Relating rhetoric to functions

Strategy for integrating DocuScope and functional coding in the CITF email analysis

Created November 13, 2026
Updated November 14, 2026

Table of contents

1 Rhetorical profiles **2**

2 Challenges **2**

2.1 Detecting rhetoric in fragments 2
2.2 Defining and varying rhetoric 3
2.3 Disentangling rhetoric's relationship with role and function . . . 3

> **Note**
>
> I wrote this document on November 13, 2026, following a comprehensive assessment of DocuScope for tagging the rhetoric of the CITF emails. The restructuring plan has since superseded this document, but its concerns about relating DocuScope tags to functional categories still stand. The plan's section on relating rhetoric, functions and actions builds on them.

The project relates how people wrote (the how) to what their emails did (the what), with sender role and workflow phase as context. DocuScope supplies the how and nothing else. Functional coding supplies the what, through codes derived from the emails without an initial codebook and grouped into functional categories. The two streams stay separate until integration, so any association between them reflects the emails themselves.

Relating the two streams needs a method that first identifies the functional units of a text and then describes their language. Corpus-based move analysis does this. It fits the project for two reasons. It defines function before it examines language, as the functional coding does. It compares functional and linguistic analyses derived independently, which matches keeping the two streams separate. The project takes the sentence as its unit, since sentences need no hand segmentation. A sentence can carry several functional categories, and it counts in each.

The central analysis is the variation within each functional category, explained by role, phase and the relationship between sender and recipient, and read closely at prototypical and peripheral cases.

1 Rhetorical profiles

We record whether each DocuScope category appears in each sentence and pool those records across the sentences in each functional category. For example, "Could you send the agreement?" holds only a few tags, so the share of each DocuScope category in it would swing with every tag. Recording only whether it contains each category gives a stable yes or no, and across all requests those answers add up to proportions, such as the share of requests that apologise. Longer sentences contain more categories, so the comparison controls for sentence length.

We compare the rhetoric of each functional category with all other sentences in the corpus. A category's profile means little on its own, and the comparison shows what sets that work apart. For example, if a third of requests contain an apology against a tenth of all other sentences, apologies characterise requests.

Before reporting a pattern, we read a sample of the relevant emails to check what the pattern means. For example, the closing courtesy "don't hesitate to contact me" should count as Facilitate, yet DocuScope tags "hesitate to" as Uncertainty. Close reading of these emails confirms whether a pattern reflects the writing or a tagging error.

2 Challenges

2.1 Detecting rhetoric in fragments

A DocuScope tag in a sentence with a functional category shows only that the two co-occur. Take the follow-up "Sorry to bother you again, could you send a copy of the agreement we signed last week?" "Could you send" is the request itself. "Sorry" is an apology that softens the request. "The agreement we signed last week" recounts an earlier event. DocuScope may tag all three, but only the first two tags describe how the writer made the request.

We considered three ways of linking a tag to the act it describes, and set each aside. Linking each DocuScope tag to the verb that performs the request, through dependency parses, would rest on single tags, such as "sorry" in the follow-up above, and DocuScope is not designed for single-tag readings. Counting the tags in the sentences before and after each sentence would catch framing such as "I know you're busy, but", yet those sentences carry their own functional categories, so counting their tags twice would blur the comparison between categories. Linking a tag to a request when its clause stands in a marked relation

to it, such as a condition ("if you have time") or a concession, works only where connectives mark the relation, and again rests on single tags.

We therefore accept co-occurrence as the measure. Some tags in a sentence describe something other than its function, as "the agreement we signed last week" does in a request, and the comparison with the corpus absorbs them where they occur evenly across functional categories. Tags that cluster in one category, as accounts of earlier events may in follow-ups, need checking in the close reading before they count as its rhetoric.

2.2 Defining and varying rhetoric

Some rhetoric defines a functional category, and some reflects how writers chose to frame the work. For example, nearly every request addresses the reader and asks for something, as in "could you send", so that rhetoric belongs to the request itself. Whether the writer apologises, sets a deadline or explains the need varies from request to request, and that variation is the kind of choice worth tracking. DocuScope tags words independently of their functional context, and it is necessary to differentiate these two kinds of rhetoric.

Within each functional category, we separate rhetoric present in nearly every sentence from rhetoric that varies across sentences. For example, nearly every request addresses the reader, while only some apologise, set a deadline or explain the need. The varying part shows how writers chose to frame the work.

2.3 Disentangling rhetoric's relationship with role and function

The rhetoric of a functional category may reflect who writes it. If CITF coordinators write most requests and study teams write most reports, a difference in rhetoric between requests and reports may come from the writers, from the work, or from the way a writer's role shapes how they do the work. Distinguishing these depends on people who write in several functional categories, and such writers may be few.

We compare the same senders across functional categories, and model sender and study as random effects. For example, a coordinator who writes both requests and reports shows whether requests read differently from reports when the writer stays the same. If too few senders write in several functional categories, the random effects carry the comparison alone, and the analysis reports the confound as a limit.