

# Game of Thrones Side-Character Dialogue Analysis (Group 62)

Nicholas Milin (261106314), Laurier Lefebvre (261143550), Yuzhe Wang (261150086)

McGill University

nicholas.milin@mail.mcgill.ca, laurier.lefebvre@mail.mcgill.ca, yuzhe.wang@mail.mcgill.ca

## Abstract

In this report, we analyze three secondary characters from the series *Game of Thrones* and examine their narrative roles and the evolution of these characters in relation to the topics they discuss. We analyzed the transcripts from all eight seasons, focusing on Petyr Baelish, Samwell Tarly and Theon Greyjoy. For each character, we manually annotated each line, classifying their discussions into eight main categories: Information/Lore, Loyalty/Honour, Orders/Demands, People, Plans/Strategy/Advice, Politics, Trivial, War/Violence. This classification revealed that dialogue is dominated by world-building and planning, not action. Globally, Information/Lore is the largest category, followed by Plans/Strategy/Advice and People. War/Violence and Politics are among the smallest categories.

## 1 Introduction

The goal of this analysis was to find the narrative roles and their evolution for three side characters in a TV series. We first looked into doing this analysis for the show Breaking Bad, however we quickly realized that the side characters in Breaking Bad have few lines. We then pivoted to trying to find a show with more significant side characters. That's when we stumbled upon Game of Thrones. Many characters in Game of Thrones have over 400 lines of dialog, which is plenty for our analysis. We then chose three of those side characters for our analysis: Petyr Baelish, Samwell Tarly and Theon Greyjoy. Our analysis suggests that, for these three characters, the show primarily uses dialogue for exposition, as well as relationship development and planning, rather than for depicting violence or overt political debate. Furthermore, each character has a clear "functional role" in the discourse. Take Samwell Tarly as an example, his frequent talk of Information/Lore and Trivial matters illustrate his role as a lore-keeper who explains the White Walker threat and the institutional world of the Night's Watch and the Citadel. Simultaneously, his trivial talk (e.g. "library", "walks", "work") grounds scenes in everyday life. Petyr Baelish and Theon Greyjoy, also display similarly distinct linguistic patterns that define their specific archetypes within the narrative.

---

Copyright © 2024, Association for the Advancement of Artificial Intelligence ([www.aaai.org](http://www.aaai.org)). All rights reserved.

## 2 Data

### 2.1 Data Acquisition and Description

We initially evaluated a comprehensive Game of Thrones dataset; however, upon further inspection, we found various flaws in the dataset, including scene descriptions and non-standardized speaker labels, such as "WILL (whispering)". While we considered cleaning the data programmatically, the variation in naming conventions increased the risk of parsing errors. To ensure data integrity, we opted for an alternate dataset from Kaggle, which was void of scene descriptions and had more consistent speaker labels. When compared to screenplays we had looked at for other TV series, a precompiled dataset was selected for its organized and complete structure, providing clear separation between speakers and dialogues, thereby eliminating the need to remove scene descriptions. The dataset contains the entire chronological transcript for all eight seasons of Game of Thrones, totalling 23,911 entries. Each entry contains the release date, season number, episode number, episode title, character and dialogue.

### 2.2 Character Selection

To ensure we collected side characters with more than 300 non-trivial lines of dialogue, we grouped and organized characters by number of lines spoken. We then chose from characters within the 400–600 lines of dialogue range to ensure we selected secondary characters while still having a sufficient amount of non-trivial lines. We selected three characters who appear across multiple seasons to examine their importance and development throughout the series:

- Petyr Baelish (449 lines): (Wiki of Westeros n.d.a) A devious opportunist, coordinating chaos within the crown to fuel his desire for power and ultimate goal of sitting on the Iron Throne.
- Samwell Tarly (556 lines): (Wiki of Westeros n.d.b) A cowardly yet highly intelligent member of the Night's Watch who discovers critical knowledge to defeat the White Walkers.
- Theon Greyjoy (455 lines): (Wiki of Westeros n.d.c) A deeply insecure and arrogant youth who is captured and psychologically tortured, eventually redeeming himself as a selfless hero.

Our character selection resulted in a final subset of 1,460 lines of dialogue throughout the series, giving us a representative and extensive sample suitable for topic analysis.

### 3 Methods

#### 3.1 Data Preparation

Data preparation began with the normalization of character names since some characters are later referred to by their abbreviated names. We corrected these to ensure consistent formatting throughout the dataset. To generate a representative sample, we used the Polars library to extract a random selection of 125 lines per character. We implemented this randomization to reduce context and selection bias, helping to ensure that the open coding process covered a representative range of topics spanning the series.

#### 3.2 Typology Development and Implementation

We employed two researchers to independently review the sampled lines and generate an initial typology of 3–8 topics through open coding. We then consolidated our typologies, merging similar topics to create a shared taxonomy. To validate the taxonomy, we re-annotated the samples blindly, refining category definitions based on topic coverage, tricky cases and overlap through the open coding process as a group. We then each used the final taxonomy to manually annotate the entire dataset. We consolidated our results by selecting the majority label and resolved split decisions through group consensus. Trivial topics were included in our analysis to see whether there was any underlying trend of a character’s importance changing over the course of the series.

### 4 Results

#### 4.1 Topics Selected

After sufficient cross-coding, we produced the following list of topics and their working definitions.

**Orders/Demands** Lines about giving orders, making demands or pleading for something. These can be lines that are literally ordering, demanding or pleading, but they can also be lines that talk about the orders and demands of other people.

**Plans/Strategy/Advice** Lines that talk about anything relating to plans, strategy or advice. For plans, these can be lines that talk about making or executing a plan, but may also simply discuss a plan. Similarly, any lines that talk about strategy—whether creating or discussing it—fit in this category. Lines relating to advice also go here; this can be literally giving advice to someone or talking about advice given by someone else.

**War/Violence** Lines relating to war or any type of violence. For war, this can be anything relating to war with the exception of planning, since that goes in Plans/Strategy/Advice. Violence includes all types of violence: physical violence such as fighting, murdering, hurting, hunting, and verbal violence such as threats.

**Loyalty/Honour** Lines relating to loyalty or honour. This can include lines that talk about loyalty or demonstrate loyalty—to the crown, to one’s master, to one’s family, etc. This category also includes lines about honour, whether it is someone’s honour, the honour of a family or of a nation. As with loyalty, lines can literally talk about honour or demonstrate it.

**Politics** Lines talking about politics, political people, or anything to do with the feudal hierarchy. This includes lines about political power struggles, the royal family, drama relating to political figures, and also lines where the speaker is trying to do something politically, such as talking to a political figure to improve their political power or talking to royalty.

**Information/Lore** Lines where the speaker is giving information or talking about information or lore. This includes lines about the transfer of information, a specific piece of information, gathering information, and other topics relating to information. This category also includes sharing lore, talking about books, history, etc.

**People** Lines talking about specific people. This includes lines where the speaker is talking about other people or themselves. Lines only fit in this category when the main topic is a person and that person is not a political figure. Lines where someone is mentioned but the main topic is something else (war, politics, orders, etc.) belong to those other categories.

**Trivial** Trivial lines or lines that are just banter. These include lines that do not talk about any substantive topic and lines that are conversational banter. We decided to keep these lines instead of cutting them from the dataset in order to analyze the distribution of trivial lines for each character.

#### 4.2 Topic Distribution

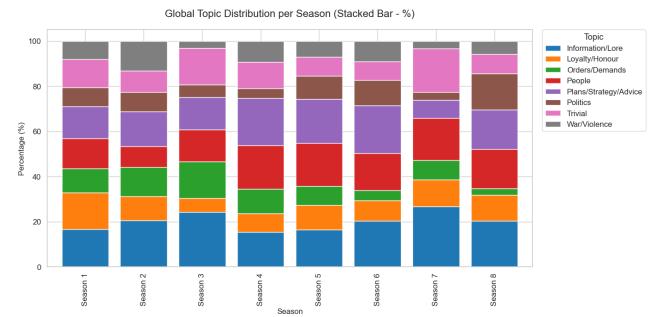


Figure 1: Global topic distribution per season.

**Global Topic Distribution per Season** Figure 1 shows that, across all eight seasons, Information/Lore is consistently the largest topic, followed by Plans/Strategy/Advice and People. The overall topic mix is relatively stable: each season displays a similar distribution pattern, with War/Violence and Politics consistently occupying a comparatively small portion. There is a noticeable mid-series increase in Politics and Plans/Strategy/Advice (around Sea-

sions 4–6), and a visible decrease in Trivial dialogue in the final seasons.

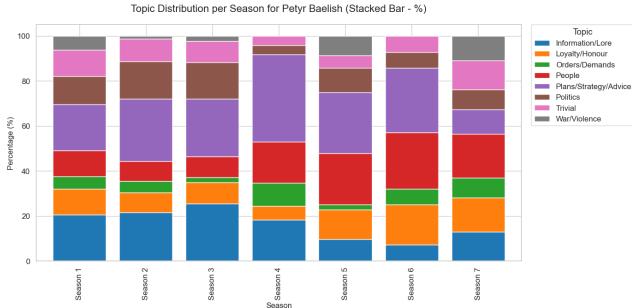


Figure 2: Topic distribution for Petyr Baelish by season.

**Petyr Baelish Topic Distribution per Season** Figure 2 shows that Plans/Strategy/Advice is a dominant topic for Petyr Baelish in most seasons, forming the largest segment in Seasons 1, 2, 4, and 5. Information/Lore briefly becomes the largest topic in Season 3. From Seasons 4 to 7, the share of People increases steadily and becomes the largest segment in Seasons 5 and 6, and roughly equal to Plans/Strategy/Advice in Season 7. Politics remains a moderate but stable percentage across all seasons, while Trivial fluctuates but never dominates a season, and War/Violence is consistently very small or absent in every season.

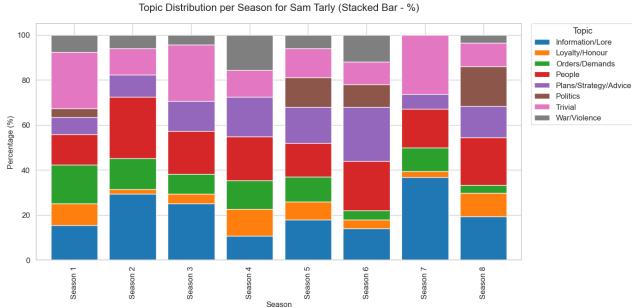


Figure 3: Topic distribution for Samwell Tarly by season.

**Samwell Tarly Topic Distribution per Season** Figure 3 shows that Information/Lore is the dominant topic for Samwell Tarly in most seasons. Trivial and People also occupy substantial portions of the distribution, making them his next most prominent topics. Politics and War/Violence are present but remain relatively minimal across the eight seasons.

**Theon Greyjoy Topic Distribution per Season** In 4, Theon Greyjoy displays a more varied topic distribution than the other two characters. While Information/Lore remains the single most significant topic overall. However, Loyalty/Honour, Orders/Demands, and War/Violence occupy noticeably larger proportions compared to Baelish and Tarly. In the early seasons, Loyalty/Honour and Orders/Demands

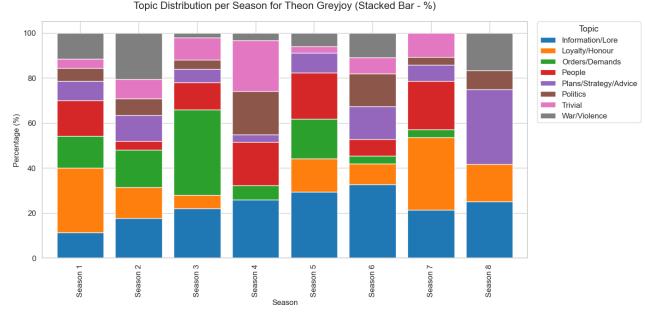


Figure 4: Topic distribution for Theon Greyjoy by season.

are particularly prominent. In the middle seasons, People becomes a larger segment, and War/Violence also increases. In the later seasons, the share of Loyalty/Honour and Plans/Strategy/Advice grows again, while Trivial dialogue becomes negligible or disappears.

### 4.3 TF-IDF Scores

To identify the most significant terms related to each category we calculated TF-IDF scores.

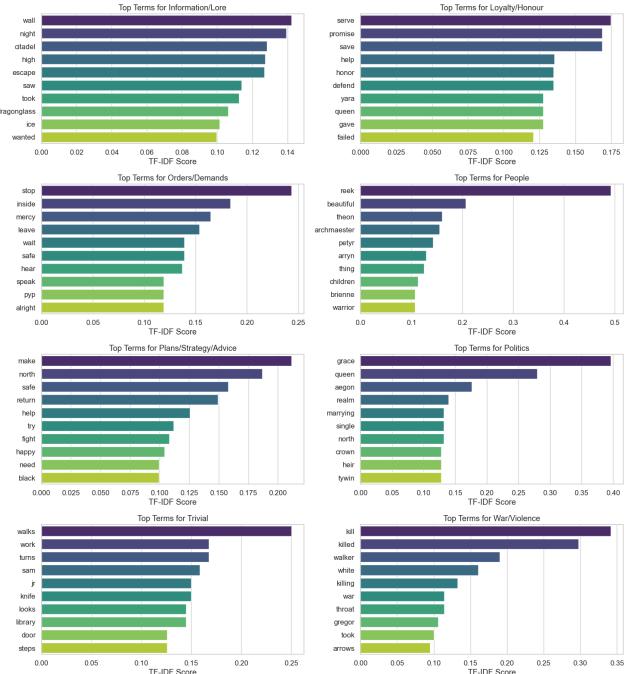


Figure 5: Top 10 TF-IDF terms for each topic.

For Information/Lore some of the high-scoring terms are “wall”, “ice” and “night”, referring to the Ice Wall and the Night’s Watch, which are mentioned in world-building conversations. The term “dragonglass”, otherwise known as obsidian, plays a crucial role in defeating the White Walkers. Within the People category, the term with the highest weight in the entire dataset is “reek”, the name Theon Greyjoy adopts when he is tortured and psychologically

abused by Ramsay. The word with the second-highest overall TF-IDF is “grace”, appearing in Politics as part of the phrase “your grace” to refer to members of the monarchy. For War/Violence, the words with the third and fourth highest TF-IDF scores are “white” and “walker”, directly referencing the main antagonists in the series.

#### 4.4 AI Topic Summary

**Information/Lore** Both schemes treat this topic as discussing information, including explaining facts, history, books, lore, or the transfer and gathering of information. Expository lines that explain what is going on also fit naturally here. Our definition explicitly includes discussing the gathering of information (for example, spying or “we need more information”). In contrast, the AI scheme sometimes leans towards Plans/Strategy/Advice when information gathering is clearly part of a plan. There is also a grey area between pure lore or exposition and strategic information about war or politics: our wording is broader in framing this as “anything about information”. At the same time, the AI description emphasizes world-building and exposition in practice.

**Loyalty/Honour** Both schemes treat this topic as lines that talk about or demonstrate loyalty and honour, including oaths, promises, fidelity, shame and similar ideas. Loyalty can be directed toward the crown, family, a group or other entities. However, when a line is mainly about war or fighting but also implies loyalty (e.g., “I’ll fight for you”), our rule allows it to be coded as Loyalty/Honour if the emphasis is clearly on devotion rather than combat. In contrast, the AI scheme tends to treat such lines as War/Violence. As a result, our definition is slightly broader in behavioural terms, while the AI one focuses more on explicit verbal markers of loyalty or honour.

**Orders/Demands** Both schemes centre this topic on directive speech: giving orders, making demands, telling someone what to do or pleading. It is not restricted to formal military commands and also includes everyday directives. Our definition explicitly includes lines that discuss other people’s orders or demands, even when the speaker is only reporting them (e.g., “He commanded me to go”). In contrast, the AI description does not highlight this, such that lines might drift into Information/Lore or Politics in that context. We also explicitly place pleading and begging here, while AI summaries sometimes associate pleas in violent contexts more with War/Violence or Loyalty/Honour, depending on what is foregrounded.

**Plans/Strategy/Advice** Both approaches define this topic in almost the same way: lines about making, executing or discussing plans, any kind of strategy, and giving or discussing advice. Planning for future actions clearly belongs here in both views. Our written rules additionally state that planning for war belongs in Plans/Strategy/Advice rather than War/Violence, which the AI also followed in principle. A key boundary issue arises with information-gathering as part of a strategy, such as spying or sending messages: the AI scheme sometimes frames these lines as Information/Lore, while our wording suggests that, when the focus is on using

information strategically, they should more often be coded as Plans/Strategy/Advice.

**War/Violence** Both schemes describe this topic as lines about war or any type of violence, including physical harm, fighting, killing and verbal threats, with planning explicitly excluded and placed in Plans/Strategy/Advice. A primary boundary issue is that the AI scheme sometimes categorizes captivity or abuse under War/Violence, whereas, depending on the emphasis, our scheme might instead consider People (when the focus is on identity) or Loyalty/Honour (when the focus is on betrayal or shame). Our definition explicitly mentions hunting as a form of violence, while AI would only necessarily treat hunting as War/Violence when it is framed as violent; otherwise, it might lean towards Information/Lore if the line is primarily descriptive.

**Politics** Both views agree that this topic covers lines about politics, political people and the feudal hierarchy, including titles, power struggles, succession, institutions and attempts to gain or exercise political power. Both emphasize political roles, such as queen, lord, or heir, as central. We explicitly include any interaction with political figures when the purpose is to change or maintain political status, even if power is only implied rather than directly named. In contrast, the AI scheme focuses more on explicit discussion of power and may not always mark subtle status-seeking as Politics. When political and personal aspects are combined in a single line, our scheme categorizes it as Politics whenever the feudal role is central. In contrast, the AI description sometimes allows such mixed lines to be treated as People.

**People** Both schemes treat this topic as lines where the main subject is a person (including self-reference), focusing on their traits, feelings, or relationships, and both attempt to maintain this label only when the line is not primarily about war, politics, orders, or similar content. We emphasize that when someone is being discussed in their political role—for example, as a lord, queen or commander—the line should normally be placed in Politics rather than People unless it clearly focuses on them as a private individual. The AI’s description of People as “talk focused on individuals” does not always sharply separate “person as human being” from “person as office-holder”, so in borderline cases where both aspects are present, our rules push more lines into Politics (or sometimes Information/Lore when the line is explaining who they are) than the AI’s broader wording might.

**Trivial** Both approaches define this topic as trivial lines or banter that do not discuss any substantive topic, capturing everyday, low-stakes chatter that does not clearly fit into another thematic category. The semantic definition is essentially the same. The main difference is methodological: we explicitly state that we retained trivial lines in the dataset to analyze how much each character uses them, whereas the AI description does not mention this motivation, although it does not alter how the category itself is understood. There is also a boundary issue when banter contains real thematic content, such as joking about war or politics: under our scheme, these lines are assigned to the substantive topic if

that is clearly dominant. At the same time, the AI description implicitly assumes a similar practice without explicitly stating it.

## 5 Discussion

In this study, our aim was to investigate the distribution of topics that our chosen side characters talk about throughout the series.

Petyr Baelish uses his dialogue very strategically, highlighting him as a cunning, manipulative character. As discussed above, he has a significantly higher percentage of lines talking about plans, strategy or giving advice than the other characters, as well as a higher distribution in Politics in the first three seasons, which then shifts to the People category later. Taken together, these categories paint a picture of Baelish as a manipulative character who starts off in politics and then shifts more toward ordinary people. This view is supported by the fact that, compared to the global distribution, he talks very little about War/Violence and also does not spend much dialogue on Orders/Demands. This supports the idea that he avoids directly ordering or threatening people and instead prefers to give advice and information, getting people to like him and act in his interests without explicit orders. We are thus left with the picture that Petyr Baelish is a cunning, manipulative character who uses dialogue very strategically.

Samwell Tarly serves as a relatable character who is a source of information and lore. As discussed in the Results section, Sam's dialogue is dominated by Information/Lore, People and Trivial. Although Information/Lore is also the largest category in the global distribution, Sam's distribution is generally above average, marking him as a source of information and lore for other characters and for the audience. This is especially visible in the first three seasons, where his Information/Lore proportion is remarkably large. Samwell also has a large and consistent portion of his dialogue in the People category, further supporting the idea that he is a source of information about others, often explaining who people are or what they have done. The Trivial category also takes up much of his dialogue. We explicitly kept trivial lines in the dataset for this kind of insight: a much larger percentage of Sam's dialogue is trivial compared to the other characters, indicating that he is in more everyday conversations. TF-IDF words for Trivial, such as "Sam" and "Jr", correspond to instances where Samwell talks to or about his son, Sam Jr., reinforcing the idea that he is a relatable, grounded character. Overall, Samwell Tarly appears as a relatable source of information and lore.

Theon Greyjoy drastically changes throughout the series, and these changes are often triggered by information. In Season 1, his most common topic is Loyalty/Honour, reflecting his situation as someone under the rule of the Starks, forced to be loyal and talking mostly about loyalty. In Season 2, War/Violence becomes much more prominent for Theon, matching his involvement in war and violent events. In Season 3, the vast majority of his dialogue falls under Orders/Demands, with another large amount in Information/Lore. This fits his status as a prisoner being tortured for information he holds. In Season 4, he is a

slave renamed Reek. He talks about himself, relays information he gathered, and is not important in conversation. This matches topics dominated by People, Information/Lore and Trivial. Season 5 is an in-between season for Theon: his topics are almost evenly spread among People, Orders/Demands, Loyalty/Honour and Information/Lore, reflecting his transition from being Reek (taking orders and talking about himself) to helping the Starks by giving information and pledging loyalty. In Season 6, People and Orders/Demands nearly disappear as he is no longer Reek but Theon again. He updates others on information he gathered while a slave and helps his sister climb the political ladder, which is reflected in his topics. In Season 7, his distribution is almost entirely Loyalty/Honour and People, reflecting his need to show loyalty to his family and to the Starks as he completes his transformation from Reek back to Theon. Finally, in Season 8, his dialogue becomes almost completely about War/Strategy/Advice and Information/Lore, with Trivial lines disappearing as he becomes a grim, duty-driven figure. This reflects his role of convincing the Starks, using information he gathered, to let him lead a small army and die on the battlefield. We therefore cannot easily classify Theon as a single archetype, since he changes so much. However, one constant stands out: he often possesses important information that changes his life. Across seasons, key pieces of information lead to his torture, his release from slavery, his sister's rise in power, the Starks' renewed trust and his final chance to fight in the war. Theon Greyjoy thus consistently carries information that drastically reshapes his fate.

## 6 Conclusion

Despite being very different characters, these three all serve a functional purpose in the exposition of the show, and they all share a common aspect: they serve as sources of information. Petyr Baelish uses information for his own benefit by manipulating others with it. Samwell Tarly is a general source of information and lore for both the audience and other characters. Many of Theon's changes throughout the series occur because of information he possesses. It is striking that in seasons where one character speaks less about information, the others talk more about it, almost compensating for each other for the benefit of the audience. This leads us to ask: How much of the show's information and lore is passed on to the audience specifically through these side characters?

## 7 Group Member Contributions

Nicholas Milin contributed to the data sampling for the open coding. Nicholas Milin and Laurier Lefebvre contributed to the open coding process. Nicholas Milin, Laurier Lefebvre annotated the dialogue and consolidated the responses. Nicholas Milin and Yuzhe Wang completed the analysis of the data. Yuzhe Wang used an LLM to produce the representative summaries of each category. Nicholas Milin contributed to the writing of the Abstract, Introduction, Data, Methods and TF-IDF Scores. Laurier Lefebvre contributed to the writing of the Introduction, Topic Description, Discus-

sion and Conclusion. Yuzhe Wang contributed to the writing of the Abstract, Introduction, Topic Distribution and AI Topic Summary. Nicholas Milin converted the document to LaTeX.

## References

Wiki of Westeros. n.d.a. Petyr Baelish. [https://gameofthrones.fandom.com/wiki/Petyr\\_Baelish](https://gameofthrones.fandom.com/wiki/Petyr_Baelish). Accessed: 2025-12-01.

Wiki of Westeros. n.d.b. Samwell Tarly. [https://gameofthrones.fandom.com/wiki/Samwell\\_Tarly](https://gameofthrones.fandom.com/wiki/Samwell_Tarly). Accessed: 2025-12-01.

Wiki of Westeros. n.d.c. Theon Greyjoy. [https://gameofthrones.fandom.com/wiki/Theon\\_Greyjoy](https://gameofthrones.fandom.com/wiki/Theon_Greyjoy). Accessed: 2025-12-01.