



**HAL**  
open science

## Measuring the construction of discorsal expertise through corpus-based genre analysis

Dacia Dressen-Hammouda

► **To cite this version:**

Dacia Dressen-Hammouda. Measuring the construction of discorsal expertise through corpus-based genre analysis. Alex Boulton, Shirley Carter-Thomas and Elizabeth Rowley-Jolivet. Corpus-Informed Research and Learning in ESP. Issues and applications, John Benjamins, pp.193-213, 2012, 9789027203571. 10.1075/scl.52 . hal-01012320

**HAL Id: hal-01012320**

**<https://uca.hal.science/hal-01012320>**

Submitted on 25 Jun 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Measuring the construction of discorsal expertise through corpus-based genre analysis**

Dacia Dressen-Hammouda

### **Abstract**

While corpus analysis has long been useful for developing genre-based teaching materials in English for Specific Purposes (ESP), somewhat less attention has been paid to how well people actually learn to use the genre features identified. This study shows how individuals' use of genre features changes over time, as a function of growing disciplinary experience. Using a measure of standard deviation, the study examines how five geologists show increasing discorsal expertise in their writing over a ten-year period. The method of analysis used allows for comparison of individual and collective uses of the generic features authors use to construct their disciplinary voice.

**Key words:** L1/L2 writing pedagogy; corpus analysis; standard deviation (SD); discorsal expertise; disciplinary voice; geology

### **1. Introduction**

Corpus analysis has long been a central practice in ESP genre-based writing pedagogy, enabling the identification of the most common features of specialized language and genres for teaching purposes. However, the success with which individuals actually learn to use the genre features identified by corpus analysis has been somewhat less addressed in discussions on corpus-based pedagogies.

This corpus-based genre study demonstrates one way in which learners might benefit from the results of corpus analysis. It examines how scientific writers' disciplinary voice shifts over time as a result of increasing expertise. "Disciplinary voice", following Matsuda and Tardy (2007) and Tardy and Matsuda (2009), is considered through a dual lens: because it is tied to specialized genre and disciplinary knowledge (Berkenkotter & Huckin 1995) its features reflect both the typified social and co-constructed dynamic of disciplinary interaction (Beaufort 1999; Ivanic 1998) and the individuality of the writer (Elbow 1999; Hyland 2008, 2010; Matsuda 2001). An individual's use of the features of disciplinary voice can be seen to evolve over time as a function of her or his place in the disciplinary community (Dressen-Hammouda 2008).

To reveal this shift, this study examines how five researchers in geology construct their disciplinary voice in English over a period of ten years, beginning with an early publication based on their doctoral dissertation. To identify how each author uses the features of disciplinary voice over time, the study proposes a method to date unused in genre-based

corpus analysis: a measure of standard deviation (Dressen-Hammouda forthcoming). This method compares changes in the authors' individual writing strategies over a period of time with norms identified in the corpus.

It is argued that standard deviation is a valuable tool for measuring how individuals move along a continuum of expertise over the course of their academic careers, developing from novice to junior to senior researchers. Writers with greater disciplinary experience are seen to rely more often on the affordances genres provide, by surpassing a genre features' normal range of use. They also seem to adhere less strictly to the more common genre conventions that often characterize less experienced writers' genre use and disciplinary voice. By gradually diverging from a perceived norm, experienced writers develop their disciplinary voice, staking their claim to their territories and reinforcing the basis of evaluation for their professional expertise by their peers.

It will be argued that such research offers an important perspective for both L1 and L2 writing pedagogy, as it highlights the social, institutional and individual features of voice which novice writers need to become aware of when learning to construct their disciplinary voice and expertise in English.

The remainder of this chapter addresses these issues, describing the methodology used, then presenting the results of the analysis, before closing with a discussion of the applications of corpus-based studies such as this for L1 and L2 writing pedagogy.

## **2. Measuring genre norms using standard deviation**

Genre analysis aims to establish the patterns and regularities – or 'norms' – that characterize the 'real-world' language used in specific settings. Some of the most common analytical methods currently used to identify genre norms include keyword frequency (Hyland 2000) and lexicostatistics (Swales 1990). Each of these approaches converges the analysis of data toward a 'snapshot' view of language use, by providing a measurement of either the most frequent use or of its average use.

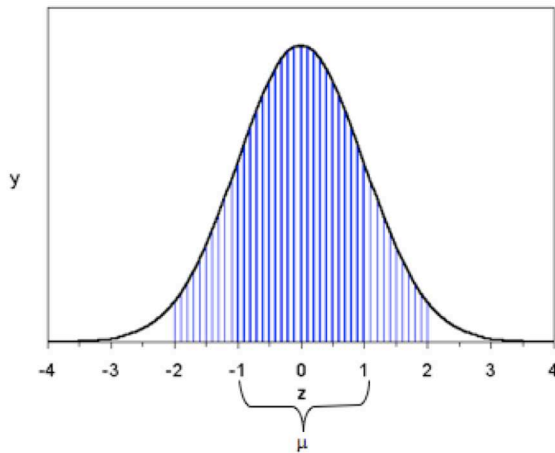
However, there is growing agreement today among language scholars that a genre norm cannot be described simply as an average number of uses, nor as a single type of use. Sinclair (2004: 289), for example, has observed that "No one would argue that frequency is other than a rough indication of the importance of a sense or phrasing." Bhatia (2002: 6) goes even further when he challenges the results of quantitative methods by asking "Is generic description a reflection of reality or a convenient fiction invented by applied linguists for pedagogical and other purposes?" In effect, while such synthetic views may be useful for teaching purposes, they do not authentically represent actual genre use.

Actual genre use is instead characterized by substantial, but equally valid, variation from perceived norms, where a norm does not represent a single value, *per se*, but an entire range of accepted values. Given that people seek to recognize patterns through prototypes (e.g. Rosch 1975), they will tolerate varying levels of variation in the expression of norms as long

as what they see resembles their idea of the prototype before deciding that the expression is no longer representative. And yet, genre analysis has yet to propose an analytical method that takes the natural variation of norms into account.

A measure of standard deviation is proposed here as a method better able to represent this linguistic reality. In a measure of standard deviation (SD), the occurrences of a particular genre variable, attested through corpus analysis, are statistically represented by a Gaussian curve (Figure 1), which is both symmetrical and ‘bell-shaped’. In this study, SD is taken to represent the probability that all genre variables have a normal distribution, or in other words, follow a specific pattern of density.

Figure 1: Example of a Gaussian curve representing the spread range for a particular genre variable, with average occurrence ‘ $\mu$ ’.



The area inside the curve represents the probability that the largest number of occurrences of the variable lie between the values that delimit that section, called the ‘spread range’. The spread range itself represents the range of ‘normal’, or most common occurrences, of that variable. The spread range’s average (‘ $\mu$ ’) lies toward the middle of the spread, and indicates the largest grouping of items that represent the variable. Outliers are also included within the spread range, although the further away they find themselves from ‘ $\mu$ ’, the lower the likelihood they will be found by the analyst in multiple situations.

What is noteworthy in this definition of a genre norm is the idea that we are no longer dealing with a single representative value to describe the norm. While SD still shows the most frequent number of uses (‘ $\mu$ ’), it also provides an entire range of values that capture the actual use of a genre norm. Whether or not a statistical tool of this nature accurately reflects the cognitive complexities of human pattern recognition remains to be seen; at the very least it does hopefully provide a metaphorical solution to the problem of describing genre norms, in a way which appears more representative of actual language use than other analytical methods currently used.

The range of variation captured by a corpus-based measure of SD can tell the analyst a number of things. For one, it can tell us how reliably an analyst may state that a particular

variable has typified as a genre norm. In point of fact, SD is essentially a measure of how much variation characterizes a particular genre variable. The larger the SD, or spread range, the less reliably one can state that a particular variable shows 'normal' behavior.

A measure of SD can also tell us how closely genre users adhere to a norm by measuring the spread of values in a data set. If, for example, the data points for a particular genre variable are all close to the average ( $\mu$ ), then the SD is close to zero and its curve is tight, implying that there is little variation attested in the corpus. If on the other hand many data points are far from the average, then the SD curve has a wider spread range, indicating that the community of writers may tolerate more variation in the expression of that particular genre variable.

Finally, a measure of SD also allows us to examine how closely individual writers conform to genre norms in their own writing. As explained above, a measure of SD tells us the most frequent use of a variable within a given genre, and suggests how much variation may be tolerated while remaining representative of the norm (its 'normal range'). An individual's use of the variables over time can then be compared to the normal ranges established in a corpus by calculating how much the individual's use of the variables deviates from or conforms to the corpus-based spread range.

SD is thus potentially useful for studying different aspects of genre variation, such as the individual expression of disciplinary voice, or the emergence of disciplinary expertise in writing. By helping writers become aware of the development of these features in their own writing, it is applicable to the needs of both L1 and L2 writing pedagogy.

### **3. Methods**

The results reported in the next section draw on two separate corpora. Both corpora are made up exclusively of 'field accounts', a part-genre found in research articles from geology. The field account, which has both an audience and set of communicative purposes that are specific to it, is an integral genre in its own right although it is embedded within the scientific research article. This explains its characterization as a *part-genre* (Ayers 1994). Geologists who carry out fieldwork use the field account to describe their fieldwork observations and interpretations to the scientific community.

The first corpus, briefly described in Dressen-Hammouda (2008), consists of 65 field accounts published in research articles between 1996-1999 in three subdisciplines of geology: geochemistry, petrology and structural geology. The corpus contains 67,312 words; complete details are given in Dressen (2002). A second, smaller corpus ( $n = 17,070$ ) consists of 19 field accounts from research articles published between 1983-2003 (Table 1). This second corpus represents the field accounts published by the five researchers in geology during the approximately ten-year span following their PhD dissertation. Three authors are native speakers of French (A-C), one is a native speaker of English (D), and one of Slovakian (E).

Table 1: *Corpus of field accounts from 5 researchers in geology (1983-2003)*

PhD	Post-doc	Publication date & journal
A France	US	A1 - 1984 Journal of Structural Geology A2 - 1989 Alpine Tectonics A3 - 1993 Geological Society of America Bulletin
B France	UK	B1 - 1992 Earth and Planetary Science Letters B2 - 1993 Tectonics B3 - 1999 Tectonics B4 - 2003 Journal of Geophysical Research
C France	UK	C1 - 1991 Comptes Rendus de l'Acad. des Sciences C2 - 1993 Journal of Geophysical Research C3 - 1996 Chemical Geology C4 - 1997 Geochimica et Cosmochimica Acta C5 - 2003 Journal of Petrology
D South Africa	Unknown	D1 - 1985 Economic Geology D2 - 1995 Mineralium Deposita D3 - 1998 Earth and Planetary Science Letters D4 - 1999 Journal of Petrology
E US	US	E1 - 1983 Earth and Planetary Science Letters E2 - 1986 American Mineralogist E3 - 1988 Contributions to Mineralogy & Petrology E4 - 1997 Geochimica et Cosmochimica Acta

The articles were targeted for inclusion in the corpus because in each case, the identified author was first author of the text. Four of the authors confirmed that they had been primarily responsible for writing the article (Authors A, B, C and E). Author D was unavailable for comment.

A previous analysis of Corpus 1 identified 13 variables that characterize the field account (Dressen 2002; Dressen-Hammouda 2008). These same variables were also used in the second corpus, in order to examine how the five authors used them at different points during the ten years following their PhD dissertation. A detailed description of these variables will be given in the next section.

In the present study, the variables were identified and counted in all articles from both corpora. The number of occurrences of each variable was entered into a standard spreadsheet program, which automatically calculated both the average and the standard deviation for each variable, across the first corpus. The range of variation thus defined in Corpus 1 was then used as a basis for comparison for Corpus 2, so as to determine the extent to which the five authors' writing strategies diverged from or resembled the trends observed in the larger corpus.

#### 4. Results

The following list details the 13 variables identified in Corpus 1. The variables each carry out one of the field account's rhetorical functions: (1) to personalize the field account by showing physical presence in the field or by demonstrating authority ('personalization cues'), (2) to prove that fieldwork was actually carried out ('doing-the-work cues'), and (3) to demonstrate

relevant research community concerns ('disciplinarity cues').<sup>1</sup> Illustrations of each cue are then given, using examples taken from Corpus 2.

Personalization cues

1. First-person pronouns/possessive adjectives
2. Evaluative adjectives and adverbs
3. Interpretive comments

Doing-the-work cues

4. Nominal/verbal markers of research activity
5. Metric, angle or direction measures
6. Locational adverbs and prepositions
7. Metadiscoursal references to visual data
8. Geographical location of the fieldwork
9. Self-citation of prior field studies

Disciplinarity cues

10. Nominal or adjectival field descriptors
11. Descriptors of geological time/age
12. Technical verbal adjectives and participles
13. Citations of others' fieldwork

As previously discussed (Dressen-Hammouda 2008), 'Personalization cues', allow writers to demonstrate explicit ownership over their field study by giving them the linguistic means to evaluate (2) and interpret (3) their observations. Authors also use personalization cues to re-invert the agency hierarchy, allowing them to overtly say "*We did this*" (1). While this strategy occurs the least frequently in the corpus, when it is used, it provides definitive proof for one's interpretation, occurring at rhetorically strategic points in the development of one's argument.

Personalization cues

- (1) First-person pronouns/possessive adjectives  
The estimated thickness of the Cretaceous from its upper contact with the Claron to the base of the sequence in which *we investigated* structures is about 570 m. (A3)
- (2) Evaluative adjectives and adverbs  
*Nowhere else ... is there such an ideal combination* of discordancy, relief, and exposure for the study of the basal contact. (D1)  
The *highly* deformed zones of fault rocks stand *in marked contrast* to the condition of the surrounding rocks which *commonly* appear *only slightly* strained or *completely* undeformed. (A3)
- (3) Interpretive comments  
*It is important to note that* the samples collected in the western half of the stock (Fig. 1) represent traverses that are essentially parallel to the roof of the stock. (E1)  
Unequivocal field relations show that B2N and B3N microgabbros form thin margins to B2 and B3 sheets, respectively, and occur as xenoliths within them, *so it is impossible for them to have intruded along the basal contact after formation of* the thin marginal zone. (D1)  
These last generally form indurated blocks, *in relief, within* the host tuff, *and it is not always clear in the field whether they represent enclaves or more indurated parts of the host tuff.* (C5)

A second set of variables also allows writers to present the results of their fieldwork. The second most frequently used type of variable, 'Doing-the-work' cues, show what the researcher actually did in the field, but without allowing the researcher to frame that activity with an explicit agent role. Thus, readers can *infer* that fieldwork has been done, due to the

presence of nouns or passive verbs that describe the researchers' activity (4), or to the various measurements that were carried out in the field (5). One can also infer such information from the text's locational adverbs (6) that imply how the researcher moves around from point to point in the field, from references to field maps (7), site locations (8), and from self-citations (9).

Doing-the-work cues

- (4) Nominal/verbal markers of research activity  
More *information is obtained from samples* from the eastern part of the stock because the magma apparently ascended to a higher level and thus these traverses were made into a deeper part of the stock. (E1)  
Thus the rocks *observed in the field* were deformed at a very superficial level under about 1 to 2 km of overburden. (A3)  
All three parts are conformable; the bottom and top *are not seen*. (B3)
- (5) Metric, angle or direction measures  
The pyroclastic deposit of Chabrières is located *2 km SE* of Mount Mèzenc. (C5)  
A projection *northward* from the segment closest to the *eastern* termination of the Elbow thrust "hits" the *south end* of the trace of yet another identically oriented sinistral strike slip fault (Fig 3) whose aerial photo expression is pronounced. (A3)  
Orientations of the axis of greatest principal compressional stress reveal a stringing fan-like pattern (fig 9) trending *south southwest 215° in the west to southeast 125° in the eastern part* of the study area. (A3)
- (6) Locational adverbs and prepositions  
The major and *upper* area of exposure is a landslip lying *close to* the SE flank of the phonolitic dome of Chabrières. Further *down*, the outcrop *extends* more than 1 km *along* the Saliouse stream. (C3)  
The felsic rocks are *crosscut* by large mafic dikes. The Miran Group is mapped as unconformably *overlain* by the Upper Proterozoic Qingbaikou system and *in fault contact with* the Middle Proterozoic Jixian system. (B3)
- (7) Metadiscoursal references to visual data  
A north northeast trending sinistral strike slip fault occurs near the eastern end of the Rubys Inn thrust and can be traced northward to where it marks the eastern termination of the north verging Pine Hill and the western termination of the south verging Elbow thrust (Fig. 3). (A3)
- (8) Geographical site location of the fieldwork  
Detailed descriptions of the sampling localities were not provided, but the location is likely *near sample 94MR355* (Figure 1), as granites intruding basic rocks are only mapped and observed *along the road in this area*. (B3)  
The fault and fold geometries revealed in *Hillsdale Canyon* and along *Highway 12* however resemble that of fault propagation folds. (A3)
- (9) Self-citation of prior field studies  
The ultimate products of fenitization of both the granitic country rocks and the mafic xenolithic material have nepheline syenitic to ijolitic mineralogy but can be clearly distinguished texturally and mineralogically from the truly magmatic components of the complex (*Author D, 1992*). (D4)

By far, the task that writers of the field account spend the most time doing is demonstrating that they master the current interpretive frames used by the community. 'Disciplinary cues' allow them to describe the field and its structures using a specialist's terminology, including nominal and adjectival field descriptors (10). In using this terminology, writers show that they are familiar with geological structures and how they are expected to co-occur (12). Writers also imply they know how the terrains should be interpreted, given their knowledge of the community's currently used scenarios and references to others' published fieldwork (13).



Finally, disciplinarity cues reify the community's interpretational frames by resituating the structures within a framework of geological time (11).

Disciplinarity cues

- (10) Nominal or adjectival field descriptors  
The *blocks* consist of: (1) local *basement rocks* (Hercynian *granites* and *metamorphic rocks*); (2) *lava clasts* (*basalts*, *trachytes* and *phonolites*); (3) various coarse-grained *rocks* displaying *cumulate textures*; (4) various *pyroclastic fragments* with more or less diffuse *boundaries* with the *host tuff*. (C5)
- (11) Descriptors of geological time/age  
Above an unconformity, strata at the latter locality include schistose *Lower Carboniferous* conglomerate and siltstone (550 – 600 m) overlain by *Upper Carboniferous* shallow marine clastic, limestone, and locally intermediate to basic volcanics (600 – 800 m). (B3)  
Further downship along the thrust moderately dipping *Cretaceous* strata in the hanging wall rest in thrust fault contact on upturned *Eocene* and *Cretaceous* beds. (A3)
- (12) Technical verbal adjectives and participles  
The south branch of the Rubys Inn thrust is marked by south *dipping* *Cretaceous* beds *resting* in thrust fault contact on footwall *Claron* Formation. (A3)  
On the other hand, the crystallization of both subzone B and the upper part of subzone A of the critical zone postdated the completion of the Steelpoort pericline, because the steeply *dipping* western limb of the structure is *onlapped* by gently *dipping* *cumulates* that overlie the lower chromitite layers south of Steelpoort (Hiemstra and Van Biljon, 1962; Cameron, 1971, Sharpe, 1981). (D1)
- (13) Citations of others' fieldwork  
*As Suppe (1985) emphasizes* such folds can lock and when this happens the fault may branch into two surfaces that propagate along the synclinal and anticlinal axial surfaces. (A3)

All three of these rhetorical functions, along with their corresponding genre variables, play an important role in the construction of an appropriate disciplinary voice, by helping experienced writers construct their authority and credibility before a community of specialists.

#### 4.1. Measuring standard deviation in the field account (Corpus 1)

A measure of SD was applied to determine the range of variation attested for each of the 13 variables in Corpus 1. Table 2 shows in column (1) the total number of variables identified in the corpus, and in column (2) the average number of variables per field account ( $\mu$ ). In column (3), the measure of each variable's SD is given. Column (4) shows each variable's spread range, or range of variation. The spread range for each variable was determined by adding and subtracting the SD from the average value ( $\mu$ ) obtained in column (2). The spread range provides a numerical representation of each variable's normal range of use, allowing the analyst to evaluate the degree of adherence to the attested genre norms by individual writers. All numbers in Table 2 have been rounded, and negative values are not given (represented by a zero).

In order to allow for comparison between the corpus and each writer's use of the variables, the actual values of Corpus 1 were normalized by dividing the number of occurrences by the total number of words in each field account (Table 3).

Table 2: *Field account (FA) cue occurrences and range of variation* (Columns 1 & 2 based on results reported in Dressen-Hammouda, 2008)

	(1) Total # of occ.	(2) Avg # per FA	(3) SD	(4) Range of variation
No. of words per FA	67312	1036	950	86 – 1986
No. of variables per FA	32959	507	472	35 – 979
<i>Personalization</i>				
1. 1 <sup>st</sup> person pronouns/poss. adj.	2723	42	41	1 – 83
2. Evaluative adj. and adv.	56	1	2	0 – 3
3. Interpretive comments	2025	31	33	0 – 64
	642	10	10	0 – 20
<i>Doing-the-work</i>				
4. Nominal and verbal activity markers	8834	136	153	0 – 289
5. Metric, angle, direction measures	1386	21	24	0 – 45
6. Locational adverbs and prepositions	1815	28	38	0 – 66
7. Metadiscoursal refs. to visual data	2629	40	46	0 – 86
8. Geographical location of fieldwork	829	13	18	0 – 30
9. Self-citation of prior field studies	1829	28	37	0 – 65
	346	5	7	0 – 12
<i>Disciplinarity</i>				
10. Nominal & adjectival field descrip.	21402	324	294	30 – 618
11. Geological age descriptors	17381	267	243	24 – 511
12. Verbal adjectives and participles	879	14	15	0 – 29
13. References to others' fieldwork	2554	39	38	2 – 77
	588	9	9	0 – 18

Table 3: *Normalized values for the field account's variables*

	Avg # of uses	Range of variation	of SD
No. of words per FA	1036	86 – 1986	950
No. of variables per FA	507	35 – 979	472
<i>Personalization</i>			
1. 1 <sup>st</sup> person pronouns/poss. adj.	.040	.023 – .057	.017
2. Evaluative adj. and adv.	.001	.000 – .003	.002
3. Interpretive comments	.030	.013 – .047	.017
	.010	.005 – .015	.005
<i>Doing-the-work</i>			
4. Nominal and verbal activity markers	.121	.087 – .155	.034
5. Metric, angle, direction measures	.020	.006 – .034	.014
6. Locational adverbs and prepositions	.024	.012 – .036	.012
7. Metadiscoursal refs. to visual data	.036	.029 – .043	.007
8. Geographical location of fieldwork	.010	.000 – .027	.017
9. Self-citation of prior field studies	.025	.008 – .042	.017
	.006	.000 – .013	.007
<i>Disciplinarity</i>			
10. Nominal & adjectival field descrip.	.328	.234 – .422	.094
11. Geological age descriptors	.272	.199 – .345	.073
12. Verbal adjectives and participles	.017	.000 – .034	.017
13. References to others' fieldwork	.038	.022 – .054	.016
	.010	.001 – .019	.009

The greater the SD in the final column of Table 3, the greater the variation attested in the corpus for that particular variable. A wider range of variation implies that some variables may be allowed a more flexible range of use than others (e.g., 2, 4, 5, 7, 8, 10, 11, and 12). In contrast, a number of other variables (e.g., 1, 3, 6, 9, and 13) show relatively little variation (less than 1%), and can thus be expected to be used with approximately the same regularity across different articles.

The preliminary results of a series of reader response studies (e.g., Paul & Charney 1995; Paul et al. 2001; Tardy & Matsuda 2009) with experienced field geologists indicate that experienced disciplinary readers are quite sensitive to the presence and/or absence of the identified variables in their evaluation of a writer’s expertise. Building on observations made by Tardy and Matsuda (2009), it is hypothesized that specialist readers are able to correlate this use with their perception of the author’s identity and level of expertise.

The following section will describe how the five individual writers use the variables in increasingly sophisticated but similar ways over the course of their career, lending credence to the validity of this hypothesis.

#### 4.2. Individual writers’ use of the variables over time (Corpus 2)

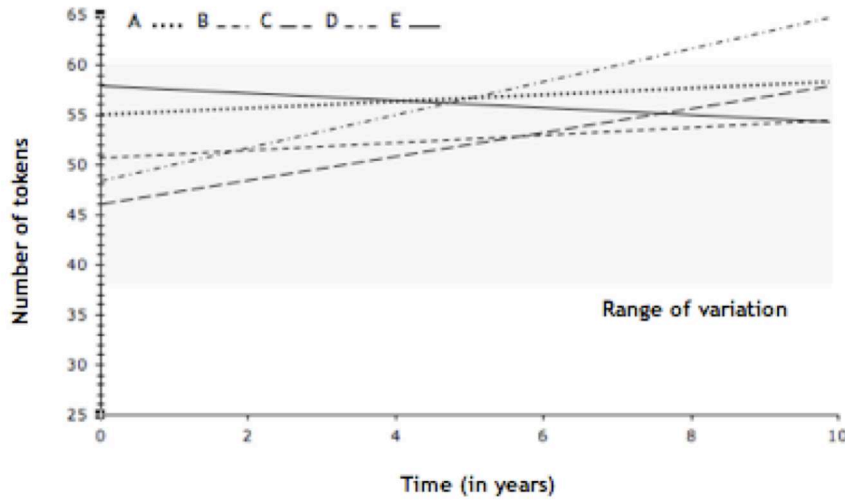
This section describes the variation observed in the use of the thirteen variables by the five writers during the ten-year period following their first research article, published at the time of their PhD dissertation, and the last published article included in the study (Table 4).

Table 4. Comparison of authors’ use of field account variables over time

	Variable density		Personalization cues		Doing-the-work cues		Disciplinary cues	
	First	Last	First	Last	First	Last	First	Last
A	.544	.580	.070	<b>.088</b>	.137	<b>.189</b>	.338	.304
B	.508	.566	.054	<b>.074</b>	.180	<b>.209</b>	.274	.282
C	.461	.582	.032	<b>.088</b>	.152	<b>.182</b>	.276	.312
D	.475	.637	.067	<b>.113</b>	.159	<b>.185</b>	.249	.339
E	.581	.557	.064	<b>.100</b>	.165	<i>.104</i>	.352	.352
SD	.386 – .594		.023	<b>– .057</b>	.087	<b>– .155</b>	.234	– .422

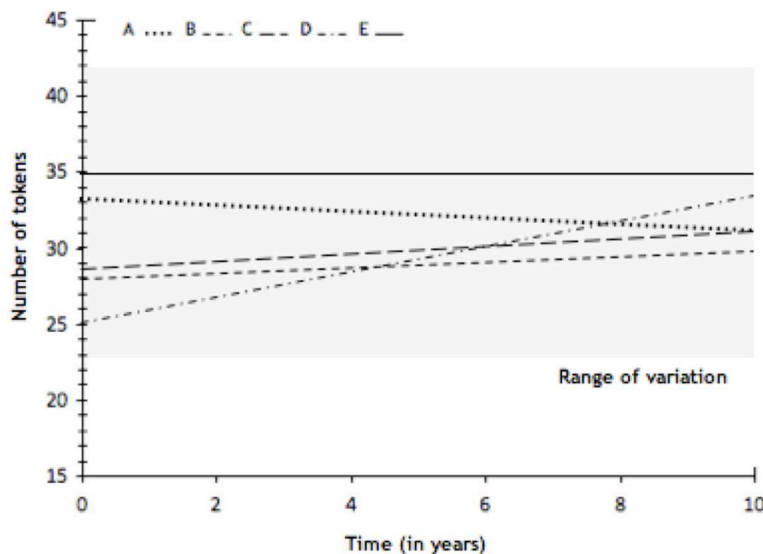
One first observation to be made is that over time, the writers’ overall use of the variables, indicated as ‘Variable density’ (Table 4, Figure 2 below), remains consistent with the corpus, with most writers staying within the corpus’ spread range (SD .386 – .594). Variable density refers to the total number of variables used per field account. This relative conformity is one advance indication that even in their earliest publications, the authors’ field writing strategies reflect their familiarity with the discursive practices of the research community, which is to be expected given their community standing earned as a result of the PhD dissertation. It is likely that less experienced or novice disciplinary writers produce a much lower density.

Figure 2. *SD analysis of variable density*



Significantly, what the consistency of density across individual writers also indicates is that the amount of text that experienced writers dedicate to describing their fieldwork does not increase radically over time. One might expect that more established field geologists would be granted more leeway – and space – in recounting the travails of their fieldwork. Based on the analysis of the writing strategies of these five geologists, however, this is clearly not the case. As seen in Figure 2, for example, while the variable density in the authors’ last article is generally higher than in the first, it remains within the normal spread range (with the exception of Author D, who consistently exceeds the spread range, except in his use of disciplinarity cues). Therefore, although the number of variables used may increase over time, it is likely that it is not the absolute number of words a writer uses to talk about the field that demonstrates and maintains her or his disciplinary expertise; but the *way* in which things are said.

Figure 3. *SD analysis of disciplinarity cues*



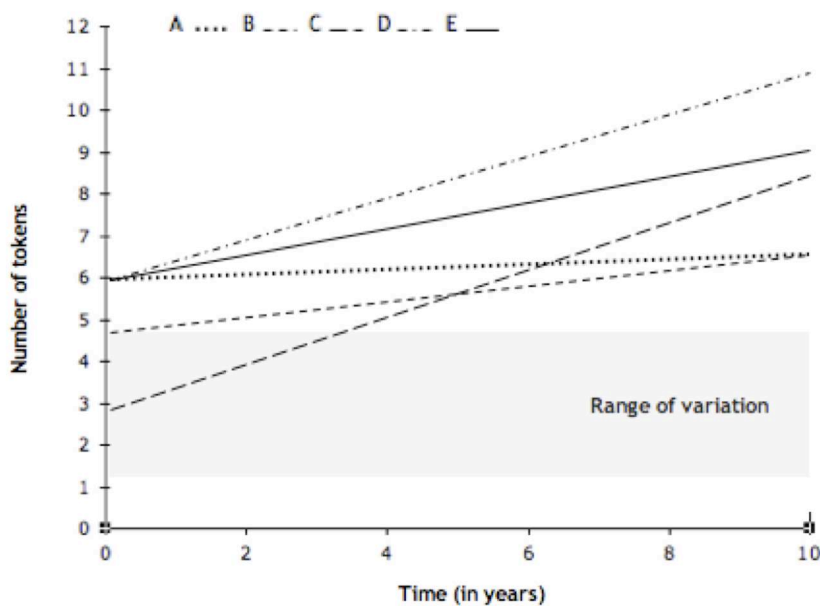
A number of other observations can be made by comparing how the individual writers use the three sets of variables over time, in comparison to Corpus 1. As can be seen in Figure 3, for example, the number of disciplinarity cues used never deviates from the spread range. This overall conformity suggests that using more disciplinarity cues over time is not how an experienced writer seeks to prove and maintain her or his credibility and authority.

A particular effort to prove credibility and authority is apparent, however, in the ways in which authors wield personalization and doing-the-work cues over time. Not only does the use of these variables tend to increase, but their use in the final text also surpasses the corpus spread range (Table 4). In effect, as more senior geologists, the writers all use more personalization cues in their writing than they did as junior geologists who had recently completed their dissertation. The same is true for doing-the-work cues, with the exception of author E, whose use of these cues decreases over time.

#### 4.3. Measuring the development of discorsal expertise?

The expression of discorsal expertise is complex, of course, and does not simply result from a writer just ‘showing himself’, which might imply, for example, the use of more personal pronouns or possessive adjectives over time. Instead, the authors’ growing demonstration of their disciplinary and discorsal expertise seems to be linked to their evolving use of the variables of field writing. We know from the previous section that the authors do not just use *more* variables in their writing because this number was seen to remain relatively constant over time. Instead, as suggested in Figures 4 and 5, the expression of their expertise may possibly result from a complex blending of variables.

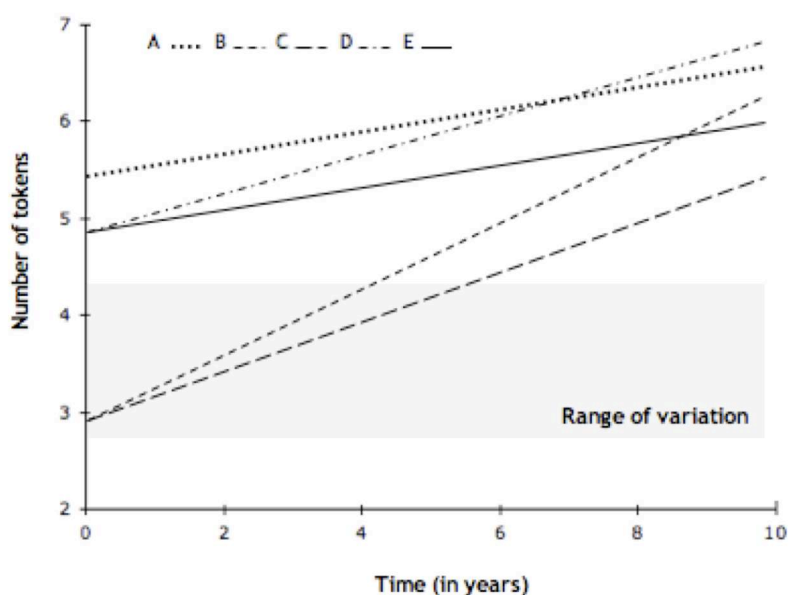
Figure 4. SD analysis of Variable 2 ‘Evaluatives’



While none of the writers showed a substantial increase in their use of personal pronouns and possessive adjectives (variable 1) or interpretive comments (variable 3), they did, however, consistently use more evaluatives (variable 2) over time (Figure 4).

Likewise, the writers consistently used a small handful of doing-the-work cues more frequently over time (e.g., 4, 5, 8). Variable 6 in particular — the locational adverbs and prepositions that indicate researcher movement in the field — shows a significant increase in frequency in the final publication in comparison to the first (Figure 5). This observation holds across the board for all authors.

Figure 5. *SD analysis of Variable 6 'Locationals'*



Based on the writing strategies of these five geologists, two strategies thus appear to differentiate senior field geologists' writing from their more junior colleagues'. First, more experienced writers will more clearly situate their position toward their field observations by marking them with evaluative adjectives (variable 2). This, combined with objective fieldwork descriptions, attests to a greater appropriation of the research subject. Second, locational adverbs (variable 6) become more frequent, allowing the specialist reader to more clearly 'see' the researcher actually in the field. In this way, experienced writers – whether consciously or not – move to more definitively situate themselves in the field, thereby providing support for their authority and expertise.

## 5. Discussion and conclusions

Several implications can be drawn from these observations. The first is that a genre corpus, if it is broadly constituted of texts chosen at random without consideration of individual writers' level of experience, effectively cancels out the effects of disciplinary expertise and individual variation in how a genre is constructed. The genre corpus reflects, in a very broad sense, what

the ‘typical’ range of writing styles for a particular genre is, without regard to particular individual writers’ level of familiarity with the genre’s discursive conventions, nor amount of experience in the discipline. Such a corpus therefore does not necessarily represent the writing strategies that characterize ‘the most experienced’ or ‘the least experienced’ writers, but merely the most frequently used strategies.

In addition, as writers gain stature in their research community, moving from junior to more senior researcher while at the same time becoming more institutionally well-established, they appear to focus more closely on just a handful of a specific genre’s set of variables. This has been seen here in the more frequent use of certain personalization and doing-the-work cues, most often extending beyond the corpus’ spread range (Table 4). It is suggested that this positive deviation from the norm may lead an informed reader to establish a writer’s level of expertise and credibility as a more senior writer – one who may therefore claim more flexibility to bend the rules of the genre (Kress & Knapp 1992).

A third point of reflection concerns the increasing attention scholars of academic discourse have been drawing to the importance of the ‘interpersonal’ nature of scientific writing (e.g. Mur-Dueñas et al. 2010), notably looking at how interpersonal markers allow writers to construct a voice of expertise. Particular attention has been paid to markers of ‘personality’ in scientific research articles, such as first person pronouns, as a means of investigating how expertise is constructed discursively,

Based on an analysis of person markers across disciplines, Lafuente-Millán (2010: 53), for example, has recently concluded that different disciplines allow for different ways of constructing authority:

In the increasingly competitive world of academia, the creation of an appropriate authorial identity by means of self-mention resources is essential for researchers in order to present themselves as competent and reliable members of the discipline, and to persuade readers about the relevance of their contributions. However, the results presented here suggest that the way writers construct this authorial self varies according to the specific epistemological and social norms of their own disciplinary communities.

Similarly, the results described in this chapter have shown that a discipline like geology provides authors with far much more than just person markers to “present themselves as competent and reliable members of the discipline.” Indeed, disciplinary practices make available a whole range of cues which writers can use to construct their credibility and authority.

Hyland (2010: 122-125) makes a similar point in his description of the markers that intervene in “proximity” building. Some of the markers he has identified across a range of disciplines include how authors handle their discussion of research methods, citation practices, grammatical ‘objectivity’, use of modality and hedging to create appropriate stance and denote a personal attitude, as well as their ability to create reader engagement within the text. The cross-disciplinary markers Hyland describes are quite similar to the results found for the geology corpus described in this study. However, as locally observed by Mur-Dueñas et al. (2010) for their own study, while some of the markers field geologists use to build credibility

in their published field accounts are used in other scientific disciplines (e.g., personalization cues, evaluatives, citation practices), others, such as locational adverbs, are clearly discipline-specific and are thus tied to the particularities of disciplinary practice. So although it is important to carry out large-scale research projects to describe scientific and academic writing in general, particular attention does need to be paid to the specificities of each community of practice.

A final topic of reflection concerns the possible pedagogical benefits of using SD. While pains have been taken in earlier parts of this chapter to discuss how using a measure of SD can broaden our analysis and description of genre norms, the question remains to be seen whether such methods may actually be useful for developing teaching materials.

As the results of this study have shown, the markers of disciplinary voice shift over time in ways that are comparable between different writers, who come to use the affordances of their disciplinary genres in ways that are strikingly alike. What this implies is not only that the resources people use to express themselves are characterized by recognizable regularities, but also that the processes at work which cause the shift in self-expression over time are in fact remarkably similar from one individual to another. This observation is all the more surprising given that we might expect for more experienced authors to gain a more idiosyncratic voice which differentiates them from other scholars. In all likelihood they do, although quantitatively speaking, measuring the idiosyncrasies that identify particular individuals is so complex that for now, at least, capturing the specificities of individual voice is still beyond our grasp (i.e., Elbow 1999; Hyland 2008).

At the same time, significant variation between individuals is also a defining aspect of disciplinary writing. Although individual writers develop remarkably similar strategies in their use of genre variables, they clearly do not all write alike. Variation is therefore an extremely important part of the equation as well, and corpus-based teaching methods could usefully draw on the types of variation revealed by SD analysis. Using the results of SD in the classroom may help students gain a better grasp of variation, showing them nuances in meaning, what counts as acceptable variation and what does not, and showing students how they can manipulate structure to create acceptable discourse. However, before engaging them in pattern analysis, it would first be useful to explain the variables in a way which goes beyond the identification of grammatical categories or lexical items, by tying the variables to disciplinary practice and socio-historical context: why is it important for experienced geologists, for example, to say they were in the field? How do they show they are competent field geologists? Linking lexical items to contextual meaning could help students become more sensitive to identifying meaning that is not marked on the page, but is nonetheless a crucial part of evaluating disciplinary competence.

Students can also work with a corpus of writing by experienced researchers, in order to gain experience in finding the variables described by SD analysis. Similarly, they could compare the use of the variables between an experienced-writer corpus and a learner corpus, to help them discern the differences in expression between novice and more experienced writers. This in turn, could help develop their awareness of their own strategies and how they might adapt them to the contingencies of their own evolving situations.



The fine-grained analyses achieved by measuring SD within a corpus of texts can thus provide practitioners with more detailed pedagogical tools to help both native and non-native English speaking writers become more aware of the gradual shifts that will need to take place in their own disciplinary voice. An awareness of the variations and norms involved in the process of developing discorsal expertise would help both student writers moving into the discipline and onto the international publishing scene, as well as non-native English researchers who have become well-established in their disciplines, but may have difficulty getting published in English.

## **Notes**

1. For the sake of clarifying the three types of rhetorical function, the cue types have been renamed from the earlier publication: here, 'Personalization cues' refer to Move 1 cues, 'Doing-the-work cues' refer to Move 2 cues, and 'Disciplinary cues' refer to Move 3 cues, all of which were described in Dressen-Hammouda (2008).

## References

- Ayers, G. 1994. *Are Abstracts Changing? A preliminary investigation through the analysis of the short texts accompanying the articles in Nature*. MA dissertation, University of Birmingham.
- Beaufort, A. 1999. *Writing in the Real World: Making the transition from school to work*. New York: Teachers College Press.
- Berkenkotter, C. & Huckin, T. 1995. *Genre Knowledge in Disciplinary Communication: Cognition, culture, power*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Bhatia, V.J. 2002. Applied genre analysis: A multi-perspective model. *Ibérica* 4: 3-19.
- Dressen, D. 2002. *Accounting for Fieldwork in Three Areas of Modern Geology: A situated analysis of textual silences and saliences*. PhD dissertation, University of Michigan.
- Dressen-Hammouda, D. 2008. From novice to disciplinary expert: Disciplinary becoming and genre mastery. *English for Specific Purposes* 27: 233-252.
- Dressen-Hammouda, D. Forthcoming. Measuring the voice of disciplinarity in scientific writing through indexicality and the implicit. *English for Specific Purposes*.
- Elbow, P. 1999. Individualism and the teaching of writing: Response to Vai Ramanathan and Dwight Atkinson. *Journal of Second Language Writing* 8: 327-338.
- Hyland, K. 2000. *Disciplinary Discourses: Social interactions in academic writing*. London: Longman.
- Hyland, K. 2008. 'Small bits of textual material': A discourse analysis of Swales' writing. *English for Specific Purposes* 27: 143-160.
- Hyland, K. 2010. Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes* 9: 116-127.
- Ivanic, R. 1998. *Writing and Identity: The discursual construction of identity in academic writing*. Amsterdam: John Benjamins.
- Kress, G. & Knapp, P. 1992. Genre in a social theory of language. *English in Education* 26: 4-15.
- Lafuente-Millán, E. 2010. 'Extending this claim, we propose...' The writer's presence in research articles from different disciplines. *Ibérica* 20: 35-56.
- Matsuda, P. 2001. Voice in Japanese written discourse: Implications for second language writing. *Journal of Second Language Writing* 10: 35-53.
- Matsuda, P. & Tardy, C. 2007. Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes* 26: 235-249.
- Mur-Dueñas, P., Lorés-Sanz, R. & Lafuente-Millán, E. 2010. Editorial for special issue on interpersonality. *Journal of English for Academic Purposes* 9: 83-85.
- Paul, D. & Charney, D. 1995. Introducing chaos (theory) into science and engineering: Effects of rhetorical strategies on scientific readers. *Written Communication* 12: 396-438.
- Paul, D., Charney, D. & Kendall, A. 2001. Moving beyond the moment: Reception studies in the rhetoric of science. *Journal of Business and Technical Communication* 15: 372-99.
- Rosch, E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology (General)* 104: 192-233.
- Sinclair, J. (ed.). 2004. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Swales, J.M. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Tardy, C. & Matsuda, P. 2009. The construction of author voice by editorial board members. *Written Communication* 26: 32-52.