

A corpus-based Comparison of Albanian and Italian Student Papers in L1 and L2: the Case of Hedges and Boosters

Vincenzo Dheskali

Fourth Semester PhD Student
Chemnitz University of Technology
vincenzo.dheskali@s2015.tu-chemnitz.de



Introduction

“[...] how writers mark their commitment to their propositions and indicate who is responsible for what claims is at the heart of skilful scientific writing.” (Hyland 1998: 79)

Methodology

- Corpus analysis
- Cross-cultural comparison with the concordance software AntConc (2014)
- Four corpora consisting of L1 and L2 writings by Albanian and Italian students
- Focus on hedges and boosters





Italian Corpus (CIAO) 	Italian English Corpus (CIAOE) 
88 Papers (2003-2015):	90 Papers (2003-2015):
- PhD Theses: 55 (2.170.146 words)	- PhD Theses: 57 (2.058.782 words)
- MA Theses: 29 (932.997 words)	- MA Theses: 26 (716.009 words)
- BA Theses: 1 (10.462 words)	- MA Term Papers: 3 (7.744 words)
- BA Term Papers: 3 (7.762 words)	- BA Theses: 4 (68.873 words)
- Males: 41 (1.481.790 words)	- Males: 45 (1.390.863 words)
- Females: 47 (1.639.577 words)	- Females: 45 (1.460.535 words)
Total No. of Word Tokens: 3.121.367	Total No. of Word Tokens: 2.851.408
Albanian Corpus (CAR) 	Albanian English Corpus (CARE) 
52 Papers (2010-2015):	41 Papers (2009-2015):
- PhD Theses: 52 (2.288.976 words)	- PhD Theses: 7 (307.419 words)
	- MA Theses: 12 (173.990 words)
	- MA Term Papers: 8 (22.214 words)
	- BA Theses: 13 (110.054 words)
	- BA Term Papers: 1 (2.620 words)
- Males: 26 (1.108.839 words)	- Males: 25 (404.492 words)
- Females: 26 (1.180.139 words)	- Females: 15 (211.805 words)
Total No. of Word Tokens: 2.288.976	Total No. of Word Tokens: 616.297

Table 1: Albanian and Italian corpora including respective word totals according to AntConc





Italian Corpus (CIAO) 	Italian English Corpus (CIAOE) 
<ul style="list-style-type: none"> - Lang. & Lit.: 40(1.435.515 words) - Social Studies: 8 (529.527 words) - Medicine: 9 (86.084 words) - Chemistry: 8 (247.867 words) - Physics: 7 (221.027 words) - Economics: 8 (395.069 words) - Math. & Inf.: 8 (206.278 words) 	<ul style="list-style-type: none"> - Lang. & Lit.: 41(1.261.708 words) - Social Studies: 8 (466.089 words) - Medicine: 9 (93.961 words) - Chemistry: 8 (204.916 words) - Physics: 8 (258.651 words) - Economics: 8 (272.188 words) - Math. & Inf.: 8 (293.885 words)
Total No. of Word Tokens: 3.121.367	Total No. of Word Tokens: 2.851.408
Albanian Corpus (CAR) 	Albanian English Corpus (CARE) 
<ul style="list-style-type: none"> - Lang. & Lit.: 8 (605.556 words) - Social Studies: 8 (483.872 words) - Medicine: 4 (103.037 words) - Chemistry: 8 (218.097 words) - Physics: 8 (334.607 words) - Economics: 8 (361.906 words) - Math. & Inf.: 8 (181.901 words) 	<ul style="list-style-type: none"> - Lang. & Lit.: 33(293.899 words) - Social Studies: 2 (105.104words) - Chemistry: 1 (69.513 words) - Physics: 1 (14.979 words) - Economics: 3 (97.952 words) - Informatics: 1 (34.850 words)
Total No. of Word Tokens: 2.288.976	Total No. of Word Tokens: 616.297

Table 2: Albanian and Italian corpora including all sections with the respective word totals

Modality (modalization): Hedges and boosters

Hedges and boosters are part of modality. Modality (modalization) builds an area of uncertainty. It is an intermediate point between positive polarity *it is* and negative polarity *it is not* which has various degrees of indeterminacy such as probability and usuality (cf. Halliday 1985; Halliday and Matthiessen 2014: 144-176).

Hedges such as *might* and *about* have the function of withholding author's full commitment towards the given information. Boosters such as *it is obvious that*, *in fact* and *definitely* have the function of emphasizing strength or author's sureness regarding the given information (cf. Hyland, 2005; Hyland 2017: 20).

Hedges and boosters: My working definition

Hedges (*possibly, almost, I think*) and **boosters** (*certainly, completely, demonstrate*) are numerous lexical and non-lexical items which express various degrees of authors' direct and indirect commitment regarding the probability and usuality of the expressed proposition.

They are modalization devices that interweave interpersonal and ideational socio-semiotic processes on a semantic level (approximators) pragmatic level (shields) and an interaction of both (shields and approximators).

They express different forms of manifestation, orientation, prosody of 'modality', as well as syntactic positioning, approximation, shields and polarity across and within different cultural and linguistic contexts of student academic writing.

(cf. Halliday 1985; Halliday and Matthiessen 2014; Lafuente Millàn 2008; Hyland 1998; 2005; 2017; Prince et al. 1980; Salager-Meyer 1994)

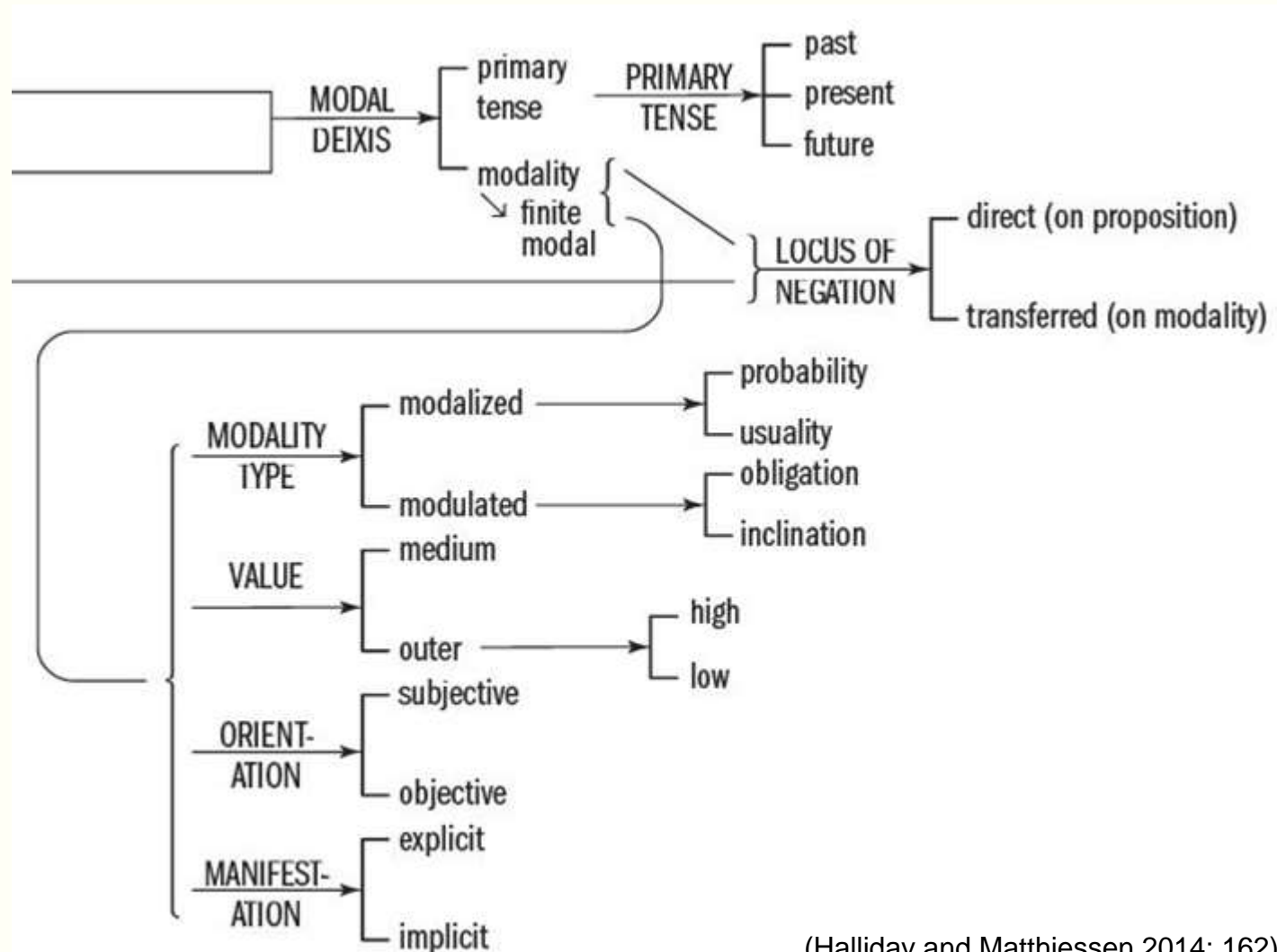
Hedging



Boosting



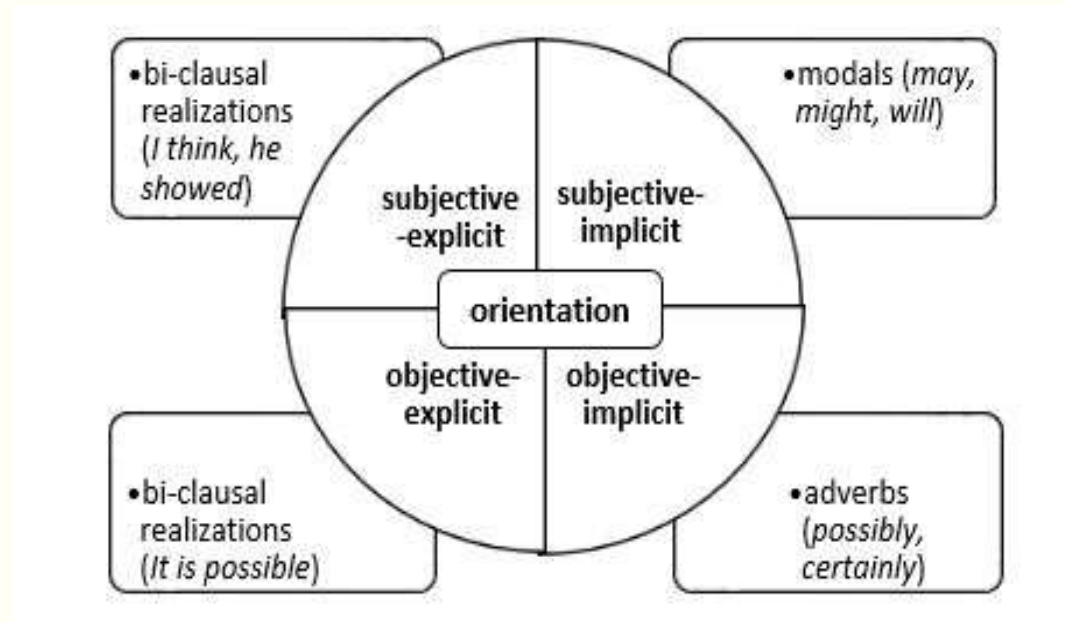
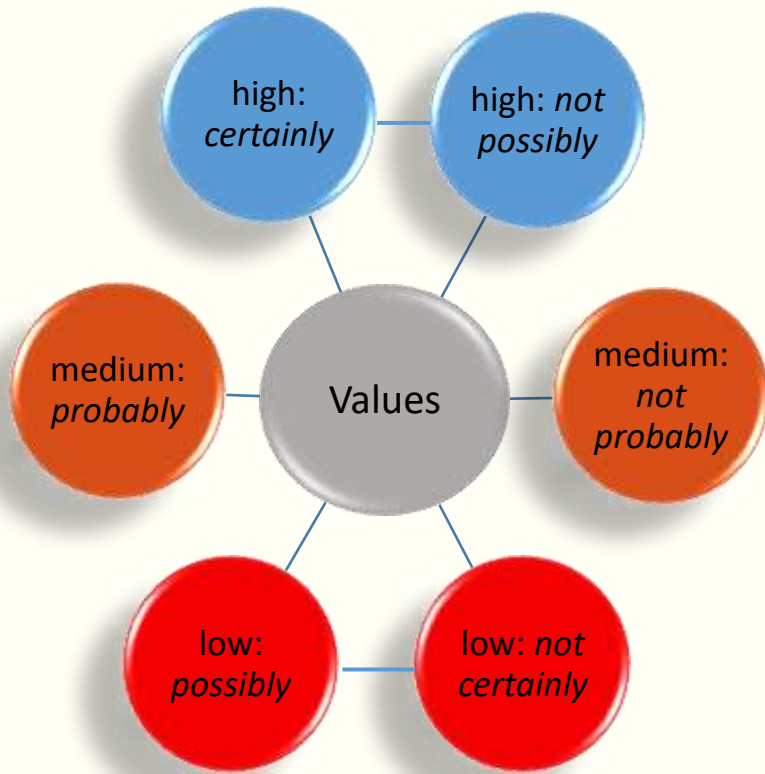
Modal deixis, modality and negation



(Halliday and Matthiessen 2014: 162)

Figure 1: The system of modality and the locus of negation

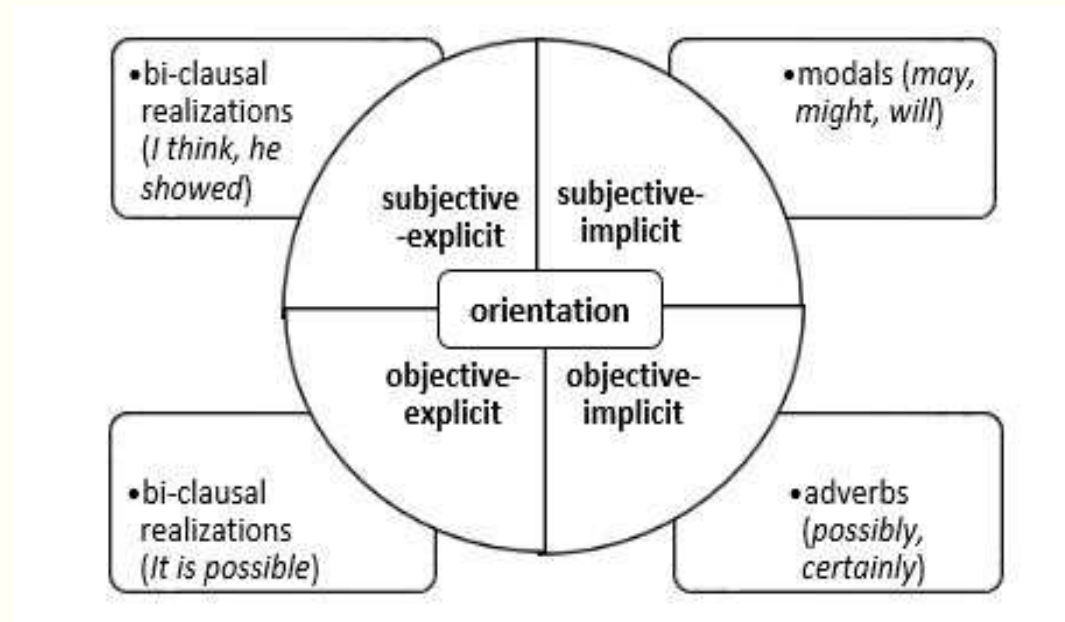
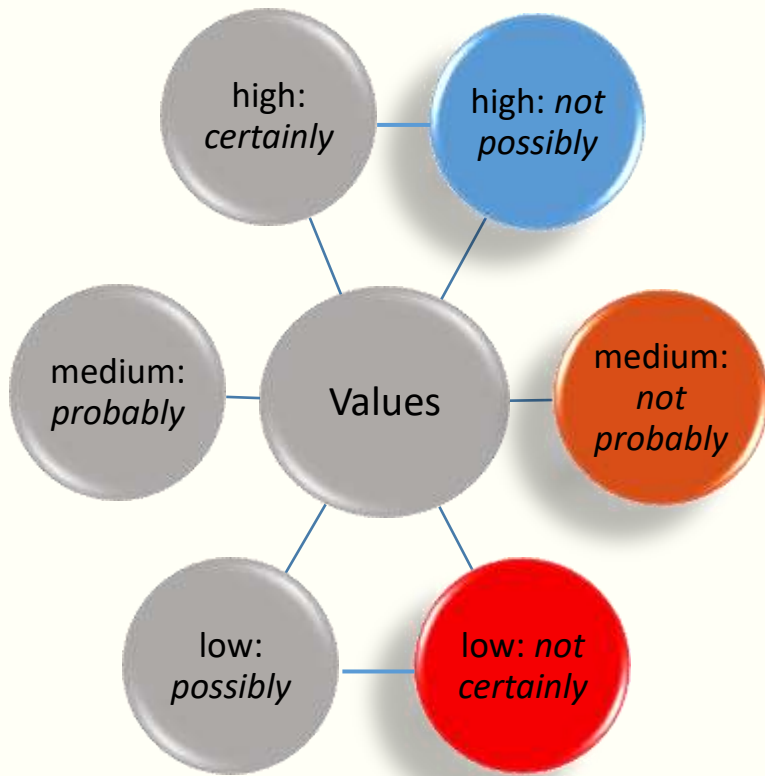
Modalization: Values, manifestation and orientation



(cf. Halliday and Matthiessen 2014)

Figures 2 and 3: My radial of values and cycle matrix of orientation and manifestation

Modalization: Values, manifestation and orientation



(cf. Halliday and Matthiessen 2014)

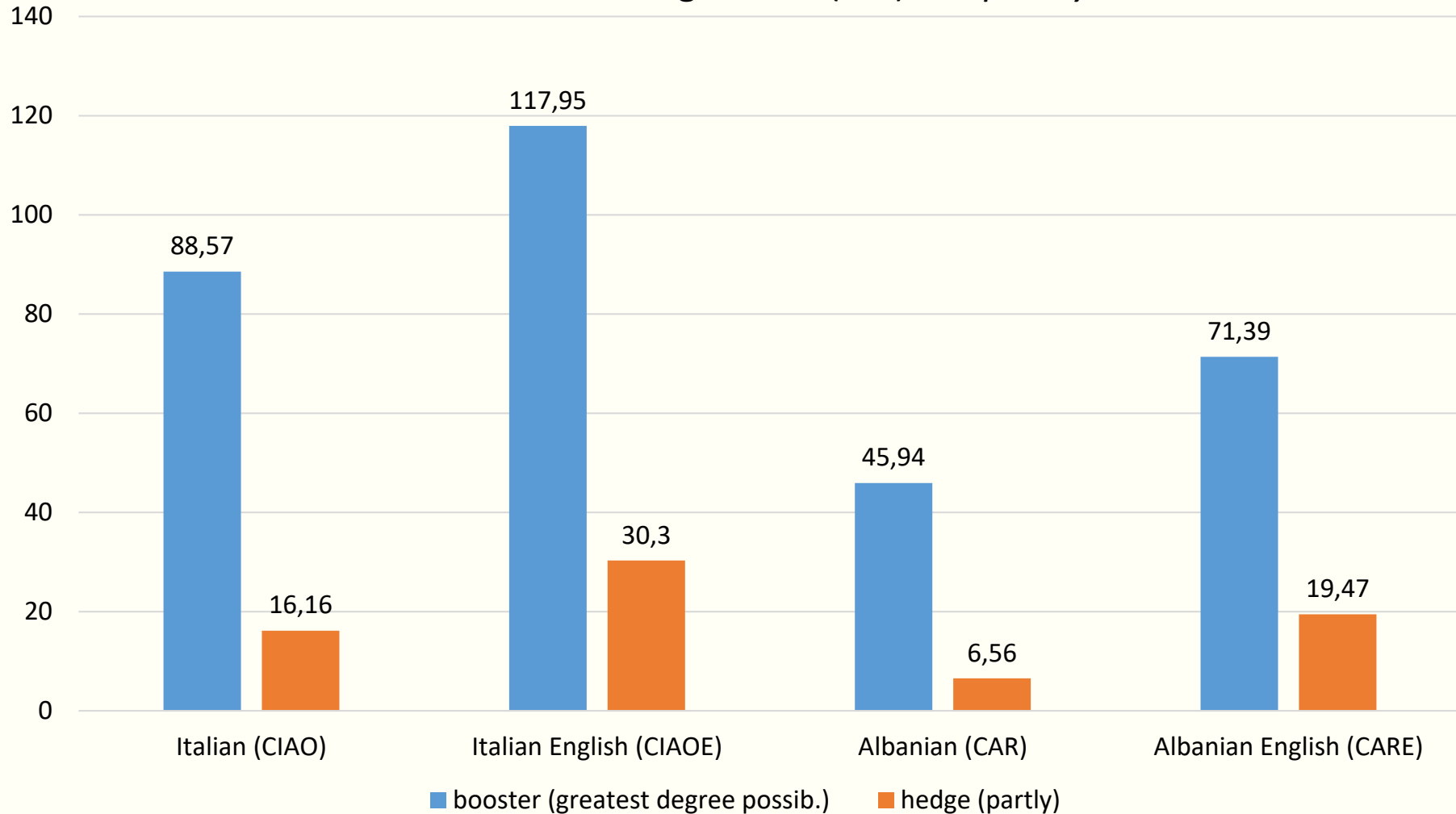
Figures 4 and 5: My radial of values and cycle matrix of orientation and manifestation

Difficulties in dividing equivalents in Albanian, English and Italian

Hedges and Boosters

English	Albanian	Italian
almost	thujajse, pothuaj, pothuajse	quasi
nearly	gati	quasi
approximately	afërsisht, përafërsisht, afersisht, perafersisht	approssimativamente, all'incirca, indicativamente
believ* (believed, I believe/we believe/ believing etc.)	beso* (besoja/besoje/besonin/besov a/do të besojë etc.)	cred* (credendo, credevo, crederanno/credette/ebbi creduto/avessero creduto etc.)
entirely/totally	tërësisht/plotësisht	Interamente/per intero/totalmente
significantly	ndjeshëm, ndjeshem, në mënyrë të ndjeshme	sensibilmente, in modo sensibile, in modo significativo, in maniera sensibile, in maniera significativa

Table 3: Equivalents of hedges and boosters in Albanian, English and Italian

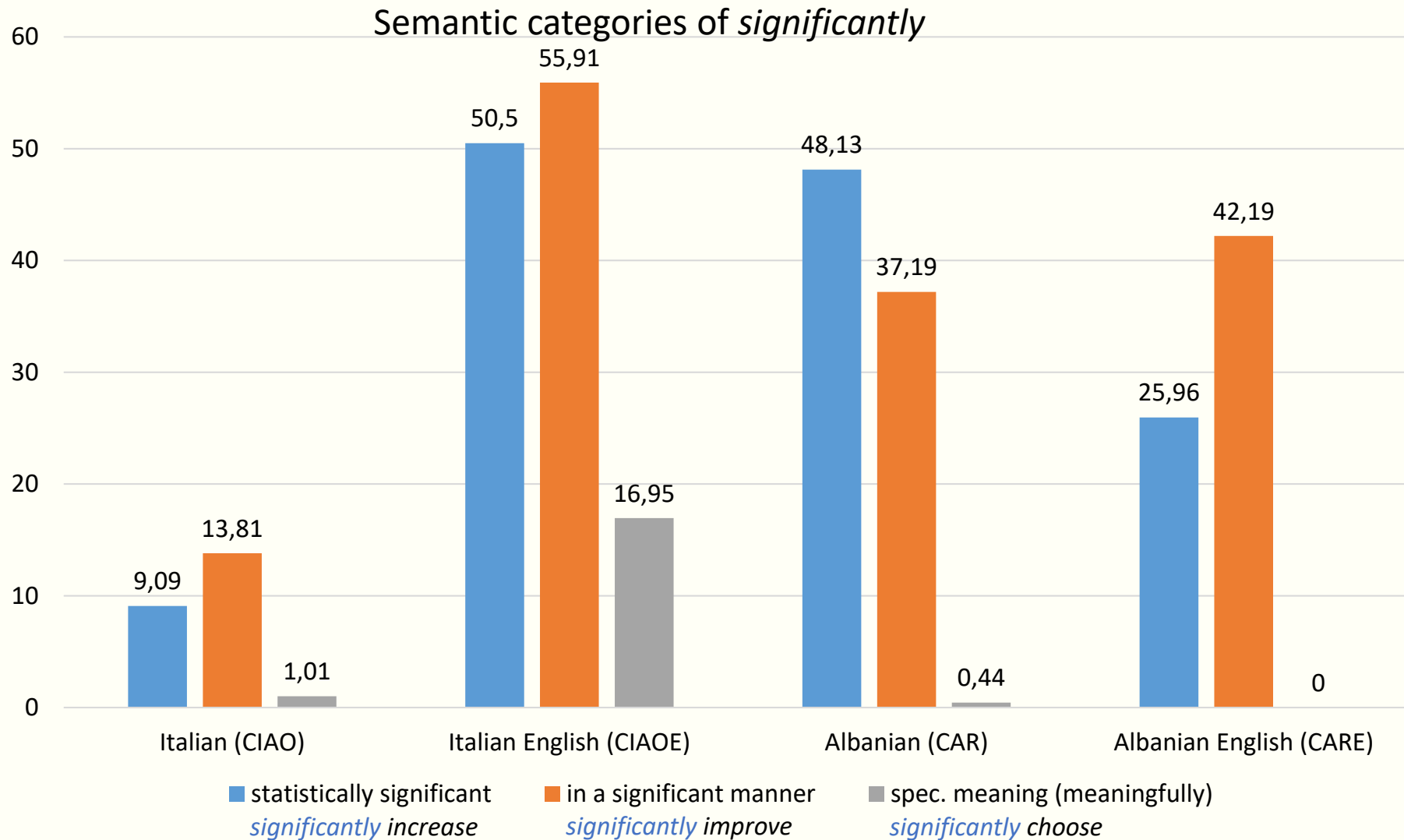
Case Study 1: *(not) completely*Modalization categories of *(not) completely*Figure 6: Frequency of *(not) completely* in all four corpora in the respective booster categories per 1.000.000 words

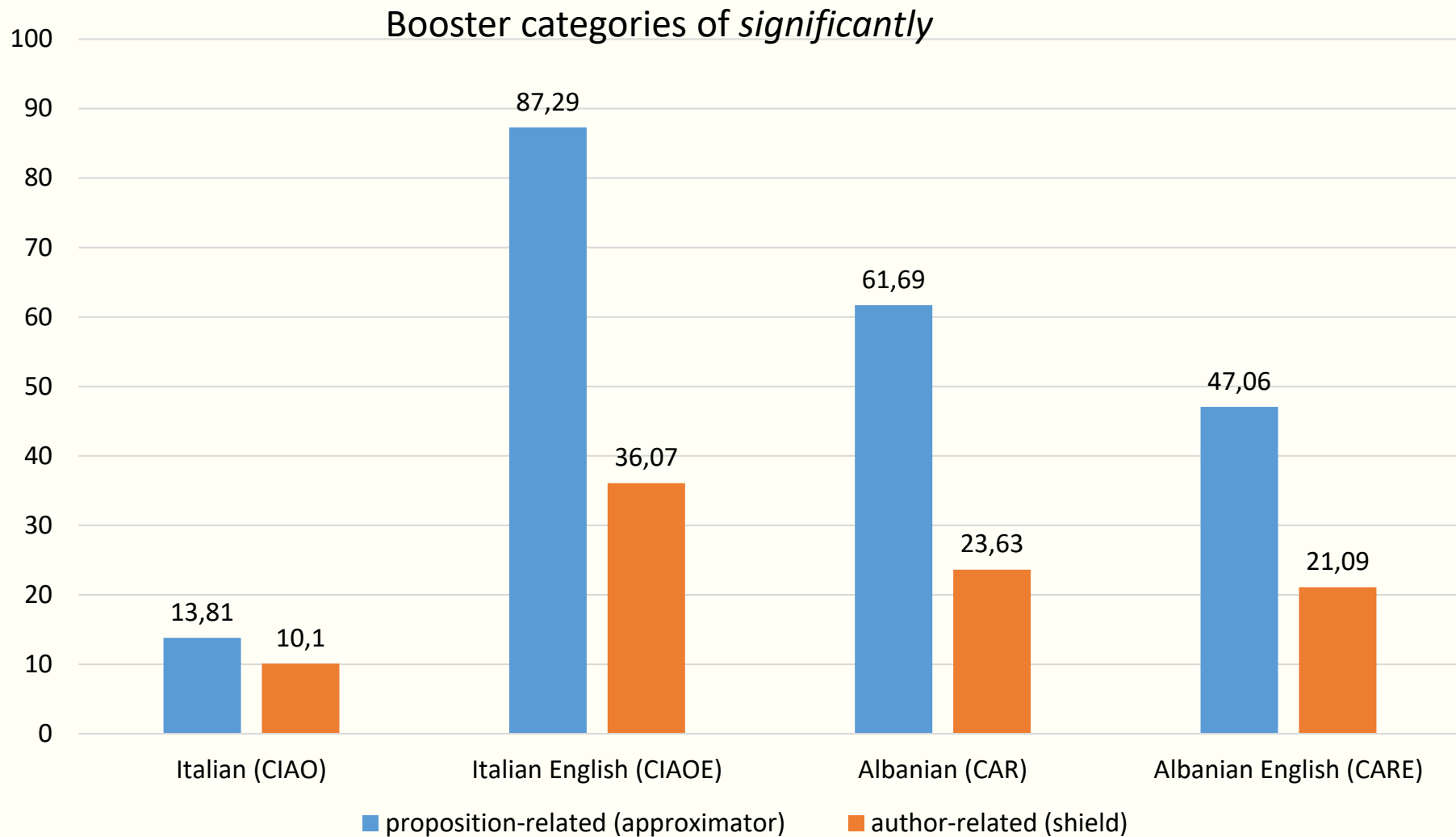
word	type	categorization	CIAO	CIAOE	CAR	CARE	
<i>almost, nearly, approximately</i>	rounder	frequency (e.g. <i>almost everyday</i>)	6.40	3.97	14	6,49	
		quantity (<i>approx. 20%</i>)	25.93	68.89	169.33	246.63	
	adaptor	degree (<i>nearly black</i>)	43.44	37.87	78.76	56.79	
		limitation (<i>almost cried</i>)	35.36	30.3	69.13	73.02	
	rounder			32.67	72.86	182.89	253.12
	adaptor			78.47	68.17	148.33	129.81
<i>probably</i>	semantic category	likely to (not) be true	96,65	106,05	56,44	81,13	
		likely to (not) happen	63,65	52,30	44,19	42,19	
<i>seem*</i>	semantic category	impression of state/characteristic	90.25	41.48	55.57	210.94	
		(cannot) seem to act	66.01	39.32	34.13	86	
		obvious/clearly seen	0	0	28.44	0	

Table 4: Frequencies of types of approximators and semantic categories of *seem** and *probably* per 1.000.000 words

word	type	categorization	CIAO	CIAOE	CAR	CARE	
<i>almost, nearly, approximately</i>	rounder	frequency	6.40	3.97	14	6,49	
		quantity	25.93	68.89	169.33	246.63	
	adaptor	degree	43.44	37.87	78.76	56.79	
		limitation	35.36	30.3	69.13	73.02	
	rounder			32.67	72.86	182.89	253.12
	adaptor			78.47	68.17	148.33	129.81
<i>probably</i>	semantic category	likely to (not) be true	96,65	106,05	56,44	81,13	
		likely to (not) happen	63,65	52,30	44,19	42,19	
<i>seem*</i>	semantic category	impression of state/characteristic	90.25	41.48	55.57	210.94	
		(cannot) seem to act	66.01	39.32	34.13	86	
		obvious/clearly seen	0	0	28.44	0	

Table 4: Frequencies of types of approximators and semantic categories of *seem** and *probably* per 1.000.000 words

Case Study 2: *significantly*Figure 7: Frequency of *significantly* in all four corpora in the respective semantic categories per 1.000.000 words

Case Study 2: *significantly*Figure 8: Frequency of *significantly* in all four corpora in the respective booster categories per 1.000.000 words

Examples from my corpora: *significantly, believe**

- (S1) [...] a hastily drawn customs tariff which *significantly* increased import duties making the cost of foreign goods on the Albanian market almost prohibitive (CARE11MPC_37). [B?]
- (S2) Familiarity is important in detecting communities, which may help improve *significantly* the design and performance of forwarding protocols in mobile environments (CARE10MPS_40).
- (S3) I had *made him believe* that I had taken his word literally (CIAOE13MML_80).
- (S4) *Arsimi i lartë i një personi na bën të besojmë* se ai person di të dallojë të drejtën nga e gabuara [...] (CAR15FPE_23). *The high education of a person makes us believe that that person knows how do distinguish between right and wrong [...]* (my translation).

 indicates the span of the hedge in S1, S2

Case Study 3: *believ**

Hedge categories of *believ**

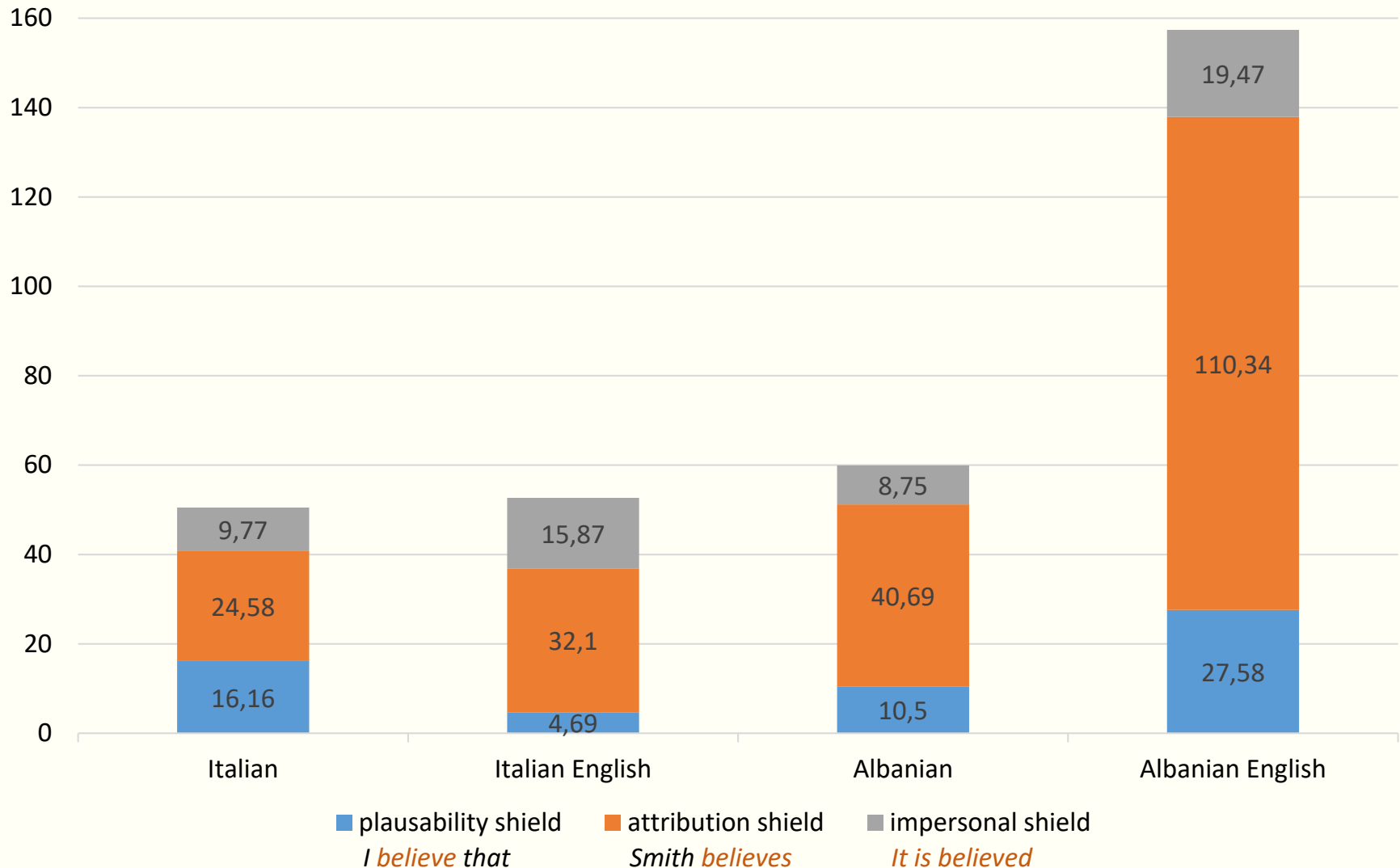


Figure 9: Frequency of *believ** in all four corpora according to its shield categories per 1.000.000 words

Examples from my corpora: *approximately, almost, seem**

- (S5) *Approximately, we can say that for $c = 10^{-5}$, the inflaton crosses ϕ_f at $\eta_f \sim 1455 \text{Mpc}$, for $c \sim 10^{-3}$ it correspond to $\eta_f = 1270 \text{Mpc}$ arriving to $\eta_f = 1456.1 \text{Mpc}$ for $c = 0$* (CIAOE14MPP_55).
- (S6) *Thuaje* në çdo kohë e thuaje në çdo vend, duket se ka pasur gjithnjë një bllok të opinionit konservator që [...] (CAR13MPL_1). *Almost in every time and almost in every place, it seems there has always been a group with conservative opinions that [...]* (my translation).
- (S7) *During the research, it was noticed that many migrants were well-informed on the economical, political, social and cultural situation in Albania, and it seemed to you as if you were discussing with individuals that were living in Albania and not with Albanian migrants living in Italy for many years* (CAR13FPS_50).
- (S8) [...] ja pse dielli **duket** I verdhë (CAR14MPP_37). [...] *that is why the sun looks* yellow (my translation).
- (S9) *Askush s'dukej* gjëkund (CAR14MPL_2). *No one was showing up* anywhere (my translation).
- (S10) *Duket* se nuk ka asnjë debat për përmbajtjen (CAR13MPS_46). *It is obvious* that there is no debate on the content (my translation)

Examples from my corpora: *seem**, *completely*, *certainly*

- (S11) *La città ti sembra un letamaio* (CIAO13MPL_48). *The city seems to you like a dung-heap* (my translation). [attribution shield?]
- (S12) *Të marrë veçmas, tiparet dukeshin në rregull.* (CAR14MPL_2). *If they are treated separately, the features seemed right* (my translation).
- (S13) *Il risultato finale (96.4% di accuratezza) ha comunque soddisfatto il pubblico e i rispeaker stessi, le cui aspettative non puntavano certamente a un simile risultato in condizioni non ideali di lavoro* (CIAO08MPL_29). *The final result (94.4% accuracy) did however satisfy the public and the respeakers themselves, who certainly did not expect such a result given the nonideal working conditions* (my translation). [transferred negative, B]
- (S14) *However, this is not completely true, as writing and speaking are not just alternative ways of doing the same thing, rather, they are ways of doing different things in order to achieve different goals* (CIAOE08FML_66). [H partly non-lexical]

Bibliography

- Biber, D., Jones, J.K. (2009). Quantitative Methods in Corpus Linguistics. In: Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics: An International Handbook. Vol.2*. Berlin/New York: Mouton de Gruyter, 987-1008.
- Fraser, B. (2010). Pragmatic competence: The case of hedging. In: *New approaches to hedging*. Bingley, UK: Emerald Group Publishing. 15-34.
- Halliday, M. A. K., Matthiessen, C.M.I.M. (2014). *An Introduction to Functional Grammar. 4th edition*. Oxon: Routledge.
- Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *TEXT 18 (3)*, 349-382.
- Hyland, K 2006, 'Medical discourse: hedging'. in K Brown (ed.), *Encyclopedia of Language and Linguistics 2nd edition*. Elsevier, Oxford, 694-697.
- McEnery, A.M. & Ostler, N. (2000). A new agenda for corpus linguistics – Working With All of the World's Languages, *Literary and Linguistic Computing, Volume 15 (4)*, 401-418.
- Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. In: J. Schmied (ed.), *Academic writing in Europe: Empirical perspectives*. Gottingen: Cuvillier, 1-22.
- Schmied, J. (2013). Academic Knowledge Presentation in MA theses: from Corpus Compilation to Case Studies of Disciplinary Conventions. *Brno Studies in English (38)*, 149-164.

Questions and Comments

1. How to better divide equivalents of English words when the meaning is the same and no formal difference exists?
2. Are impersonal shields subjective explicit and fronted hedges or boosters, objective explicit (see final example sentence), differently from Halliday's initial system?
3. Can the hedge categories of shields and approximators be used ranking boosters or my suggested equivalent categories of author-related and proposition-related boosters?
4. Is the following instance to be ranked as direct negative?

*Certamente, non essendoci nelle aree linguistiche sarde separazione tra sistema italiano e sistema locale, si crea una varietà linguistica caratterizzata dall'interferenza “[...] (CIAO05FML_78), **Certainly**, not having a separation between the Italian and local system in the Sardinian linguistic areas, a linguistic variety characterized from ‘inference’ [...] emerges,*

A corpus-based Comparison of Albanian and Italian Student Papers in L1 and L2: the Case of Hedges and Boosters

Vincenzo Dheskali

Fourth Semester PhD Student

Chemnitz University of Technology

vincenzo.dheskali@s2015.tu-chemnitz.de

