# "And why not?" Exploiting the Enron email corpus in the business English classroom

Evan Frendo

# EnronSent Corpus

"The EnronSent corpus is a special preparation of a portion of the Enron Email Dataset designed specifically for use in Corpus Linguistics and language analysis. Divided across 45 plain text files, this corpus contains 2,205,910 lines and 13,810,266 words."

"96,107 messages from the "Sent Mail" directories of all the users in the corpus."

http://verbs.colorado.edu/enronsent/

(Styler, 2011)

# Why is this corpus important?

"a rich example of how real people in a real organization use e-mail—full of mundane lunch plans, boring meeting notes, embarrassing flirtations …"

"nothing else like it in the public domain"

(Leber, 2013)

# Giving personal news

Just a quick note to let you all know that I got engaged this weekend.  And to all you single fellas out there, your life just got a whole lot easier now that I have taken myself off the market.

EnronSent 30

# Gossip

Also, I hooked XXX up with one of my customers (A divorced brother from XXX). Girlm she called me this morning, and she was so giddy.  She said they really had a good time talking on the phone, they talked two hours.  They are meeting for lunch on tomorrow. I'll be anxious to hear what he thinks of her after meeting her.  I know she'll like him, because he is a nice looking brother, and nice conversation.  I just hope he thinks she's cute.

Hmm, so you hooked XXX up with a customer. Yeah, I hope he thinks she's cute also. She dresses nice and keeps herself well maintained, but sometimes when I look her dead in the face, she's not the cutest person. Keep me posted on that.

# Communication problems

To all of you who are without the luxury of a dial tone this morning:

XXX is working on the malfunctioning phone circuit. We don't know when the service will be reinstated; but in the meantime you can reflect on what life was like before the darn thing was invented.

EnronSent 40

# Stress at work

Thanks for the note.  Things are pretty crazy around here, but it is clear that lawyers will be around here for quite some time.  Anyway, hope all is well with you.

EnronSent 14

# Work-related

Ok - you got the ball. Regards John Thanks.

# Work-related

Please plan to attend a meeting scheduled for June 20, 2001 @ 10:00AM to 12:00, 3AC 0301, to discuss procedures for commissioning measurement facilities. If you can't personally attend, please have someone attend in your place. We would like to have a good cross section of representation from Operations, Measurement Technology, Measurement Services, Gas Control, Market Services, and  Marketing.

Thanks

# Work-related

Ok - you got the **ball**. Regards John Thanks.

# Why is this corpus important?

- Research


- BE trainers – offers easy access to "real" emails


- LTOL

# Outline

- summarize some of these research findings,

- demonstrate just how easy it is to access the corpus, and

- discuss the implications for business English teaching.

# Outline

- **summarize some of these research findings,**
- demonstrate just how easy it is to access the corpus, and
- discuss the implications for business English teaching.

# Recommending Recipients in the Enron Email Corpus

Vitor R. Carvalho
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

William W. Cohen
Machine Learning Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

## ABSTRACT

Email is the most popular communication tool of the internet. In this paper we investigate how email systems can be enhanced to work as *recipient recommendation systems*, i.e., suggesting who recipients of a message might be, while the message is being composed, given its current contents and given its previously-specified recipients. This can be a valuable addition to email clients, particularly in large corporations. It can be used to identify people in an organization that are working in a similar topic or project, or to find people with appropriate expertise or skills. Recipient recommendation can also prevent a user from forgetting to add an important collaborator or manager as recipient, preventing costly misunderstandings and communication delays.

In this paper we present the first study of recipient recommendation in a real large-scale corporate email collection, the Enron Email corpus. We begin by defining the problem as a large multi-class multi-label classification task, where

## Keywords

## 1. INTRODUCTION

One important use of work-related email is negotiating and delegating shared tasks and subtasks, and more generally, communication between task-oriented working groups. Previously we have argued [9] that providing intelligent assistance for this use of email is important because the cost of errors in task management is high: for instance, deadlines can be missed or opportunities wasted because of such errors.

Here we consider an automated technique, called *recipient recommendation*, that is designed to avoid a specific type of high-cost email error: errors that result when a message is not sent to all intended recipients. An example of such an error would be forgetting to CC an important collaborator, or manager, on a message to a working group;

LEE DRUTMAN
*The Sunlight Foundation and
University of California, Washington Center*
DANIEL J. HOPKINS
*Georgetown University*

# The Inside View: Using the Enron E-mail Archive to Understand Corporate Political Attention

For decades, scholars have debated the role of corporations in American politics. To date, they have relied on either interviews or publicly disclosed spending and lobbying reports. This article presents new methods and data that enable us to consider the internal processes of corporate political attention instead. Aided by automated content analysis, this article uses more than 250,000 internal e-mails from Enron to observe its political attention between 1999 and 2002. These e-mails depict Enron's employees as focused on monitoring and formally participating in political processes, including bureaucratic processes. Only a small fraction of their political attention focused on elections.

# KNOWLEDGE DISCOVERY IN CORPORATE EMAIL:

## THE COMPLIANCE BOT MEETS ENRON

by

**K. Krasnow Waterman**

Submitted to the Alfred P. Sloan School of Management on May 12, 2006
in Partial Fulfillment of the Requirements for the Degree of

**Master of Science in Management of Technology**

## ABSTRACT

I propose the creation of a real-time compliance "bot" – software to momentarily pause each employee's email at the moment of sending and to electronically assess whether that email is likely to create liability or unanticipated expense for the corporation. My thesis describes the confluence of historical events making such a product necessary and desirable – increase in corporate regulation, explosive growth of email, acceptance of email as evidence in litigation. The cautionary tale of Enron provides the backdrop for the thesis. The government released hundreds of thousands of Enron management emails and they have become research fodder for those interested in "Knowledge Discovery," a computer science discipline that gleans meaningful information from data otherwise indecipherable due to its sheer size. CEO's and other C-level corporate managers are my intended audience, so I have attempted to counter the weightiness of the technical topics by focusing on the search for readily understandable management headaches such as the loss of productivity due to high participation in the fantasy football pool or the potential for dirty jokes to become evidence in an employment law claim.

# The Best Kept Secrets with Corpus Linguistics

*Neil Cooke*
CCSR
University of Surrey
*n.cooke@surrey.ac.uk*

*Dr. Lee Gillam*
Computer Science Department
University of Surrey
*l.gillam@surrey.ac.uk*

*Prof. Ahmet Kondoz*
CCSR
University of Surrey
*a.kondoz@surrey.ac.uk*

## Abstract

This paper presents the use of corpus linguistics techniques on supposedly "clean" corpora and identifies potential pitfalls. Our work relates to the task of filtering sensitive content, in which data security is strategically important for the protection of government and military information, and of growing importance in combating identity fraud. A naïve keyword filtering approach produces a large proportion of false positives, and the need for more fine-grained approaches, suggests the consideration of using corpus linguistics for such content filtering.

We present work undertaken on the Enron corpus, a collection of emails that has had various tasks undertaken on various portions and versions of it by other researchers. We have made some efforts to reconcile differences between the different versions by

Squib

# Virtual business: An Enron email corpus study

Greg Kessler

*Department of Linguistics, Ohio University, 383 Gordy Hall, Athens, OH 45701, USA*

## Abstract

The word virtual is an exceptional technology oriented neologism deserving explanation. It is commonly accepted as referring to computer simulation or indicating lack of authenticity. However, the increasing polysemic nature of the word virtual has resulted in a much broader range of use within natural language. Many of these meanings appear to increasingly conflict with previously recognized meanings. The author explores contemporary natural language use of the word virtual in the context of the Enron email dataset. This dataset has been selected for three reasons: it represents the largest authentic collection of email available to date, Enron was a tech savvy company with high proclivity for use of neologisms, and Enron has been referred to by some as a virtual company.

Examples of virtual in this study indicate broad reference to nearly anything computer related, including computer business, computer based simulation and varied forms of computer mediated communication. It is also observed illustrating inauthenticity, hyperbole and potential deception. This study demonstrates a lack of confidence that suggests an underlying concern about the potential ambiguity of the word virtual, resulting from the increasing polysemic nature of this word.

# Email Formality in the Workplace: A Case Study on the Enron Corpus

**Kelly Peterson, Matt Hohensee, and Fei Xia**
Linguistics Department
University of Washington
Seattle, WA 98195
{kellypet, hohensee, fxia}@uw.edu

## Abstract

Email is an important way of communication in our daily life and it has become the subject of various NLP and social studies. In this paper, we focus on email formality and explore the factors that could affect the sender's choice of formality. As a case study, we use the Enron email corpus to test how formality is affected by social distance, relative power, and the weight of imposition, as defined in Brown and Levinson's model of politeness (1987). Our experiments show that their model largely holds in the Enron corpus. We believe that the methodology proposed in the paper can be applied to other social media domains and be used to test other linguistic or social theories.

has shown that senders of email exhibit a wide range of language and form choices which vary in different social contexts (Orlikowski and Yates, 1994). Through various theories of sociolinguistics, it is proposed that these changes take place in a predictable manner.

Brown and Levinson (1987) have proposed a model where in order to save the "face" or public self image of the hearer of a message, a speaker can employ a range of verbal strategies. Their model of politeness states that in social situations there are three factors which are considered in a decision whether or when to use communication techniques such as formality:

# Communication Networks from the Enron Email Corpus
## "It's Always About the People. Enron is no Different"[1]

JANA DIESNER
TERRILL L. FRANTZ
KATHLEEN M. CARLEY
*Center for Computational Analysis of Social and Organizational Systems (CASOS), Institute for Software Research International (ISRI), School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, 1327 Wean Hall, Pittsburgh, PA 15213, USA*
*email: jdiesner@andrew.cmu.edu*
*email: terrillf@andrew.cmu.edu*
*email: kathleen.carley@cs.cmu.edu*

### Abstract

The Enron email corpus is appealing to researchers because it represents a rich temporal record of internal communication within a large, real-world organization facing a severe and survival-threatening crisis. We describe how we enhanced the original corpus database and present findings from our investigation undertaken with a social network analytic perspective. We explore the dynamics of the structure and properties of the organizational communication network, as well as the characteristics and patterns of communicative behavior of the employees from different organizational levels. We found that during the crisis period, communication among employees became more diverse with respect to established contacts and formal roles. Also during the crisis period, previously disconnected employees began to engage in mutual communication, so that interpersonal communication was intensified and spread through the network, bypassing formal chains of communication. The findings of this study provide valuable insight into a real-world organizational crisis, which may be further used for validating or developing theories and dynamic models of organizational crises; thereby leading to a better understanding of the underlying causes of, and response to, organization failure.

**Keywords:** Enron, email corpus, communication networks, social network analysis, dynamic network analysis, organizational crisis, organizational hierarchy

# Requests and Commitments in Email are More Complex Than You Think: Eight Reasons to be Cautious

**Andrew Lampert**
CSIRO ICT Centre
Locked Bag 17
North Ryde 1670
Australia
Andrew.Lampert@csiro.au

**Robert Dale**
Centre for Language Technology
Macquarie University 2109
Australia
rdale@ics.mq.edu.au

**Cécile Paris**
CSIRO ICT Centre
Locked Bag 17
North Ryde 1670
Australia
Cecile.Paris@csiro.au

## Abstract

Many workplace tasks are managed through email communication, involving the exchange of requests and commitments. Our aim is to build a tool that can automatically identify and manage such requests and commitments. A detailed analysis of real data, however, reveals a range of interesting edge cases that make even human annotation of training data difficult. In this paper, as an important step in the development of annotation guidelines for wider use in the growing email processing community, we identify eight categories of problematic data and propose how they should be handled in annotation and extraction tasks.

## 1 Introduction

Our aim is to create tools that assist email users by automatically detecting requests and commitments in incoming and outgoing email. The motivation for

or response because they are buried in their email (Whittaker and Sidner, 1996). More recent studies of task-focused email usage have also identified problems with "keeping track of lots of concurrent actions: One's own to-dos and to-dos one expects from others" using existing email clients (Bellotti et al., 2003).

To provide support here, we are working to augment existing email clients with features such as action-oriented summaries of email messages and threads; task-based navigation and visualisations; and dashboards that provide overviews of the state of an email inbox or collection with much greater fidelity than is possible with current tools.

In working towards this goal, we have conducted a series of manual annotation experiments, exploring the level of human agreement that is achievable in identifying requests and commitments in email messages. Previous work has often relied on canonical examples as the basis for simple definitions of

# Identifying speech acts

"Many workplace tasks are managed through email communication, involving the exchange of requests and commitments. Our aim is to build a tool that can automatically identify and manage such requests and commitments."

(Lambert, Dale & Paris, 2008, p.64)

# Requests

We consider a request to be an utterance from an email sender that places an obligation on an email recipient to:

1. Schedule an action, often by adding an entry to a calendar or task list;

2. Perform an action; or

3. Respond with some speech act.

(Lambert, Dale & Paris, 2008, p.65)

# Commitments

We consider a commitment to be an offer or promise made by an email sender for future action or response from some specified agent. The agent who is committed is often, but not always, the sender.

(Lambert, Dale & Paris, 2008, p.66)

# Meeting request

Today's Prebid meeting will take place in EB32c2 at 3pm.

In this case the announcement functions as a request to attend

(Lambert, Dale & Paris, 2008, p.68)

# Example of request + commitment

My assistant, XXX, will send you an updated version of my bio.

"When XXX is a recipient of the email, this commitment also functions as a request to XXX. In contrast, the same email sent without XXX as a recipient does not function as a request, since it is not communicated to XXX. It does still function as a commitment for a third party."

(Lambert, Dale & Paris, 2008, p.71)

# A classification scheme for annotating speech acts in a business email corpus

*Rachele De Felice, University College London*
*Jeannique Darby, University of Oxford*
*Anthony Fisher, University of Nottingham*
*David Peplow, Sheffield Hallam University*

## Abstract

*This paper reports on the process of manual annotation of speech acts in a corpus of business emails, in the context of the PROBE project (PRagmatics of Business English). The project aims to bring together corpus, computational, and theoretical linguistics by drawing on the insights made available by the annotated corpus. The corpus data sheds light on the linguistic and discourse structures of speech act use in business email communication. This enhanced linguistic description can be compared to theoretical linguistic representations of speech act categories to assess how well traditional distinctions relate to real-world, naturally occurring data. From a computational perspective, the annotated data is required for the development of an automated speech act tagging tool. Central to this research is the creation of a high quality, manually annotated speech act corpus, using an easily interpretable classification scheme. We discuss the scheme chosen for the project and the training guidelines given to the annotators, and describe the main challenges identified by the annotators.*

# Speech acts

*Table 1:* Speech act categories used

| Speech act | Tag | Example |
|---|---|---|
| Direct request | DR | *Please send me the files.* |
| Question-request | QR | *Could you send me the files?* |
| Open question | QQ | *What time is the meeting?* |
| First person commitment | FPC | *I will attend the meeting.* |
| First person expression of feeling | FPF | *I am uncertain about the agenda.* |
| First person other | FPO | *I am an employee of this company.* |
| Other statements | OT | *The meeting is at 8 tomorrow.* <br> *You always work so hard.* |

(De Felice, Darby, Fisher, & Peplow, 2013, p.80)

# Identifying Speech Acts in E-Mails: Toward Automated Scoring of the *TOEIC®* E-Mail Task

**Rachele De Felice**

**Paul Deane**

September 2012

We developed a computational model for automated speech act identification and tested it on a corpus of TOEIC responses, achieving up to 79.28% accuracy.

# Comparison of TOEIC Email data and Enron – differences possibly due to:

- Email culture at Enron (eg  whether it was considered more appropriate to express opinions and commitments in person or over the phone rather than over email). (p. 43)
- L1 seems to show a tendency to prefer indirect requests to direct requests (p. 44)
- L1 e-mails tend to include a larger amount of supporting information and background details (p. 44)
- Power and social relations between the sender and receiver (p. 45)
- Stylistic choices (p. 45)

# A CORPUS STUDY OF EMAIL WRITING IN A BUISNESS SETTING AND ITS PRACTICAL

# APPLICATION IN TEACHING ENGLISH AS A SECOND LANGUAGE

A CREATIVE PROJECT

SUBMITED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTERS OF ARTS

BY

NAOMI LANCASTER

# Greetings

"The first conclusion which can be drawn from the data is that persons in a business setting tend to prefer not to open their emails with the traditional "dear" and rarely even use the more informal "Hey," "Hi" or "Hello." In fact, the traditional "dear" did not show up a single time in this dataset of 400 emails."

"A non-native speaker wishing to sound natural in a business setting could probably safely leave off any sort of greeting in correspondence."

(Lancaster 2010, p.22)

Dear - 2602 times

# Profanity

"Another unexpected phenomenon which occurred in the study was the presence of profanity in the emails. While this language occurred in only 2% of emails total, it was present. It is possible that the emails it occurred in were of a personal as opposed to professional nature, but they were regardless all from a cooperate email account …"

(Lancaster 2010, p.42)

# Closings

| Closing | Number of Occurrences (out of 400, 40 subjects) | Percentage of Occurrences (out of 400) |
|---|---|---|
| Thanks | 71 | 18% |
| Thank you | 9 | 2% |
| Take care | 4 | 1% |
| Keep in touch | 2 | 0.5% |
| Talk to you soon | 3 | 0.75% |
| See you then / tonight | 3 | 0.75% |
| Regards | 10 | 3% |
| God Bless | 1 | 0.25% |
| Thx | 1 | 0.25% |
| None | 296 | 74% |

(Lancaster 2010, p.43-44)

# Signatures

| Signature | Number of Occurrences (out of 400, 40 subjects) | Percentage of Occurrences (out of 400) |
|---|---|---|
| Full name | 29 | 7% |
| First only | 80 | 20% |
| Last only | 9 | 2% |
| Initials only | 30 | 8% |
| Nickname (obvious) | 1 | 0.25% |
| None | 251 | 63% |

(Lancaster, 2010, p.47)

# Teacher survey
# (11 teachers at a university)

Q3 How do you teach students to open emails?

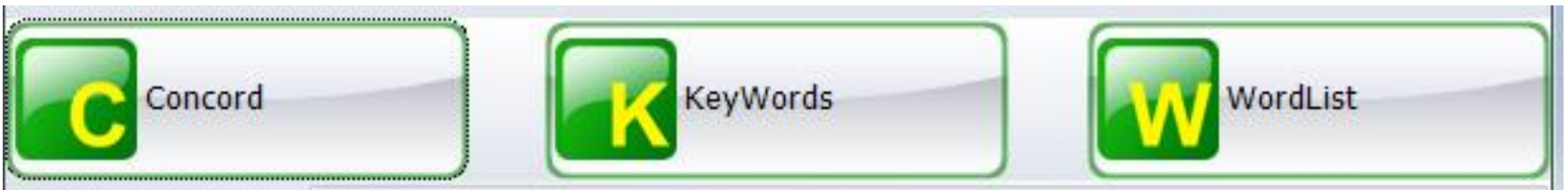"All of the teachers agreed that starting an email with "Dear/Mr./Ms./Miss/Mrs./Professor Last Name" (though they did not all mention all of the options) was the way to teach openings to students."

(Lancaster, 2010, p.55)

# Outline

- summarize some of these research findings,

- **demonstrate just how easy it is to access the corpus**, and

- discuss the implications for business English teaching.

# WordSmith Tools

**Concord**

| N | Concordance |
|---|---|
| 1 | 01/05/2001 12:12 PM Hello Kate- El paso just called to verbal some deals and he had |
| 2 | nothing new on the subject of El Paso. Just curious if you knew where we are at in |
| 3 | the " rerun" dance at the edf man pary just over a year ago!!! I ordered 3 different |
| 4 | mean that we would pay a third party, just someone else at Enron. I agree with Rod |
| 5 | t-shirts that say "class of '94 - still partyin'". just tell your girlfriends we are going to lake |
| 6 | contacts at OneOk in the recent past. Just this morning I talked with a Jerry Knight |
| 7 | to the ISO. Chris 503-464-8219 Pat, Just a reminder that I will need the MW diff/ |
| 8 | your office. All he needs is your passwords. Just wondering if you had a chance to |
| 9 | with the issues, or head him off at the pass. Just let me know. Russell Diamond/ |
| 10 | . If you have forgotten your password, just go to the login screen, hit forgot |
| 11 | What: 6th annual John, Robert & Mo's Party (just 5th annual for me). 2kegs, assorted |
| 12 | would free up legal time for both parties. Just food for thought. Kent <> - Package |
| 13 | an equal amount for his prospective partner just across the street. Enron 's lawyers were |
| 14 | and books out to the respective parties. Just contact Michael directly about getting this |
| 15 | for terminated or non-terminated parties; just tweak the letter. Sara Shackleton Enron |
| 16 | I heard in a conference call of many parties just now from the large oil co rep (Evie |
| 17 | Hi! I'd like to invite you to: Super Bowl Party Just click on the link below, and it will take |
| 18 | weekend. Who is giving you a pool party?? Just hope you hadn't flaked out yet. How |
| 19 | Hi! I'd like to invite you to: Super Bowl Party Just click on the link below, and it will take |
| 20 | an equal amount for his prospective partner just across the street. Enron 's lawyers were |
| 21 | Hi! I'd like to invite you to: Super Bowl Party Just click on the link below, and it will take |
| 22 | Ward Manager, West Gas Origination Patti, Just a reminder that we are no longer |
| 23 | this request. Thanks. Sara Shackleton Paul: Just checking to see if you are the lawyer |
| 24 | .hotmail.com we will have a list of patterns just in case people don't know what they are, |
| 25 | remove you from the list. Patricia: Just checking to see if you received this |

| N | Concordance |
|---|---|
| 1 | 01/05/2001 12:12 PM Hello Kate- El paso just called to verbal some deals and he had |
| 2 | nothing new on the subject of El Paso. Just curious if you knew where we are at in |
| 3 | the " rerun" dance at the edf man pary just over a year ago!!! I ordered 3 different |
| 4 | mean that we would pay a third party, just someone else at Enron. I agree with Rod |
| 5 | t-shirts that say "class of '94 - still partyin'". just tell your girlfriends we are going to lake |
| 6 | contacts at OneOk in the recent past. Just this morning I talked with a Jerry Knight |
| 7 | to the ISO. Chris 503-464-8219 Pat, Just a reminder that I will need the MW diff/ |
| 8 | your office. All he needs is your passwords. Just wondering if you had a chance to |
| 9 | with the issues, or head him off at the pass. Just let me know. Russell Diamond/ |
| 10 | . If you have forgotten your password, just go to the login screen, hit forgot |
| 11 | What: 6th annual John, Robert & Mo's Party (just 5th annual for me). 2kegs, assorted |
| 12 | would free up legal time for both parties. Just food for thought. Kent <> - Package |
| 13 | an equal amount for his prospective partner just across the street. Enron 's lawyers were |
| 14 | and books out to the respective parties. Just contact Michael directly about getting this |
| 15 | for terminated or non-terminated parties; just tweak the letter. Sara Shackleton Enron |
| 16 | I heard in a conference call of many parties just now from the large oil co rep (Evie |
| 17 | Hi! I'd like to invite you to: Super Bowl Party Just click on the link below, and it will take |

ing to zero ----- Original Message ----- nice. I just talked to fayez and he said it will be up

questions, please call me. <> Steve, FYI, I just talked with Jo and she will take a look a

ing to zero -----Original Message----- nice. i just talked to fayez and he said it will be up

just check in and see how you are. Also, I just talked to Gale Ford, massage lady, and

f the 58 million by PG&E is a CPUC issue. I just talked to Becky and I don't have a good

PL I don't think she will go for it. I just talked to Jennifer. Can you do lunch on

and paid! I appreciate it! Gloria Gloria, I just talked to Trang and she said that they

o me on your approval. Thanks! Jeff and I just talked and Jeff is going to go back to

info about this position? Vince Hi Esther, I just talked to John Sherriff about his "Enron"

Bill Shelley Corman 07/18/2000 11:56 PM I just talked to Cynthia Sandherr about this an

eopold@dowjone= s.com =20 <> =20 Sue: Just talked to the Chamber. Add them to the

> yah, let's get it out of the way. Just talked to Rene. He says he'll be home a

PM - John M Energy Team Request.doc <> Just talked to Shapiro. He wants Lay, Alex

Michael: Just wanted to let you know that I just talked to Don Miller, the lead guy from

and Houston Video Conference Rooms I just talked to Tino Monaldo, Dennis Langley's

from my group to this meeting. Chris I just talked with Dan Rothfuss of CG&E and it

| N | Word | With | L5 | L4 | L3 | L2 | L1 | Centre | R1 | R2 | R3 | R4 | R5 |
|---|------|------|----|----|----|----|----|--------|----|----|----|----|----|
| 1 | A | just | 222 | 228 | 190 | 94 | 18 | | 1,446 | 855 | 461 | 375 | 356 |
| 2 | WANTED | just | 20 | | 3 | 2 | | | 1,321 | | 13 | 7 | 5 |
| 3 | LET | just | 12 | 17 | 8 | 73 | | | 902 | 86 | 317 | 62 | 35 |
| 4 | TO | just | 343 | 374 | 259 | 221 | 262 | | 677 | 3,165 | 789 | 733 | 537 |
| 5 | GOT | just | 14 | 5 | 1 | 3 | 4 | | 619 | 4 | 7 | 17 | 26 |
| 6 | WANT | just | 20 | 32 | 32 | 42 | 11 | | 446 | 12 | 64 | 13 | 20 |
| 7 | IN | just | 137 | 176 | 109 | 54 | 104 | | 424 | 236 | 173 | 242 | 176 |
| 8 | THE | just | 435 | 540 | 538 | 379 | 20 | | 343 | 1,192 | 897 | 877 | 822 |
| 9 | AS | just | 47 | 41 | 49 | 32 | 7 | | 337 | 29 | 122 | 48 | 60 |
| 10 | NEED | just | 26 | 26 | 34 | 22 | 3 | | 331 | 3 | 49 | 29 | 34 |
| 11 | KIDDING | just | | 3 | 1 | | 1 | | 282 | | | | 2 |
| 12 | RECEIVED | just | 2 | 1 | | 3 | 1 | | 265 | 2 | 1 | 6 | 17 |
| 13 | LIKE | just | 48 | 26 | 36 | 46 | 18 | | 264 | 31 | 21 | 34 | 30 |
| 14 | HAVE | just | 68 | 81 | 88 | 38 | 381 | | 243 | 18 | 82 | 117 | 113 |
| 15 | SPOKE | just | 1 | | | | | | 243 | | | | 4 |
| 16 | ONE | just | 32 | 12 | 21 | 23 | 8 | | 187 | 54 | 65 | 40 | 29 |
| 17 | CHECKING | just | 1 | | | 2 | | | 178 | 1 | | 1 | 5 |
| 18 | SENT | just | 6 | 9 | 4 | 1 | | | 178 | 5 | 2 | 3 | 9 |
| 19 | TALKED | just | 2 | | 3 | | | | 165 | | 3 | 5 | 7 |
| 20 | ABOUT | just | 26 | 48 | 33 | 14 | 22 | | 164 | 152 | 59 | 86 | 67 |
| 21 | CALLED | just | 5 | 4 | 1 | 5 | 3 | | 164 | 1 | 5 | 1 | 5 |
| 22 | THOUGHT | just | 9 | 10 | 36 | 8 | 2 | | 161 | 83 | 15 | 11 | 7 |
| 23 | FOR | just | 149 | 141 | 114 | 76 | 75 | | 160 | 163 | 188 | 213 | 225 |
| 24 | DON | just | 25 | 25 | 107 | 14 | | | 156 | 10 | 21 | 21 | 20 |
| 25 | GIVE | just | 15 | 6 | 4 | 1 | 1 | | 156 | 17 | 93 | 23 | 20 |

concordance   collocates   plot   patterns   clusters   timeline   filenames   source text   notes

**WordList**

| N | Word | Freq. | % | Texts | % |
|---|---|---|---|---|---|
| 1 | # | 1,047,28 | 7.26 | 57 | 100.00 |
| 2 | THE | 619,457 | 4.30 | 46 | 80.70 |
| 3 | TO | 428,044 | 2.97 | 46 | 80.70 |
| 4 | AND | 280,432 | 1.95 | 46 | 80.70 |
| 5 | I | 242,346 | 1.68 | 49 | 85.96 |
| 6 | OF | 237,608 | 1.65 | 46 | 80.70 |
| 7 | A | 220,627 | 1.53 | 50 | 87.72 |
| 8 | YOU | 197,647 | 1.37 | 46 | 80.70 |
| 9 | IN | 180,887 | 1.25 | 46 | 80.70 |
| 10 | FOR | 166,619 | 1.16 | 46 | 80.70 |
| 11 | IS | 157,322 | 1.09 | 46 | 80.70 |
| 12 | THAT | 142,203 | 0.99 | 46 | 80.70 |
| 13 | ON | 125,236 | 0.87 | 46 | 80.70 |
| 14 | THIS | 120,814 | 0.84 | 46 | 80.70 |
| 15 | WE | 115,535 | 0.80 | 45 | 78.95 |
| 16 | BE | 108,110 | 0.75 | 45 | 78.95 |
| 17 | HAVE | 104,644 | 0.73 | 46 | 80.70 |
| 18 | WITH | 99,321 | 0.69 | 46 | 80.70 |
| 19 | IT | 95,364 | 0.66 | 46 | 80.70 |
| 20 | S | 90,176 | 0.63 | 51 | 89.47 |
| 21 | WILL | 87,910 | 0.61 | 46 | 80.70 |
| 22 | ARE | 83,964 | 0.58 | 46 | 80.70 |
| 23 | IF | 75,231 | 0.52 | 45 | 78.95 |
| 24 | AT | 71,299 | 0.49 | 46 | 80.70 |
| 25 | AS | 70,059 | 0.49 | 46 | 80.70 |

# Intertextuality

XXX, **just spoke to** XXX and **agreed** to take the CSA out and include some changes in the credit provisions. **He will be communicating with you** to include these changes. I would need you to let me know once you have finished **the document**, so as to fax me the **modified pages of the Master**.

We intend to meet Capex tomorrow morning to present them the ISDA. Pls let me know if you can't have it ready for this afternoon.

# Intertextuality

As per your request …

As you recall …

As earlier requested …

As discussed …

As previously reported …

As amended by Change order #1 …

As a result of these transactions …

As you and I discussed last week …

doc will work). Thanks, Karen. Jonathan, Per Jay's request I am forwarding you the

efforts in scheduling this call. John Per your request, I have scheduled a

Joan Veselack 03/23/2000 11:50 AM John, per voicemail. Reduced deliveries to Allied

elative to Elec. compression projects. John, Per our discussion, attached is a draft form

D Trevino 01/12/2000 11:20 AM Julie, Per our conversation yesterday, the deals/

original contract $2.29-$0.15 = $2.14 July Per original contract $2.62-$0.15 = $2.47

original contract $2.29-$0.20 = $2.09 July Per original contract $2.62-$0.20 = $2.42

if you have any question. Regards, Julian Per my email, you will need Julian on the

Sustaita 01/04/2001 08:56 AM Joya, Per the request below. (1.) I will need to

the printer (May 5). Thanks. Joyce Joyce: Per my earlier e-mail message, I have had a

the printer (May 5). Thanks. Joyce Joyce: Per my earlier e-mail message, I have had a

Kopetzy (cc: on this e-mail). Thanks. John, Per our phone conversation, here is how to

dated supply demand analysis. John John, Per your request this fall for updated Ontario

for light microwave popcorn. PL FYI John- per our conversation about getting all calc

til we know VAR, capital, etc. Mitch John, Per Kevin, we've already sent our requests.

ou!!! I'm free from 4:00-5:00 today John, Per your and Jeff's request, Research and

olf 212-648-8231 212-648-5125 fax John: Per our conversation, attached is the copy of
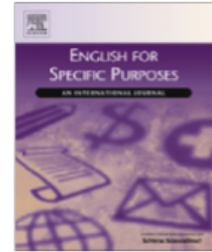
# "*Just spoke to …*": The types and directionality of intertextuality in professional discourse

## Martin Warren

Department of English, The Hong Kong Polytechnic University, Hong Kong

**ARTICLE INFO**

**ABSTRACT**

E-mail communication in business and professional contexts has been analysed with a focus on textual and communicative complexity, stylistic conventions, and the relation between e-mails and professional culture. The purpose of the present study is to compare the professional literacy in two professions by studying the use of intertextuality in the e-mail messages two professionals read or wrote in a number of discourse flows. Intertextuality is examined in terms of types and directionality of use. The findings show that while the use of intertextuality is prevalent across all of the e-mails, the types of intertextuality, and their relative frequencies of use, and the directionality of intertextuality are influenced by profession-specific communicative contexts and goals.

# Idioms and metaphors

- Normally taught as they come up, or in lists. Rarely taught in context.

- What is not common in published materials are examples of creative use of metaphors.

# Ball – 602 instances

On the ball
Get the ball rolling
Keep the ball rolling
Move this ball forward
Ball park
Ball games
Ball in my court
Drop the ball
Play hard ball

Whole ball of wax
A wrecking ball
Have a ball

# Get the ball rolling

- I thought I'd **get the ball rolling** and try to get …
- In order to **get the ball rolling**, I ask that you …
- Hopefully this will at least **get the ball rolling**. It is then up to …
- … a draft of the agreement to "**get the ball rolling**" again. I believe I have included …
- I had to call today and **get the ball rolling** because it was …
- … and would like to **get the ball rolling** as soon as possible.
- Anyway, just some thoughts to **get the ball rolling**. I'm open to just about anything.

# Have a ball

- Sounds like you guys **had a ball** at dinner.
- Hope you **had a ball**
- I **had a ball** at XXX's party Saturday
- Re looking forward to **having a ball**

# Play hard ball

- XXX is trying to **play hard ball**
- XXX indicated that he was going to **play hard ball** with his interruptible shippers
- They wanted an answer today, but I **played hard ball** and told them that I had to have more time
- XXX is trying to act right, but I **am playing hard ball**.

# On the ball

… but at least from June 1, we will be totally **on the ball**.

Let's talk tomorrow about how we get everyone **on the ball** and thinking.

# Keep the ball rolling

Just want to **keep the ball rolling**.

… I want to **keep the ball rolling at light speed**.

# Ball in court

Is this **ball in my court**? Let me know.

But overall I believe that the **ball is still ' in your court'.**

# Ball park

They asked for a **ball park** and I could not help them.

Do you recall a **ball park** figure on the costs.

No sense in offering something up if we don't know what the numbers look like, at least in a "**ball park**" sense.

Could you review from XX's perspective and see if you come up with numbers in the same **ball park**?

I will pretty much leave it in your **ball park** about Friday night. (court???)

The baseball will begin at 3:05, and the **ball park** opens at 1:00. (no metaphor)

# Drop the ball

You can count on us not **dropping the ball**.

They really **dropped the ball** on that one, although I wish I had …

XXX **dropped the ball**, unfortunately.

Sorry if I **dropped the ball**

Someone REALLY **dropped the ball** here!

If we **dropped the ball** on this, I apologize.

While I am not sure who **dropped the ball** on this item, the fact remains …

I owe you an apology for **dropping the ball**

# Drop the ball

You can count on us not **dropping the ball**.

They really **dropped the ball** on that one, although I wish I had …

XXX **dropped the ball**, unfortunately.

<span style="color:red">Sorry</span> if I **dropped the ball**

Someone REALLY **dropped the ball** here!

If we **dropped the ball** on this, I <span style="color:red">apologize</span>.

While I am not sure who **dropped the ball** on this item, the fact remains …

I owe you an <span style="color:red">apology</span> for **dropping the ball.**

# Drop the ball

Just wanted to let you know that I've received all the … stuff and that you can count on me to completely **drop the ball**.

Kidding!  I'll take care of it.

# Red

A copy of the agreement is kept in a red file (near Sam's desk)

Attached is the red-line version of the …

So it's a red-letter day for us …

I will raise a red flag …

He might be barking at a red herring.

My answers below in red.

Take the red-eye flight home on Monday night.

So let me know if I have any money coming to me or if I am in the red.

… investors fear more red ink splashed over the balance sheets …

I drank more red wine than beer that weekend.

The paperwork and red tape have seemed unending …

I've already marked up w/ red pen …

# Outline

- summarize some of these research findings,
- demonstrate just how easy it is to access the corpus, and
- **discuss the implications for business English teaching**.

# Implications

- Publishers need to become more aware of this corpus– too many books are B2C
- Teachers need to become more aware of this corpus, and learn how to exploit it
  - Great examples of company internal mails
  - Great examples of relational talk
  - Great examples of creative language

# References

Carvalho, V.R., & Cohen, W.W. (2007). Recommending recipients in the Enron email corpus. Technical Report, Carnegie Mellon University. Retrieved from http://www.cs.cmu.edu/~wcohen/postscript/cc-predict-submitted.pdf

Cooke,N., Gillam,L., & Kondoz, A. (2007, July). The best kept secrets with corpus linguistics. In *Proceedings of the 4th Corpus Linguistics Conference*, Birmingham, England. Retrieved from http://www.cs.surrey.ac.uk/BIMA/People/L.Gillam/downloads/publications/TBKSWCL.final.pdf

De Felice, R; Darby, J; Fisher, A; Peplow, D; (2013). A classification scheme for annotating speech acts in a business email corpus. *ICAME Journal, 37*, 71-105. Retrieved from http://icame.uib.no/ij37/Pages_71-106.pdf

De Felice, R. & Deane, P. (2012). Identifying speech acts in e-mails: Toward automated scoring of the TOEIC® e-mail task (ETS Research Report ETS RR–12-16). Princeton, NJ: ETS. Retrieved from http://www.ets.org/Media/Research/pdf/RR-12-16.pdf

Diesner, J, & Frantz, T.L. & Carley, K.M. (2005). Communication networks from the Enron email corpus, "It's always about the people. Enron is no different". *Computational and Mathematical Organization Theory, 11*(3), 201-228.

Drutman, L., & Hopkins, D.J. (2013). The inside view: Using the Enron email archive to understand corporate political attention. *Legislative Studies Quarterly. 38*(1):5-30.

Lampert, A., Dale, R., & Paris, C. (2008). Requests and commitments in email are more complex than you think: Eight reasons to be cautious. In *Proceedings of the Australasian Language Technology Association Workshop* (pp. 64-72). Hobart, Australia. Retrieved from http://www.alta.asn.au/events/alta2008/proceedings/pdf/ALTA2008_09.pdf

Lancaster, N. (2010). A corpus study of email writing in a business setting and its practical application in teaching English as a second language. (Master's thesis, Ball State University). Retrieved from http://cardinalscholar.bsu.edu/bitstream/123456789/193693/1/Lancastern_2010-2_BODY.pdf

Leber, J. (2013). The immortal life of the Enron e-mails. Retrieved from http://www.technologyreview.com/news/515801/the-immortal-life-of-the-enron-e-mails/

Styler, W. (2011). The EnronSent Corpus. Technical Report 01-2011, University of Colorado at Boulder Institute of Cognitive Science, Boulder, CO.

Warren, M. (2013). "Just spoke to …": The types and directionality of intertextuality in professional discourse. *English for Specific Purposes, 32*(1), 12-24.

Waterman, K.K. (2006). Knowledge discovery in corporate email: the compliance BOT meets Enron. (Master's thesis, Massachusetts Institute of Technology). Retrieved from http://dspace.mit.edu/bitstream/handle/1721.1/37574/85813548.pdf?sequence...

# "And why not?" Exploiting the Enron email corpus in the business English classroom

Evan Frendo