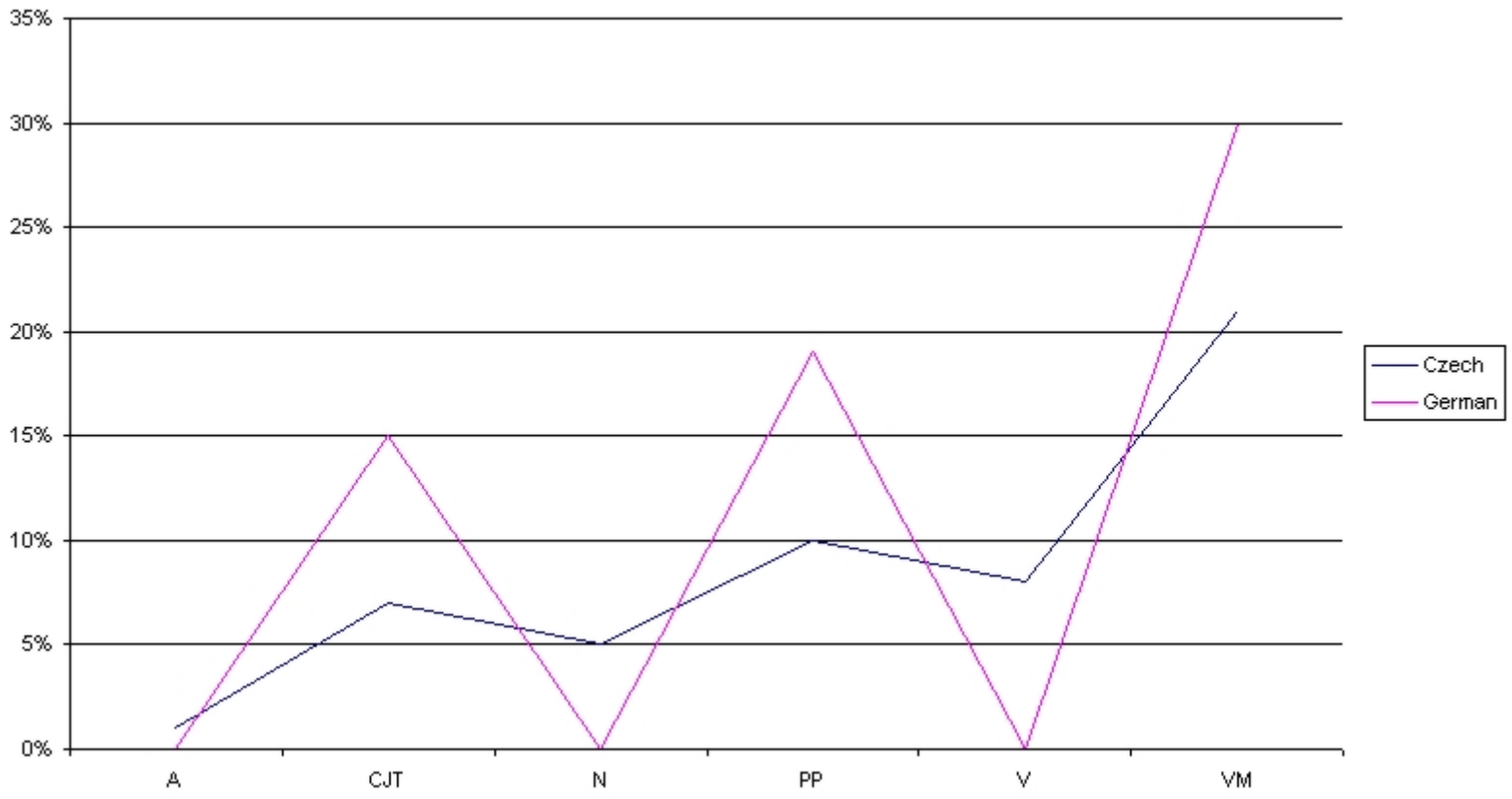
3.10 Test B, gap 1

Suntan lotion primes the skin's defences

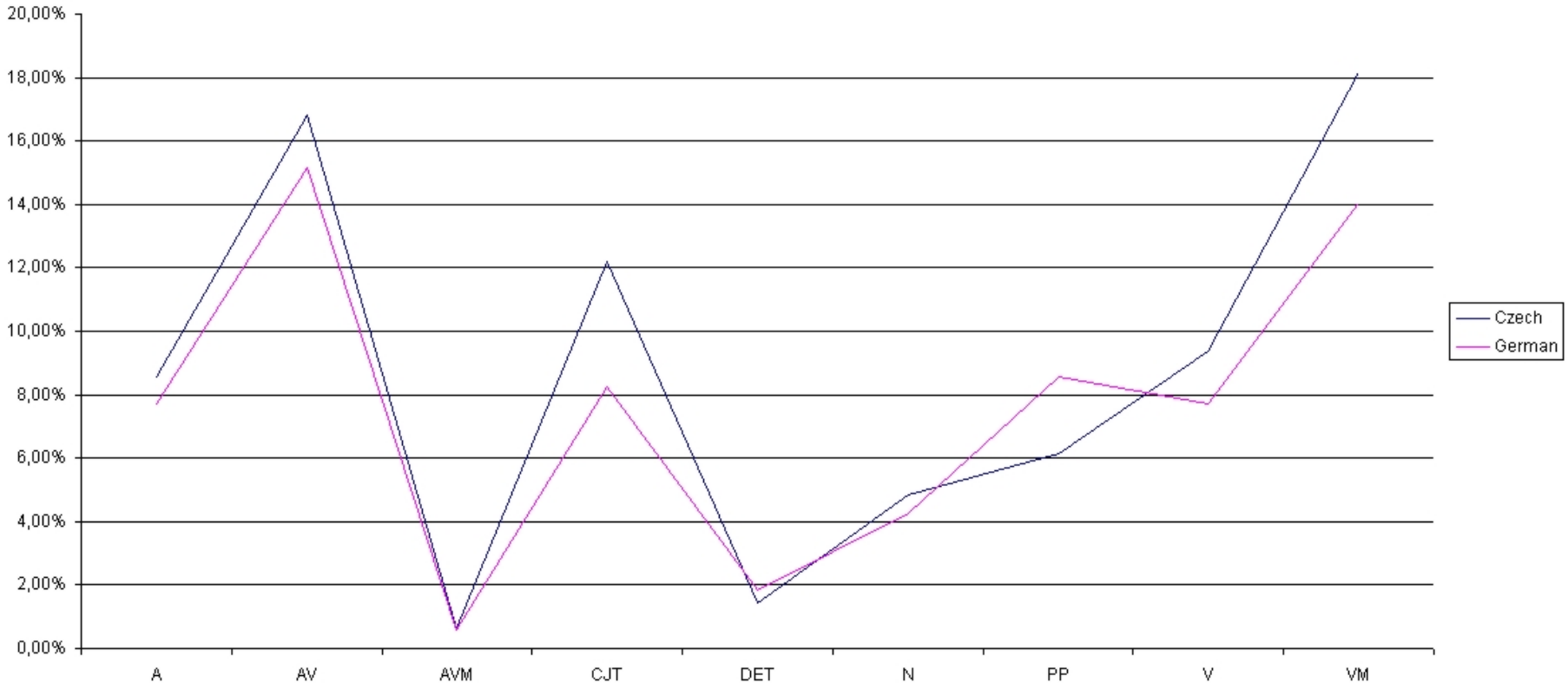
It **MODAL**_____ possible to develop suntan lotions that kick-start….



3.11 Test A, gap 3, POS distribution



3.12 Test A, Overall POS distribution



3.13 Findings

Czech learners

- are lexically more diverse
- use more modal hedging
- use more clause linking

German learners

- are more idiomatic
- use less modal hedging (exception: must/must not)

use more periphrastics (PPs etc.)

4. Lexical Gradience

subcorpus	frequency list (20)
arXiv 0001AX-0046AX	mass, energy, time, number, quantum, length, hole, stars, case, data, scale, density, state, probability, terms, model, order, code, field, value
New Scientist-physics 0001NS-0046NS	quantum, universe, energy, theory, time, space, light, matter, gravity, particles, physicists, years, Earth, holes, idea, issue, page, stars, physics, magazine
PNAS 0047PN-0107PN	cells, cell, data, DNA, gene, species, table, rate, time, analysis, results, control, stress, number, group, levels, expression, effects, sequences, mice
New Scientist-biosciences 0047NS-0107NS	cells, genes, team, years, researchers, fields, species, field, farmers, water, DNA, gene, people, cell, human, primates, work, way, core, animals

4.1 Vagueness vs. Specialization – Subjectivity vs. Objectivity?

High specialization leads to use of very few generic terms
= ontologically "deep" lexical items (**WordNet ontology**)

	academic text **0007AX**	**popular academic text** **0007NS**
markers of specialization	*conjectures, compactification, coalescence, planetesimals, angular, mesoscopic, gauge field, accretion, radial drag*	*dead stars, cloud of gas, hot star, proto-planetary disc, rogue comets*
markers of vagueness	*suggest X may have, should detect Rc, deviations are weak, may be turbulent*	*it may be hard, can be slow, they probably rebound, could charge up*

4.2 Semantic complexity analyzer statistics

	AX	NS01-46	PN	NS47-107	NS all	AX+PN	all
semantic parameters							
mean complexity	23.61	19.11	26.28	19.79	19.50	25.06	22.37
SD complexity	2.12	1.26	1.28	1.24	1.25	2.17	3.31
Max. degree of sem. spec.	13.69	12.78	15.53	13.84	13.38	14.69	14.06
Degree of sem. spec.	8.08	8.09	8.19	8.26	8.19	8.14	8.16
syntactic parameters							
SD unknown	7.49	3.85	5.38	3.34	3.59	6.87	11.92
Max. noun length	24.17	22.83	39.58	17.38	19.72	32.55	26.34
Mean noun length	6.54	6.89	6.82	6.78	6.83	6.69	6.76
number of commas	191	42	463	28	34	339	191
Max. commas/sentence	11	4	23	3	3	18	11
lexical parameters							
mean length	3113	812	4359	500	634	3790	2262
mean vocab. unknown	30.05	12.03	34.59	10.94	11.41	32.52	22.30

4.3 Mean Complexity (Complexana Scores)

Type/Domain	Physics/Astrophysics	Biosciences	Psychology
Specialized academic	23.61	26.28	22.37
Popular academic	19.11	19.79	19.36

all popular versions show ontologically “shallower” semantic items

5. Source/Target domains

5.1 Metaphors in physics and psychology

arXiv:astro-ph/0201342 v1 21 Jan 2002

Gamma Ray Bursts from the First Stars: Neutrino Signals
Raffaella Schneider, Dafne Guetta, Andrea Ferrara
If the first (PopIII) stars were very massive, their final **fate** is to collapse into very massive black holes. Once a proto-black hole has formed into the stellar core, accretion continues through a disk. It is widely accepted, although not confirmed, that magnetic fields **drive** an energetic jet which produces a burst of TeV neutrinos by photon-meson interaction, and eventually **breaks out** of the stellar **envelope** appearing as a Gamma Ray Burst (GRB). Based on recent numerical simulations and neutrino emission models, we **predict** the **expected** neutrino diffuse **flux** from these PopIII GRBs and compare it with the capabilities of present and planned detectors as AMANDA and IceCube. If **beamed** into 1% of the sky, we **find** that the rate of PopIII GRBs is = 4 × 106 yr-1. High energy neutrinos from PopIII GRBs could dominate the overall **flux** in two energy **bands** [104-105] GeV and [105 - 106] GeV of neutrino telescopes. The enhanced sensitivities of forthcoming detectors in the high-energy band (AMANDA-II, IceCube) will provide a fundamental **insight** on the characteristic explosion energies of PopIII
GRBs

Acta Psychologica
Volume 118, Issue 3 , March 2005, Pages 245-260

Save the last dance for me: unwanted serial position effects in jury evaluations
Wändi Bruine de Bruin,
Whenever competing options are considered in sequence, their evaluations may be affected by order of appearance. Such serial position effects would **threaten** the **fairness** of competitions using jury evaluations. Randomization cannot reduce potential order effects, but it does **give** candidates **an equal chance** of being assigned to preferred serial positions. Whether, or what, serial position **effects emerge** may depend on the cognitive demands of the judgment task. In end-of-sequence procedures, final scores are **not given** until all candidates have performed, possibly **burdening judges' memory**. If judges' evaluations are based on how well they remember performances, serial position effects may resemble those found with free recall. Candidates may also be evaluated **step-by-step**, immediately after each performance. This procedure should not **burden memory**, though it may produce different serial position effects. Yet, this paper reports similar serial position effects

→ both text types use metaphors to make argumentation transparent

- salience of visual metaphors in physics
- hedge marking in psychology

What are the functions of verbs of perception?

5.2 Manner classification: English as a manner language

Classic approach: Influential two-way typology by Talmy, 1985

- parameters of a motion event: FIGURE, MOTION, PATH, GROUND, MANNER, CAUSE

(1)	The boat	floated	under	the bridge
	FIGURE	MOTION+MANNER	PATH	GROUND

manner languages (English, German, Swedish, Russian)

- MANNER is typically encoded in verb, path info in nonverbal elements (PP)

path languages (Greek, Spanish, Japanese)

- verb encodes direction of MOTION, MANNER in gerunds, PPs or omitted

5.3 Manner classification: English as a manner language

Framing function of PATH parameter:

- verb-framed languages: conflate MOTION + PATH
- satellite-framed languages: conflate MOTION + MANNER

(2) a. to swim across the river
b. cruzar el río nadando

→ motion/space is an empirically rich domain (Papafragou, 2002)

although: manner languages have path verbs
path languages have manner verbs

but: the preferred lexicalization differs

5.4 Circular Motion in science texts

ac-burr - 199
If the orbit passes over the poles the Earth will spin under it every 24 h, so that any point on the surface will pass below the orbit every 12 h.

ac-burr - 470
The satellites have a cylindrical form. They are stabilised to spin about an axis parallel to the Earth's axis at a rate.

ac-burr - 211
If the orbit radius is chosen at a certain value, then the period of the orbit can be exactly 24 h which means that the Earth will rotate beneath at precisely the same angular velocity. So the satellite will remain above the same point over.

ac-davies - 294
To return the particle to its initial state it is necessary to rotate it through 720°.

tou-newc - 47
It swivels about its centre point on huge roller-bearings and is operated by hydraulic pressure. The High Level Bridge, which is further

→ Circular motion, cf. *spin/rotate/turn/twist/pivot/swivel...*

5.5 “Manners of perception”

Manner: emerges in preferred lexicalisation patterns

→ substantial manner of perception in the visual field expected

advantages:	manner typology is metaphorical expected to be different in the “two cultures”
assumptions:	Humanities are more logical because terminology-driven? Sciences are more visual because empirically driven? Can a specific metaphorical use in popular sciences be isolated?

5.6 Verbs of perception

visual		**auditory & olfactory**	**tactile/haptic**
admire	*observe*	*hear*	*feel*
discover	*peek*	*listen*	*sense*
focus	*peer*	*savor*	*touch*
gaze	*perceive*	*smell*	
glance	*recognize*	*sniff*	
glimpse	*see*	*taste*	
goggle	*spot*		
inspect	*squint*		
leer	*stare*		
look at	*watch*		
notice	*witness*		

→ all refer to default, cognitive standard situation

- all accept complements (objects, prepositional objects, clausal complements)

+HUMAN V_{percep} DO

5.7 Source and Target domains

„one of the reasons why different historical periods produce large numbers of items from similar conceptual metaphors is due to **the way we perceive our environment**. The very essence of **our forms of perception appears not only to be figurative in nature** but also inevitably so…" (Trim, 2007)

Source domain: conventional, bodily perception

Target domains: complements of $V_{percept}$ (direct objects, subjects of clausal complements)

(3) **0004AX** NNS mechanic . SENT. Here RB here we PP we focus VVP focus
a DT **a pilot-wave** NN pilot-wave **analogue**

Source: put something in the focus of attention
Target: abstract representation of a measured phenomenon

5.8 Verbs of perception in sciences and humanities

total numbers from POS-tagged corpus:

Verbs	Physics 001AX-046AX	BioSciences 047PN-106PN	NS Physics 001NS-046NS	NS BioSciences 047NS-106NS	Psychology
visual perception					
discover	20	17	7	15	12
focus	21	21	2	4	3
glance	0	0	3	0	2
inspect	0	1	0	0	7
look at	3	0	0	0	2
notice	13	4	10	3	16
observe	123	182	18	1	246
peer	0	1	1	2	2
perceive	2	25	1	1	101
recognize	4	23	0	3	8
see	305	264	74	28	219
spot	0	1	13	4	30
stare	0	0	1	0	11
watch	0	0	1	0	16
auditory perception					
hear	0	6	2	3	22
listen	0	0	1	0	51

5.9 Corpus Examples

V_{visual}	Example
discover	**0104PN** discussion We PP we have VHP have discovered VVN discover unprecedented JJ unprecedented variation NN varia
focus	**0004AX** NNS mechanic . SENT. Here RB here we PP we focus VVP focus a DT a pilot-wave NN pilot-wave analogue
inspect	**0088PN** be evaluated VVN evaluate by IN by inspecting VVG inspect probability NN probability plots NNS plot display
notice	**0028AX** We PP we notice VVP notice that IN that , , , in IN in spite NN spite of IN
observe	**0017AX** Mmin . SENT. We PP we observed VVD observe 10 CD @card@ events NNS event in IN in the DT the
peer	**0032NS** allowing VVG allow us PP us to TO to peer VV peer inside RB inside , he PP he speculates
perceive	**0027AX** low-energy probes NNS probe perceive VVP perceive as IN as Minkowski NP (flat JJ fla
recognize	**0004AX** evaporate if IN if one PP one recognises VVZ recognize that IN that our PP$ our universe NN universe is
see	**0047PN** density map NN map , , , we PP we see VVP see at IN at least JJS least four CD four to TO to fi
spot	**0008NS** should be VB be able JJ able to TO to spot VV spot the DT the terminuses NNS terminuses . SENT. A D
stare	**0004NS** Right now RB now we PP we 're VBP be staring VVG stare into IN into a DT a sort NN sort of IN of quantum

5.10 Totals Visual Perception

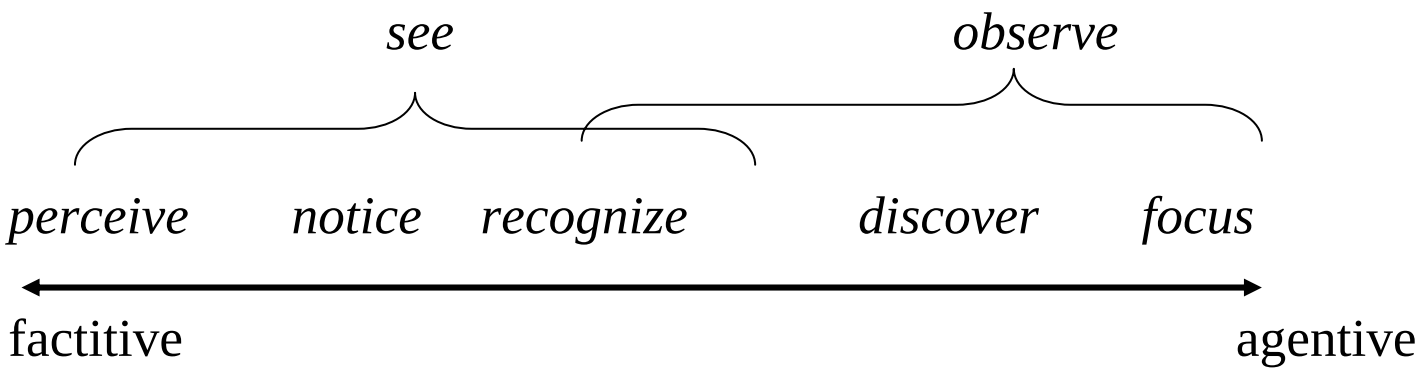
Subcorpus	words	V_{visual}	V_{visual} per 100k words
Physics	202.558	622	307,1
BioSciences	297.604	600	201,6
Psychology	217.254	676	288,5

→ physical sciences have highest relative share of $V_{percept}$

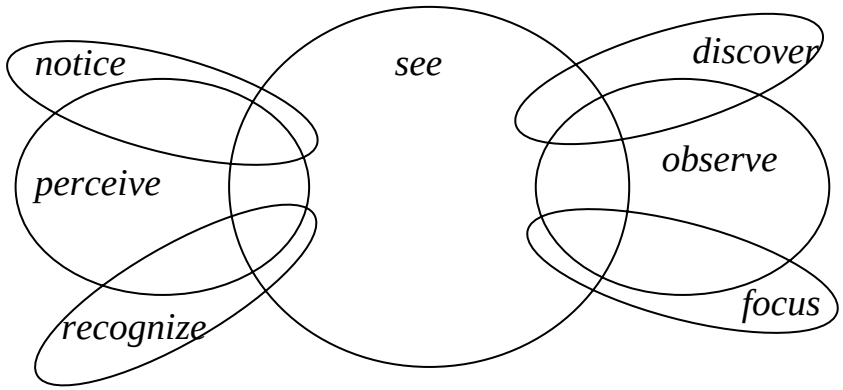
- low counts for auditory/olfactory $V_{percept}$
- low/zero counts for diversified manners of perception
- lexical variability highest for psychology
- marked difference for popular-academic physics

Subcorpus	V_{visual} per 100k words
Physics – academic texts	303,3
Physics – popular-academic texts	320,0
BioSciences – academic texts	201,8
BioSciences – popular-academic texts	200,0

5.11 Continuum of perceptual agency



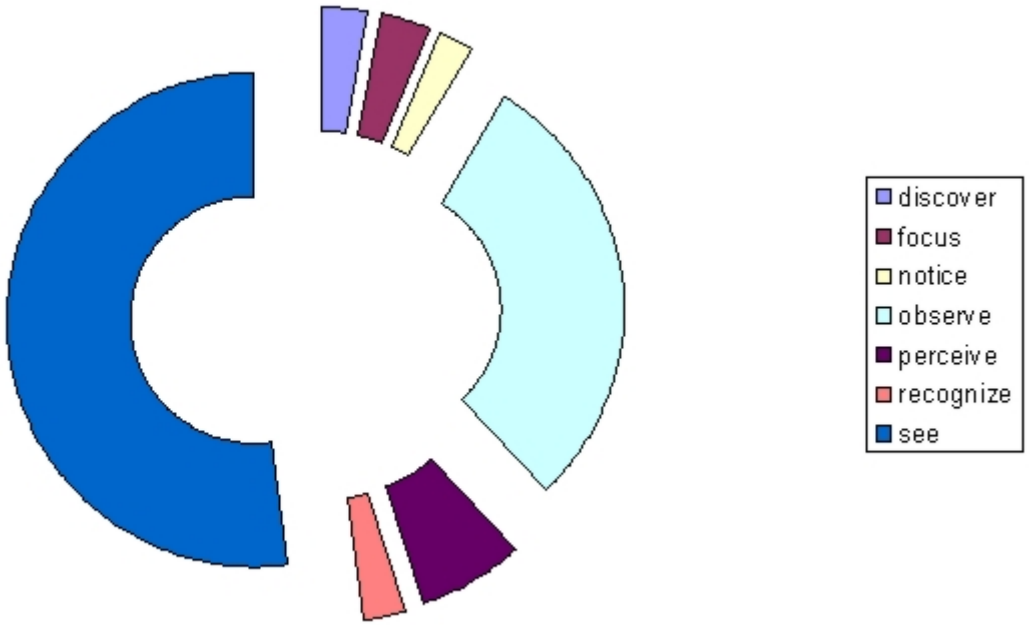
Groupings: top-down view

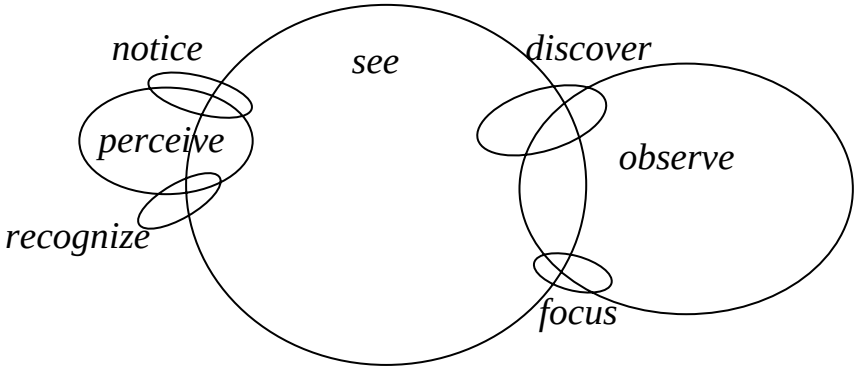


5.12 Subtotals for most frequent V_{visual}

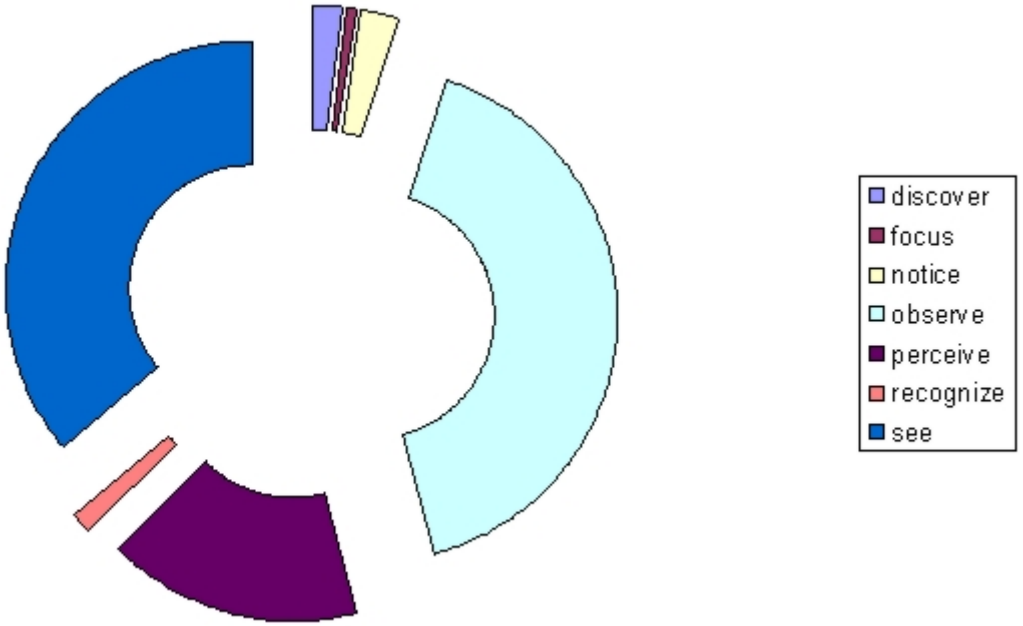
	Physics		subtotal	BioSciences		subtotal	Psychology	total
verb	**popular**	**academic**		**popular**	**academic**			
discover	7	20	27	15	7	22	12	61
focus	2	21	23	4	21	25	3	51
notice	10	13	23	3	4	7	16	46
observe	18	128	146	1	128	129	246	521
perceive	1	2	3	1	25	26	101	130
recognize	0	4	4	3	23	26	8	38
see	74	305	379	28	264	292	219	890

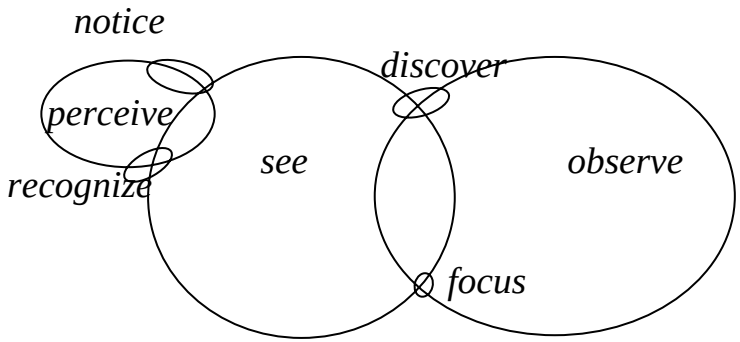
5.13 Groupings: bottom-up view; total



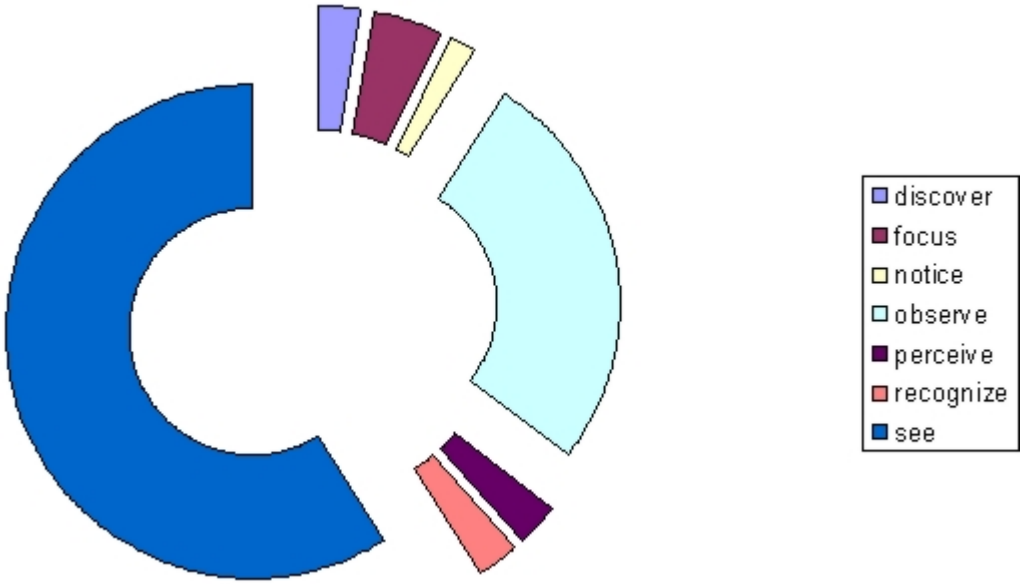


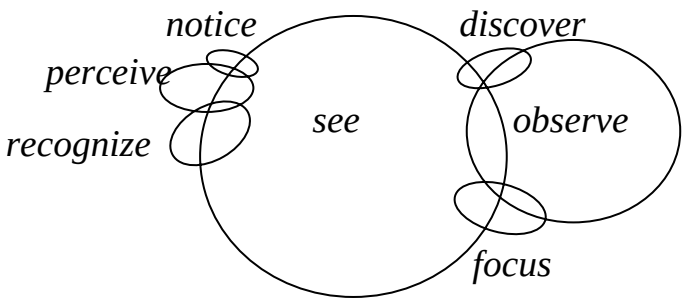
5.14 Groupings: bottom-up view; psychology



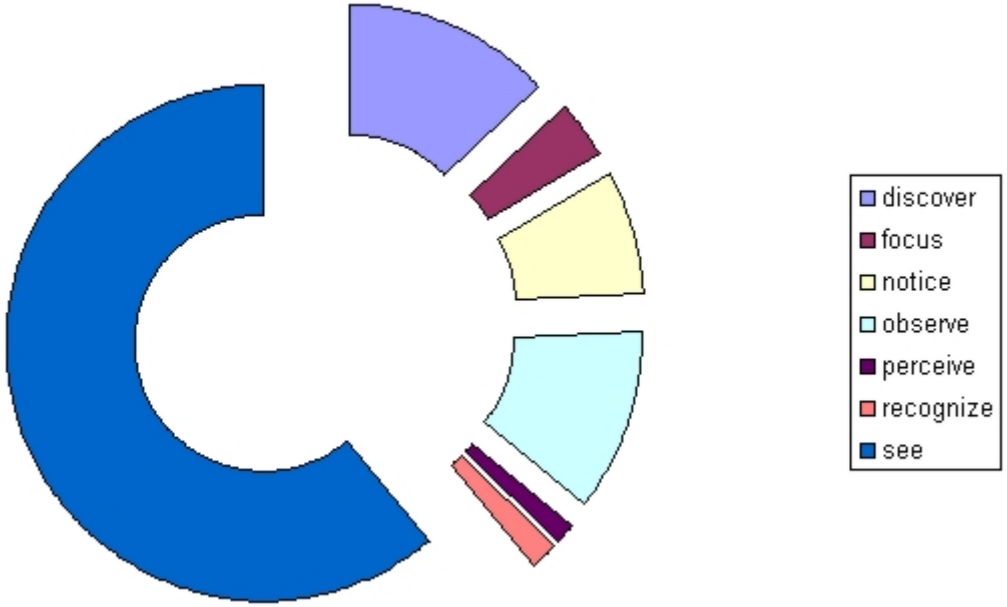


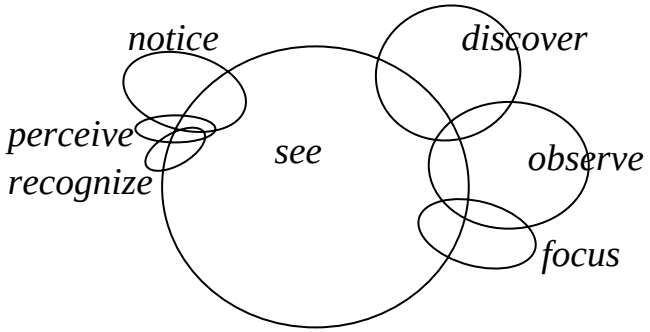
5.15 Groupings: bottom-up view; sciences (academic physics & biosciences)





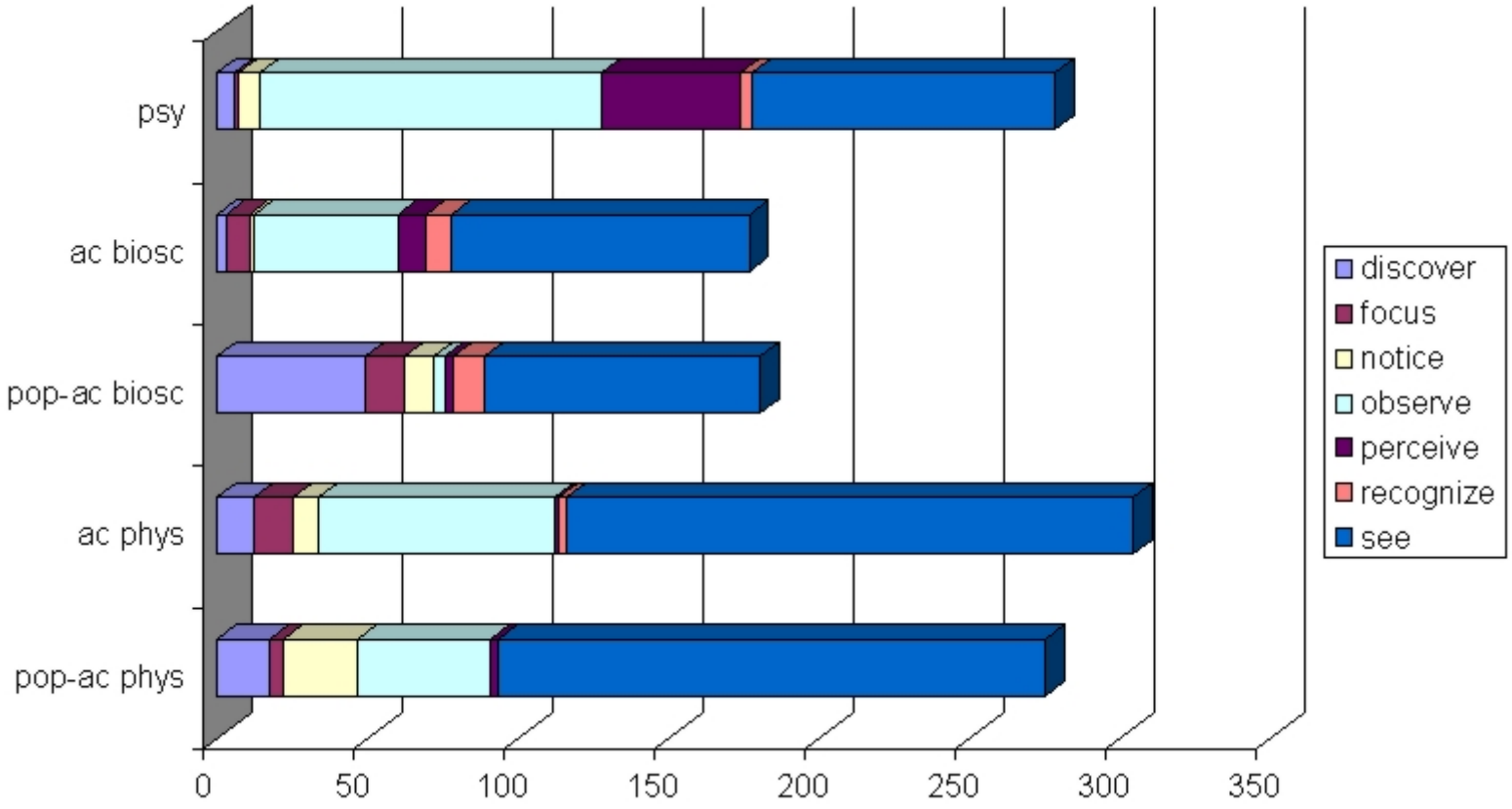
5.16 Groupings: bottom-up view; sciences (popular physics & biosciences)





5.17 Findings

Signatures for all subcorpora (distribution per 100.000 words)



5.18 Findings

Sciences and psychology show distinct signatures

psychology poles: **perceive – see – observe**

academic sciences poles: **see – observe**

popular science poles: **see – agentive V_{visual}**

→ popular science signatures reflect their origins in the academic sciences
→ popular science discourse: the “true” mediator?

5.19 Perceptual fields/domains for V_{visual}

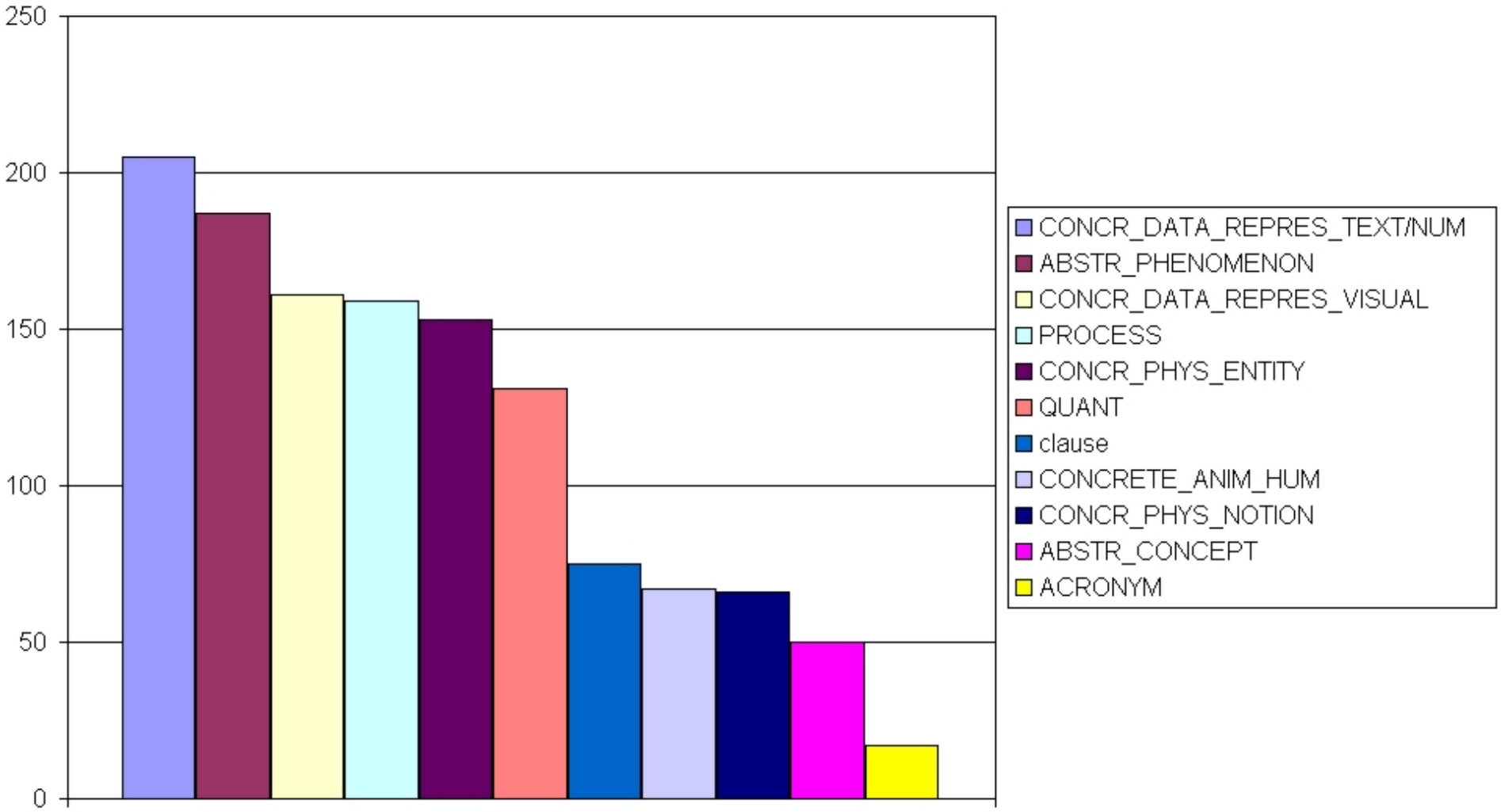
rank	freq	complement	rank	freq	complement
1	75	*that*-clause	21	8	*signal*
2	71	*figure*	22	8	*whether*-clause
3	46	*reference*	23	7	*discussion*
4	27	*effect*	24	7	*hole*
5	27	*table*	25	7	*level*
6	17	*behaviour*	26	7	*pattern*
7	16	*commentary*	27	7	*rate*
8	14	*text*	28	6	*event*
9	13	*method*	29	6	*information*
10	13	*result*	30	6	*number*
11	12	*section*	31	6	*particle*
12	11	*change*	32	5	*amount*
13	11	*expression*	33	5	*graphic*
14	10	*appendix*	34	5	*increase*
15	10	*difference*	35	5	*region*
16	10	*star*	36	5	*variation*
17	9	*material*	37	4	*cluster*
18	8	*cell*	38	4	*courtship*
19	8	*diagram*	39	4	*detail*
20	8	*equation*	40	4	*distribution*

5.20 Target domain ontology for V_{visual}

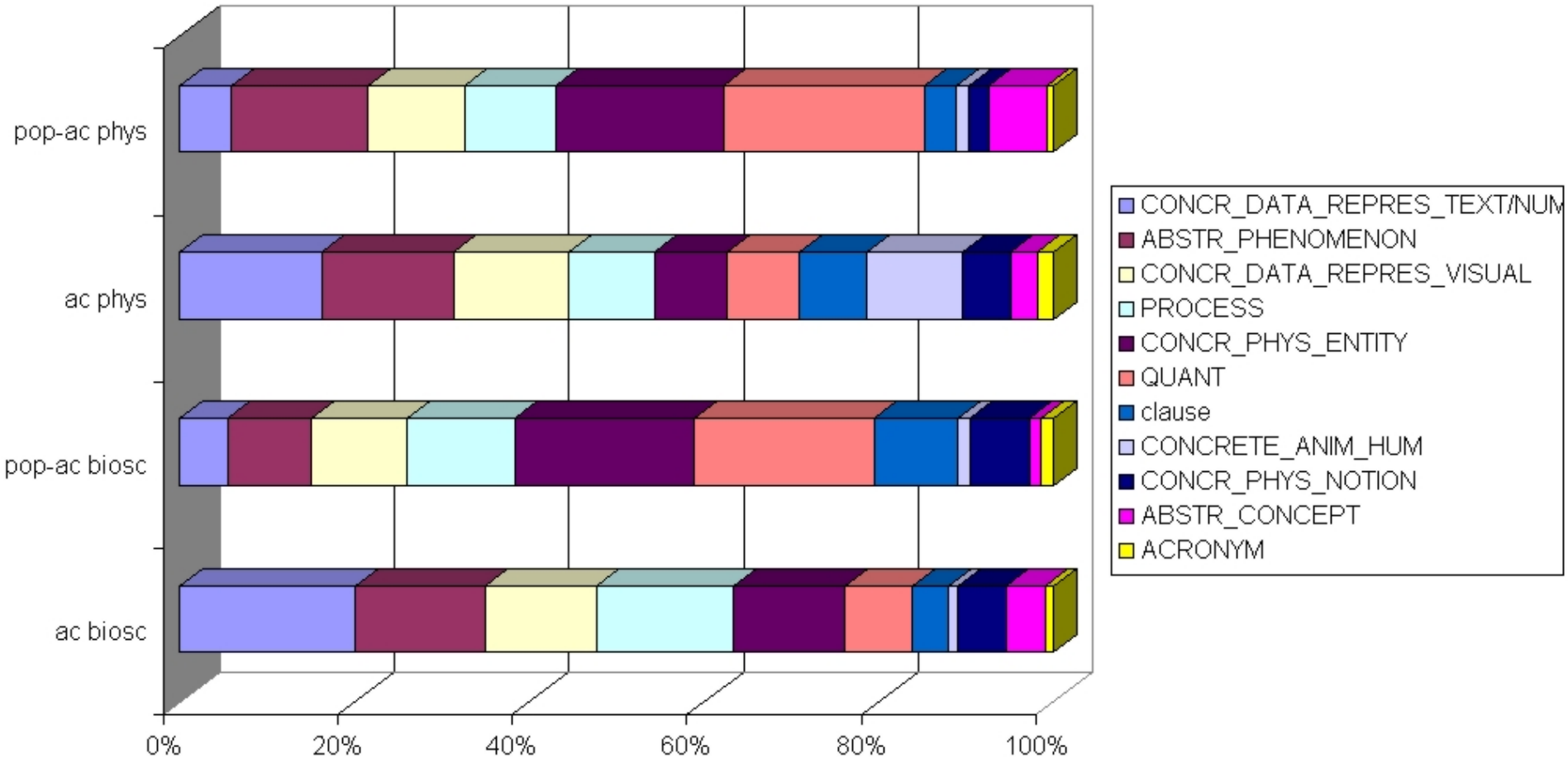
rank	freq	complement	target domains
55	4	*plaque*	CONCR_PHYS_ENTITY
56	4	*structure*	ABSTR_CONCEPT
57	4	*term*	CONCR_DATA_REPRES_TEXT/NUM
58	4	*theory*	CONCR_DATA_REPRES_TEXT/NUM
59	4	*Valentini*	CONCRETE_ANIM_HUM
60	3	*acceleration*	CONCR_PHYS_NOTION
61	3	*affect*	ABSTR_PHENOMENON

target domains	ac biosc	pop-ac biosc	ac phys	pop-ac phys
ABSTR_CONCEPT	24	1	16	9
ABSTR_PHENOMENON	79	7	80	21
ACRONYM	5	1	10	1
clause	22	7	41	5
CONCR_DATA_REPRES_TEXT/NUM	107	4	86	8
CONCR_DATA_REPRES_VISUAL	68	8	70	15
CONCR_PHYS_ENTITY	68	15	44	26
CONCR_PHYS_NOTION	29	5	29	3
CONCRETE_ANIM_HUM	6	1	58	2
PROCESS	84	9	52	14
QUANT	41	15	44	31

5.21 Domain distribution – total values



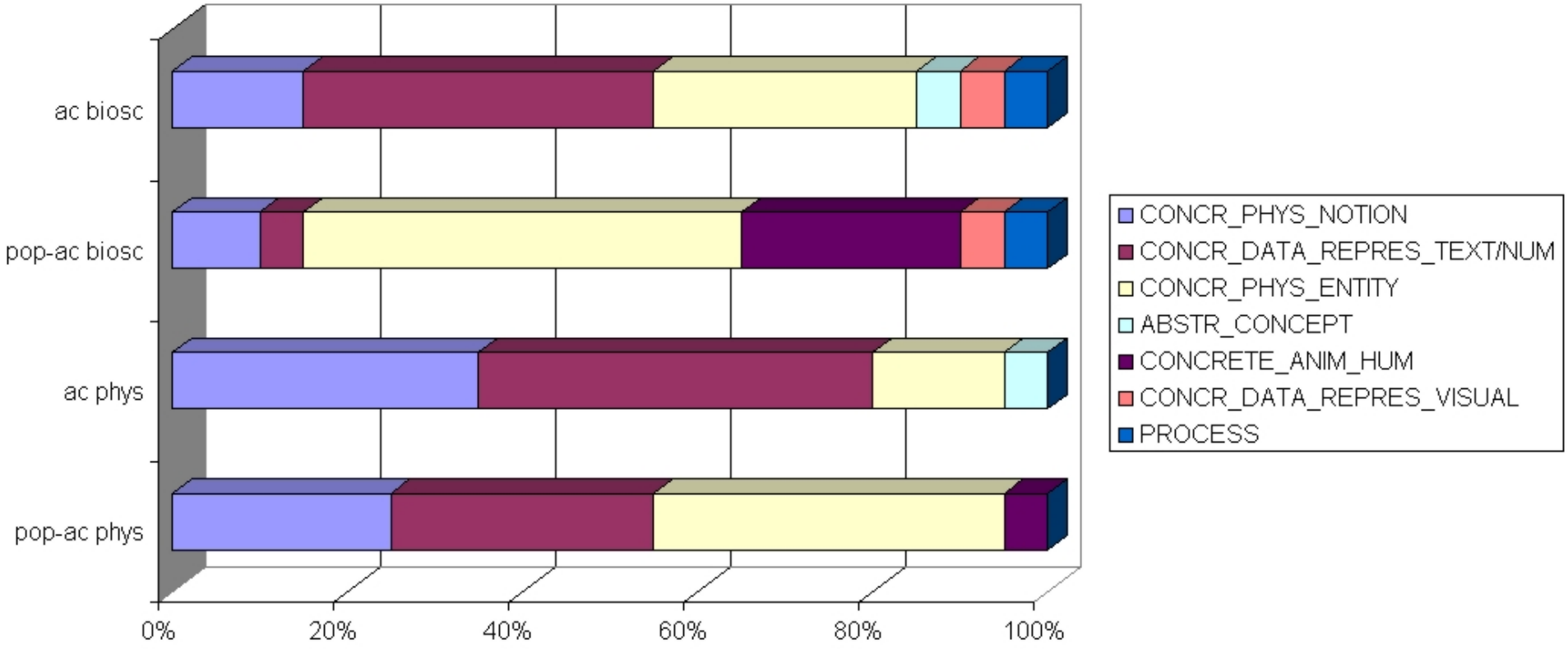
5.22 Domain distribution – subcorpora



5.23 Domains: Frequency view

subcorpus	frequency list (20)
arXiv 0001AX-0046AX	*mass, energy, time, number, quantum, length, hole, stars, case, data, scale, density, state, probability, terms, model, order, code, field, value*
New Scientist-physics 0001NS-0046NS	*quantum, universe, energy, theory, time, space, light, matter, gravity, particles, physicists, years, Earth, holes, idea, issue, page, stars, physics, magazine*
PNAS 0047PN-0107PN	*cells, cell, data, DNA, gene, species, table, rate, time, analysis, results, control, stress, number, group, levels, expression, effects, sequences, mice*
New Scientist-biosciences 0047NS-0107NS	*cells, genes, team, years, researchers, fields, species, field, farmers, water, DNA, gene, people, cell, human, primates, work, way, core, animals*

5.24 Domains: Frequency distribution

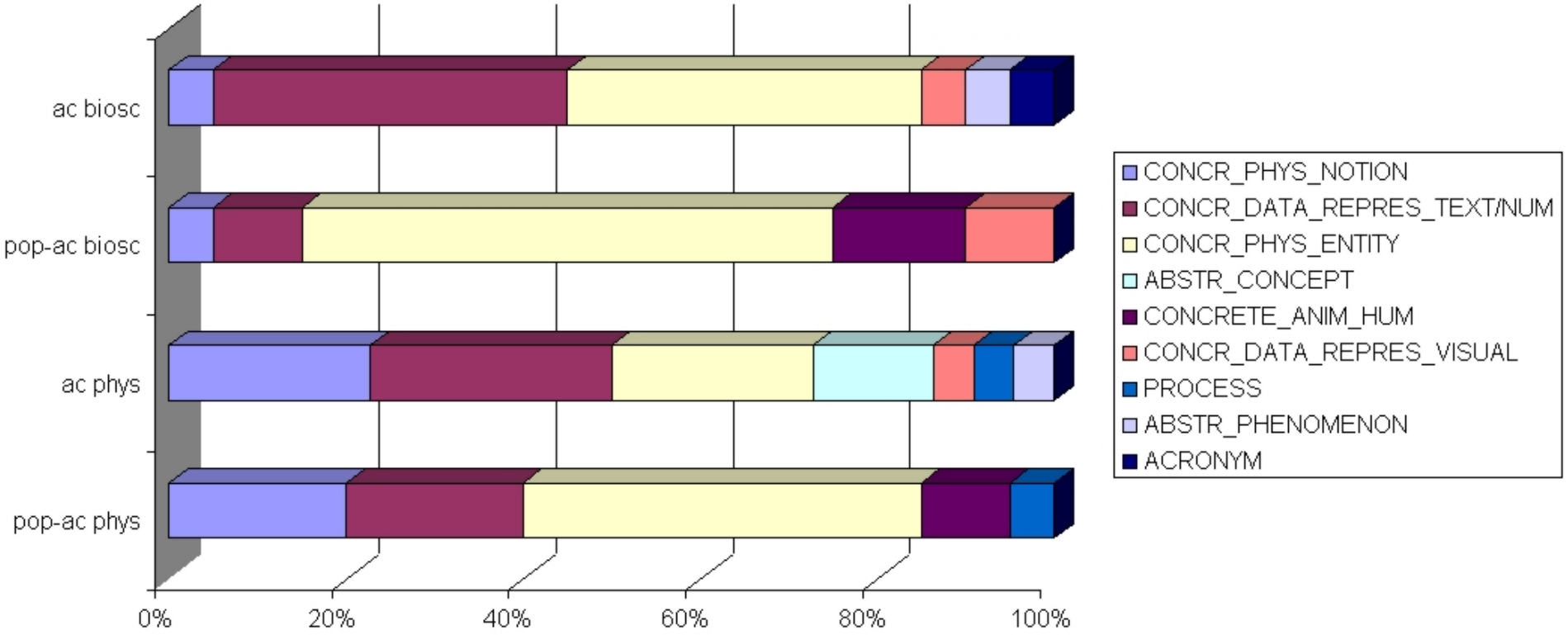


5.25 Domains: Keyness view

subcorpus	keyness list (20)
arXiv 0001AX-0046AX	energy, data, mass, quantum, Pioneer, scale, spacecraft, Earth, model, hole, density, distribution, state, number, theory, probability, constant, acceleration, value, length, results, spin
New Scientist-physics 0001NS-0046NS	quantum, Universe, energy, theory, University, gravity, particles, space, physicists, Earth, physics, relativity, hole, particle, matter, expansion, Einstein, mechanics, alpha, photons
PNAS 0047PN-0107PN	cells, data, DNA, gene, species, Table, analysis, males, www, rate, stress, sequences, levels, University, protein, studies, Science, values, effects, GFP
New Scientist-biosciences 0047NS-0107NS	cells, University, genes, issues, researchers, species, fields, farmers, DNA, gene, primates, page, cell, core, diatom, Beard, ions, shells, Asia, crops

- are V_{visual} relevant, i.e. do they deal with the salient concepts?

5.26 Domains: Keyness distribution



- more diversified
- V_{visual} not significantly used for entities that can be seen (low salience of concreteness)

→ V_{visual} mediate between abstract and concrete

6. Conclusion

- versatile determinants of the linguistic shape of commitments
- subjectivity can be lexicalised in different ways according to
  1. origin (academic or popular academic) and
  2. readership (scientists or educated laypersons)
  3. English as a first or second language

- higher semantic complexity of academic texts
- sensitivize students to the pragmatic and semantic markers of academic English
- investigate and quantify the results of their own academic work
- metaphor relevant in the visual perception
- manners of perception generally underused in academic prose
- usage of verbs of perception in psychology and sciences: differences in agency and factivity