

## **LOOKING AT LANGUAGE IN THE HOSPITALITY INDUSTRY**

**Margaret Healy**  
**Mary Immaculate College**  
**Limerick, Ireland**  
[Margaret.Healy@mic.ul.ie](mailto:Margaret.Healy@mic.ul.ie)

**Kristin Onderdonk-Horan**  
**Shannon College of Hotel Management**  
**Shannon, Ireland**  
[kristinhoran@shannoncollege.com](mailto:kristinhoran@shannoncollege.com)

## ABSTRACT

This paper looks at the discourse of the hospitality industry, specifically the use of English in training students to become future hotel managers. Shannon College of Hotel Management has participated in a unique, year-long, collaborative research project, in partnership with Cambridge University Press and Mary Immaculate College Limerick, in order to capture the language used by both native speakers and non-native English speakers in the context of hotel management training. The resultant CLAS Corpus, a one-million word spoken corpus, when complete, will provide an up-to-the-minute, invaluable resource for both lecturers and students in terms of their English language use, the industry's lingua franca.

The current research trend in corpus linguistics is towards the development of industry-specific and spoken corpora. The CLAS corpus incorporates both these features. Audio recordings of various speech events at SCHM were collected, providing a balanced matrix of lectures, language classes, practicals and student presentations across the three student years. Scott's (2009) WordSmith Tools 5.0 will be used to examine the discourse in detail, providing wordlists, keyword lists and concordances which will be analysed using Wenger's Community of Practice (1998) framework.

Specifically, the data will be analysed for 'hotelspeak' - the keyword lexis specific to our industry - and various layers of meanings in the shared repertoire (Wenger, 1998). Through error coding, investigation of non-native speakers' use of language will identify typical learner patterns, providing significant insights on common errors shared within various first language groups, thus facilitating tailored teaching practices. Additionally, these data will inform the development of more specialised and pragmatic teaching materials, based on the corpus results. Some preliminary findings from the corpus illustrating some features of the hotel discourse will be presented.

## **INTRODUCTION**

The development of the Cambridge, Limerick and Shannon (CLAS) Corpus<sup>1</sup> is significant in that it is the first substantial corpus comprised of English language used in the hospitality industry. Language forms an essential part of success in hospitality, and this corpus will offer insights into the discourse of both native and non-native speakers of English, as well as providing researchers with the data to map the progression of language use of speakers from an introductory level through to expert within the industry.

This information is increasingly important for educators in the hospitality industry, hotel managers and managers responsible for staff training as successful interactions with clients are often based upon efficacious use of language, including the ability to master politeness, formality, and hospitality-specific lexis. Evaluation of the CLAS corpus will provide insights into the language currently used in education within the industry, learner language development for non-native English speakers, and language data for materials creation. These aspects of learner language acquisition will be beneficial to teaching English for Hospitality as currently there is a dearth of English language materials with a hospitality focus, and more specifically a hotel focus, above Intermediate level (see Appendix 1 for an overview of textbooks available). In the broader spectrum, the CLAS Corpus will be used to help further clarify the reference-level language descriptors currently used for English in the Common European Framework of Reference for Languages (CEFR) and will feed into the English Profile Programme.

## **BACKGROUND TO THE STUDY**

### **Shannon College of Hotel Management**

Shannon College of Hotel Management (SCHM) was approached in 2008 by researchers from Cambridge University Press (CUP) and Mary Immaculate College (MIC), Limerick, and was invited to participate in this study. The suitability of SCHM for this research was determined as a result of numerous factors which included: the diversity of the student body, the occurrence of certain recordable speech acts (such as: lectures, practical classes, presentations, interviews, meetings) and, specifically, its setting as a specialised academic location which had not previously been studied.

---

<sup>1</sup> A corpus is a collection of spoken (i.e. recordings) or written language which is stored on a computer and available for quantitative or qualitative analysis (see Biber et al 1998).

SCHM is a specialised college which offers two main programmes of study: a Bachelor of Business Studies Degree in International Hotel Management (BBS) and a Bachelor of Commerce Degree with a National University of Ireland (NUI) Diploma in International Hotel Management (BComm). Both are Level 8 degrees awarded by the NUI, with the BBS being taught entirely in Shannon and the BComm having fourth year students leaving the Shannon campus to complete the final year in NUI, Galway. A brief overview of the programme structure can be seen below in Table 1. The first year of the degree has a high number of contact hours with an average of 23 hours per week, divided between both lectures and practical classes. Years Three and Four have a more theoretical focus with fewer contact hours. For a more detailed synopsis of the course, including course content, see Appendix 2.

<b>Year</b>	<b>BBS</b>	<b>BComm</b>
1	Lectures in SCHM	Lectures in SCHM
2	12 month work placement	12 month work placement
3	Lectures in SCHM	Lectures in SCHM
4	Lectures in SCHM	Lectures in NUI Galway
4 <sup>+</sup>	9 month trainee management placement	9 month trainee management placement

**Table 1: Progression of students on degree programmes in SCHM.**

One striking feature of SCHM, which makes it ideal for this study, is the number of international students studying in the college. Table 2 shows a breakdown of the number of Irish and International students in each of the four years for the academic year 2009- 2010.

<b>Year</b>	<b>Irish</b>	<b>International</b>	<b>Total</b>
1	38	56	94
2	39	63	102
3	37	28	65
4*	38	36	74

\* This data includes the students who are in Galway on the BComm.

**Table 2: Students by nationality (Irish and International) in 2009-10.**

The international student body is quite diverse with 23 nationalities represented, however, the largest percentage of overseas students come from China (54%). Linguistically, 47% of the student body are native English speakers and 53% non-native. The minimum English language

entry requirement for non-native speakers is a B2-C1 (IELTS 6.0- International English Language Testing System).

### **English Profile**

The English Profile network is a collaborative research project, the purpose of which is to develop reference level descriptors for English to be referenced as part of the CEFR. These descriptors will “provide detailed information about the language that learners can be expected to demonstrate at each level, offering a clear benchmark for progress that will inform curricula development as well as the development of courses and test material...” for use in teaching English as a foreign language (<http://www.englishprofile.org/>). The CLAS corpus’s role in the English Profile Project is to provide learner data related to the hospitality industry and to clarify reference level descriptors mainly around the B2, C1, C2 areas. For more information about the English Profile Project see Appendix 3.

### **RELATED STUDIES**

The field of Corpus Linguistics has been gaining momentum over the last thirty years, in line with the development of computers and software. From the start, these corpora were used not only by the researchers involved in their creation, but also by publishers, who recognized the value of using real data, particularly in dictionary production. The first major machine-readable English corpus, the Brown Corpus, was created in the U.S. in the 1960s by Kucera and Francis. This corpus covered a variety of areas and consisted of one million words. Following the publication of *Computational Analysis of Present-Day American English* (1967) by Brown and Kucera, the data was then used by Houghton-Mifflin Publishers to publish the first corpus-based dictionary in 1969. The next major corpus project was the creation of the Collins Birmingham University International Language Database (COBUILD) corpus, called The Bank of English, by Sinclair et al. (1980) in conjunction with Collins Publishers. The analysis of this corpus formed the basis for the publication of the Collins’ COBUILD English dictionary, the first monolingual learner’s dictionary based on corpus data. This trend has continued with other major projects including the Cambridge International Corpus (CIC), International Corpus of English (ICE), and many others – see Appendix 4 for details of a small number of current corpora consulted in our research. Although the corpora mentioned here were primarily used for creating dictionaries, O’Keeffe, McCarthy and Carter (2007) suggest that the main uses for corpora are lexicography, grammar, stylistics and translation.

Most corpora are still focused on written language, although more spoken corpora are now

being compiled. The largest corpus currently covering spoken English is the British National Corpus (BNC) with ten million spoken words. While the focus now has shifted from written to spoken corpora, the tendency currently is also to create and develop new spoken corpora in specific vocational and professional areas, called specialised corpora.

### **Specialised corpora**

A key feature of the successful hotel, as previously mentioned, is often determined by direct face-to-face contact with the client, the personal touch in dealing with people while providing the service. This service encounter model has been investigated in relation to several other sectors, for example in travel agents (Coupland and Ylanne-McEwen, 2000), petrol stations/small shops (Binchy, 2000), bookshops (Aston, 1988). In the academic world too, where the CLAS corpus is also situated, specific small purpose-built corpora are being developed examining the professional discourse of that metier, for example, Vaughan (2007) investigated teacher talk within the context of a community of practice, a parallel framework which can also be applied to the hospitality sector.

From the early 1960s, Gumperez (1962, 1968, 1971, 1972) did considerable and seminal work on speech communities and Swales (1985, 1988) developed further on these concepts in terms of discourse communities. More recently Lave and Wenger (1991) and Wenger (1998) have expanded on the framework of communities of practice, taking as the concept of *legitimate peripheral participation* as the basis for learning and induction into the relevant professional and vocational arena. This framework is particularly appropriate for the research being carried out at SCHM, a location where apprentices in the hospitality industry are trained through practical and legitimate participation in the work of the industry, as well as gaining the theoretical knowledge in an academic setting. Wenger (1998) identifies three key elements which constitute a community of practice and by which membership can be recognised, namely:

- there must be a *joint enterprise* which the members negotiate collectively, a common purpose,
- there must be *mutual engagement* of the members in the enterprise and its furtherance,
- there must be a *shared repertoire* of expression by the members.

Training people (whether native or non-native speaker) in specialised areas requires the learner to “control that subset of the English language which the particular community has extracted to

form its particular style” (O’Toole: 1994). The relationship between communication skills and success in the hospitality industry cannot be underestimated. “...Strong communication both verbal and non-verbal between staff and guests is an essential component of customer service... helping employees learn English produces a more cohesive and self-confident staff, better guest experiences and higher perceptions of value and professionalism.” (Sheridan: 2004: 22). Blue and Harun (2003: 77) noted that the hospitality register in English is vast and that a number of functions of hospitality language such as greeting a guest, requesting information and salutations necessitate some level of knowledge of the pertinent phrases before someone can operate successfully in guest-contact positions. Blue and Harun (ibid) go further to describe the minimum ‘cluster’ of language skills that staff dealing with guests should have acquired, namely:

- how to address a person;
- how to solicit and give the necessary information;
- how to respond to questions/requests;
- how to use prompts;
- how to use gestures;
- how to deal with difficult customers;
- how to appease complaints.

Some hotels do offer training programmes for employees, however, many do not have the funding to provide such programmes, and tailoring courses for each department, to provide the necessary spoken skills for that department, can prove expensive and time consuming. Thus, most hotels would expect that their staff members are trained in language skills before being employed, underlining the need for linguistic training in hospitality-focussed educational institutions.

### **Using the data**

O’Keeffe et al. (2007, 5/6) point out in their book *From Corpus to Classroom* that a spoken corpus is more time-consuming to create than a written corpus because, first of all, the speech has to be transcribed and one hour of speech can take approximately two working days to transcribe. In addition to this, there is also the requirement to tag the recorded speech for many other elements which will allow for qualitative assessment of the data at a later stage. These tags would include many of the non-linguistic features of spoken language, such as laughter, overlap, hesitation, interruption, and other such vocalization traits.

Once a specialized corpus, such as the CLAS corpus, has been built with carefully selected materials, the next step is to analyse the data. As noted by Biber et al. (1998) one of the benefits of using a corpus is that both quantitative and qualitative techniques can be used. Methods such as discourse analysis and conversation analysis have become widely used in investigating corpus data. Sinclair (1991: 3) also points out that the growth in corpora, and more specifically the investigation of collocation, semantics and pragmatics, has changed the way we study and teach language. Examination of these elements will be investigated and analysed based on the corpus findings and the discourse of this community of practice at SCHM will be primarily evident through its shared repertoire (Wenger: 1998) as demonstrated by the findings of the CLAS corpus.

In the past, examples of language use were created by teachers and writers to demonstrate grammar points and vocabulary use, but now actual speech events are used, providing authenticity and the opportunity to maintain currency with language change. Newer texts are now corpus-based, focusing on actual examples of spoken interaction, and grammar rules now cater for descriptive use rather than just the prescriptive rules. One of the outcomes of this research is to develop further materials in the Hospitality sector, in particular English language training.

In practical terms, data analysis will be constructed using computer software especially designed for corpus analysis, in this case the latest version issued in 2009 of Mike Scott's software suite called Wordsmith Tools (Scott 1999) will be used. Details of the major features of this software will be explained later when discussing some of the preliminary findings.

## **RESEARCH METHODOLOGY**

The first step in the compilation of the CLAS corpus entailed the project leaders from CUP and MIC meeting the faculty and students at SCHM and presenting the project to them. Full co-operation in the endeavour was forthcoming and practical arrangements for data collection were put in place. This began in November 2008 and continued over a twelve month period, covering two academic years. Over one hundred separate academic speech events, amounting to one hundred and twenty hours of data, were recorded using a digital audio recorder and stored in MP3 sound file format. All students and teaching related staff members at the college participated in the project. This involvement necessitated the completion of certain documentation by the participants, such as the Speaker Information Sheet which elicited some

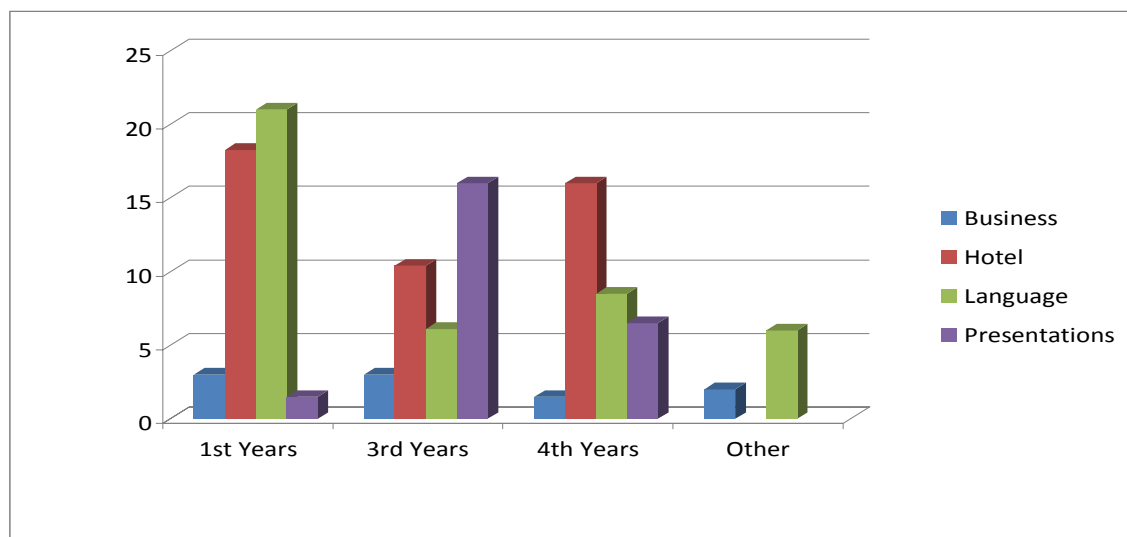
general personal details, information about educational background, and the acquisition and use of English and other languages in day-to-day life. That data has been transferred into a database with almost four hundred entries listing students in 1<sup>st</sup>, 3<sup>rd</sup> and 4<sup>th</sup> years over two separate academic years, the teaching staff and any other occasional participants. The corpus will be linked to this database, thus enabling the identity and personal profile of each speaker to be accessed and attributed to each speaker turn. Furthermore, each participant signed a Consent Form in respect of the data being collected, the copyright of which resides with CUP - with access to it by members of SCHM and MIC, and anonymity is guaranteed by the end users.

For each individual recording event, additional documentation included a Speaker Sign-in sheet; an Individual Recording Details sheet; a Speaker Order sheet; a classroom layout diagram indicating students' seating positions; and the attachment of lecture hand-outs, presentations and any other class materials. The template for all these documents formed part of the Sub-Corpus Agreement contract which was signed on behalf of Cambridge University Press/Cambridge ESOL, Mary Immaculate College and Shannon College of Hotel Management in October 2009 and which covered all legal considerations and ethical guidelines in respect of the CLAS corpus.

As previously indicated, Shannon College was selected as a specialised and discrete locus for this linguistic research offering the opportunity to capture a broad spectrum of academic English across a range of business topics and specific hotel-oriented modules. These data can be subsequently analysed, compared and contrasted with data from other similar and dissimilar corpora, for example, spoken corpora in the academic and business environments (MICASE, CANBEC, LIBEL) and general spoken conversational corpora (CIC, LCIE, CANCODE) – see Appendix 4 for details of these corpora. Furthermore, the contribution of non-native English speaker data in this corpus will facilitate comparisons with other non-native language in other corpora, such as Cambridge ESOL corpus, and, in addition, will feed into the English Profile Programme (see Appendix 3 for details).

The challenge in the CLAS corpus was to ensure a comprehensive matrix of speech events that represented a typical academic year in SCHM's life. All types of recording events representative of daily college life needed to be reflected in the data collected. In summary, the distribution across the four main categories of recordings – general business modules, hotel

specific modules, language for non-native speakers, and student presentations (native and non-native speakers jointly or separately) – across the three student years was well represented, as demonstrated in Table 3.



**Table 3: Summary breakdown of CLAS recordings by year and content.**

Recordings of first year students (37 in number) accounted for almost 44 hours of speech; third years contributed over 35 hours in 31 recordings; fourth years provided in excess of 32 hours in 30 recordings and miscellaneous meetings and CELS<sup>2</sup> language classes amounted to another 8 hours, totaling almost 120 hours overall in 107 recorded speech events. It must be noted that the higher number of recordings and time for first year students reflects the significantly increased teaching contact hours in the first academic year. A full analysis of the recordings by year and content is attached in Appendix 5.

Recordings were made using a digital audio recorder throughout, with the exception of one video recording of a student group presentation. Audio recording only was selected for this corpus as it was generally believed to be the least intrusive method of data collection for all the participants. In reality, the participants became so inured to being recorded that they conducted their participation in a most natural and uninhibited way. Nevertheless, it must be noted that there were the occasional incidents when the speaker(s) became conscious of the fact that they were being recorded. This usually happened after the particular utterance or exchange which caused them then to remember the recorder. Nonetheless, consideration had

<sup>2</sup> CELS, Certificate in English Language Studies, is a year-long preparatory course offered to non-native speakers who hope to enrol in SCHM's degree course or other third-level colleges the following year. Students prepare for the Cambridge IELTS exam at the end of their course.

to be given to the observer's paradox, a term coined by William Labov (1972) to express the notion that "the observation of an event or experiment is influenced by the presence of the observer/investigator". Labov (1972: 3) explained that "...the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation..." In the data collection for the CLAS corpus, the observer's paradox did not generally become an inhibiting factor overall.

The CLAS corpus is a work in progress. Having completed the data collection, the next phase is the transcription process which is currently in train. This too presents its own challenges, particularly in terms of transcription conventions and tags to be used. To-date, approximately 200,000 words have been transcribed – see Appendix 6 for the recordings now transcribed. Using this sample from the expected one million words, certain linguistic features of this environment and microcosm of industry are beginning to emerge which will be discussed hereafter. It is important to point out at this juncture that the sample corpus used here is not a finalised, proofed document so there may be some slight discrepancies in the data presented here when compared with the final, definitive version of the corpus.

## **DISCUSSION OF PRELIMINARY FINDINGS**

One of the main purposes of the CLAS corpus is to identify the linguistic indices within the community of practice of the hotel management training sector. The transcribed corpus will be analyzed using WordSmith Tools 5.0 software (Scott, 2009), the three main features of which are wordlists, keywords and concordances. Wordlists provide a large amount of statistical information regarding the variety and frequency of word token in the corpus. Keywords are words which occur unusually frequently in this corpus in comparison with their occurrence in some kind of reference corpus (ibid). By comparing two corpora, the linguistic features of the hotel management training sector will become apparent and these can be analysed both quantitatively (wordlist frequencies) and qualitatively using a variety of approaches, including Discourse Analysis (DA) and Conversation Analysis (CA). This analysis will provide a synchronic overview of hotel training discourse and, furthermore, will facilitate the process of initiation into this specific community of practice as demonstrated through the language use, in other words, the 'hotelspeak', across the three student years viewed diachronically.

Many linguistic features will emerge from this corpus – anticipated ones such as acronyms (C & B, F & B, SOPs, PMS, FIEs), culinary terminology (velouté, béchamel, posset), specific hotel technology (Picasso, Opera, Fidelio), jargon (rack rate, on the pass, par stock), and indeed other culturally associated linguistic features (trade-names, place names, organisations). For the purposes of this paper and based only on the available transcriptions to date, specific keywords are already identifiable. The reference corpus used for comparative purposes is the Limerick Corpus of Irish English (LCIE) (Farr et al., 2002), a one-million word corpus of casual Irish English conversations.

### Keywords

An investigation of the first 200 keywords (there are 776 keywords in this list) shows that 84 of them are linguistically oriented to the hotel training sector and these include both hotel-specific and general business terminology. Table 4 records the first twenty keywords, showing their order and frequency on the list. The majority of these lexical words present no great surprises in this particular environment but there are a few words which may be unexpected to the casual observer – words such as Opera (no. 14) and Micros (no. 18) both hotel software packages; PMS as the acronym for Property Management Systems; and the word ‘linen’ which, although a predictable hotel word, is here shown to be remarkably high up the list of keyness at no. 15 when compared with the reference corpus LCIE. The Line Order column in the following table refers to the ranking of each word in the full keyword list. For further reference, Appendix 7 contains a list of the first 50 keywords.

No.	Line Order	Word	Frequency
1	4	<b>hotel</b>	596
2	7	<b>management</b>	232
3	9	<b>information</b>	233
4	10	<b>system</b>	307
5	18	<b>room</b>	345
6	19	<b>website</b>	134
7	20	<b>guest</b>	125
8	22	<b>staff</b>	146
9	26	<b>tourism</b>	129
10	27	<b>restaurant</b>	138
11	28	<b>software</b>	112
12	35	<b>guests</b>	86
13	37	<b>rooms</b>	116
14	45	<b>Opera</b>	71

15	46	<b>linen</b>	64
16	47	<b>hotels</b>	100
17	49	<b>training</b>	133
18	50	<b>Micros</b>	58
19	52	<b>tourist</b>	85
20	53	<b>PMS</b>	57

**Table 4: List of the first twenty keywords in a sample from the CLAS corpus.**

### Concordances

Within a corpus, a concordance is the list of all the occurrences of a particular word or word clusters in the total text. Concordances also display the words that surround a particular, selected word; therefore, word patterns can be noted and explored. This is a particularly useful feature from a pedagogical point of view, especially with non-native speakers who are still acquiring the language. From the concordance toolkit, the surrounding source text can be viewed in context for further information. Examining the various texts in which the particular word is embedded allows for a greater expansive understanding of the word with its many variations of meaning, nuance and indeed genre. For example, the word ‘*suite*’ is at no. 35 in the keyword list of this sample (ref. Appendix 7). Investigating the source texts here reveals that in the first lines the word *suite* means the bedroom and its ancillary facilities in a hotel, one of the most common meanings of the word. These examples are taken from first year student presentations on designing a hotel bedroom as an assignment in the Accommodation module.

N Concordance	
22	For example this is the bedroom in the honeymoon <b>suite</b> . As you can see with the um headboard they've
23	lake view and golf side bedroom five one bedroom <b>suite</b> two penthouse and one presidential suite. Only
24	bedroom suite two penthouse and one presidential <b>suite</b> . Only twenty-five bedrooms have been
25	a at a the theme of it was a New York penthouse <b>suite</b> . And um the people that are in the group are
26	a walk-in wardrobe and a beige rug. We have an en- <b>suite</b> with the standard shower sink toilet and now
27	the room so the furniture and the fittings in our <b>suite</b> suits these very simple but smart. Um first we

The second set of meanings for *suite*, used to indicate a group of related software programmes suitable for hotel use, comes from fourth year students who were required to assess various hotel software packages as an assignment in their Hotel Management Information Systems module.

N Concordance

8 the best uh most appealing feature of the Executive Suite is that they have a high speed of operation and  
 9 you go with? We we we're chosen the Executive Suite and cause it's all in one and you can upgrade it  
 10 restaurant or worldwide hotel chain. there's a Execu Suite Hotel PMS and uh it may improve work  
 11 they go uh Touch POS and the another they go uh Suite PMS. And the ones that you would  
 12 over here. We have two softwares Hotellinx Suite PMS and uh Executive Suite hotel PMS. And  
 13 softwares Hotellinx Suite PMS and uh Executive Suite hotel PMS. And the other features of both this  
 14 is a short description of Hotellinx and Executive Suite again. Uh the hotel sta= this suites this is

This is a particularly effective example of the purpose and benefits to be acquired from corpus linguistic studies. In this specific instance, the difference in meaning and usage of the same word by two different levels of students exemplifies one of the hypothesis within this research, namely that community of practice is demonstrated, supported and enhanced by the linguistic elements of this environment.

**Wordlists**

Without dwelling on the myriad of statistical information available from the wordlist generated by the sample under review, a few brief statistics can be noted here. The total number of tokens (i.e. running words) in the text is 198,905, with 8,381 types (distinct words). As is typical in most corpora, the first 100 words are mainly function words and very few have any great lexical content, except perhaps the very basic verbs (be, have, do, know, get). Hotel specific content words are very scarce in the same 100 words. It is revealing to compare the position of some keywords in this sample to their position in the wordlist. Table 5 illustrates this point.

<u>Wordlist Position</u>	<u>Frequency</u>	<u>%</u>	<u>Texts</u>	<u>Word</u>	<u>Keyword Position</u>
1	9,743	4.90	53	<b>the</b>	13
2	6,844	3.44	54	<b>and</b>	12
3	5,043	2.54	54	<b>to</b>	137
4	4,928	2.48	53	<b>you</b>	
5	3,975	2.00	53	<b>of</b>	41
6	3,827	1.92	53	<b>a</b>	
7	3,214	1.62	54	<b>it</b>	732
8	3,157	1.59	53	<b>in</b>	355
9	2,805	1.41	53	<b>that</b>	625
10	2,673	1.34	53	<b>is</b>	131
54	596	0.30	34	<b>hotel</b>	4
128	232	0.01	26	<b>management</b>	7
91	307	0.15	19	<b>system</b>	10
86	345	0.17	18	<b>room</b>	18

208	125	0.06	16	<b>guest</b>	20
192	138	0.07	18	<b>restaurant</b>	27
464	51	0.03	9	<b>suite</b>	96
587	40	0.02	6	<b>reservations</b>	100
550	42	0.02	12	<b>hospitality</b>	101
475	49	0.02	16	<b>friendly</b>	125

**Table 5: Comparison of word order in Wordlist and in Keyword list in sample from CLAS corpus.**

The example word ‘suite’ used previously, while being at no. 96 in the keyword list, is ranked no. 464 in the wordlist. An examination of source text context would reveal whether this word is used more often in regard to hotel rooms and household furniture, or what percentage of its usage is consistent with technical phraseology.

This brief analytical glimpse of the data from the CLAS corpus at this initial stage provides a foretaste of very interesting, exciting and pedagogically informative data that is expected to emerge. Deciphering wordlists, keywords and concordances in the context of current hotel management training discourse will enhance the overall academic and vocational training of the students at the SCHM. Transcription of the corpus is underway and will take some time to be completed. Once done and proofed, the CLAS corpus will take its place on the global arena of corpus linguistics as an invaluable and practical resource to many in the hotel management training sector.

## **CONCLUSIONS AND RECOMMENDATIONS**

The resource that the CLAS corpus is anticipated to become will complement and increase the store of corpus linguistics research worldwide. The specific contribution of this corpus results from the fact that it has been situated and developed in a discrete, defined area of professional and vocational endeavour, an arena that has a pedagogical and, indeed, a commercial imperative. Through their participation and involvement in this project, SCHM has had the opportunity to archive its discourse over the one-year data collection period, significant at the beginning of a new millennium and perhaps altogether more relevant in the current difficult economic times.

There are several benefits from this project for the participant institutions involved. By opening its doors to this international project, SCHM now has a record of its academic discourse in hotel management training. This collection of data can be accessed in the future

and exploited as a resource for self-directed reflective professional development for faculty members. In the context of its own community of practice, the College can interrogate its own discourse and examine how its language contributes to the inherent joint enterprise (Wenger, 1998) for both staff and students alike.

One of the principal objectives of the project was to record and investigate the English language itself, particularly as articulated by the non-native speakers, and to chart their language learning over the four-year course of study. The purpose here is to design and develop future teaching materials more suited to their language needs, specifically in respect of the language requirements of the hotel industry. As mentioned previously, there is an acute scarcity of suitable and advanced-level language teaching materials and, by examining this discourse and specialised hotel lexis in detail, the resources and empirical data will now become available for such materials development.

Complementary to this area of specialist English language teaching materials, the CLAS corpus will also provide an experiential resource of how language is used by native speakers (faculty and students alike) versus non-native speakers in the context of hotel management education and training. This will offer the opportunity to educationalists to present the English language in a way that is linguistically adapted to the specific needs of their profession and which can further supplement the discursal and lexical norms of General Business English as provided by the Bachelors of Business Studies degree offered by SCHM to all its students.

The English Profile Programme has been mentioned earlier and how the CLAS corpus will dovetail into it. This corpus will provide a further resource for language research into competency descriptors for non-native English speaking students at different levels. To date, these competency descriptors have been largely defined by data obtained from written exams at the various proficiency levels. The CLAS corpus data will enhance and refine these standard definitions because much of its learner English has been gathered in what may be regarded as a less daunting environment, in other words, in general language classes and oral presentation and not in an exam scenario.

Through its participation also in this project, MIC, especially through its Inter-Varietal Applied Corpus Studies (IVACS) research centre, has extended its boundaries in terms of

inter-institutional collaborative research and, as a major establishment devoted to teacher training, the benefits from this partnership will resound in the future.

While the CLAS corpus is still only a work in progress, as evidenced from this preliminary discussion of the earliest findings, many advances in hotel management training, linguistic understanding and ultimately appropriate materials development can be eagerly awaited.

## REFERENCES

- Aston, G. (ed) 1988. *Negotiating Service: Studies in the discourse of bookshop encounters*. Bologna: Cooperativa Libraria Universitaria Editrice Bologna.
- Biber, D., S. Conrad and R. Reppen, 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Binchy, J., 2000. *Relational Language and Politeness in Southern-Irish Service Encounters*. (unpublished thesis) University of Limerick.
- Coupland, N. and V. Ylanne-McEwen, 2000. Talk about the weather: small talk, leisure talk and the travel industry, in: Coupland, J. (ed), *Small Talk*, 163-182, Essex: Pearson Education.
- English Profile, <http://www.englishprofile.org/> (accessed 12/05/2010).
- Farr, F., B. Murphy and A. O’Keeffe. 2002. ‘The Limerick Corpus of Irish English: design, description and application.’ *Teanga*, 21: 5-29.
- Gumperz, J. 1962. ‘Types of linguistic communities.’ *Anthropological Linguistics*, 4(1): 28-40.
- Gumperz, J. 1968. ‘The Speech Community.’ In: *International Encyclopedia of the Social Sciences*. London: Macmillan, 381-6.
- Gumperz, J. 1971. *Language in Social Groups*. Stanford: Stanford University Press.
- Gumperz, J. 1972. ‘The Speech Community.’ In: P.P. Giglioli (ed.) *Language and Social Context*. London: Penguin, 219-231.
- Labov, W. 1972. *Sociolinguistic Patterns*. Oxford: Blackwell.
- Lave, J. and E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- O’Keeffe A., M. McCarthy and R. Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- O’Toole, M. 1994. ‘Training in a Second Language Environment.’ *Journal of European Industrial Training*, 18(1): 4-9.
- Scott, M. 1999. *Wordsmith Tools*. Software. Oxford: Oxford University Press.
- Scott, M. 2009. *Wordsmith Tools version 5*. Software. Oxford: Oxford University Press.
- Sheridan, M., 2004. The Ten-Minute Manager’s Guide to Improving Language Skills. *Restaurant and Institutions*. March 15, 2004: 22-23.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Swales, J.M. 1985. 'A genre-based approach to language across the curriculum.' In: Tickoo, M.L. (ed.) *Language Across the Curriculum*. Singapore: RELC, 10-22.

Swales, J.M. 1988. 'Discourse communities, genres and English as an international language.' *World Englishes*, 7(2): 211-220.

Vaughan, E., 2007. 'I think we should just accept ... our horrible lowly status': Analysing Teacher-Teacher Talk Within the Context of Community of Practice. *Language Awareness*, 16/3, 173-189.

Wenger, E. 1998. *Communities of Practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.