5th International Conference on Corpus Linguistics (CILC2013)

Corpus Analysis and Register Variation: A Field in Need of an Update

Rosa Giménez-Moreno[a]*, Hanna Skorczynska[b]

[a]Universitat de València-Estudi General, Valencia 46022, Spain
[b]Universitat Politècnica de València, Valencia 46022, Spain

Abstract

This study reviews the development of research on register variation (RV) over the last century to the present, emphasizing the influence of corpus analyses on its greatest advances and also on its major weaknesses and ambiguities. In search of practical and useful methods to analyse language registers, in the second part of the paper, the authors sketch a different approach to RV which has been used over the last ten years in language teaching at university level and professional communication training.

Keywords: Language registers; corpus analysis; English varieties; professional communication; language education

1. Introduction

Progress in Register Variation (RV) during the late 1980s and 1990s modified and improved the design of corpora to analyse language registers; however, at the same time, these new corpus analyses (CA) conditioned further advances and feasible applications of RV into language education and professional communication. New insights often alter the development of emerging corpora and vice versa. As in many research issues, this seems to be the case of RV. The present study will revise how the main insights within RV have been marking the way CA have developed and how the design of these analyses has marked the limitations and ambiguities this field of research currently has. While pointing out these limitations and weak points, we will also highlight the main areas of

---

*Corresponding author Tel.: +34 963864262; fax:+34 963864161.
E-mail address: rosa.gimenez@uv.es

consensus within this field, as the second aim of this paper is to present an all-inclusive and learner-friendly approach to RV which allows researchers to explore ways to avoid previous limitations and to seek out methods to analyse registers in a more practical and useful manner.

2. Theoretical overview throughout the C20[th] up until the present time

In the first half of the C19[th] specialists such as Malinowski (1923) and Firth (1935) already talked about language varieties dependent on the "context of situation". Looking at it retrospectively, within the world of language research and linguistics, developments on contextual variation were simultaneous -or even prior- to those on textual variation. However, throughout the last century, text and genre variation have expanded with greater intensity and effectiveness, almost engulfing or absorbing contextual variation.

At that time, specialists claimed the value of collecting and analysing natural texts (Boas, 1940) and these initial – mainly theoretical – works on contextual variation were based on the researcher's own intuition, observation and exemplification of common English usage (Quirk, 1960). Within them the word "register" hardly appeared and it was not used as a separate type of language variety, being widely identified as "style"; therefore, researchers were analysing variation dependent on individual choices (style) together with contextual variation (register). The result was quite a restricted typology which only distinguished between formal, neutral and informal variation. This overgeneralising "trio" is still popular in social and professional communication and widely used in the teaching world.

Two remarkable references in this early period were the works by Ferguson (1959) and Joos (1967). Ferguson emphasised the notion of "scale" within any language's contextual variation and the relevance of providing a clear distinction between both ends of these scales: their "high" and "low" varieties. Eight years later Joos published his book *The Five Clocks* (1967) where "register" was still identified as "style" and illustrated with fragments from every day general English. However, his conceptualisation was based on "context" as key parameter which helped him break with the famous trio and identify five main contextual types: frozen, formal, consultative, casual and intimate. This well-known typology is still used for educational purposes and in the last decade we can find studies which still include this typology (e.g. Takahashi, 2006).

The term "register", within the area of Linguistics, was first used by Thomas B. Reid in 1956 (in his article "Linguistics, structuralism and philology") and in the 60s register theory burst onto the research scene at the same time as the take-off of Systemic Functional Linguistics. RV consolidated as a field of language variation by a group of linguists interested in variation according to the use, diatypic variation, in contrast to diatopic variation (Halliday et al., 1964; Ure, 1968; Davies, 1969). In fact, the term "register" often alternated with the term "diatype" (Gregory, 1967). Their approach to RV was based on the renowned three dimensions: "field", "mode" and "tenor" (the latter often alternated by "style of discourse") which were at that time new and marked the corpora design tendencies of the following decade. However, within RV they identified –and often also overlapped- styles, functions, topics and modes. For example, Ure (1971) used "mode" as a parameter to differentiate "registers" and her corpus is made of 34 spoken and 30 written texts with a wide span of diverse sources, from a conversation of two women cleaning a room to a radio interview, and from a school essay to a publisher's blurb. Chiu in 1973 based her study on register variation on "mode" and "field" to compile and analyse what she called "the register of administrative writing". With such an unclear concept of register, the notion of a reliable corpus to analyse RV was still weak and, at this time, these two previous studies based on a fairly large corpus contrasted with other works such as the famous book by Gregory and Carroll's (1978) - *Language and Situation: Language Varieties and their Social Contexts* - based on just one tape-recorded conversation between an "on campus café employee and his customer".

In the 80s, RV gained impetus thanks to the group led by Douglas Biber (see Biber, 1988, 1993, 1995; Biber & Finegan, 1994; Ghadessy, 1988, 1993). These specialists define register as "general cover term for situationally defined varieties (…) associated with particular contexts or purposes" (Biber, 1995:1), which are dependent on their users' language communities (Ure, 1982:5). Although this concept is still broad and vague, they already see registers as clusters of associated features which have a tendency to co-occur (Halliday, 1980, 1988). Therefore the "analyses of these register features are necessarily quantitative, because the associated register distinctions are based on differences in the relative distribution of linguistic features" (Biber, 1995: 29). They encourage the analysis of large on-line dictionaries and computer-based corpora (e.g. Brown Corpus, Lancaster-Oslo-Bergen Corpus, London-Lund

Corpus), with reliable concordancing programs and statistical techniques. However, these analyses were aimed at the identification of dimensions ("Multidimensional Analysis", MA) within certain "registers" (which were mainly "genres") and therefore they constricted the notion of register to those dimensions and text-types with the result of very complex set of register typologies. For example, the corpus for Biber (1995) analyses 65 linguistic features in more than 30 different types of "registers". The heading "Press" includes "registers" such as reportage, editorials, and reviews, or the heading "Academic prose" includes "registers" such as natural sciences, medicine, mathematics and "humanities. In these MA, "registers" are analysed and contrasted in terms of a wide and heterogeneous list of dimensions: involvement versus exposition (regarding the text's structural elaboration), on-line versus planned/integrated production (regarding lexical elaboration, argumentative versus reported presentation of information, narrative versus non-narrative discourse organization, attitudinal versus authoritative discourse, informational versus interactional focus, rhetorical manipulation versus structural complexity, informal interaction versus explicit elaboration, discourse chaining versus discourse fragmentation, etc.

At the beginning of this millenium Biber's RV theory develops including more sophisticated computer software and larger corpora which consolidates the term "register" as a distinct linguistic variety (see Reppen et al. 2002; Biber, 2006; Biber et al. 2007; Biber & Conrad, 2009; Biber, 2012). Some of their key theoretical insights are that:

(a)   Register, genre and style are seen as the three fundamental functional varieties of language: three distinct but complementary concepts (also Giménez-Moreno, 1997).

(b)   Situational features of registers are presented as more basic than the linguistic characteristics. The main challenge is to set up the relationship between situational characteristics of register and its specific functional relationship with linguistic features.

(c)   Effective register analyses need to follow three pre-requisites: follow a comparative approach, be based on a representative sample of texts and use quantitative analysis.

One of Biber's recent insights has inspired the current research. According to him (Biber, 2012) the difference between Variationist Corpus Approaches, which give priority to linguistic features, and Text-Linguistic Corpus Approaches, focused on genres, has significant consequences for the overall research design, the chosen variables and the statistical techniques applied. As a result, the importance of register has been more apparent in text-linguistic studies than in studies of linguistic variation. As mentioned in the introduction, new insights on genre variation and text linguistics have altered the development of emerging corpora on RV and also emerging corpora on genre variation have altered new insights on RV. The result is a quite delicate situation. Within the world of Linguistics, MA's contribution to RV is priceless; however its impact is not equivalent within the world of Language Acquisition, Professional Communication or Modern Languages for Specific Purposes. A reason might be because their conceptualization of register is still too imprecise and still remains overlapped with the other parameters of analysis such as genres, styles, dimensions and functions of language. Ambiguous and unclear exemplifications of theoretical frameworks appear in Biber and Conrad (2009) including registers such as "casual conversation", "service encounters", "newspapers" "research article introductions" and "fictional writing". "Conversations" and "academic writing" are "registers" contrasted in Biber (2012). As Bhatia (2012: 560) points out: "boundaries between register and genre seem to be blurred. (…) One may find this overlapping or blurring across register and genre, and perhaps to some extent also style, somewhat puzzling." Another constraint is that their theory's practical application seems too complex to be useful for language students, professional practitioners and even also for language researchers.

Nowadays ambiguity and dispersion seems constant in RV. The ISO (International Organization for Standardization) released in 1999 a standard for registering linguistic terms (Data Category Registry) which identified "registers" such as: bench-level register, dialect register, facetious register, formal register, in house register, ironic register, neutral register, slang register, taboo register, technical register and vulgar register. We find studies on the "register of netspeak" (Omar, 2012), the registers of "politics", "science and academic language" and "mathematic"s (Franke, 2012; Silva et al. 2012; Berggvist & Österholm, 2012), "natural English register" (Bardelle, 2011), or the "register of parentese" (Corthals, 2012). Santos et al. (2012) analyse the following registers: figurative, popular, informal, pejorative, disparaging, vulgarism, slang and ironic. RV current coverage can hardly become

more heterogeneous and unpractical in terms of conceptualization, systematic analysis and language teaching. It is not surprising to see articles published asking for meaningful contributions to clarify this area, such as Sampson (1997) "'Genre', 'style' and 'register'. Sources of confusion?", Lee (2002) "Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle" or Frog and Koski (2012) "Register. Intersections of Language, Context and Communication".

In order to help students, a group of language lecturers are currently trying to cast some light into this issue from different perspectives. For example, from a MD's perspective, Aguado et al. (2012) have recently tried to explore "the use of multidimensional analysis of learner language to promote register awareness". They collected a corpus from interviewing 59 nonnative students and 28 native speakers in order to analyse dimensions such as personal narrative, interpersonal communication and picture description. The study follows Biber (1988) and Conrad and Biber (2001) analytical procedure and uses an interesting research tool developed by their own team called CRAT (Corpus Research Analysis Tool). On these lines, one of the main future challenges of RV's theory is to find more practical methods to promote register awareness among professional communication trainees and general language learners.

3. Searching for an all-inclusive but practical approach to RV

The approach to RV analysis that we propose here is all-inclusive, but at the same time practical in its application. The term 'all-inclusive' refers to a set of parameters and criteria that should be considered if we want to choose a comprehensive approach, built on the previous research conducted in the field. In this sense, we refer to the key defining parameters of register: settings and roles, keeping in mind that register variation depends on the context of situation and on the roles that participants play when they communicate with each other in a particular context. The all-inclusive approach should also aim at covering all possible communicative contexts: from family to professional situations. Moreover, its practicality should imply using a clearly defined range of registers and the corresponding co-occurring language features, which can be scaled in terms of the proportion of use. The approach proposed also requires an appropriate methodology of analysis, and in our view it should combine native speakers' observation and collaboration, as well as the use of computer applications in corpus creation and searches.

In an attempt to elaborate the all-inclusive approach to register variation, Giménez-Moreno (2006) proposed four basic registers, which belong in two principal domains of our lives: personal or private, and public (see Table 1), and which can be usefully exploited in the EFL classroom.

Table 1. Basic register varieties (Giménez-Moreno, 2006: 100).

Registers	Participants Roles	Communicative Settings
Personal/Private life:		
Family	Relatives	Family settings (homes and public places but exclusive/closed)
Amicable	Friends/Acquaintances	Friendly settings (homes and public places but exclusive/closed)
Public life:		
Social	Neighbours/Citizens	Social open settings
Professional	Colleagues Social services: police, doctors, clerks, etc.	Institutions of public services and work environment

The family register would be used in family settings, at home or in public places, among relatives in exclusive or closed groups. A similar exclusivity is characteristic of the amicable register, which would be used in the private domain of one's life with friends and acquaintances, and both in public and more private settings. Regarding the public life, Giménez-Moreno distinguished between the social and the professional register. The former would be preferred in social open settings by participants playing the role of citizens or neighbours. The professional register, in turn, is frequent in the workplace and institutions of public services where members of social services, such as police officers, doctors and clerks, would use it.

Table 2. Correlations between registers and spoken genres (Giménez-Moreno, 2006: 104).

Registers	Tones	Spoken Genres
Family	Intimate	Pillow talk
	Neutral	Daily meal talk
	Ritual	Christmas dinner, and other family celebrations
Amicable	Intimate	Confessions, and "have a laugh" talk
	Neutral	Dinner party talk
	Ritual	Open gathering, and street encounter with acquaintances
Social	Casual	Warning, requests and street/lift encounters with neighbours
	Neutral	Social open assembly (with diverse neighbours and other citizens)
	Ceremonial	Religious ceremonies (with all sorts of members)
Professional	Casual	Coffee break
	Neutral	Business meeting
	Ceremonial	Awards ceremony

As Table 2 shows, each of the registers mentioned can be expressed in at least three communicative tones or frequencies: intimate, neutral and ritual for the family and the amicable register, on the one hand; and casual, neutral and ceremonial for the social and the professional register, on the other. Examples of the spoken genres in which the listed registers and tones are typically used are included in Table 2.

In an application of this approach to register variation to a corpus of 224 emails sent by 54 native-speaker English business practitioners to one recipient, Giménez-Moreno (2011a) used ten native informants, who classified those professional emails according to the register and tone: professional formal, professional neutral and professional casual. As a result, the author identified ten linguistic parameters of register variation in professional communication, which are shown in Table 3 below.

Table 3. Ten linguistic parameters of register variation in professional communication (Giménez-Moreno, 2011a: 19).

A	B
(+) Informal/casual	(+) Formal/ritual
(Showing commitment, involvement and closeness)	(Showing deference, neutrality and objectivity)
1. Personal expressions	1. Impersonal expressions
2. Active verbs/expressions	2. Passive verbs/expressions
3. Direct speech	3. Indirect speech
4. Ordinary reporting verbs (e.g. say)	4. Specific reporting verbs (e.g. mention)
5. Ordinary connectors (e.g. so)	5. More elaborate connectors (e.g. furthermore)
6. General terms/expression (e.g. man)	6. Precise terms/expressions (e.g. technician)
7. Emotive/subjective/attitudinal terms (e.g. guess)	7. Neutral/objective terms (e.g. inform)
8. Phrasal verbs and informal idiomatic expressions	8. Latin terms and standard formal expression
9. Use of contractions, abbreviations and "fast language"	9. Detailed and concrete expressions without contractions using nominalization and modifiers
10. Straight statements and direct commands	10. Politeness, caution and mitigation markers

The study showed that register fluctuation will not be determined by the type of chosen linguistic features, but also by their proportion and combination according to the conventional and intentional roles performed by the participants in a specific context. In this corpus register variation fluctuated around the professional casual tone (around 60%).

Further applications and development of the approach proposed here included tone fluctuation in phone conversations and meetings, as well as register variation in intercultural business communication. In the first study (Giménez-Moreno, 2011b), the transcriptions of 15 business meetings and 20 phone conversations from business English textbooks were examined by native informants in order to identify linguistic features of the formal, neutral and casual tone of the professional register. The results showed that within the same text/discourse, the tone used by the participants not only fluctuates on the scale from professional formal to casual but also incorporates instances of other neighbouring registers (e.g. social or amicable). In the second study (Giménez-Moreno & Skorczynska, 2013),

the choice of the formal or casual tone of the professional register used by British, Polish and Spanish company employees in emails written in English was described as varying in correlation with the national culture of the writer. All of the writers used similar moves in writing their replies to enquiries, however Polish and Spanish employees tended to reinforce the formal tone in comparison to the British writers. In view of the results obtained, the need to increase English learners' awareness of a wide range of tones used by native speakers was pointed out.

The present approach has been adapted and used in the last ten years to teach English to university students with different language levels (from B1 to C2) in diverse undergraduate and postgraduate degrees (e.g. English Studies, International Business Studies, Tourism Studies and International Legal Studies). Up until now the students continue to show interest and satisfaction in the way they improve their learning and usage of register variation in English.

4. Future research

In view of the research presented here, register variation is still a field of study that requires further advancement based on the use of corpora and the refinement of parameters in register description. Obviously, register variation research has immediate applications to foreign language teaching and intercultural communication, and this type of perspective that the field offers should attract scholars and communication practitioners. Future lines of research should consider other communicative contexts, such as amicable and family settings, as no research concerned with these situations has been conducted to the best of our knowledge. There are two aspects of research on register variation that are especially important in our view: a close team work between scholars and native speakers, and a rigorous design of corpora, in which a precise identification of context and participants' roles should allow for the creation of fully representative corpora, and ensure reliable research results.

References

Aguado-Jiménez, P., Pérez-Paredes, P., & Sánchez, P. (2012). Exploring the use of multidimensional analysis of learner language to promote register awareness. *System 40*, 90-103.

Bardelle, C. (2011). Student understanding of the negation of statements with universal quantifier. *Proceedings of the 35th Conference of the International group for the Psychology of Mathematics Education 2*, 97-104.

Bhatia, V. K. (2012). Book review: *Register, Genre, and Style (by D. Biber and S. Conrad,* Cambridge University Press, NewYork). *Journal of Sociolinguistics, 16*(4), 558−561

Bergqvist, E., & Österholm, M. (2012). Communicating mathematics or mathematical communication? An analysis of competence frameworks. In T.Y. Tso (Ed.) *Proceedings 36th Conference of the International Group for the Psychology of Mathematics* (pp. 67-75). Taipei: PME36.

Biber, D., & Finegan, E. (Eds.) (1994). *Sociolinguistic Perspectives on Register*. New York: Oxford University Press.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, *19*: 219-41.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University Language. A Corpus-based Study of Spoken and Written Registers*. John Benjamins: Amsterdam, Philadelphia.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory 6*(1), 9-37.

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins

Biber, D., & Conrad, S. (2009). *Register, Genre and Style*. New York: Cambridge University Press.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins P.C.

Boas, F. (1940). *Race, Language, and Culture*. New York: Macmillan

Chiu, R. (1973). Measuring register characteristics. *IRAL*, *11*, 51-68.

Conrad, S., & Biber, D., (2001). *Variation in English. Multi-Dimensional Studies. Pearson Education, Harlow. CRAT: Corpus Research Analysis Tool*. http://perezparedes.blogspot.com/search/label/CRAT.

Corthals, P. (2012). Speaker gender, child age, syntax, and the prosody of parentese. *American Speech-Language Hearing Association (ASHA), Annual Convention*, Atlanta USA 15th-17th November 2012

Davies, A. (1969). The notion of register. *Educational Review*, *22*(1), 64-77.

Ferguson, C. A. (1959). Diglossia. *Word*, *15*, 325-40.

Firth, J.R. (1935). The technique of semantics. *Transactions of the Philological Society,* 36-72.

Franke, J. (2012). *The Potential of Teaching Politics in English*. From http://books.google.es [20/01/2013]

Frog, Koski, K. (Eds.) (2012). *Register. Intersections of Language, Context and Communication.* Pre-Print Materials of the International Colloquium 23rd−25th May 2012 Helsinki, Finland. *Folkloristiikan toimite 17*. Helsinki: Folklore Studies, University of Helsinki.

Ghadessy, M. (Ed). (1993). *Register analysis: theory and practice*. London: Pinter Publishers.

Ghadessy, M. (Ed.). (1988). *Registers of written English: situational factors and linguistic features*. London: Pinter Publishers.

Giménez-Moreno, R. (1997). The Boundaries of the Concepts of Genre, Register and Style in Academic Discourse. In J. Piqué and D. J. Viera. *Applied Languages: Theory and Practice in ESP* (pp.37-45). Valencia: Publicacions de la Universitat de València.

Giménez-Moreno, R. (2006). A new approach to register variation: the missing link. *Ibérica, 12*, 89-110.

Giménez-Moreno, R. (2010). English for higher education studies. In R. Giménez-Moreno (Ed.), *Words for Working. Professional and Academic English for International Business and Economics* (pp. 295-334). Valencia: Publicacions de la Universitat de València.

Giménez-Moreno, R. (2011a). Register variation in international business correspondence. *International Journal of English Studies, 11*(1), 15-34.

Giménez-Moreno, R. (2011b). Role fluctuation in professional English register variation. In M. L. Carrió, J. Contreras, F. Olmo, H. Skorczynska, I. Tamarit, & D. Westall (Eds.), *La Investigación y la Enseñanza Aplicadas a las Lenguas de Especialidad y a la Tecnología* (pp. 71-79). Valencia: Universitat Politècnica de València.

Giménez-Moreno, R., & Skorczynska, H. (2013). Business communication across three European cultures: a contrastive analysis of British, Spanish and Polish email writing. *Ibérica 26* - Forthcoming special issue on *Intercultural and International Business Communication* (Fall 2013).

Gregory, M., & Carroll, S. (1978). *Language and Situation. Language Varieties and their Social Contexts.* London: Routledge & Kegan Paul Ltd.

Gregory, M. (1967). Aspects of varieties differentiation. *Journal of Linguistics, 3*, 177-197.

Halliday, M. A. K. (1988). On the Language of Physical Science. In M. Ghadessy, (Ed.) *Registers of Written English: Situational Factors and Linguistic Features* (pp. 162-178). London: Pinter Publishers.

Halliday, M. A. K., McIntosh, A., & Strevens, P. (1964). *The Linguistic Sciences and Language Teaching.* London: Longman.

Halliday, M. A. K. (1980). Register Variation and the Identity of a Text. *Sophia Linguistica, 6*, 60-79

Joos, M. (1967). *The Five Clocks*. New York: Harcourt Brace Jovanovich.

Lee, D. (2002). Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle. *Teaching and Learning by Doing Corpus Analysis, Language and Computers: Studies in Practical Linguistics, 42*, 247-292.

Malinowski, B. (1923). The problem of meaning in primitive languages. Supplementary essay in C.K. Ogden & I.A. Richards (Eds.), *The meaning of meaning*, (pp. 296-336). New York: Harcourt, Brace.

Omar, A. (2012). Impact of technology on teens' written language. *International Journal of Advanced Trends in Computer Science and Engineering 1*(1): 9-17.

Reid, T. B. (1956). Linguistics, structuralism, philology. *Archivum Linguisticum 8*.

Reppen, R., Fitzmaurice, S. M., & Biber, D. (Eds.). (2002). *Using corpora to explore linguistic variation.* Amsterdam: John Benjamins.

Sampson, J. (1997). "Genre", "style" and "register". Sources of confusion? (Proposing consistent meanings for the terms in linguistics) *Revue Belge de Philologie et d'Histoire, 75*(3): 699-708.

Santos, A. P., Oliveira, H. G., Ramos, C., & Marques, N. C. (2012). The Role of Language Registers in Polarity Propagation. In H. Caseli, A. Villavicencio, A. Teixeira & F. Perdigao (Eds.) *PROPOR 2012, LNAI 7243* (pp. 235–240). Berlin: Springer-Verlag.

Silva, C., Weinburg, M., Malloy, R., Smith, K. H., & Marshall, J. N. (2012). Toward Integration: An Instructional Model of Science and Academic Language Childhood Education. *Childhood Education, 88*(2), 91-95.

Takahashi, K. (2006). A study of register variation in the British National Corpus. *Literary and Linguistic Computing*, *21*(1): 111-126.

Ure, J. (1968). Practical Registers. *English Language Teaching, 23*, 107-215.

Ure, J. (1971). Lexical Density and Register Differentiation. In G. E. Perren & J. L. M. Trim (Eds.) *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics.* Cambridge.

Ure, J. (1982). Introduction: approaches to the study of register range. I*nternational Journal of the Sociology of Language, 35*, 5-23.