The Discourse of Outsourced Call Centres:
A Corpus-Based, Multi-Dimensional Analysis

Eric Friginal[1]

Abstract

This study explores the discourse of outsourced call centres involving Filipino call-takers (or "agents") and American customers engaged in various types of technical and service support transactions. The specific goals of this study are to establish the patterning of linguistic features in call centre discourse and to examine how the speakers use features and patterns of speech across differing transactions. The data for analysis come from a corpus of call centre texts collected in the Philippines (N of texts=364, with approximately 365,630 words). The research design follows a quantitative multi-dimensional framework developed by Biber (1988) for the extraction and interpretation of linguistic co-occurrence in the corpus. Three linguistic dimensions are extracted and interpreted microanalytically: (1) Addressee-Focused, Polite, and Elaborated Information vs. Involved and Simplified Narrative; (2) Planned, Procedural Talk; and (3) Managed Information Flow. Results show variations in the linguistic and textual composition of the discourse of agents and callers across the three dimensions.

1. Introduction

The influx of outsourced call centres from the United States (U.S.) to the Philippines since the late-90s has paved the way for the creation of jobs for Filipino professionals who are able to communicate in English and provide telephone-based customer services to American clients. The Philippines has become one of the major centres for U.S.-based outsourcing, second only to India, because of its tradition of English education, affinity to the American culture, and overall cheap labor market ("Service Alert," 2004; Teves, 2003). The Philippines is producing over 400,000 English-speaking university graduates every year. Of these, around 80,000 are in the fields of information technology, computers, and engineering. Another 110,000 come from business-related fields, such as commerce, finance, and accounting (BPAP, 2007). This broad educational base means that the Philippines has a workforce that is waiting to be tapped by the marketplace to provide knowledge-based processing and back office outsourcing to many international businesses.

The $5 billion-a-year call centre industry in the Philippines employs more than 150,000 individuals and is projected to have a yearly 75% increase in the coming five years ("As India Gets Too Costly," 2006; Tuchman, 2006). The Philippine government continues to invite U.S. companies to relocate their business process operations into the

---

[1] Department of English, Northern Arizona University
   e-mail: eric.friginal@nau.edu

country's major cities by providing tax incentives, improving technology architecture, and focusing on the marketability of its human resources (Friginal, 2004; Magellan Alliance, 2005; Uy, 2004). In a recent statement, the country's Secretary of Labor and Employment said that Filipinos' intelligence, adaptability, industry, and proficiency with the English language have made the Philippines one of the world's top hubs for call centre investments and operations.

Currently, language-based research focusing on outsourced call centre interactions in the Philippines is still limited but its importance has always been recognized by many stakeholders. There is an understandable, urgent demand for effective, high-level language and phone handling skills for Filipino professionals engaged in assisting American callers. Because of this demand, several call centre companies in the Philippines invest a considerable amount of money to train their employees and support measures to acquire data and information that would lead to the production of language training and assessment materials. More linguistic data from various approaches in discourse analyses provide far-reaching descriptions of outsourced call centre interactions that could improve agents' language use and overall quality of service.

1.1 Corpus-Based Research on Spoken Discourse

Although quantitative, corpus-informed methodologies remain under-exploited in the analysis of spoken discourse (Reaser, 2003), Biber's (1988, 1995, 2001, 2006) works that examine the frequency distribution and statistical co-occurrence of linguistic features from various registers suggest a myriad of possibilities in the exploration of spoken data. Similar corpus-based discourse studies, e.g., Aarts and Meyer, 1995; Baker and McEnery, 2005; Leech and Smith, 2005, also offer directions for empirical investigations that attempt to generalize factors explaining the linguistic patterning in corpora. A methodical description of specific register characteristics has been achieved through corpus analysis. Biber's multi-feature, multi-dimensional analytical framework has been applied in the analysis of spoken registers and used in the interpretation of linguistic phenomena. MD data come from Factor Analysis (FA) which considers the sequential, partial, and observed correlations of a wide-range of variables (*cf.* Biber, 1988; Biber and Conrad, 2001 for an extensive discussion of MD analytical procedures). Various MD studies of spoken registers have covered topics such as stance and dialects (Precht, 2000), gender and diachronic speech (Biber and Burgess, 2001; Rey, 2001), sub-registers of conversation (Quaglio, 2004), televised interaction/debates (Connor-Linton, 1989; Scott, 1998), and job interviews (White, 1994).

This study adds data to the body of research in spoken, professional discourse, particularly in the context of outsourced call centre transactions in the Philippines. The extraction of co-occurring features of call centre discourse through MD analysis has not been conducted in previous research. The identification of linguistic dimensions through the statistical co-occurrence of lexico/syntactic items in the corpus offers unique information not extensively surveyed by researchers connected within the call centre industry. As a direct application of MD results, a thorough description of the linguistic usage relative to social or demographic groups, e.g., agent vs. caller, male vs. female agents, in the register possibly helps in understanding the dynamics of talk inherent in

call centre transactions. Subsequent interpretations of data could facilitate the creation of measures in the assessment of service quality and the design of language training materials for agents who are non-native speakers of English.

1.2 Research Questions

The overarching research question investigated in this paper is: What are the patterns of linguistic variation among outsourced call centre transactions? The patterning of linguistic features in the corpus creates interpretable dimensions which correspond to salient functional distinctions within this register. These dimensions help distinguish the discourse of outsourced call centres from other kinds of conversations or spoken interactions. The underlying hypothesis for the research question is that many social categories, e.g., speaker role and gender, nature of service transaction, affect the overall flow of talk and influence linguistic choices in the interactions. To test this claim in future studies, it is important to establish the linguistic characteristics of outsourced call centre interactions and the extent to which interactants differ across dimensional scales. In addition, significance testing of group differences will also provide useful data that prove whether or not variation in the use of features in the corpus is due to chance. Results will help in further interpreting the linguistic patterns in the discourse and in setting forth additional guidelines for future related research.

2. The Corpus of Outsourced Call Centre Transactions

The corpus of call centre transactions was collected in the Philippines over a period of four weeks in July, 2006. A U.S.-owned call centre company shared data for research and sponsored the corpus collection and transcription. The transactions were retrieved following the list of audio files cued in the database of recorded calls for a particular work shift. However, files that were either too long or too short were dropped. The calls that qualified in the corpus used for this paper (N=364) ranged from five to twenty-five minutes in duration. Convenience sampling of audio files was done to ensure a comparable number of files per account and achieve a relatively balanced number of male and female agents. The calls were transcribed into machine readable text documents following conventions used in the collection of the service encounter corpus of T2K-SWAL (TOEFL 2000 Spoken and Written Academic Language, *cf.* Biber, 2006 for a description of this corpus). Personal information about the callers, if any, e.g., names, addresses, phone numbers, credit card or social security numbers, etc., was consistently and scrupulously replaced by different proper nouns or a series of numbers in the transcripts. No attempt was made to transcribe phonetically, however, some comments about pronunciation, whenever they resulted in misunderstanding were added in the texts. The transcribed texts were manually checked for format and accuracy. Table 1 shows the number of texts, approximate number of words, and the types of service transactions ("accounts") in the corpus. Eight accounts divided into two major kinds, technical (TECH) and customer support (CS), comprise the corpus of transactions. Technical

support calls involve troubleshooting machines and equipment while customer support calls are concerned with order placement and product purchase or inquiry.

Description of Transactions/Accounts	Code	Number of Texts	Approximate Number of Words
Troubleshoot Office Equipment	TECH 1	61	70,780
Internet Service (Home)	TECH 2	45	45,403
Internet Service (Business)	TECH 3	43	33,549
Troubleshoot Machines/Heavy Equipment	TECH 4	46	44,489
Order/Check Order Status (Home Products)	CS 1	42	40,729
Purchase Mobile Phone Minutes	CS 2	52	50,150
Inquire/Order (Equipment)	CS 3	40	40,549
Inquire/Order (Tools and Spare Parts)	CS 4	35	39,981
Total		364	365,630

Table 1: Description of accounts and composition of the corpus.
N male agents = 181; N female agents = 183; N male callers = 184; N female callers = 180

3. Multi-Dimensional Analysis of the Outsourced Call Centre Corpus

According to Tabachnick and Fidell (2001:455), the purposes of FA are "to summarize patterns of correlations among observed variables, to reduce a large number of observed variables to a smaller number of factors, and to provide an operational definition (a regression equation) for an underlying process by using observed variables." These purposes of FA support the overall focus of this project which aims to describe statistically correlating linguistic features and group them into interpretable sets of linguistic dimensions in the corpus.

Initial data processing for FA required an automatic segmentation of the text documents of transactions into groups of agents' and callers' texts in order to analyze the language of agents and callers separately. A total of 728 segmented files (from 364 transcripts of transactions) of callers' and agents' turns comprise the corpus for MD analysis. The segmented texts of the transactions were tagged for parts of speech and semantic categories using Biber's tagging program. Next, the tagged features in the corpus were counted and normalized per 1,000 words by a tag-count program also developed by Biber. The composition of the tag-counted linguistic features used in the study was based primarily on prior studies, especially White (1994) and Biber (1988). Additional features not captured by the tagging program but relevant to telephone-based service transactions, e.g., filled pauses, politeness markers, length of turns, were included in the dataset. A combination of computational tools developed for the study was utilized in order to extract the normalized frequency counts of these supplementary items.

There was a need to run several FAs piloting various combinations of over seventy tag-counted features in order to finalize the list of items comprising the dataset. Linguistic features that correlated below .250 in communality values after extraction and

did not load in any of the factors were excluded. After a series of tests, thirty-six lexical and syntactic features, shown in Table 2, were used in the final MD analysis.

Linguistic Features	Description/Example
1. Type/Token	Number of words occurring in the first 400 words of texts
2. Word Length	Mean length of words in a text (in letters)
3. Word count	Total number of words per agent/caller texts
4. Private Verbs	e.g., *anticipate, assume, believe, feel, think, show, imply*
5. *That* Deletion	e.g., *I think [Ø] he's gone*.
6. Contractions	e.g., *can't, I'm, doesn't*
7. Present Tense Verbs	All present tense verbs identified by the tagging program
8. Second Person Pronouns	*you, your, yours, yourself* (and contracted forms)
9. Verb *Do*	*do, does, did* (and contracted forms)
10. Demonstrative Pronouns	*that, those, this, these*
11. First Person Pronouns	*I, me, my, mine, myself* (plural and all contracted forms)
12. Pronoun *It*	Instances of pronoun *It*
13. Verb *Be*	Forms of *Be* verb
14. Discourse Particles	e.g., *oh, well, anyway, anyhow, anyways*
15. Possibility Modals	*can, could, might, may*
16. Coordinating Conjunctions	*and, or, but*
17. WH Clauses	Clauses with WH (*what*, *which, who*) head
18. Nouns	All nouns identified by the tagging program
19. Prepositions	All prepositions identified by the tagging program
20. Attributive Adjectives	e.g., *the <u>small</u> chair*
21. Past Tense Verbs	Past tense verbs identified by the tagging program
22. Perfect Aspect Verbs	Verbs in perfect aspect construction
23. Nominalizations	Words ending in *–tion, -ment, -ness,* or *–ity* (and plurals)
24. Adverb Time	Time Adverbials, e.g., *nowadays, eventually*
25. Adverbs	Total Adverbs (not Time, Place, Downtoners, etc)
26. Prediction Modals	*will, would, shall*
27. Verb *Have*	*has, have, had* (and contracted forms)
28. Average Length of Turns	Total number of words divided by number of turns
29. Filled Pauses	*uhm, uhh, hmm*
30. Respect markers	*Ma'am, Sir*
31. Politeness markers – *Thanks*	*Thank you*, *thanks*, *[I] appreciate [it]*
32. Politeness markers – *Please please*	
33. Discourse Markers – *OK*	*OK* (marker of information management)
34. Discourse Markers – *I mean I mean* and *You know* (marker of participation)	
35. Discourse Markers – *Next/Then*	*next, then* (temporal adverbs)
36. Discourse Markers – *Because*	*because, 'coz, so* (marker of cause and result)

Table 2: Linguistic features used in the analysis.

After finalizing the dataset for analysis, initial data screening using SPSS v.14.0 was conducted to test for multivariate outliers, multicollinearity, singularity, and normality in the distribution of variables. Results indicated that the dataset met relevant assumptions of FA. The Kaiser-Meyer-Olkin Measure for Sampling Adequacy

(KMO=.724, middling) and Bartlett's Test for Sphericity (Approx. Chi-Square=13101.705, df=666; p<.0001) also indicated that partial and observed correlations in the data were sufficient for FA. SPSS Principal Axis Factoring with Promax rotation was used in the extraction of factors. Figure 1 shows the resulting scree plot of the analysis.
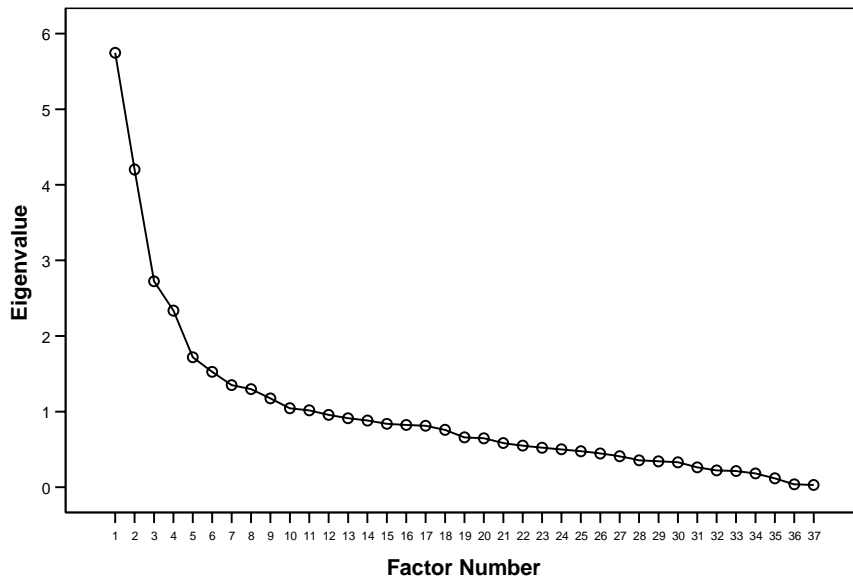

Figure 1: Scree Plot of Eigenvalues.

Results from a three-factor solution, listed in Table 3, were deemed to be the most interpretable merging of features after running tests that included four and five-factor solutions. The three-factor solution reported a 34.29 cumulative percentage of Initial Eigenvalues (Total Variance Explained). With a cut of +/- .30 for inclusion of a variable in interpretation of a factor, eight out of thirty-six features did not load on any factor. Tabachnick and Fidell (2001) claim that the failure of variables to load on a factor reflects heterogeneity of items.

Factor 1		Factor 2	
Second Person Pronouns	.674	Word Count	.801
Word Length	.608	Length of Turns	.688
Please	.519	Type/Token	.640
Nouns	.512	Second Person Pronouns	.501
Possibility Modals	.415	*Next/Then*	.469
Nominalizations	.399	Word Length	.424
Length of Turns	.381	Adverb Time	.404
Thanks	.312	Prepositions	.383
Ma'am/Sir	.309	*Please*	.368
-------------------------------------		Present Tense Verbs	.347
Pronoun *It*	-.671	Nominalizations	.329
First Person Pronouns	-.656	*Because/So*	.315
Past Tense Verbs	-.610		
That Deletion	-.518	--	
Private Verbs	-.439	Discourse Particles	-.308
WH Clauses	-.395		
Perfect Aspect Verbs	-.358		
I mean/You know	-.344	Factor 3	
Verb *Do*	-.326		
		Discourse Particles	.944
		OK	.878
		Adverbs	.866
		--	
		Length of Turns	-.341

Table 3: Summary of the linguistic features of the three factors extracted in the corpus.

4. Discussion

The factor scores of the 728 subjects were computed using the standardized scores of the features that loaded in the three factors. For each of the 728 texts, the standardized scores of the variables were added together to obtain the factor score of every subject. The texts with high or low factor scores helped in interpreting the meaning of the co-occurrence of features. The following sub-sections show the analysis and interpretation of the co-occurrence of linguistic features in the three extracted factors.

4.1 Dimension 1: Addressee-Focussed, Polite, and Elaborated Information vs. Involved and Simplified narrative

A total of eighteen linguistic features comprise Factor 1 with nine features on each of the positive and negative sides of the factor. Positive features include politeness and respect markers, e.g., *thanks*, *please*, *ma'am* and *sir*, markers of elaboration and information density, e.g., long words and turns, nominalizations, and more nouns, and second person

pronouns, e.g., *you*, *your*, which indicate "other-directed" focus of talk. Possibility modals (*can*, *could*, *may*, *might*) also loaded positively on this factor. The features on the negative side of this factor, especially pronoun *it*, first person pronouns, *that* deletion, private verbs, WH clauses, and verb *do*, resemble the grouping in the dimension "Involved Production" identified by Biber (1988) and White (1994). These features are typical of spoken texts and generally contrast with written, informational, and planned discourse. Also on the negative side of the factor are past tense verbs, perfect aspect verbs, and the use of discourse markers *I mean* and *You know*. These elements point to an accounting of personal experience or narrative that tries to explain the occurrence of a particular situation. *I mean* and *You know* are considered by Schiffrin (1987) as markers of information and participation; *I mean* marks speaker orientation toward the meaning of one's own talk while *You know* marks interactive transitions.

The features of Factor 1 appear to distinguish primarily between the unique characteristics of information exchanged across the different types of transactions in the corpus. Similarly, Factor 1 distinguishes between caller or agent roles based on how they communicate a concern or provide a response. In other words, the merging of features in this dimension seems to represent the contrast between the dominant objectives of utterances. Participants who use more positive features are likely aiming to give details and solutions. In the process, these interactants use more nouns, nominalizations, and longer utterances to deliver the information. In addition, the information density of utterances is high because of the higher average word lengths in the texts (Biber, 1988). The turns of participants are elaborated and they also hint at giving explanations, likelihood, or risks though the use of a significant frequency of possibility modals, e.g., *"A reboot **can** cause problems when you connect to the external modem."* Second person pronouns indicate that the transfer of information is highly addressee-focused.

The grouping of features on the negative side of the factor appears to illustrate personal narrative and experiences, and highly simplified information. The combination of past tense verbs, private verbs, pronoun *it*, and discourse markers *I mean* and *You know* demonstrates the specific goal of the utterances to provide a personal account on how a situation happened. The involved production features, e.g. first person pronouns, WH clauses, verb *do*, and *that* deletion, and *I mean, You know* serve a communicative purpose in the maintenance of the interaction, establish personal orientation (White, 1994), and purposely ask for response or assistance. Turns are not elaborated and respect markers are not frequently used. The majority of utterances on the negative side of the factor have smaller word counts and are significantly shorter in length.

The consistent use of addressee-focused politeness and respect markers on the positive side of Factor 1 characterizes the overall nature of outsourced call centre transactions. Service encounters commonly allocate for courteous language and the recognition of roles, and call centre agents are expected to show respect and courtesy when assisting customers (D'Ausilio, 1998). In this factor, the frequency of politeness and respect markers clearly differentiates the discourse of callers and agents. Although callers also use polite words, e.g. *please*, *thanks*, *appreciate*, and some *ma'ams* or *sirs*, agents have very high frequencies of these features across the board. More research is called for to explore the use of *ma'am* and *sir* by Filipino agents in the corpus. The frequency of these markers may distinguish Filipinos from Indian or American call centre agents. It is likely that Filipino agents overuse *ma'am* or *sir* because of their

interlanguage background and the way service is typically conducted in the Philippines. Directing respect markers towards customers is highly expected during service encounters in this country. It is possible that this expectation transfers to the way Filipino agents interact with their callers. To summarize, the combination of positive and negative features in Factor 1 shows a linguistic dimension (Dimension 1) that differentiates between addressee-focused, polite, and elaborated information and involved and simplified narrative portraying how informational content is produced in the discourse. The text sample below highlights some of the positive features of Dimension 1, especially politeness and respect markers, in the agent's turns:

> Agent: **Thank yo**u for calling **Phone Company Payment Services**, my name is **Jane**, how **can** I help **you**?
> *Caller: Yes, uh, when are you guys gonna go back telling us when how much time is left on these phone cards? I mean on these phones?*
> Agent: I **apologize** for the **inconvenience sir**, I'll, let me explain on that ok? **Please**, give me your **cell phone number** so I can check on **your minutes**.
> *Caller: 333-333-3333, I think it has run out because I wanted to use it but it said it didn't have enough time.*
> Agent: Ok, let me just verify the **charges** at the **moment sir**, **please** give me **your name** and **address** on the **account please.**

> **Text Sample 1**: Call excerpt from CS 2 (Purchase Mobile Phone Minutes).

4.2 Dimension 2: Planned, Procedural Talk

The linguistic items that load on the positive side of Factor 2 include lexical specificity and information density features (type/token ratio, average word length), complex and abstract information features (word count, length of turns, and nominalization), temporal adverbs (*next/then*) and specific time adverbials, e.g., *eventually, immediately*, second person pronouns, prepositions, cause and result discourse markers (*because/so*), politeness marker *please*, and present tense verbs. Only discourse particles, e.g., *oh*, *well, anyway*, loaded on the negative side of the factor. The merging of features that indicate lexical specificity and complexity and abstraction of information differentiates call centre discourse from general conversation or other forms of purposeful oral interactions. Biber (1988) states that these features are more common in academic written texts and less observed in spoken texts because of production circumstances. In typical, on-line conversations, general topic shifts allow for the occurrence of more common words and phrases and limited complex or abstract vocabulary. Shorter words based on average word lengths are often used with familiar vocabulary repeated over during the interaction. Therefore, information-packaging in the call centre discourse is similar in this sense to written, planned texts because of the presence of many features that are not commonly produced on-line. Due to the amount of information exchanged in transactions, more diversity in key words is used in the utterances which increases type/token ratio and word

count. More nouns, nominalization, and longer word length suggest that the information is technical and specialized.

The positive side of the factor signifies a one-way (addressee-focused) transfer of a large amount of abstract and technical information. In this case, the information appears to be "real-time," procedural or process-based due to the presence of temporal adverbs combining with prepositions, e.g., *in, on, below, above*, and, especially, present tense verbs. The frequent occurrence of present tense verbs in the texts illustrates the use of directives/imperatives in utterances, e.g., "..then **click** next"; "..now, **change** your password to XX." Instructional language is expressed through a series of directions marked by second person pronouns (especially *you* and *your*), succession between steps (*next/then*) and progression through the discourse (*now*). It appears that the instructional tenor of the turns also includes explanations through the use of cause and result discourse markers (*because/so*) common in the factor. The recurrent use of *please* shows that the delivery of directive or procedural information is still mostly polite as in Dimension 1. Discourse particles, used very sparingly in this factor, perhaps indicate that the utterances are somewhat prepared or organized, and produced with limited hesitations or tentativeness. It follows that participants who provide directive or instructional information should know their content and how to best facilitate its transfer to the receiver. In sum, the linguistic dimension (Dimension 2) shown here seems to capture the major characteristics of planned information that primarily intends to give directives and procedures in the transactions.

Dimension 2 clearly makes a distinction between agents and callers' use of linguistic features; agents use more of the features on the positive side of the factor, and predictably engage in directive, procedural talk more than the callers. Agents' speech in this dimension is produced on-line but covers a wide-range of topics and makes use of a variety of terms and jargon that comprise their set spiels. In a way, agents' utterances in giving directions and steps are planned because they have clear expectations about the questions directed to them. The moves in assisting a caller are well-defined for many agents, and their series of procedures are commonly established from the time they started training. Many agents have memorized procedural spiels and are constantly reminded of them by accessible notes and tools during the calls. The callers use limited positive features of Dimension 2 and relatively more discourse particles. The use of more discourse particles in the callers' utterances conceivably indicates a communicative shift to clarify, ask follow up questions, or express uncertainty. Several positive features of Dimension 2 appear in some caller texts whenever they repeat or echo a directive. The differences in accounts for Dimension 2 are brought about by the nature of service given in each of the transactions. Most accounts, especially those that deal with troubleshooting and real-time processes, are typically procedural while some provide service though direct question-answer sequences with very limited instructions. Text Sample 2 illustrates the procedures and steps in the agent's talk for a troubleshooting transaction:

Agent: So **then please** go ahead and hit the "No"
*Caller: Hey **well** uhh we require a 9.*
Agent: Oh, **you** require 9?
Caller: Yes.
Agent: **Then** go ahead and **please** type in "Yes" and then hit 9.

*Caller: **Ok**, and then enter again?*
Agent: Yes, uh huh.
*Caller: **Well** it just says dialing.*
Agent: Uh huh, by the way Sarah just give me an **update** whenever the **message on** the **screen** changes so that I could uhh put down **documentation** here.
*Caller: **Ok** [long pause] It says "connect phone cord and press," then it says "done press enter"*
Agent: Hmm, it, it actually means Sarah that uhm the only reasons that the **postage machine** would say connect the "**connect phone cord message"** is **because** it's not detecting a dial tone **because** it's connect, it's hooked up to a wrong type of **phone line** or the **phone cord** itself is defective. **Now** we need a **connection**, uhm since this is a brand new **postage machine** uhh there's a big **possibility** that the **phone line** that it's hooked up to is not correct, so uhm Sarah is it ok if I get the **phone number** where you have the **postage machine** hooked up to so that I could check if uhm if it's dialing out or not?

Text Sample 2: Call excerpt from TECH 1 (Troubleshoot Office Equipment).

4.3 Dimension 3: Managed Information Flow

The linguistic features on the positive side of Factor 3 are discourse particles, the discourse marker *OK*, and adverbs – any adverb form occurring in the dictionary, or any form that is longer than five letters and ends in –*ly* (Biber, 1988). The adverbs comprising this list do not include time and place adverbials and those counted as amplifiers or downtoners. The positive features in this factor are very common in conversation. Discourse particles are regarded as necessary for conversational coherence (Schiffrin, 1994) and in monitoring the flow of information in talk (Biber, 1988; Chafe, 1985). *OK* is also regularly used in conversation and purposeful interactions like service encounters, and serves as either a marker of information management (Schiffrin, 1987) or an apparent backchannel. The combination of discourse particles and backchannels could be interpreted as a conversational devise to maintain and monitor the overall flow of transactions. More of these features emerge because the interactions are conducted over the telephone with defined turns and adjacency pairs. It is possible that backchanneling through *OK* and discourse particles that initiate turns are preferred by participants to avoid dead air and long pauses. Adverbs typically indicate a form of expansion of ideas because they provide additional descriptive information in the texts, and are used as modifiers or adverbials attached in clauses (Biber, Johansson, Leech, Conrad, and Finnegan, 1999). However, this interpretation is not supported by the lone negative feature in the factor. Length of turns on the negative side of the scale signals that discourse particles, *OK*, and adverbs co-occur with shorter utterances. Elaboration of information with adverbs does not hold up in this case because the turns tend to be shorter in length. It is possible that adverbs in this dimension are used for quick responses, e.g., *absolutely*, *exactly*, and as epistemic stance adverbials, e.g., *actually, basically, really*, that have been included in the agents' repertoire of conversational devices.

The grouping of linguistic features in Factor 3 signifies the speaker's attempt at managing the flow of information (Dimension 3). This dimension separates callers and agents in their use of discourse particles, *OK*, and adverbials (positive features) intended to facilitate and monitor the transaction. Agents have substantially used more of these features than the callers. The use of the positive features by the agents could be related to common conventions in the register such as establishing rapport, avoiding dead air, and backchanneling to show attentiveness and focus. Filipino agents undergo skills training in phone-handling, and some of the topics covered in many training sessions include backchanneling and providing confirmatory responses to control the flow of transactions. Peltzman and Fishburn (2006) report that there is a need to explicitly teach Filipino agents to backchannel or provide verbal feedback to constantly remind the caller that someone is on the other end of the line, listening, and ready to provide service. Filipinos have traditionally observed turns in conversations avoiding interruptions and overlaps. Children are taught that politeness in conversations involves respecting turns and speaking only when recognized. Peltzman and Fishburn suggest that in the context of outsourced call centres, limited backchanneling and echoing of customer concerns, i.e. "reflective listening" could imply that the agents are not enthusiastic in serving the caller, or not personalizing the call. They also mention that American callers are accustomed to hearing short responses and confirmatory utterances in conversations. Because of this emphasis in language training in Philippine call centres, it is likely that the agents consciously use features like discourse markers to provide verbal feedback during the calls.

It is important to note here that in Dimension 2, planned, procedural features co-occur with limited frequencies of discourse particles. In relation to Dimension 3, this result could imply that management of information flow for agents is slightly less when they provide straightforward directives. In this case, agents' utterances are more organized and there are limited provisions for backchannels because the callers do not control the information exchange. Furthermore, because discourse particles may also indicate signals of uncertainty and tentativeness, it could be interpreted that agents engaged in procedural transactions may use more discourse particles when they are quite uncertain about the procedures they provide, but less when they are in control of the procedural steps in the transactions.

Some of the communicative markers in Dimension 3 are likely to have been overused by many agents and have become part of their mannerisms in handling calls. The use of *OK*, *actually*, *basically*, *well*, and *anyway* in the turns (as shown by Text Sample 3 below) is common across accounts. In managing the flow of information and trying to control the transaction through the features in Dimension 3, it appears that the agents are serving three unique purposes: (1) direct management, i.e. avoiding dead air, confirming the message, initiating the turn; (2) indirect management though mannerisms acquired while supporting American callers; and (3) making use of the positive features to supplement fillers to "buy thinking time" before a response.

Caller: Two of them was on the 25th and one of them was on the 21st of June
Agent: **Ok** I'll just go ahead and check [long pause] **ok** [hold 22 seconds] the other one I believe was on the two you have **actually** won three recruits right?
Caller: Yes

Agent: **Ok** you have three recruits so let me just check [long pause] **ok** so it is here that since you recruited them just last 25th they **supposedly** [long pause] **ok** let me just go ahead and check on this, I'll call you back because I **actually** checked your sales record and that coupon is not loaded in your, there **ok**?
Caller: [unclear] I don't see it there
Agent: Yes, yes and uh **ok**, you know the start is **actually** June 30 **well** but **anyway** you have until the end of this month to redeem this coupon **basically**, so whatever, let me just go ahead and check why the coupon is not loaded

Text Sample 3: Call excerpt from CS 1 (Order/Check Order Status - Home Products).

5. Conclusions and Limitations, and Directions for Future Research

The exploration of spoken interactions in outsourced call centres using multi-dimensional analysis has revealed several interesting and unique characteristics of the discourse. The wide-range of information exchanged by participants appears to be described by the statistical co-occurrence of different linguistic features in texts. The three extracted dimensions seem to have exposed marked attributes that distinguish the functional characteristic of speakers' discourse and the linguistic patterns prevailing in the transactions. Specific foci on the amount of transactional information required to be exchanged; the overall objective of the exchange; and ways of facilitating the exchange are interpretable through the linguistic dimensions. It would be very relevant to apply the same dimensions to parallel call centre corpora, and examine how the Filipino agents' manner of handling and facilitating the transactions compare with, for example, Indian or American call-takers.

The next phase of the study will involve a thorough comparison of how identified social categories, e.g., speaker roles, gender, nature of accounts, determine and predict the use of the features identified in the three dimensions. Clearly, the present analysis is exploratory in nature, both in the composition of the linguistic features for FA and the identified social or subject categories. For example, success or failure of the transactions associated with language ability and use, accuracy of service, and workflow compliance/efficiency are not included in the current research. Results of evaluations regarding quality of support in the transactions which are relevant in the application of FA results to pedagogy and performance assessment are not reported in this study. Other demographic data, especially about the agents, e.g., age, experience, educational background are also not integrated into this analysis. The study has limited discussion and interpretation regarding ESL, Philippine English, and specific outsourced call centre training practices. More importantly, issues about pronunciation and oral speech production variables are not considered in this paper. Features of verbal communication not encoded in the transcribed texts of transactions are not captured by corpus tools. Corpus-based approaches to spoken discourse are still limited in accounting for phonetic factors and oral production circumstances primarily because orthographic text transcriptions are mostly confined to lexical and structural features (Baker, 2006).

References

Aarts, B. and C. Meyer. (eds.) (1995) The Verb in Contemporary English: Theory and Description. Cambridge: Cambridge University Press.

As India Gets Too Costly. (October 10, 2006) *Philippine Daily Inquirer* B21.

Baker, P. (2006) Using Corpora in Discourse Analysis. New York: Continuum.

Baker, P. and T. McEnery. (2005) 'A Corpus-Based Approach to Discourses of Refugees and Asylum Seekers in UN and Newspaper Texts'. *Language and Politics 4*, 2–23.

Biber, D. (1988) Variation Across Speech and Writing. Cambridge: Cambridge University Press.

Biber, D. (1995) Dimensions of Register Variation: A Cross-Linguistic Perspective. Cambridge: Cambridge University Press.

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

Biber, D. and J. Burges. (2001) Historical Shifts in the Language of Women and Men, in S. Conrad and D. Biber (eds.) Variation in English: Multi-Dimensional Studies, pp. 157–70. London: Longman.

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finnegan. (1999) Longman Grammar of Spoken and Written English. Harlow, Essex: Pearson Education.

[BPAP] Business Processing Association Philippines. (2007) [Homepage on the Internet] 'BackOffice Outsourcing in the Philippines', retrieved 2007 January 12 from http://www.bpap.org

Chafe, W.L. (1985) Linguistic Differences Produced by Differences Between Speaking and Writing, in D.R. Olson, N. Torrence, and A. Hildyard (eds.) Literacy, Language, and Learning: the Nature and Consequences of Reading and Writing, pp. 222–41. Cambridge: Cambridge University Press.

Connor-Linton, J.T. (1989) Crosstalk: A Multi-Feature Analysis of Soviet-American Spacebridges. Unpublished Ph.D. Dissertation, University of Southern California.

D'Ausilio, R. (1998) Wake Up Your Call Center: How to be a Better Call Center Agent. Indiana: Purdue University Press.

Friginal, R. (2004, March 18) 'Call Center Industry to Supply 50% of World Demand', retrieve October 5, 2004 from http://www.malaya.com.ph

Leech, G. and N. Smith. (2005) 'Extending the Possibilities of Corpus-Based Research on English in the Twentieth Century: A Prequel to LOB and FLOB'. *ICAME Journal 29*, 34–56.

Magellan Alliance. (2005) The Philippines as an Offshore Service Delivery Location. Manila: John Clements Consultants, Inc.

Peltzman, R. and J. Fishburn. (2006) 'Silence is Not an Option', paper presented at the Talking across the World: English Communication Skills for the ITES Industry Inaugural Conference. February 2006 Manila, Philippines.

Precht, K. (2000) Patterns of Stance in English. Unpublished Ph.D. Dissertation, Northern Arizona University.

Quaglio, P. (2004) The Language of NBC's Friends: A Comparison with Face-to-Face. Unpublished Ph.D. Dissertation, Northern Arizona University.

Reaser, J. (2003) 'A Quantitative Approach to (Sub) Registers: the Case of Sports Announcer Talk'. *Discourse Studies 5*, 303–21.

Rey, J. M. (2001) Changing Gender Roles in Popular Culture: Dialogue in Star Trek Episodes from 1966 to 1993, in S. Conrad and D. Biber (eds.) Variation in English: Multi-dimensional studies, pp. 138–56. London: Longman.

Schiffrin, D. (1987) Discourse Markers. Cambridge: Cambridge University Press.

Schiffrin, D. (1994) Approaches to Discourse: Language as Social Interaction. Oxford: Blackwell.

Scott, S. (1998) Patterns of Language Use in Adult, Face-to-Face Disagreements. Unpublished Ph.D. Dissertation, Northern Arizona University.

Service Alert. (2004, April 8). *Malaya* B7.

Tabachnick, B.G. and L.S. Fidell. (2001) Using Multivariate Statistics, 4th ed. Boston: Allyn and Bacon.

Teves, R. (2003, December 14) 'The Business of Call Centers', *Manila Standard* A23.

Tuchman, M. (2006) Updates and Downdates: Call Center 2010 in the Asia-Pacific Region. Manila: ITES 44–40 Special Publication.

Uy, V. (2004, October 8) 'Call Centers Join Job Fair in Manila; 20,000 Workers Wanted. *Philippine Daily Inquirer* A12.

White, M. (1994) Language in Job Interviews: Differences Relating to Success and Socioeconomic variables. Unpublished Ph.D. Dissertation, Northern Arizona University