

Analysing Speech Acts in Organisational Communication

Marcus Grattan

Computer and Information Science
Linköping University, Sweden
margr792@student.liu.se

Andrea Fried

Management and Engineering
Linköping University, Sweden
andrea.fried@liu.se

Arne Jönsson

Computer and Information Science
Linköping University, Linköping, Sweden
arne.jonsson@liu.se

Abstract

The research presented in this paper addresses the challenges of analysing organisational stakeholder communication using speech act theory, supported by a novel analytical model. We present a large, automatically annotated dataset focused on present- and future-oriented speech acts in Swedish organisational communication, specifically with regard to the information security standard ISO/IEC 27001. To construct and evaluate the dataset, we employ a hybrid annotation pipeline combining quantised large language models, retrieval-augmented prompting, and a fine-tuned multilingual XLM-RoBERTa classifier, which is evaluated against a manually annotated gold standard. The classifier is subsequently applied to a large corpus of corporate texts from Swedish companies' websites, enabling comparative analysis between general corporate communication and ISO-focused discourse. The results show that ISO/IEC 27001 communication is dominated by assertive speech acts, with commissives playing a secondary role, reflecting an emphasis on factual compliance rather than aspirational or promotional claims. Beyond its empirical findings, the study contributes a reusable dataset and classifier to the CLARIN infrastructure, demonstrates how CLARIN K-centres can support social science research, and provides insights into the use of quantised large language models for speech act classification in mid-resource language settings.

1 Introduction

In a joint project between the Swedish SMS CLARIN K-centre¹ and management researchers, we investigate stakeholder communication regarding information security on Swedish corporate websites. Rogers (2002) argues that preventive innovations, such as those aimed at ensuring information security, are often adopted slowly because their benefits to organisations are intangible and long-term. Clear communication with stakeholders that emphasises the benefits of information security is therefore essential, particularly given the regulatory context that demands accountability for managing not only social and environmental risks, but also technical risks.

Recent work suggests that stakeholder communication of organisations is influenced by two circumstances. First, it is influenced by the fact that preventive innovations not only generate indirect benefits in the form of social capital, such as reputation and shared understandings, but can also yield direct benefits in the form of economic capital gains (see Bourdieu (1986), Kuhn (2008), and Nahapiet and Ghoshal (1998)). For instance, companies can monetise safety training and consultancy services around information security (Hu et al., 2022; Khando et al., 2021; Siponen, 2000) generating revenue streams and increasing economic capital (Bourdieu, 1991). Yet, such economic interests may weaken stakeholder communication and potentially undermine social capital in the longer run (Kuhn, 2008). The assumption is that the pursuit of direct economic benefits encourages organisations to make more promises, relying heavily on commissive utterances, which can crowd out dialogic, trust-building communication and,

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹<https://sprakbanken-clarin.lingfil.uu.se/centra/sms-en.html>

if unmet, erode credibility, mutual learning, and durable social ties among stakeholders (see Cooren & Taylor, 1997; Searle, 1985).

Secondly, we can assume that organisations with standard certifications (here: ISO/IEC 27001 certification) for information security tend to focus more on their achievements when communicating with stakeholders. Upon certification, they undergo an external audit in which their achievements in information security were recognised. Therefore, they presumably rely more than non-certified organisations do on assertive utterances to demonstrate their credibility, compliance with regulations and leadership in information security. This could potentially result in a tone that is less self-promotional and less promising on their websites (see Cooren and Seidl, 2020; Kuhn, 2008; Schoeneborn et al., 2019; Taylor, 1993).

The ISO/IEC 27001 standard requires organisations to establish, implement and audit the rules and authorities relating to the legal and ethical aspects necessary for the proper functioning and accountability of the entire data and algorithm lifecycle (Janssen et al., 2020, p.3), see also (Culot et al., 2021; Disterer, 2013; Mirtsch et al., 2021). ISO/IEC 27001 certification ensures information security and encourages good corporate behaviour by requiring compliance to be audited by external groups.

We therefore ask: How do Swedish companies communicate information security to stakeholders on their corporate websites? Specifically, we are interested in understanding if the speech acts in these communications are influenced by (a) the pursuit of social versus economic capital gains and (b) ISO/IEC 27001 certification. The objective is to determine if explicit ISO/IEC 27001 communication reflects distinct, more assertive speech acts that emphasize descriptive content and speech act logic, as opposed to commissive speech acts that express future-oriented intentions to ensure information security. Methodically, the paper aims to develop a novel approach to computational text analysis that is sensitive to specific organizational topics by analysing the content of corporate websites to reveal differences in speech act distributions.

2 The corpus and previous analyses

A specialised corpus of corporate texts from Swedish company websites was created in 2020, using a web crawler, to analyse the discourse surrounding information security with a focus on the ISO/IEC 27001 standards (Jönsson et al., 2024). The corpus comprises 441,625 sentences, filtered for duplicates. This method of data collection introduces significant noise, characterised by the presence of links, button labels, dates, and other extraneous elements. In addition, some sentences are incomplete, fragmented, or consist of multiple unrelated sentences, and many web pages are in English, or both English and Swedish. English text, both web pages or paragraphs in English and Swedish web pages, is filtered out. We further filtered out sentences with more than 250 words and paragraphs with more than 500 sentences, see (Jönsson et al., 2024). The final corpus contains data from 291 companies, a total of 5,960,328 tokens, with a vocabulary size of 204,312 and an average sentence length of 24.3 tokens.

Following Mirtsch et al. (2020)'s suggestion, we first identified companies that were certified for ISO/IEC 27001. These companies adopted the standard and underwent a related audit. We call these companies *Compliers*². Mirtsch et al. (2020) also discovered companies that implemented the requirements without being certified for ISO/IEC 27001, which we refer to as *Stewards*. Beyond these findings, we identified a third group of companies that do not claim to have implemented the ISO/IEC 27001 standard requirements. Rather, they refer to their suppliers and other stakeholders to inform of their secure handling of information, we call this group *Advocates*. Furthermore, our dataset revealed that some companies have integrated ISO/IEC 27001 consulting or training into their business models, creating an opportunity for economic capital gains. Other companies recognize only the preventive nature of adhering to ISO/IEC 27001 requirements and focus on social capital gains.

Based on these categorisations, we partitioned the corpus into subsets. The first denotes the three approaches to standard adherence *Compliers* (C), *Stewards* (S), and *Advocates* (A). The second signifies social capital (S) or economic capital (E) respectively. These categories help to understand the varying approaches of companies to information security communication and are added manually to the dataset.

²Compliers were previously referred to as Adopters (Grattan et al., 2025).

We can also form six distinct types of adoption of preventive innovation, referred to as CS, CE, SS, SE, AS, AE³, in which each company belongs.

Our previous analyses of the corpus included word frequency counts, word clouds, and content analysis along with sentiment analysis to uncover differences in communication about preventive cyber innovations (Jönsson et al., 2024). To capture future-versus-present-oriented communication in relation to InfoSec standards, we conducted a speech acts analysis, which is presented in this paper.

We will compare all paragraphs, Section 5.1, with only those paragraphs containing ISO sentences, Section 5.2. This allows for analysis of separate general patterns of organizational website communication from communication that is explicitly focused on ISO/IEC 27001.

Furthermore, looking at all paragraphs captures how information security is embedded in broader corporate self-presentations, where ISO/IEC 27001 is only one of many topics and is often mentioned indirectly or in passing. In contrast, the ISO paragraphs subset focuses on moments where organizations explicitly foreground the standard and therefore articulate more deliberate and institutionally framed meanings of information security. Comparing the two makes it possible to examine whether organizations change what they say and how they position themselves—for example, by emphasizing verified achievements, compliance, or formal responsibilities—when ISO/IEC 27001 becomes the explicit topic, as opposed to being part of general corporate discourse. This sheds light on whether ISO communication reflects a distinct content logic (e.g. legitimizing, evidencing, or distancing responsibility) rather than merely mirroring overall communication styles.

3 Model building

Our working assumption was that organisations express their adherence to standards through intentional statements (Commissives) or descriptive statements (Assertives). This approach aligns with the concept of direction of fit (Searle, 1985), where Commissives represent a world-to-word fit, expressing future intentions and commitments that an organisation plans to fulfil. Conversely, Assertives represent a word-to-world fit, stating current or past conditions within the organisation. For the purposes of streamlined analysis and because of their overlap with Assertives in expressing current realities, Declaratives were grouped under the Assertive class. Furthermore, Expressives and Directives were deemed outside the scope of this analysis and thus grouped, along with unclear and fragmented data, under Other. A summation of the classes used is provided in Table 1.

Class	Speech Act
Commissive	Commissive
Assertive	Assertive & Declarative
Other	Directive & Expressive & Unclear data

Table 1: Classes used in the text classification task of this paper.

The construction of the Speech Act classification model began with selecting a large language model (LLM) for an initial round of annotation. Specifically, only quantised models were considered due to computing restraints. The selection was based on the performance on a small manually curated data set, sampled from the ISO/IES 27001 corpus (Jönsson et al., 2024), containing approximately 100 manually selected instances per class. Initially, two models were considered. That is, Mixtral8x7B-Instruct quantised to an average of 3.5 bits per weight (BPW), and Llama3-70B-Instruct with 2.4BPW. The models were then prompted using the same prompt, which included Role-, Chain-of-Thought-, and Few-Shot prompting strategies. The metrics considered for model selection were Accuracy and per-class F1 score. Llama3 showed superior scores, as per Table 2, across the metrics tested and was therefore selected.

After selecting Llama3, an optimal hyperparameter configuration was sought. The hyperparameters considered were *Top P*, *Top K*, and *Temperature*. Other parameters were, given the exponential increase in computation time, not considered.

³In a previous paper Jönsson et al., 2024, these were referred to as 11, 12, 21, 22, 31 and 32, respectively.

Model selection		
Metrics	Mixtral	Llama3
Accuracy	0.695	0.786
Commissive F1	0.785	0.826
Assertive F1	0.645	0.767
Other F1	0.675	0.774

Table 2: Model selection results

Using the dataset mentioned above, each parameter permutation (with a defined range and step-size) was tested in a two round process. The first round tested a broad range of configurations to identify a potential subset of parameters that yielded task-optimal results. This was achieved by defining a wide range and large step sizes for each parameter, with performance assessed on the basis of accuracy and per-class F1 score.

The first round of testing resulted in a set of permutations that indicated increased performance on the dataset, compared to results from the model selection process. The configuration that resulted in the highest accuracy was picked from the first round for the second round of tuning. The second round of testing aimed to fine-tune the model hyperparameters around the most promising configurations identified in the first round. Thus, this round of tests involved narrowing the range and reducing the step size to fine-tune the parameters, seeking possible optimal configurations within a narrower range.

Despite observing a slight decrease in overall accuracy in the second round of testing, the decision to proceed with the parameters from the second round was driven by a significant improvement in the F1 score for the Commissive class, which increased from 0.766 to 0.851. This improvement indicated that the model more accurately identified commissive instances. Furthermore, the improved accuracy from the first round of testing seemed to stem from better performance on the Other class. The results from the chosen hyperparameter configuration from each round during the tuning process are summarised in Table 3.

Hyperparameter tuning: Results		
Metrics	First-round	Second-round
Accuracy	0.792	0.789
Commissive F1	0.766	0.851
Assertive F1	0.768	0.762
Other F1	0.854	0.764

Table 3: Hyperparameter Tuning: Results.

After the hyperparameter tuning process, the selected model was deployed to annotate a larger section of the corpus in order to create a demonstration dataset for Clue and Reasoning Prompting (CARP) (Sun et al., 2023). The demonstration dataset is used for few-shot example retrieval in later classification steps and consists of a text, its classification, supporting clues (e.g keywords, phrases, and/or syntactic features), as well as a supporting diagnostic reasoning for the classification given the clues. The supported clues and diagnostic reasoning are generated by the LLM.

Generated annotations for the demonstration dataset were manually validated, and correctly labelled instances were organised into class-specific datasets, each containing the original text and its corresponding Speech Act label. Ultimately, each class contained 550 validated entries. To ensure diversity and mitigate model bias, 50 instances per class were manually curated by reviewing and correcting the model’s

misclassifications. Due to data sparsity in the minority class, Commissive, ChatGPT-4 was employed for data augmentation. Forty carefully selected examples were expanded to better capture under-represented syntactic and semantic structures.

To complete the process of constructing a demonstration dataset the LLM was prompted, following the procedure described in Sun et al. (2023), to generate supporting clues and a diagnostic reasoning for each entry. This process yielded a 1,650-entry demonstration dataset, evenly distributed across classes.

To support classification-sensitive demonstration retrieval in the final annotation round, an XLM-RoBERTa model (Conneau et al., 2020) was fine-tuned on the CARP demonstration dataset. Fine-tuning was essential for aligning the model not only with sentence embeddings but also with class annotations, enabling accurate retrieval of semantically and label-relevant examples for unseen sentences. The model was used to compute text embeddings, which were appended to each dataset entry.

In the final annotation step the Llama3 model was provided with task instructions, class definitions, and retrieved demonstrations via the fine-tuned retrieval model. Demonstrations were added up to the model’s token limit, sorted in ascending order of cosine distance to the input sentence, as described by Sun et al. (2023).

This resulted in an automatically annotated dataset, comprising a random 56% subset of the full corpus. A final XLM-RoBERTa classifier was fine-tuned on this dataset. The model and dataset are available to the CLARIN community⁴. Training employed a linear learning rate scheduler with warm-up steps corresponding to 10% of batches per epoch and used the AdamW optimiser. To address class imbalance, the loss function incorporated class weights. Early stopping was triggered if validation loss increased for three consecutive evaluations. Weight decay was set to 0.01, and the learning rate was set to a conservative value of 4×10^{-5} to prevent premature convergence. A batch size of 16 was chosen to encourage generalisation. An 80/20 train/validation split was used, with evaluation occurring three times per epoch. Model selection was based on the weighted F1 score. The final model was trained for 2.33 epochs. Entries labelled "No match found" were excluded from the training data.

To evaluate the classifier, a test set of 5,000 entries was manually annotated. An average of 1,000 sentences were labelled per day, split across two sessions. The final test set comprised 665 Commissive, 3,228 Assertive, and 1,107 Other instances. Following Abercrombie et al. (2025), intra-annotator agreement was measured on 10% of the test set one week after completion, yielding an agreement score of 0.8, indicating strong consistency.

For more detail and discussion related to the model-building process, see Grattan (2024).

4 Classifier model results

The automatically annotated dataset comprised 138,052 entries excluding instances for which no annotation was found. In total the automatically annotated dataset included 32,458 (24%) commissive entries, 80,377 (58%) assertive entries and, 25,215 (18%) entries annotated as other. The model returned no annotation 301 times. Dataset metrics are presented in Table 4

Automatically Annotated Dataset				
Metric	Commissive	Assertive	Other	Total
Count	32,458	80,377	25,215	138,052
Percentage	24	58	18	

Table 4: Automatically annotated dataset metrics

The classifier model achieved an accuracy of 0.73 and a weighted F1-score of 0.75 on the test dataset, outperforming the most frequent class baseline, which had an accuracy of 0.65 and a weighted F1-score of 0.51. The model’s weighted precision and recall were 0.80 and 0.73, respectively. For the Commissive class, the model achieved precision, recall, and F1-scores of 0.34, 0.68, and 0.45, respectively. In the

⁴<https://huggingface.co/MarcusGrattan/classifier-model/tree/main>

Assertive class, the model performed better with precision of 0.88, recall of 0.74, and an F1 score of 0.80. The Other class achieved a precision of 0.84, recall of 0.74, and F1 score of 0.79. Per class metrics are summarised in Table 5.

Per class model metrics			
Metric	Commissive	Assertive	Other
Precision	0.34	0.88	0.84
Recall	0.68	0.74	0.74
F1	0.45	0.80	0.79

Table 5: Per class results

5 Applying the model

The model was applied to the corpus of corporate websites of Swedish companies that communicate with their stakeholders (e.g. clients, customers, government, non-profit organisations and other) about their ISO/IEC 27001 implementation efforts, see Section 2. As in previous studies between the management researchers and the SMS CLARIN K-centre we looked at various categorisations and subsets of the corpus. In this case, we use the whole corpus as described in Section 2, called All paragraphs, and contrasted that with the subset comprising only paragraphs containing ISO-related terms, i.e. if a paragraph has at least one sentence with an ISO-related term the whole paragraph is included. The datasets are further divided into groups according to the various governance approaches, capital benefit dimensions, and preventive innovation adoption, see Section 2.

We performed statistical analyses on the data, including pairwise comparisons, using Cramér’s V, and no difference showed an effect. However, the primary goal of our study is pattern recognition and hypothesis generation rather than hypothesis testing. Traditional significance testing adds rigour to confirmatory research aiming to generalise the findings. We have chosen to focus on effect sizes and qualitative validation through illustrative examples that demonstrate the substantive meaning behind the numerical patterns. This approach acknowledges the exploratory nature of our analyses.

5.1 All paragraphs

The corpus comprising all paragraphs consists in total of 81,555 paragraphs with 441,625 sentences.

The results when data are categorised according to governance approaches are shown in Table 6 and visually in Figure 1, revealing little difference between the analysis categories Compliers, Stewards, and Advocates.

	Assertive	Commissive	Other
Compliers	100,547 (51)	47,805 (24)	48,673 (25)
Stewards	81,372 (52)	33,746 (21)	42,305 (27)
Advocates	51,404 (59)	18,568 (21)	17,205 (20)

Table 6: Speech act distribution for all paragraphs according to governance approaches. Number of sentences with percentage in parenthesis.

Table 7 and Figure 2 show the result when using the social/economic capital dimension. The results show very little difference between the dimensions.

Table 8 and Figure 3 show the result when combining the dimensions into six categories that represent the three governance approaches and the social/economic capital dimension. The results show very little difference between the six dimensions⁵.

⁵Here we have the largest value on Cramér’s V, 0.16, for CE vs AS, which is considered weak.

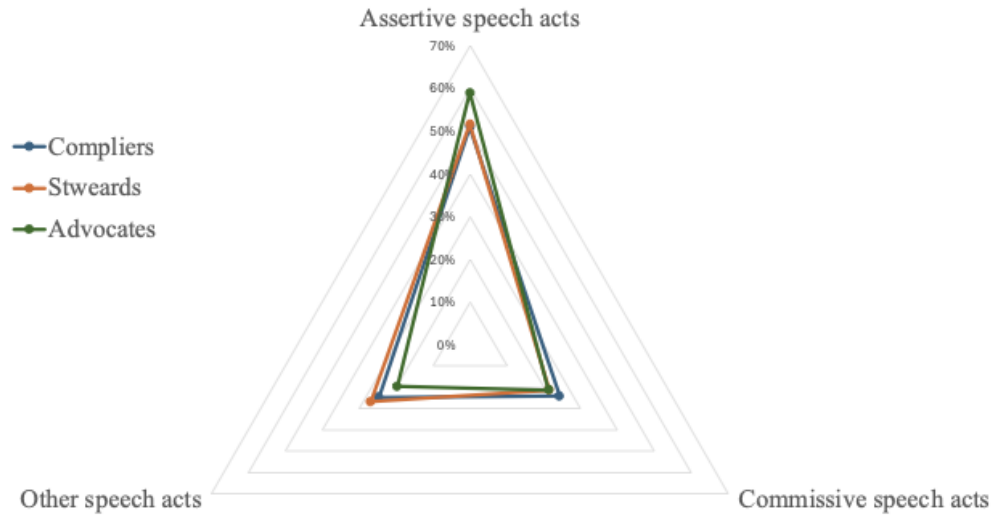


Figure 1: Diagram of the distribution for all paragraphs according to governance approaches, percentage.

	Assertive	Commissive	Other
Social capital	180,096 (52)	77,127 (22)	86,037 (25)
Economic capital	53,227 (54)	22,992 (23)	22,146 (23)

Table 7: Speech act distribution for all paragraphs for the social/economic capital dimension. Number of sentences with percentage in parenthesis.

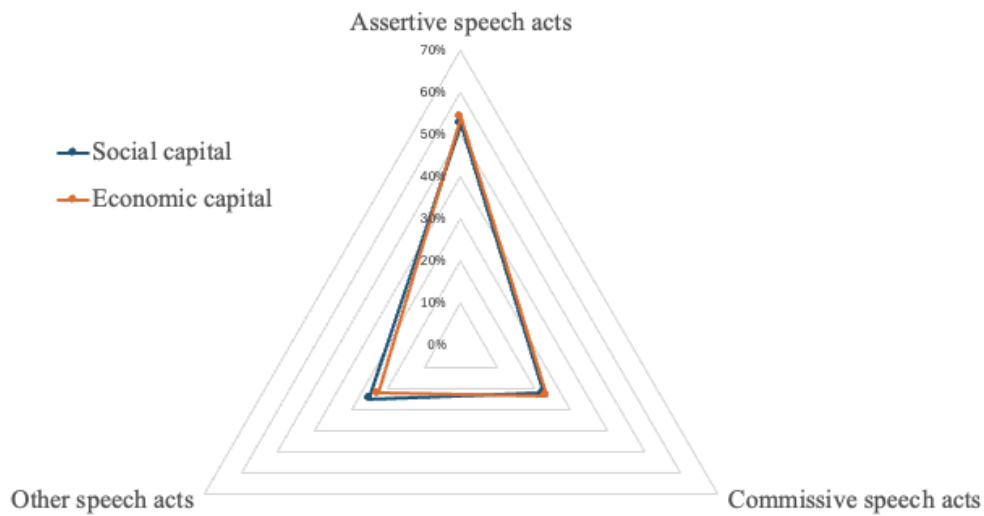


Figure 2: Diagram of the distribution for all paragraphs for the social/economic capital dimension, percentage.

5.2 ISO paragraphs

The corpus comprising ISO-related paragraphs consists in total of 1,492 paragraphs with 31,440 sentences.

The results when data are categorised according to governance approaches are shown in Table 9 and visually in Figure 4, revealing little difference between the analysis categories Compliers, Stewards, and

	Assertive	Commissive	Other
CS	95,015 (51)	44,779 (24)	46,000 (25)
CE	5,532 (49)	3,026 (27)	2,673 (24)
SS	65,454 (51)	27,090 (21)	35,454 (28)
SE	15,918 (54)	6,656 (23)	6,851 (23)
AS	19,627 (67)	5,258 (18)	4,583 (16)
AE	31,777 (55)	13,310 (23)	12,622 (22)

Table 8: Speech act distribution for all paragraphs using the six types of preventive innovation adoption. Number of sentences with percentage in parenthesis.

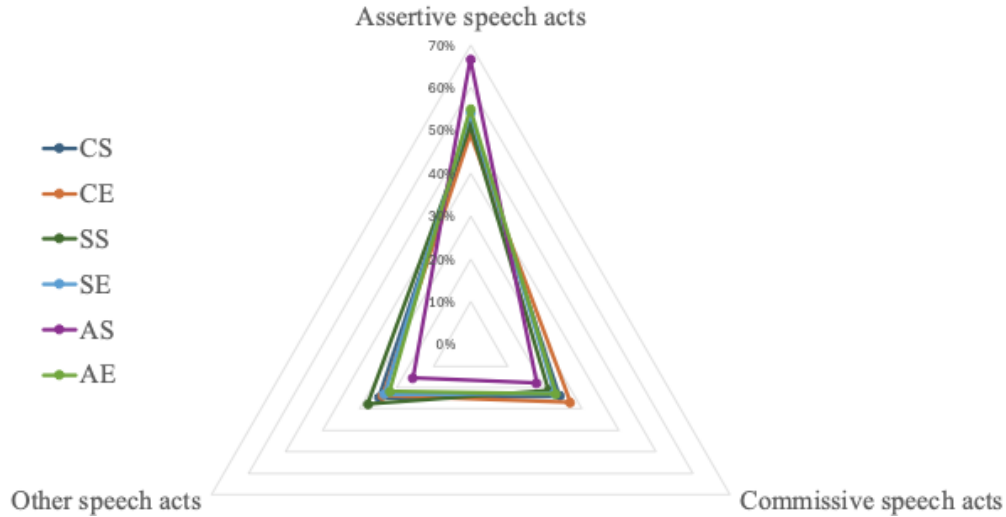


Figure 3: Diagram of the distribution for all paragraphs using the six types of preventive innovation adoption, percentage.

Advocates.

	Assertive	Commissive	Other
Compliers	4,887 (59)	2,199 (27)	1,174 (14)
Stewards	3,868 (58)	1,983 (30)	824 (12)
Advocates	11,430 (69)	3,492 (21)	1,583 (10)

Table 9: Speech act distribution for ISO paragraphs according to governance approaches. Number of sentences with percentage in parenthesis.

Table 10 and Figure 5 show the result when using the social/economic capital benefit dimension. The results show very little difference between the dimensions.

Table 11 and Figure 6 show the result when combining the dimensions into six categories that represent the three governance approaches and the social/economic capital dimension. The results show very little difference between the six dimensions.

5.3 Discussion

The performance of the classification model revealed that it often confused Commissives with Assertives, and to a lesser degree, vice versa. The precision and recall metrics for the Commissive class suggest that

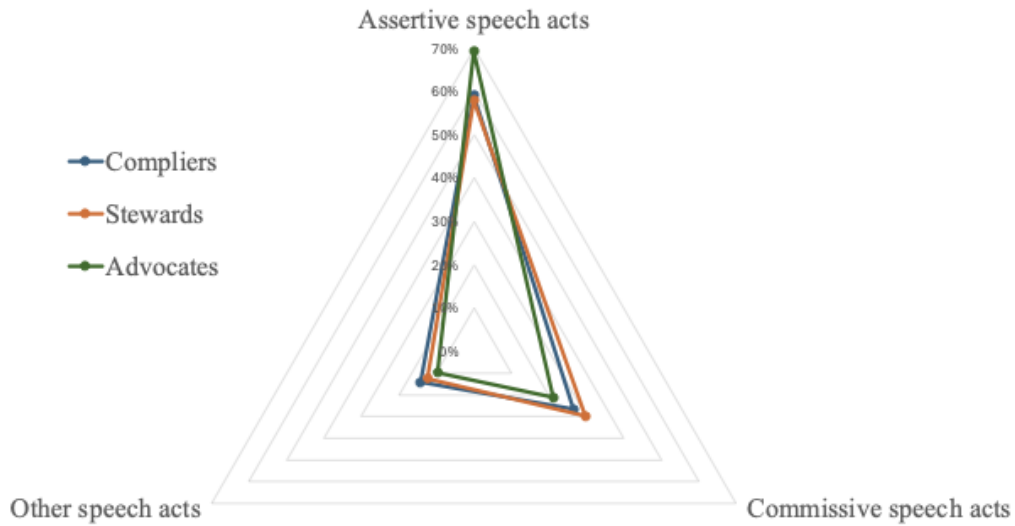


Figure 4: Diagram of the distribution for ISO paragraphs according to governance approaches, percentage.

	Assertive	Commissive	Other
Social capital	7,070 (58)	3,362 (28)	1,663 (14)
Economic capital	13,115 (68)	4,312 (22)	1,918 (10)

Table 10: Speech act distribution for ISO paragraphs for the social/economic capital dimension. Number of sentences with percentage in parenthesis.

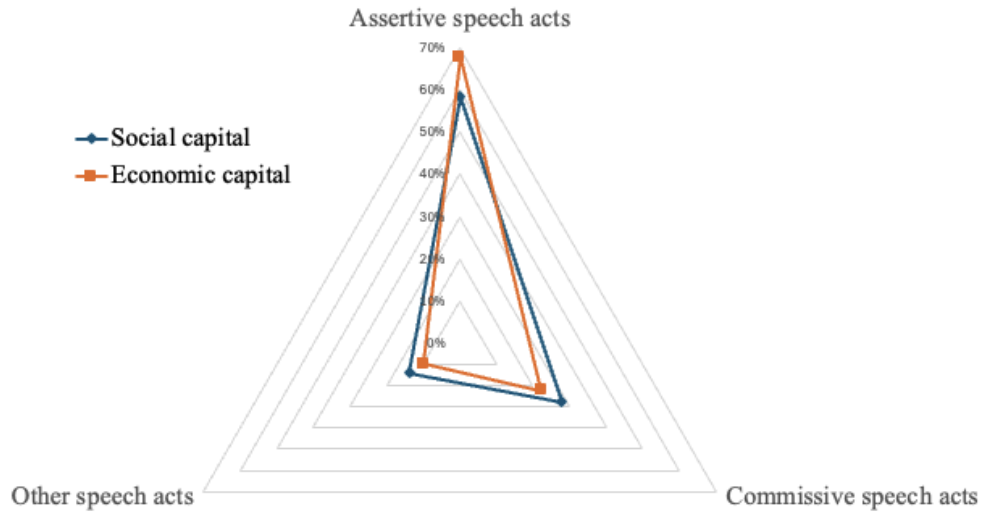


Figure 5: Diagram of the distribution for iso paragraphs for the social/economic capital dimension, percentage.

the model had difficulty accurately identifying Commissive speech acts. Specifically, the precision was low (0.34), indicating that many instances labelled as Commissives were incorrect. On the other hand, the recall for Commissives was higher (0.68), suggesting that the model was relatively effective at identifying true Commissive statements, though not always accurately.

	Assertive	Commissive	Other
CS	4,214 (59)	1,848 (26)	1,045 (15)
CE	673 (58)	351 (30)	129 (11)
SS	2,564 (58)	1,327 (30)	536 (12)
SE	1,304 (58)	656 (29)	288 (13)
AS	292 (52)	187 (33)	82 (15)
AE	11,138 (70)	3,305 (21)	1,501 (9)

Table 11: Speech act distribution for ISO paragraphs using the six types of preventive innovation adoption. Number of sentences with percentage in parenthesis.

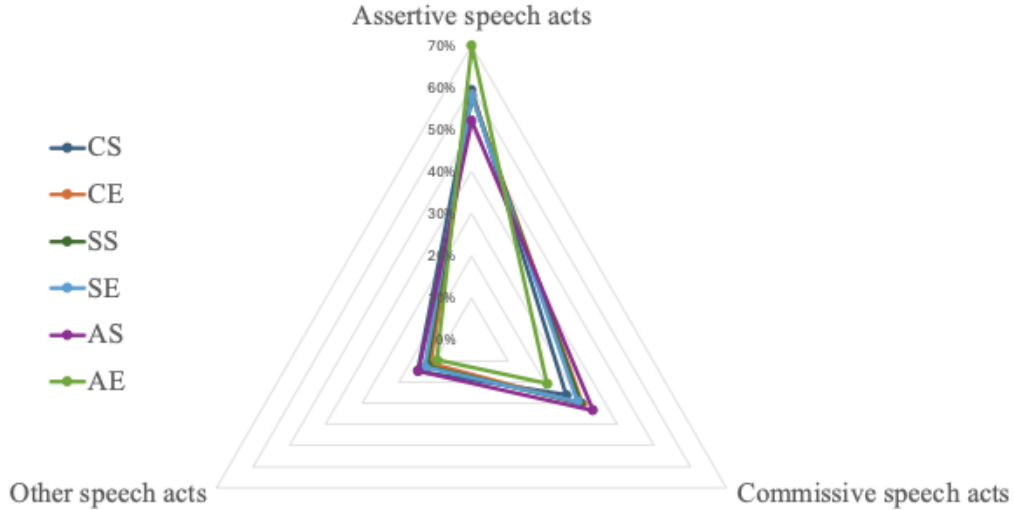


Figure 6: Diagram of the distribution for ISO paragraphs using the six types of preventive innovation adoption, percentage.

During manual annotation, it was noted that few Commissive instances were explicit commitments; instead, many were implicit guarantees or commitments. For instance, *The contracted supplier must be certified according to ISO 27001, which is a globally renowned information security standard.* is formulated as assertive, but not for the speaker itself, but for others, in this case suppliers.

This made it particularly challenging to differentiate between Assertives and Commissives. It is likely that the LLM used for automatic annotation faced similar difficulties in distinguishing between these two categories. Additionally, the indirect nature of Commissives may have contributed to ambiguities in the demonstration dataset, influencing the model’s performance.

One general result when applying the model, as shown in the visualisations in Figures 1, 2, and 3, compared with Figures 4, 5, and 6, is that in the ISO paragraph sub-corpus the proportion of utterances classified as Other is much smaller; instead, most utterances are classified as Assertives or Commissives. The dataset comprising all paragraphs includes substantial web clutter and generic content, which inflates the proportion of statements classified as Other and can mask meaningful communicative patterns, whereas the ISO-focused subset concentrates on thematically relevant paragraphs where organisations explicitly discuss information security.

By contrasting the two, it becomes possible to assess how topic filtering affects speech act distributions and to demonstrate that ISO-related communication is more strongly characterised by Assertive and Commissive speech acts. This comparison therefore strengthens the validity of the findings by showing which results are robust across general corporate communication and which emerge only when analysing

communication that is directly tied to the ISO standard.

Contrary to expectations, the various analysis dimensions of stakeholder communication: preventive innovation benefits, Figures 2 and 5, governance approaches, Figures 1 and 4, and both dimensions combined, Figures 3 and 6, show very similar patterns. For instance, contrary to expectations, Advocates exhibit a similar frequency of Assertive speech acts as the other categories, see Figures 1 and 4. Manual analysis of these Assertive acts reveals that they primarily describe the actions of third parties. This poses a challenge for the model, as it struggles to determine whether a statement reflects the organization's own reality or that of others, which impacts classification results.

The ISO paragraphs dataset reflects a formalised, institutionally anchored mode of corporate communication. By explicitly referencing ISO/IEC 27001, these paragraphs focus on compliance-related meanings such as certification status, audits, roles and responsibilities, and the governance of information security practices. As a result, this communication is less promotional and more oriented toward legitimacy, accountability, and alignment with externally defined rules. ISO-focused paragraphs therefore primarily express what is (assertive speech acts) and what is formally required or guaranteed (often implicit commissives), rather than aspirational or marketing-oriented claims. In terms of content, this suggests that, when organisations discuss ISO/IEC 27001 directly, they portray information security as a structured, rule-bound field overseen by standards and third-party verification rather than as an open-ended strategic objective or competitive pledge.

As shown in Section 5.2, ISO/IEC 27001 communication is dominated by assertive speech acts, with commissives playing a secondary but meaningful role, Tables 9 and 10. Surprisingly, this indicates that organisations primarily use ISO-focused paragraphs to state facts about existing structures, certifications, procedures, and third-party requirements. In doing so, they present information security as an established and verifiable organisational reality. At the same time, the presence of commissive speech acts across both the social/economic capital dimension, Table 10, and the governance approaches of Compliers, Stewards, and Advocates, Table 9, suggests that ISO communication also functions to signal ongoing responsibility and future-oriented assurance, even when these commitments are implicit rather than explicit. The relatively stable distribution of Assertives and Commissives across categories implies that ISO/IEC 27001 is framed less as a strategic aspiration and more as a matter of factual compliance. This means, credibility is established by describing what is already in place and reinforcing expectations of continued conformity, regardless of standard adherence or benefit orientation.

6 Conclusions

The research presented in this paper demonstrates how large language models (LLMs) can be effectively leveraged to reduce reliance on manually annotated datasets in the domain of Speech Act classification. Utilising techniques such as Clue and Reasoning Prompting (CARP) (Sun et al., 2023) and Retrieval-Augmented Generation (RAG) (Lewis et al., 2021). The research successfully generated a substantial dataset of 138,052 entries. This dataset facilitated the training and evaluation of a classifier, which achieved an accuracy of 73% and a weighted F1-score of 0.75.

An expanded LLM selection process, incorporating a wider array of models, more data and focused tests specifically aimed at indirect and/or context dependent commissives, could improve the selection process and extend to better Commissive classification results. Additionally, while LLM selection was restricted to quantised models, using un-quantised models while keeping similar model parameter size, would likely improve performance.

The results of applying the model did not exhibit differences in how the adoption of the ISO/IEC 27001 standard was expressed by Swedish companies depending on the various dimensions of the analysis, e.g. the governance approach or the capital dimension. However, for the management researchers, the results revealed other interesting properties of the way companies present themselves, such as the role of assertive speech acts for ISO communication.

This work builds upon existing frameworks and proposes novel approaches for dataset generation in text classification tasks, particularly within the context of organisational communication related to ISO/IEC 27001 standards. Beyond its technical contributions, the study demonstrates how CLARIN K-

centres can support management researchers and add value to the CLARIN infrastructure and to the broader field of organisational communication research. Furthermore, it highlights the potential of quantised LLMs for Speech Act classification, offering valuable insights into their performance, especially in mid-resource language settings.

References

- Abercrombie, G., Dinkar, T., Cercas Curry, A., Rieser, V., & Hovy, D. (2025, November). Consistency is key: Disentangling label variation in natural language processing with intra-annotator agreement. In G. Abercrombie, V. Basile, S. Frenda, S. Tonelli, & S. Dudy (Eds.), *Proceedings of the 4th workshop on perspectivist approaches to nlp* (pp. 63–74). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.nlperspectives-1.6>
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood Press.
- Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020, July). Unsupervised cross-lingual representation learning at scale. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 8440–8451). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.747>
- Cooren, F., & Seidl, D. (2020). Niklas luhmann’s radical communication approach and its implications for research on organizational communication. *Academy of Management Review*, 45(2), 479–497.
- Cooren, F., & Taylor, J. R. (1997). Organization as an effect of mediation: Redefining the link between organization and communication. *Communication Theory*, 7(3), 219–260. <https://doi.org/10.1111/j.1468-2885.1997.tb00151.x>
- Culot, G., Nassimbeni, G., Podrecca, M., & Sartor, M. (2021). The ISO/IEC 27001 information security management standard: Literature review and theory-based research agenda. *The TQM Journal*, 33(7), 76–105.
- Disterer, G. (2013). ISO/IEC 27000, 27001 and 27002 for information security management. *Journal of Information Security*, 4(2), 92–100.
- Grattan, M. (2024). *Now or later: Classifying future- and present-oriented speech acts in organizational communication* [Bachelor’s Thesis]. Linköping University.
- Grattan, M., Fried, A., & Jönsson, A. (2025). Challenges of analysing speech acts in organisational communication. *Proceedings of the CLARIN Annual Conference 2025*, 148–152.
- Hu, S., Hsu, C.-H., & Zhou, Z. (2022). Security education, training, and awareness programs: Literature review. *Journal of Computer Information Systems*, 62(4), 752–764.
- Janssen, M., Brousa, P., Estevez, E., Barbosad, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, 37. <https://doi.org/https://doi.org/10.1016/j.giq.2020.101493>
- Jönsson, A., Bandyopadhyay, S., Dragisic, S. P., & Fried, A. (2024). Analyses of information security standards on data crawled from company web sites using SweClarin resources. *Selected papers from the 2023 CLARIN Annual Conference*. <https://doi.org/https://doi.org/10.3384/ecp210>
- Khando, K., Gao, S., Islam, S. M., & Salman, A. (2021). Enhancing employees’ information security awareness in private and public organisations: A systematic literature review. *Computers & Security*, 106, 102267.
- Kuhn, T. S. (2008). A communicative theory of the firm: Developing an alternative perspective on intra-organizational power and stakeholder relationships. *Organization Studies*, 29, 1227–1254.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). Retrieval-augmented generation for knowledge-intensive nlp tasks.

- Mirtsch, M., Blind, K., Koch, C., & Dudek, G. (2021). Information security management in ict and non-ict sector companies: A preventive innovation perspective. *Computer & Security*, 109. <https://doi.org/10.1016/j.cose.2021.102383>
- Mirtsch, M., Kinne, J., & Blind, K. (2020). Exploring the adoption of the international information security management system standard ISO/IEC 27001: A web mining-based analysis. *IEEE Transactions on Engineering Management*, 68(1), 87–100.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2), 242–266.
- Rogers, E. M. (2002). Diffusion of preventive innovations. *Addictive behaviors*, 27, 989–993.
- Schoeneborn, D., Kuhn, T. R., & K’arremann, D. (2019). The communicative constitution of organization, organizing, and organizationality. *Organization Studies*, 40(4), 475–496.
- Searle, J. R. (1985, November). *Expression and meaning*. Cambridge University Press.
- Siponen, M. T. (2000). A conceptual foundation for organizational information security awareness. *Information Management & Computer Security*, 8(1), 31–41.
- Sun, X., Li, X., Li, J., Wu, F., Guo, S., Zhang, T., & Wang, G. (2023). Text classification via large language models. *Findings of the Association for Computational Linguistics: EMNLP 2023*.
- Taylor, J. R. (1993). *Rethinking the theory of organizational communication: How to read an organization*. Norwood, NJ: Ablex.