#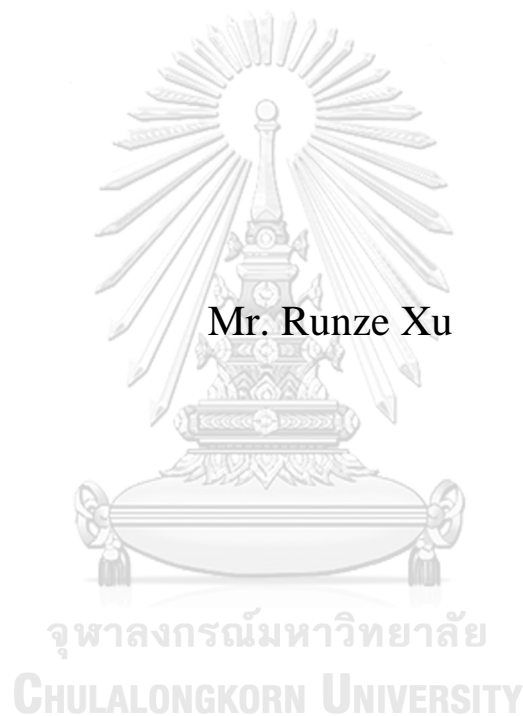 LEXICAL BUNDLES IN MAINSTREAM MOVIE'AND NOAH BAUMBACH'S SELECTED FILM SCRIPTS: A CORPUS-BASED STUDY OF TELECINEMATIC DISCOURSE

Mr. Runze Xu

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in English as an International Language
Inter-Department of English as an International Language
GRADUATE SCHOOL
Chulalongkorn University
Academic Year 2021
Copyright of Chulalongkorn University

กลุ่มคำเรียงในบทภาพยนตร์ของภาพยนตร์กระแสหลักกับในบทภาพยนตร์ คัดสรรของโนอาห์ บาวม์บาค: การศึกษาวาทกรรมหนังและโทรทัศน์ตามแนวภาษาศาสตร์คลังข้อมูล

นายหรุ่นเจ๋อ ซู

Thesis Title   LEXICAL BUNDLES IN MAINSTREAM MOVIE'AND NOAH BAUMBACH'S SELECTED FILM SCRIPTS: A CORPUS-BASED STUDY OF TELECINEMATIC DISCOURSE

By   Mr. Runze Xu

Field of Study   English as an International Language

Thesis Advisor   Associate Professor Raksangob Wijitsopon, Ph.D.

Accepted by the GRADUATE SCHOOL, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Arts

.... Dean of the GRADUATE SCHOOL

(Associate Professor YOOTTHANA CHUPPUNNARAT, Ph.D.)

THESIS COMMITTEE

.... Chairman

(Associate Professor WIROTE AROONMANAKUN, Ph.D.)

.... Thesis Advisor

(Associate Professor Raksangob Wijitsopon, Ph.D.)

.... Examiner

(Associate Professor Issra Pramoolsook)

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

หรุ่นเจ๋อ ซู : กลุ่มคำเรียงในบทภาพยนตร์ของภาพยนตร์กระแสหลักกับในบทภาพยนตร์ คัดสรรของโนอาห์ บาวม์บาค: การศึกษาวาทกรรมหนังและโทรทัศน์ตามแนวภาษาศาสตร์คลังข้อมูล. (LEXICAL BUNDLES IN MAINSTREAM MOVIE'AND NOAH BAUMBACH'S SELECTED FILM SCRIPTS: A CORPUS-BASED STUDY OF TELECINEMATIC DISCOURSE) อ.ที่ปรึกษาหลัก : รศ. ดร.รักสงบ วิจิตรโสภณ

-

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

| สาขาวิชา | ภาษาอังกฤษเป็นภาษานานาชาติ | ลายมือชื่อนิสิต .... |
| ปีการศึกษา | 2564 | ลายมือชื่อ อ.ที่ปรึกษาหลัก .... |

6288010020 : MAJOR ENGLISH AS AN INTERNATIONAL LANGUAGE
KEYWORD: MAINSTREAM MOVIES NOAH BAUMBACH TELECINEMATIC DISCOURSE CORPUS LINGUISTICS LEXICAL BUNDLES

Runze Xu : LEXICAL BUNDLES IN MAINSTREAM MOVIE'AND NOAH BAUMBACH'S SELECTED FILM SCRIPTS: A CORPUS-BASED STUDY OF TELECINEMATIC DISCOURSE. Advisor: Assoc. Prof. Raksangob Wijitsopon, Ph.D.

The present study aims to investigate and compare language use in film scripts of American mainstream movies and in Noah Baumbach's, focusing on lexical bundles and their functional contributions to meanings and aesthetic values of the given texts. Two corpora were compiled as the datasets: Mainstream Film Corpus (MFC), which consists of 100 screenplays of top Hollywood blockbusters during 2005-2019, and Noah Baumbach Corpus (NBC), which contains five film scripts written by Noah Baumbach alone. Frequent four-word lexical bundles were extracted and approached qualitatively in terms of their functions, following Biber et al. (2004) and Biber (2006) as analytical frameworks.

Results demonstrate that lexical bundles in MFC and those in NBC are similar in overall distributional patterns of lexical bundle types, with stance bundles topping the lists, followed by referential, special-function and discourse-organizing bundles, respectively. This is argued to reflect shared features of the film script as a register. However, a number of differences between the two corpora were also observed. Crucially, while lexical bundles in MFC highlight dialogue and narrative descriptions of scene, movement and body language in film scripts of mainstream movies, those in NBC feature dialogues, which serve purposes of characterization and character development in Noah Baumbach's storylines. The study thereby contributes to the growing fields of telecinematic discourse and telecinematic stylistics. Limitations of the study and recommendations for future research are discussed.

CHULALONGKORN UNIVERSITY

Field of Study:	English as an International Language	Student's Signature ....
Academic Year:	2021	Advisor's Signature ....

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Page

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I INTRODUCTION

This chapter aims to provide an overview of the present study. I first give background information about telecinematic discourse and Noah Baumbach's film scripts to provide rationales for the study. Then, I describe research questions, objectives and significance of the study.

1.1 Background of the study

Recent years have witnessed a growing academic interest in "telecinematic discourse", the term made popular by Piazza, Bednarek and Rossi (2011), which refers to "research that deals with a linguistic analysis of films and television series" (Zago, 2020, p. 168). According to Zago (2020), the telecinematic discourse is worth scholarly attention because it has been instrumental to our understanding of the relationship between speech and writing, as the dialogues are written and later the texts are presented orally and through acting by actors (Gregory, 1967). Also, telecinematic dialogues exhibit complexity in fictional discourse; they are spoken by fictional characters to one another but mainly aimed at audiences, who watch them talk. This means that telecinematic utterances have greater illocutionary forces than those in real conversations, which are studied in pragmatics. As Bednarek (2018) illustrates, when the protagonist in a story receives an invitation to a party, this could mean more than an invitation to audiences; for example, it can be an act that strengthens or destroys relationship between the two characters, provoking emotional reactions among the viewers. In previous studies on telecinematic discourses, various genres of mainstream films have been investigated qualitatively, such as romantic comedy (Nuryani, 2016), young adult adventure films (Nursanti, 2015), and gangster drama (Statham, 2015). However, to the best of my knowledge, no work has been done on film scripts in general, not limited to a particular film genre. One of the objectives of the present study, therefore, is to further an insight into the language of film scripts as a register in language use.

Another gap found in the current literature of telecinematic discourse is that it has not

touched on language use in independent films, those often produced by young directors, upon a low budget, whose filmmaking tends to deviate from mainstream Hollywood blockbusters. A prime example of American independent films include those produced by Noah Baumbach, one of the leading directors of 21ˢᵗ century American independent comedy-dramas (Duralde, 2019; Grobar & Grobar, 2017; Hornaday, 2015; *Noah Baumbach*, 2019). In the film industry, Baumbach is recognized as an independent writer-director, i.e., he writes his own screenplays for his directed films (Ryan, 1998). Two of his works, *The Squid and the Whale* and *Marriage Story*, have been nominated for Best Original Screenplay in the Academy Award. Currently, he has won 47 awards and 121 nominations from various events(See *Noah Baumbach Awards*), He has also garnered 72% for career average score (which is high among film directors, see *Noah Baumbach's Scores*, 2021) (which is high among film directors, see "Noah Baumbach's Scores," 2021) in an authoritative film scoring website Metacritic. Having earned these accolades in the past three decades, Noah Baumbach has become a defining voice of American independent cinema.

Baumbach's films consistently observe the characters' mental impulses in a satirical manner, with stories largely focusing on intellectual families in New York City (Jones, 2016). The protagonists in his films are often writers or filmmakers aged 30 to 40, many of whom have unlikeable personalities such as "unapologetically plac[ing] themselves at the center of the world" and "more […] immature than the children" (O'Meara, 2014). These characteristics have made their conversations with others full of contradictions and confrontations, becoming the distinctive features of Baumbach's film dialogues.

The distinctiveness of his film dialogues significantly contribute to Baumbach's success (Laplante, 2013). His dialogue lines help audience see realistic and terrifying kinds of human contradictions, and they tend to be close to real conversational contexts and nuances, particularly the rhythms, speed, and tone (Hellerman, 2019). Unlike other film characters who are often shown to be keen to listen to others' talks and patiently provide feedback until the following turn (Nerdwriter1, 2017),

Baumbach's characters tend to interrupt when other characters are speaking. Baumbach's film scripts highlights the problems in characters' communication through conversational interruptions. Hence, in his films, people tend to talk over each other, which seems greatly different from Hollywood films in general where classical narration strategy is employed such as that of Robert Altman, an American director (Sutton, 2015). By adapting this storytelling strategy, Baumbach presents the difficulty of human communication.

Furthermore, while Hollywood mainstream films aims to create likable characters and fascinating plots to hook audiences to the film (Hsu, 2008), the scriptwriting of Baumbach's films mostly presents non-entertaining stories and unlovable characters in comparison. As Baumbach himself states; "In the layout of the script, I actually try to show where things should overlap so that the actors have that key". This is to present "a certain compartmentalization in the family" (Grobar & Grobar, 2017). Baumbach's films are dissimilar from those in the drama or comedy genres because he likes to combine pains and humor in the same story (*Noah Baumbach shows 'Greenberg' how he sees it*, 2010). The above stylistic distinctiveness of his filmscripts, including dialogues and emotional depth of his characters, is a significant reason for an investigation into the language of his screenplays.

To provide a description of telecinematic language of film scripts in general and that of Noah Baumbach, the present study employs the concept of 'lexical bundles' in corpus linguistics. A lexical bundle is frequent repeated sequences of words extracted by computer. Corpus processing software can show how frequent a bundle is presented in a discourse and in how many texts in a corpus. Because lexical bundles are derived through parameters of frequency and distribution in a corpus of texts, they can be drawn upon to describe language of a particular register or writer. Compared to other corpus approaches such as keywords, lexical bundles are continuous units that reflect the preferred expressions of writers (Altenberg, 1998). A lexical bundle as a repeated string "may have little (or no) psychological reality for speakers" (Scott & Tribble, 2006, p. 19), indicating that scriptwriters' unique writing styles. A large number of studies that utilize lexical bundles look at the language of different

established registers, such as business emails(Siricharoen & Wijitsopon, 2020), academic papers (Shirazizadeh & Amirfazlian, 2021) and learner English writing (Chen & Baker, 2010). This means that much has been done on lexical bundles in academic and professional discourse. However, scant literature remains on lexical bundles in telecinematic discourse, with an exceptions of Bednarek (2012), which looks at fictional television series; and Freddi (2011), which looks at film dialogues only without narrations in the scripts.

Taking into account all these gaps, the present study aims to describe the language of mainstream film scripts in general and the style of Noah Baumbach's screenplays through an examination and comparison of lexical bundles in the given texts. It is expected that the study can make a contribution to the growing field of telecinematic discourse and telecinematic stylistics, as well as combining corpus linguistics, discourse analysis, stylistics, and film studies, thus serving the academic community's interest in digital humanities and multidisciplinary research.

1.2 Research Questions

1   What lexical bundles characterize American mainstream film scripts? What meanings and/or functions do they have in film scripts?

2   What lexical bundles characterize Noah Buambach's film scripts? What meanings and/or functions do they have in his film scripts?

3   What lexical bundles contribute to stylistic differences between Noah Baumbach's and mainstream film scripts? What different meanings do they create in Noah Buambach's film scripts?

1.3 Objectives of the Study

1.   To describe the language of mainstream film scripts and its functions

2. To describe the style of Noah Baumbach's film scripts and its relationship with meanings and functions in his screenplays

3. To examine stylistic differences between lexical bundles in American mainstream film scripts and Noah Baumbach's film scripts

1.4 Scope of the Study

1. A corpus used to represent mainstream film scripts in this study consists of those ranked top 5 during 2005 to 2019.

2. The study investigates the five screenplays written by Noah Baumbach alone, which include *The Squid and The Whale* (2005)*, Margot at the Wedding* (2007)*, While We're Young* (2015), *The Meyerowitz Stories (New and Selected)* (2017) and *Marriage Story* (2019) .

3. The study only looks at 4-word lexical bundles as a stylistic feature in both datasets.

1.5 Definition of Terms

1. *Mainstream Films*:

The mainstream films in this study refer to those films that are produced by entertainment conglomerates and are very popular in America. The criterion of popularity is measured with American domestic box-office, the films included in the study must have earned over 50 million dollars in the U.S.

2. *American Independent Cinema*

"The independent filmmaking consists of low-budget projects by (mostly) young

directors, they have strong personal vision away from the mainstream influence"
(Tzioumakis, 2006, p. 1).

3. *Screenplays*

A screenplay is a collection of texts which normally conclude everything that
occurred in a feature film, including the audio elements such as dialogues, voiceover,
and background music. The visual elements are also included in a screenplay, like the
character's action, gesture, and environmental description.

4. *Lexical Bundles*

Lexical bundles are frequent recurring lexical sequences in a register. Also, it is
sometimes called clusters, multiword sequences, n-grams.

1.6 Significance of the Study

The study aims to make contributions to the field of telecinematic discourse and
stylistics as well as to film studies. Firstly, because the study examines common
lexical sequences in mainstream film scripts, it can contribute to telecinematic
discourse studies. Second, because the study compares the independent filmmaker
Noah Baumbach's works with abundant mainstream films in terms of phraseology, it
can shed light on stylistic differences between Baumbach's film scripts and those of
mainstream ones. At the same time, because Noah Baumbach's films belong to the
American independent film genre, the study can provide an overview of linguistic and
stylistic differences between mainstream and indie telecinematic discourse. Finally,
findings of the research can enrich academic interests in film studies because it
examines the screenplay as part of Noah Baumbach's filmmaking, which has been
overlooked by scholars compared to other independent filmmakers such as Wes
Anderson and David Lynch.

CHAPTER II LITERATURE REVIEW

This chapter aims to contextualize the present study both in theoretical and analytical aspects. It provides a comprehensive discussion on works related to film studies, telecinematic discourse and lexical bundles.

2.1 Film Studies

The term 'film studies' refers to an academic discipline that views film as an art form (Hill & Gibson, 2000), making it equally significant with other art forms such as literatures, music and paintings, etc. This is because a film can influence people's consciousness by involving them deeply (Bacon; Pepperell & Punt, 2006). Film studies often pay attention to film history, film theories, film production, theoretical and critical analysis of films (*Film Studies*, 2012). In addition to traditional films, video streaming such as YouTube has nowadays received attention and incorporated into contemporary courses in films studies.

Film studies emerged in 20th century. Since then, it has undergone serious developments where theories were introduced in the 1920s, including those developed by Bela Balazs (2010), and were used to investigate the relationship between films and society (Williams & Gledhill, 2000). This has transformed film studies into an institutionalized academic subject. Of the theories introduced, the most influential is *auteur theory*, espoused by Truffaut (1957), which appeared in the monthly journal *Cahiers du Cinema*. The theory gave importance to film directors. It suggested that a director owns the authorship of a film, and one film reflects the artistic vision of its director. One notable debate in that era was that of Truffaut's attack on the '*cinema de papa*' (also called *'daddy's cinema'*), a theory opposing 'auteur film' which referred to those commercial films produced in post-war French (Roberts & Wallis, 2001). A representative of *cinema de papa* is *Le Rouge et le noir* (The Red and the Black, Claude Autant-Lara, 1954). Autant-Lara insisted that the key aesthetics of the major French cinema was 'tradition of quality', which was considered a technique in productions over the innovation, and put the literary value of screenplay over the

director's style (Hill, 2008). Truffaut claimed '*cinema de papa*' was prefabricated and presenting merely the taste of bourgeois. According to Truffaut (1957), the *cinema de papa* mainly relied on:

1. high production values
2. reliance on stars
3. genre traditions
4. consider story's importance over film director (cited in Buckland, 2010, pp. 83–84)

Later, many editors from *Cahiers du Cinema* became renowned directors, like Francois Truffaut, Jean-Luc Godard, Eric Rohmer, and Claude Chabrol. These directors were considered as contributors of 'French new wave' (Marie, 2003), one of the most influential movement in film history (Neupert, 2007). They popularized the *auteur theory* worldwide and hence the personal style of film director has become the core content of film studies. Beyond film, the *auteur theory* became influential to film marketing. Some American directors were promoted as superstars over the other components of their films, such as Steven Spielberg, Quentin Tarantino and Martin Scorsese, etc. (Casey Benyahia & Mortimer, 2012, p. 23).

In the 1960s and 1970s, the British Film Institute (BFI) took up residence in academia by importing concepts from established disciplines like psychoanalysis and linguistics to film theories. BFI developed the academic curriculum through "combination of seminars, lectures and screenings" (Casey Benyahia & Mortimer, 2012, p. 24). In the meantime, the theory of narration and myth made it possible for film scholars to discuss popular films in an intellectual way (Mulvey & En, 2008). The BFI emphasized the necessity to connect academic theories to popular culture.

According to Casey, Benyahia and Mortimer (2012), popular films rarely appeared in academic studies or curriculum despite its marketability, and blockbusters were judged critically by film scholars. This situation developed over two stages in accordance with development of film studies:

Before films studies emerged, films were primarily labelled as a part of business, industry, or popular culture. They were considered to have less artistic value than other art forms such as literature and theatre.

With the development of film studies, film scholars started to distinguish between films that were worth studying academically and those that were not. The former was regarded as *film canon*. Interesingly, popular films were usually excluded from the *film canon*.

However, shortly after 2000s, the popular culture earned more academic recognition, such as Jenkins's influential study (2006) on fandom, Chew's (2019) investigation on the utopian politics of *Zootopia*, Teo's (2019) study on the most top-grossing Chinese film *Wolf Warrior 2* and Chinese film market's transformation, and Mayer's (2017) study on feminism inside science fiction film *Arrival* and Disney musical *Into the Forest* etc. These studies contributed to the growing academic landscape in film studies on American mainstream films.

2.1.1 American Mainstream Films

In relation to theoretical concepts in film studies introduced above, mainstream films are very similar to Film de papa in that they are "more polished, expensive and conservative films produced and distributed by the conglomerated Hollywood majors" ("Introduction: problems of definition and the discourse of American independent cinema," 2017, p. 2). The concept of American mainstream films (also labelled 'major films' and 'blockbuster movies') has been seen as on the opposite of the American 'independent film' (also labelled 'arthouse film') (Meyer et al., 2016), which can be associated with the concept 'film auteurs' in the abovementioned auteur theory (See also 2.1.2 below). As Sarris (1963) mentioned, American independent film directors could be categorized as *Film Auteur*, because they are film directors who have "creative dominance over his/her work". According to Zuckerman and Kim's (2003) classification, the critical difference between the two labels of films

mainly came from their 'market identities'; the mainstream ones aimed at the mass market while the independent films at niche market. Following Zuckerman and Kim (2003), Gemser, Van Oostrum, & Leenders (2007) classified the two by the following criteria:

> "if the film is released in film theatres that predominantly show art house films, the film is classified as art house."
> "if the film is released in film theatres that show above all mainstream films, it is coded as mainstream."
> (Gemser et al., 2007, p. 45)

In an era of streaming services and Covid-19, the criteria may be criticized as outdated. However, the present study would like to argue that the criteria still work especially in the context of American films:

1. Most of American films are still released in cinemas, mainly on a streaming service. For example, Netflix's films like *Roma* (2018), *The Irishmen* (2019), and Baumbach's *Marriage Story* (2019) had robust theatrical release in America (Brueggemann & Brueggemann, 2019; Vivarelli & Vivarelli, 2019).

2. Under the influence of Covid-19, cinema keeps its popularity. After the mass vaccinations in America, the domestic box-office number started to be back to normal (*'A Quiet Place Part II' Makes Serious Memorial Day Noise With A $48.4 Million Three-Day Bow; 'Cruella' Is Solid In Second With $21.3 Million*, 2021).

One of the representative study on mainstream films is *Movie Blockbusters* by Stringer (*Movie blockbusters*, 2013), which focused on blockbusters as a phenomenon. Stringer stated that blockbuster movies were often considered as of lower quality and remained neglected in the field of film studies. The study further distinguished the concept of blockbuster movies in terms of spectator's reception, industrial production scale, studio release scale and theatre exhibition. Newman

(2011), however, suggested that mainstream American movies can be explained in terms of aesthetics and styles of production as well. For example, a crucial narrative convention of mainstream cinema is 'goal-oriented storytelling' (Thompson, 2003), i.e., all plots are served for solving a problem (e.g., defeat the villain; rescue a person; destroy the bomb, etc.). In terms of telecinematic discourse, verbal styles in mainstream cinema dialogues tend to be formulaic and unnatural. King (2005, p. 64) concluded three unrealistic features of conventional characters in mainstream films:

1. **Omniscient:** showing a wide range of knowledge that is not related to a film character's profession.
2. **Highly communicative:** giving rather than withholding relevant information.
3. **Unselfconscious:** a film character is not shy or embarrassed to bare one's heart.

McIntyle (2012) conducted a corpus linguistic research to examine blockbuster movies in terms of its dialogue style. The study investigated male and female dialogues in selected major films by analyzing keywords, key domains, and n-grams. Drawing on the 'prototypical theory' (Rosch & Lloyd, 1978), the study revealed that the selected movies shared prototypical stylistic characteristics, i.e. linguistic elements that were more central than other elements (labelled as peripheral/secondary). In terms of key domains, the domain 'in power' was the most frequent one in male speech but not common in female speech, which in turn reflected the notion of masculinity (e.g., men should fight for power, heroes against authority) in Hollywood blockbusters. The domains like 'sailing, swimming, etc.', 'warfare, defense, and army' and 'weapon' were also frequent in blockbusters, because those movies (e.g., *Titanic* (1997), *Jaws* (1974)) were likely to put its protagonist in danger. Furthermore, the significance of 'speech act' as a key domain related to its plot-advancing function, the n-grams like *get me out of here* and *help me* reflect typical 'life-or-death' scenarios in those films. The study unveiled linguistic similarities in blockbusters,

which highlight gender stereotypes, featuring females as being in need of help and males as saviors and ones who fight for power.

2.1.2 American Independent Film

For a person familiar with pop culture or American films, the label *American independent film* refers to the films that had low budget and out of the Hollywood majors. However, this definition can be problematic because many well-known independent film companies were subsidiaries of entertainment conglomerates, e.g., the Focus Feature, which released Noah Baumbach's film *Greenberg,* has been owned by Times Warner. Tzioumakis (2017) stated that the concept of the American independent films was somehow blurred with mainstream films, so *independent cinema* is a relative concept. Despite that, he noted that an independent film "depart[s] from some or all conventions associated with classical narrative and film style" Tzioumakis (2006. p7). Based on this observation, independent filmmakers were broadly viewed as those who produce films against the mainstream conventions noted above in 2.1.1.

However, a study by O'Meara (2018) suggested that American indie films can be distinguished from its mainstream counterparts particularly in the dialogue. In the study, six directors were illustrated as *film auteurs* or filmmakers of American independent films: Noah Baumbach, Richard Linklater, Jim Jarmusch, Hal Hartley, Whit Stillman, and Wes Anderson. Overall, these directors were referred to as *Verbal-driven filmmakers*. O'Meara (2018) mentioned that these filmmakers like to write scripts on their own. They were usually screenplay writers of their films. Their films usually consisted of abundant dialogues and the characters, especially male characters, were more talkative in a near-monologue way, rather than for solving problems like those in mainstream films. This corresponded with Bihlmeyer's (2005) observation that male characters in mainstream films did not tend to be talkative as talkativeness is a feature closely related to femininity. On the other hand, male

characters in independent filmscripts tend to detail their thoughts and feelings in long meandering conversations, e.g., those in Baumbach's or Wes Anderson's films. Furthermore, verbal-driven conversation in American independent films can be seen as relatively more natural, compared to mainstream cinema talk. It had "regular interruptions, hesitations, or avoidances" (O'Meara, 2018, p. 220). In conclusion, the verbal-driven discourse was a major characteristic of American independent films that contrast with mainstream ones, in which case the former was more naturalistic and highly stylized speech. These, however, were intuitively observed features. The present study systematically examined and compared the nature of mainstream and independent film scripts through the concept of lexical bundles.

2.1.3 Noah Baumbach

Noah Baumbach is one of the most renowned American independent film directors. He is regarded as a director who makes "some of the most personal and influential films in American cinema" (Sharf & Sharf, 2021) and has established his reputation as one of the most respected and stylish phenomenal *auteurs* (Brigham, 2021). Before entering the film industry, Baumbach majored in English Literature at Vassar University and worked on prose writing in New York. Indeed, his works are under the influence of his literary background. One strong inspiration for him comes from writer Saul Bellow, especially his work *Herzog* (1964), who created ironic contrasts between the childlike intellectuals and their social background. Over the course of his career, he has blended wit and trauma in his works in a deeply personal approach as Baumbach himself stated: he was interested in "a cinematic world where the pain and humor exist simultaneously" (Horn, 2010). These facets of his works are discussed in detail below.

2.1.3.1 Filmography

Noah Baumbach started his career with *Kicking and Screaming* (Baumbach, 1995), the low-budget film depicts a bunch of new graduates who resisted entering the adult world. He became a rising filmmaker then with the critical acclaims from this debut. After two 1990s' failures *Mr. Jealousy* (1997) and *Highball* (1997), (the latter one he refused to own credit) (Pfefferman, 2005), he did not direct any single feature film for eight years. In 2004 Baumbach co-wrote the script of *The Life Aquatic with Steve Zissou* (2004) with Wes Anderson, and the following year he earned many accolades with *the Squid and the Whale* (Baumbach, 2005). After this film was nominated for the Best Original Screenplay in Oscar, Baumbach entered a stable period of filmmaking.

Baumbach directed *Margot at the Wedding* in 2007, which starred famous actors such as Jack Black and Nicole Kidman. It got mixed reviews like 'the characters are too unlikable to enthrall the audiences' (Margot at the Wedding (2007)).In *Greenberg* (2010), Baumbach collaborated with the mainstream star Ben Stiller as the titular character who experienced a nervous breakdown. The film was nominated for Golden Bear in Berlin Film Festival.

Frances Ha (2013) was shot in black and white. Scott (2013) acclaimed the film for its "swift, jaunty rhythms and sharp, off-kilter jokes". Moreover, Kohn (2013) pointed out that the film masterfully depicted the aimlessness of young generation. The leading actress Greta Gerwig co-wrote the script of *Frances Ha,* by which she was also nominated for The Golden Globe Award for Best Actress in a Comedy or Musical Motion Picture. In 2015, Gerwig and Baumbach co-wrote the script of *Mistress America* and the film was praised for its acerbic script and Gerwig's brilliant performance (Hornaday, 2015; Taubin, 2015).

While We Were Youngs (2015) and *The Meyerowitz Stories* (2017) are two familial comedies about marriage, artistic ambition and aging. Many critics (Brody, 2015;

Erbland & Erbland, 2017; Kemp, 2015; Lane, 2017; Myers, 2015) stated that Baumbach started working with more famed Hollywood stars (e.g., Dustin Hoffman, Ben Stiller, Naomi Watts, Emma Thompson, etc.), and tended to narrate his story in a more cheerful way compared to the previous traumatized tales.

Baumbach reached the peak of his career after *Marriage Story* (2019)*,* his most widely acclaimed film to date (according to the average score 94 in the website metacritic.com, his highest score). The film starred Scarlett Johansson and Adam Driver; it captured genuine happiness and sorrow of marriage life and brought actress Laura Dern an academy award for Best Supporting Actress.

2.1.3.2 Styles

Production Style: Baumbach has a flexible filmmaking process to make the budget as low as possible (Parker, 2013). For example, he would ask actors to have the shares of profit, instead of getting salary before shooting. However, as an indie director, he collaborated with many mainstream actors, such as Nicole Kidman in *Margot at the Wedding* (2007)*, Ben Stiller* and Adam Sandler in *Meyerowitz Stories* (2017). He provided opportunities for these Hollywood celebrities for new trying and higher praises.

Storytelling Style: Feller (2020) pointed out that audiences could easily capture the similarities between Baumbach's film and his own life. For example, the divorce plot in *the Squid and the Whale* (2005) was very similar to his parents' (Genzlinger, 2019), who separated in 1980s in New York. The autobiographical approach in his films created sensitive atmosphere and absorbing emotional depth. Instead of mimicking the flow of daily conversation, Noah Baumbach focused on miscommunication and emotional dysfunction.

Baker (2013) argued that Baumbach's film *Frances Ha* is a romance but not a romantic comedy, it's 'not a boy-girl romance' or 'girl-girl romance' but a romance between the title character (Frances) and her capital-S (alternative) Self'. In his films, the unlovable characters could find a way to accept themselves in the end of the story.

Characterization Style According to Handler (2019) and Foundas (2010), Baumbach's characters had two major characteristics. First, all his films depicted an immature protagonist. As a line in Greenberg revealed: 'hurt people hurt people', Baumbach's characters usually suffered from past traumas, and their internal pains made them treat others badly. Secondly, a key topic of Baumbach's film was the self-struggle of anxious men and women. In most of his works, males were usually immature and sensitive, and females had explicit manners. These struggling characters shared some behavioral pattern, i.e. they are usually "noxious narcissists" and "stubborn self-aggrandizement" (O'Meara, 2018, pp. 372-385).

In addition, O'Meara (2018b) laimed that Baumbach's films have been influenced by his Jewish background. His films show correspondence with those of other Jewish comedians such as Groucho Marx, Woody Allen, Ethan, and Joel Coen, etc. She argued that those films, e.g., *Annie Hall* (1979) and *A Serious Man* (2009)), established a typical Jewish male image. His Jewish characters have these characteristics:

1. **Fast-talking**: In align to Mast's depiction (1990), the Jews in comedies like Woody Allen, usually had a 'Mouth as nervous brain'. O'Meara refered it to Baumbach's *Greenberg* (2013), the titular character endlessly talking his own topics, and answering his own questions before anyone else.
2. **Effeminate:** Abrams (2012) noted that the Jewish film characters have two typical self-images: one is hyper-masculine/militarized, the other is queer-alike and inadequate masculine. Baumbach's characters are closer to the lateral, which presented as avoid responsibilities, complain frequently, more intellectually strong than physically, etc.

3. **Neurotic**: The neurotic characters in Baumbach's films usually presented by their anxieties, which was related to childishness. In *Greenberg,* Roger would put a Chapstick to his mouth whenever he felt anxious. It is similar to *Hannah and Her Sisters* (1988), the protagonist kept seeing doctor to confirm a non-existed tumor.

Wilkins (2021) concluded that Baumbach maintained his artistic autonomy in film industry, by asserting his identity as 'literary filmmaker'. Baumbach's literary scaffolded the intellectual characterizations, the majority of his films centered on characters who are either academics, or well-educated creatives. The elitist nature indicated by overwhelmingly occurred cultural references such as niche books, films, music (e.g., French film *The mother and the whore* in his *The Squid and the Whale*).

2.2 Telecinematic Discourse

The term 'telecinematic discourse' has only recently been highlighted as an independent term of scholarly pursuit. Two edited volumed have recently published: Telecinematic Discourse (Piazza et al., 2011) and Telecinematic Stylistics (Hoffmann & Kirner-Ludwig, 2020). The *International Journal of Corpus Linguistics* has dedicated a special issue in 2021 to corpus-based studies of telecinematic language. Before these, earlier linguistics studies tended to overlook film/television language(Bednarek et al., 2021). In spite of numerous linguistic investigations on drama (Calvo, 1990; Culpeper, 2009) and fiction (Lambrou & Stockwell, 2010 Chapter 6; Takala, 1994), the language of audiovisual medium had been underrepresented during the past decades.

2.2.1 Definition and Characteristics of Film Discourse

The term 'telecinematic discourse' was initially proposed by Piazza, Bednarek and Rossi (2011), who used this term to cover all studies on the language of film and television. In detail, they defined telecinematic discourse as spoken and written languages in narrative film and television. To explain precisely, it is:

> an exploration of spoken and written language used in fictional/narrative film and television from various perspectives and discussing different kinds of data. [... It] attempts to understand, describe, and define such language in its relation to real life and in consideration of its functions within the fictional narrative: how special if at all is the language of cinema and television.
>
> (Piazza et al., 2011, p. 1)

The term 'telecinematic discourse' treats film and television as two similar media genres. However, while there were many similarities between film and television, the two forms were mainly different in technological aspects. According to Bednarek (2010, p. 20), the film production usually paid more attention to camera language, such as shot type, camera angles, composition and lighting, etc. However, Androutsopoulos (2012, p. 149) claimed the clear-cut boundary between film and television is hard to identify, because the productions of TV series "adopt film narratives and visual aesthetics", and both film and television are increasingly transferred to the Internet.

Although there is a long-established public impression on film and television that they both are pure visual media that focus on images, cinematic sounds have come to acquire a crucial status. They consist of dialogue, monologues, voiceover, soundtrack, etc. The dialogues spoken by characters, which originate from film scripts that are usually written before the film shooting, are the central component of film sound because it "has much to contribute not only during audience reception but likewise in film production"(Hoffmann & Kirner-Ludwig, 2020, p. 1). Therefore, film dialogues

can be investigated in terms of its audio (the speech actors uttered in film) and textual medium (the lines writer crafted in screenplays), the latter of which is the focus of this research.

However, the language of television and film is not limited to monologue and dialogue. Bateman and Schmidt (2012) argued that the telecinematic discourse includes all types of cinematographic and semiotic resources utilized in the production, such as editing, lighting, color grading, etc. Furthermore, recent research (e.g. Hoffmann, 2020) tends to be against the original definition from Pizza, Bednarek and Rossi (2011) that it has limited the telecinematic discourse as the study of "scripted talk in fictional/narrative film and television". A newly proposed definition of telecinematic discourse is presented as:

> Telecinematic discourse refers to the use (and interplay) of both (aural) film discourse and (visual) cinematic discourse. While film discourse refers to the use of verbal language in all of its possible forms, shapes and shades of expression (spoken and written, monologue and dialogue, diegetic and non-diegetic), cinematic discourse describes the manifold (visual) techniques and semiotic resources (apart from aural language) which directors strategically apply to create comprehensive, complex telecinematic experiences for a given audience at home or in cinemas.
>
> (Hoffmann, 2020, p. 5)

According to Zago (2020), the telecinematic discourse is a linguistically interesting object to study because:

1. The discourse in cinema and TV presents the dynamics between writing and speaking. As the tv/film dialogues are written in text then verbally narrated with actors (Gregory, 1967), the investigation of telecinematic discourse brings about problematizing differences between the verbal and the written. The problematization is similar to research in other linguistic

fields like 'computer-mediated communication or colloquialization of English'.

2. The telecinematic discourse is pragmatically and discoursally complex. Though the dialogues are likely to behave in a true-to-life way, it is addressed to audiences. This means that the illocutionary force (Searle, 1976) of dialogues in film/TV series is greater than those in real life conversation.

According to Forchini (2020), studies on cinematic discourse can be categorized into four types, in terms of the data selected:

1. authentic movie discourse (e.g., original script written by scriptwriter)
2. Non-authentic movie discourse studies (e.g., on web script)
3. Authentic and (4) unauthentic audiovisual translation studies. (e.g., dubbing and film subtitles)
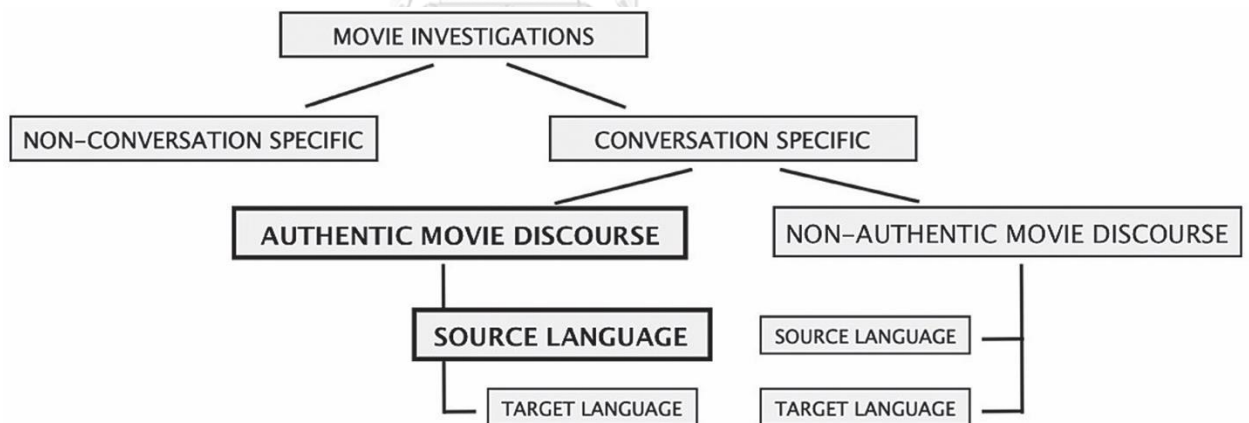
Figure 1 Classification of movie investigation within linguistics (Forchini, 2020)

2.2.2 Film scripts

A film script provides information about a film which normally includes dialogues, monologues, soundtrack, and voiceover, etc. Such a script is written and used for films. Beyond its use in films, the script is significant for budget management, casting, and further adjustment on the story (Gregory, 1967), and almost everything that happened on the screen. In terms of accessibility, American scripts were made publicly available through their official websites or through a website like scriptslug.com.

The film script has its own distinctive features and structure, as illustrated in Figure 1 below.
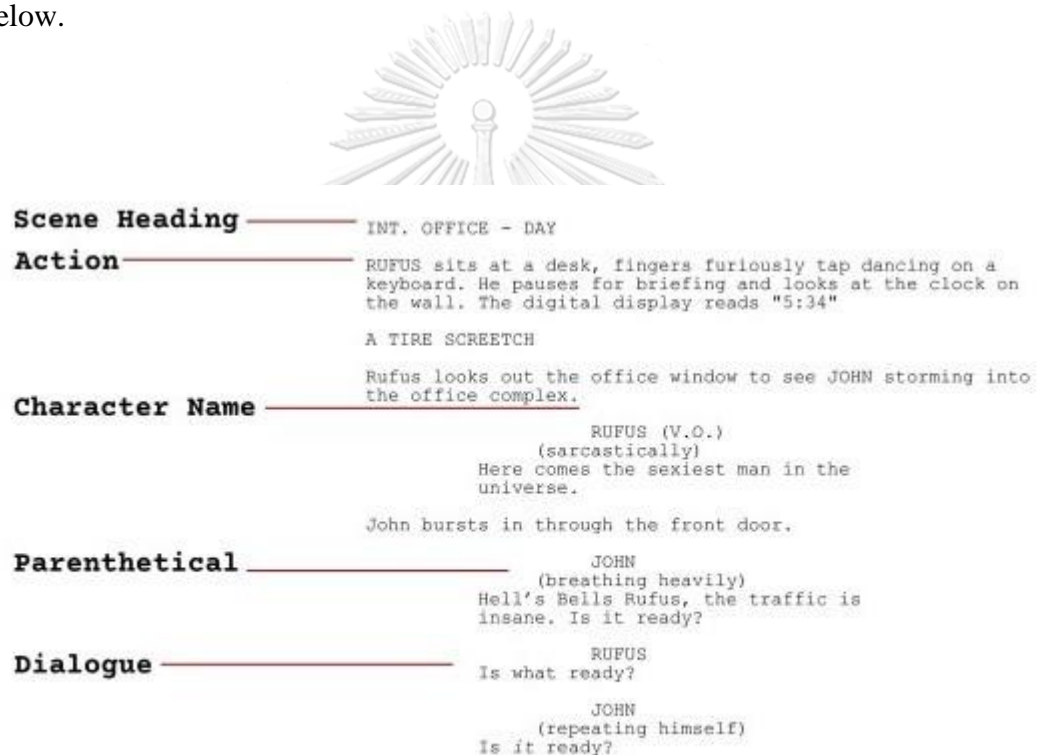
Figure 2 *A sample of the standard format of a film script, adopted from "Elements of Screenplay Formatting - ScreenCraft," (2015)*

These texts were scripts that closely resembled the standard film script format (Cole, 2002), which includes necessary elements enumerated in the subsection as follows:

2.2.2.1 Sluglines

A slugline is a scene heading presented in full capitalization. It tells readers the location where the scene takes place. The line must distinguish the scene as inside

(INT.) or outside (EXT.), which refers to an indoor or outdoor scene. A period followed the location, which indicated the time of the day (morning, evening, afternoon, etc.). Sluglines are usually presented in the following way:

    2 INT. JOSH AND CORNELIA'S APARTMENT. DAY (*While We're Young,* 2015)

In this example, '2' refers to the scene number in the script. Accordingly, it is the second scene of the film "While We're Young", which takes place indoor at Josh and Cornelia's apartment during the daytime.

2.2.2.2 Action Lines

Action lines are right below the sluglines; they provide information about a character's physical actions. Below is an example of action lines:

The kid launches into his pitch. She listens intently and starts writing down her phone number.

2.2.2.3 Dialogue

Dialogue consists of utterances of characters. In dialogue, a character's name is capitalized, and utterances are placed beneath the character's name. This is illustrated below:

<div align="center">

ELIZA

Do you realize, eating meat

is worse than driving an SUV

for a year.

</div>

2.2.2.4 Extensions

An extension is normally placed next to the character's name in parentheses. It tells the readers the manner of the spoken dialogues presented, such as:

Voice over (V.O.): A narrator talks outside the scene, like she/he is directly communicating with the audience. It can also be an internal thought.

Off-Screen (O.S.): A dialogue can be heard by other characters but can't be

seen by

audiences.

2.2.2.5 Transitions

Transitions tell the film editor how to edit between two scenes, such as:

CUT TO: the normal transition is usually indicated at the end of the scene.

DISSOLVE TO: a scene dissolves to the other, usually indicated the time

passing.

INTERCUT: editor jumps back and forth between two scenes, usually

used in a

phone conversation.

2.2.3 Previous studies on telecinematic discourse

Because telecinematic discourse features dialogues, several telecinematic stylistic/ discourse studies apply pragmatic perspectives into their investigation. For example, Bousfield and McIntyre (2018) examined the film *Full Metal Jacket* (1987) with the prototype-based taxonomy of impoliteness and rudeness(Bousfield, 2010). Their study focused on the character Sergeant Hartman, a military trainer who treats his soldiers with severe cruelty and verbal insults. It revealed that this character adopted creative impoliteness as the opposition of expected response, by doing so to enhance his verbal aggression for remodeling recruits' identities into the military model. Nursanti (2015) analyzed the maxim flouting of young adult franchise *Hunger Game* (2012-2015) with relevance principle(Grice, 1975). The study focused on all the maxim floutings in character's utterances, and it found that the characters flouted the maxim of relation most frequently, which meant that the characters were likely to make utterance sway from the topic of conversation.

A sociolinguistic perspective has also been taken in telecinematic stylistics. For example, Lee (2018) investigated a non-human character Japanese anime series *From*

the New World (2013), using Androutsopoulos's (2012) three-level film analysis framework and Bucholtz and Hall's (2005) identity framework. Lee proposed that observed from the keywords list of the character's sub-corpus, it shows that the character identity is established from his language, including "strong in-group affiliation, low social status and overall presentation of subserviency". Gregori Signes (2020) explored the victim naming of the series *Twin Peaks* with Goffman's participant framework. The study focused on the respective name-calling on the central victim Laura Palmer by other characters.

Recently, a large and growing body of telecinematic stylistic studies have investigated telecinematic discourse through a corpus-assisted approach. (Bednarek, 2011; Jautz & Minow, 2020; Pavesi, 2020; Reichelt, 2020). Those studies examined keywords, collocations, concordances to investigate linguistic patterns underlying the films and TV series. Pavesi (2020), for instance, studied demonstratives in English and American films and found that demonstrative pronouns have occurred frequently in the cinematic discourse, and proximal demonstratives (this/these) are more frequent in film dialogue in comparison to other spoken registers of English. This was because the demonstrative "this" has a narrative function that can intensify the attention of audience toward the current character/object/event.

Jautz and Minow (2020) explored the formulaic nature of soap operas. They studied the trigram We need to talk in SOAP corpus (Davis, 2011–). This lexical bundle had been proven as 'unusually frequent' in the previous study of Bednerak (2012). The analysis showed that the phraseology "[name], we need to talk" occurs more in soap opera than in COCA, this utterance is usually spoken before a problem-oriented conversation, and male characters use it more than females. The frequent occurrence of this trigram was due to the necessity of entertainment since soap operas showrunners were required to create dramatic and romantic scenes, especially in mixed-gender conversations. (e.g., male protagonist tries to confess to his girlfriend after misunderstandings).

The corpus-based approach was also adopted to examine telecinematic characterization from the diachronic perspective. In an investigation on the stability of *Gilmore Girls'* characters across 8 seasons of the series, Bednarek (2011) examined key words and lexical bundles in the character Lorelai's dialogue. The researcher found that the character had been diachronically stable through the whole program, i.e. there is no apparent style shift in her dialogue.

Although there seems to be quite a number of recent corpus-based studies on film scripts, most of them focus on the dialogue part of the film script even though, as demonstrated above, the text of a film script consists of several other components. Moreover, those studies tend to highlight specific aspects of language use in a particular movie/ series, e.g. characterization, rather than general linguistic patterns that occur in full texts of film scripts across movies. The present study, therefore, examines the whole texts of film scripts, covering all elements in the text of film scripts (see 2.2.2).

2.3 Lexical Bundles

Lexical bundles are frequent sequences of words that repeatedly occur across several texts in certain varieties (Biber et al., 1999). Because they are automatically extracted from a corpus, they can be everyday utterances (e.g., 'I don't know) and do not have to be idiomatic expressions (e.g., 'kick the bucket'). Conrad and Biber (2005) stated that lexical bundles are essential elements of discourse that can fulfill communicative purposes. Biber et al. (2004)pointed out that their functions in discourse can be categorized into four categories:

2.3.1   Stance expressions
are bundles that present the degree of attitudes or certainty of expression after the bundles. There are five sub-groups of stance expressions:

- **Epistemic stance** bundles present the text-producer's (un)certainty, e.g. *I'm not sure, I don't think so, you know what I.*
- **Desire** presents the wishes and desires of the text-producer, or ask information about the other person's desire, e.g. *I just wanted to, if you want to.*
- **Obligation/directive** bundles present obligations or direct the order that speaker wants the interlocuter to accomplish, usually ordered with second person pronoun e.g., *you don't have to, you have to be.*
- **Intention/prediction** contains lexical bundles that describe the writer's intention to do certain future action, e.g. *I'm going to, I was going to.*
- **Ability/effort** bundles indicate ability, e.g., *to be able to*, *it is possible to*

2.3.2 Discourse organizers

can indicate the logical relation between the prior and following discourse.

- **Topic introduction/focus** serves the function of introducing a new topic. e.g., *I would like to, I'm sorry to.*
- **Topic elaboration/clarification** refers to the bundles which elaborate or clarify the topic, e.g., *know what I mean, nothing to do with.*
- **Conditions** bundles normally contains a complementizer 'if', e.g., *if you have time, if you do not*

2.3.3 Referential expressions

refers to the bundles indicating a reference to the physical/abstract units or to the textual context.

- **Identification/focus bundles** emphasize a single important feature or identify a noteworthy part of something, e.g., *as one of the*; *those of you who.*

- **Tangible** bundles are related to concrete attributes, e.g., *in the size of, in the form of*;

• **Intangible** refers to lexical bundles which related to abstract attributes, e.g., *in nature of the, on the basis of*

• **Time reference** refers to lexical bundles that describe temporal attributes or time periods**,** e.g., *at the same time, in the meantime.*

2.3.4 Special-function lexical bundles

are those that do not belong to the above three major categories. They are often used to fulfill pragmatic functions. They are identified by Biber *et al*. (2004), Biber (2006), and Siricharoen and Wijitsopon (2020), for instance:

- **Politeness:** lexical bundles usually involve terms to perform polite acts, e.g. 'thank (s)' to serve text-producer's politeness strategy, e.g., *thank you for your, thanks for your help*

- **Request** refers to lexical bundles that ask the audience politely to conduct an action. They usually contain the word 'please', e.g., *please help me with*

- **Provoke a further communication** concerns the lexical bundles that provide

  chance for the audience to contact the text-producer fatherly, e.g., *let me know if*

- **Offer** concerns the lexical bundles that provide help, suggestion in the latter proposition, e.g., *let me know if*

- **Expectation** contains lexical bundles that present the text-producer's expectation,

  e.g., *I look forward to, look forward to seeing*

- **Hybrid function** concerns the lexical bundles that merge two functional types. e.g., *let me know if*

In this study, the lexical bundle approach aims to examine the formulaic characteristics of both mainstream American films and Baumbach's. Using stylistic

evaluations of their scripted film language, the study compares the writing styles of both registers as revealed by observed lexical bundles.

CHAPTER III METHODOLOGY

This chapter describes the methodology adopted in the present study, including corpus compilation, software, and analytical frameworks.

3.1 Corpus compilation

To answer the three research questions, two corpora were manually compiled: a corpus of mainstream film scripts and a corpus of Noah Baumbach's film scripts.
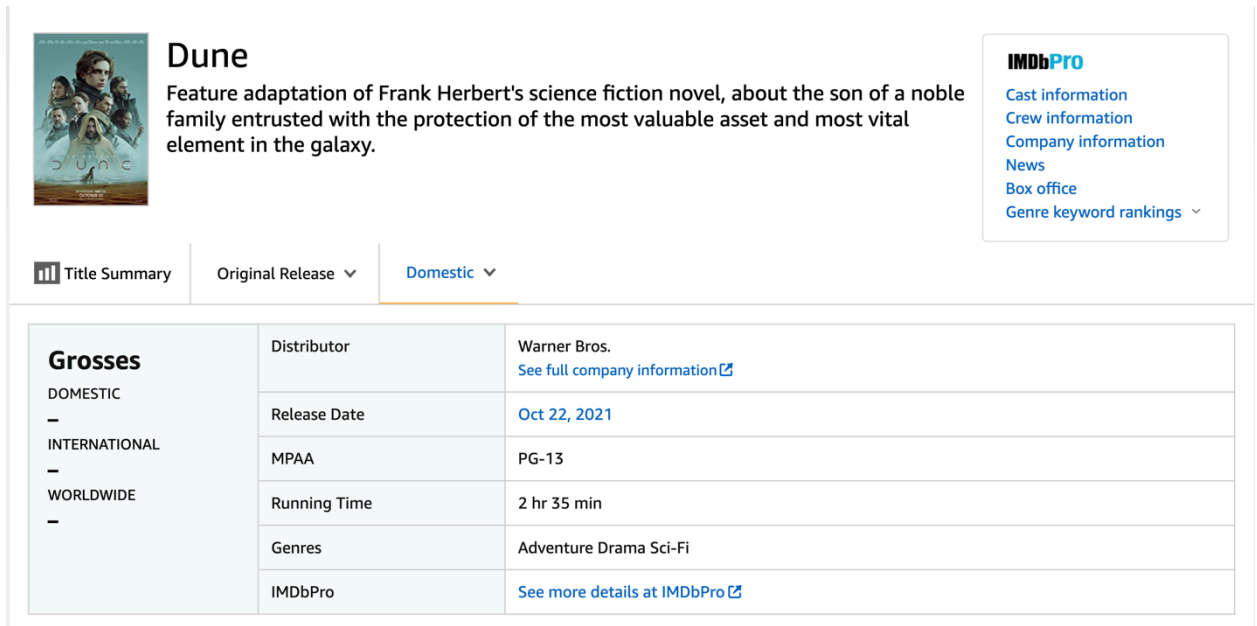
3.1.1 Mainstream film script corpus (MFC)

In this study, a corpus of mainstream box-office film scripts (henceforth MFC) contains 100 film scripts with 2,383,551 words tokens and 61,068-word types (as illustrated in appendix B). There were three criteria used for the compilation of MFC:

1. The release date of the films is in parallel with Noah Baumbach's selected works, namely 2005 to 2019.

2. The top 5 ranking films in the U.S. domestic box-office chart of each year during 2005 and 2019 were selected. If the scripts of the top 5 films were unavailable online, the next place was selected. For example, in 2014, the film in 5th place was *Transformer: Age of Extinction,* but its script was not found, so the script of the 6th-place film *Maleficent* was used instead).

3. Since every film on NBC is drama genre, an additional 25 dramas were chosen to increase the MFC's viability and make it more comparable to NBC. All 25 films rank in the top 40 on the annual domestic box office chart for the United States from 2005 to 2019.

It must be noted that by prioritizing movie ranks in the compilation of MFC, balance of movie genres in the corpus cannot be fully achieved. In other words, movie genres are not evenly distributed in MFC. This can affect types of lexical bundles that turn up in the analysis, which in turn will have implications for the way patterns lexical bundles in MFC are interpreted. However, since this study is interested in linguistic

patterns in mainstream movies, use of ranking as the major criterion in movie selection is inevitable. Moreover, movie genre is not a clear-cut category. Most of the films can be described as belonging to two or more genres, as indicated by the metadata 'genre' on boxofficemojo.com. As illustrated by Figure 3 below, the film *Dune* belongs to three movie genres, i.e. adventure, drama, and sci-fi.

Figure 3 Metadata of the film Dune in boxofficemojo.com

Nevertheless, attempt has been made to create transparency in data collection of MFC. The proportion of movie genres in MFC, according to boxofficemojo.com, is listed below in Table 1:

Genres	Raw Numbers of Films	Proportion in MFC
Adventure	53	53%
Drama	42	42%
Action	39	39%

Comedy	30	30%
Fantasy	30	30%
Family	26	26%
Sci-Fi	24	24%
Thriller	19	19%
Animation	17	17%
Biography	14	14%
Romance	12	12%
Crime	11	11%
Music	10	10%
Mystery	8	8%
Musical	5	5%
Sport	5	5%
History	4	4%
War	4	4%

Table 1: The proportion of film genres in MFC

The texts of all 100 film scripts were retrieved from scriptslug.com, and the downloaded data was manually checked in terms of spelling correction.

3.1.2 Noah Baumbach's film script corpus (NBC)

A corpus of Noah Baumbach's film script (henceforth NBC) consists of 5 texts of his screenplays, namely *The Squid and The Whale (2005), Margot at the Wedding (2007), While We're Young* (2015), *The Meyerowitz Stories (New and Selected)* (2017) and *Marriage Story* (2019). They were chosen to create a corpus because they were

written by Noah Baumbach alone so his writing style can be ensured. Therefore, works like *Frances Ha* (2013), which has another co-writer, were excluded from the selection.

In total, NBC consists of 118,573 tokens. The scripts were downloaded from scriptslug.com and manually checked for spelling correction. The scripts in NBC are final drafts used in film shooting.

The process of extracting the script texts in Noah Baumbach scripts included three steps:

1. The researcher copied all the PDF script files or their web page;

2. Those texts were pasted onto a text editor; and those texts were converted into plain texts and saved as those with the .txt surname.

3. After converting the texts, the files were renamed as 'release year + film title' and then were moved to a new folder named 'Noah Baumbach Scripts

3.2 Software

This research employs the software *Antconc* (Anthony, 2020) to extract lexical bundles in MFC. The corpus was loaded to *Antconc*, and lexical bundles were extracted via the n-gram/cluster function. It must be noted that via *Antconc*, the apostrophes will not be identified in the result, and hence such expressions with apostrophes as *I'm* or *don't* were identified as two-word lexical bundles like *I am* and *do not*, respectively. This means that the bundle *I don't think* is counted as a four-word bundle (amounting to *I do not think*) like the cluster *I think that you*. For this study, I decided to follow *Antconc*'s system of automatic identification of lexical bundles to avoid the impact of spelling differences between straight and curly apostrophes in different film scripts (e.g., *don't* and *don't*). Hence, a list of lexical bundles in the study contains those with clearly distinct four words, e.g. *I think that*

you, and those with three words in an extracted form with the apostrophe, e.g. *I don't think*, which are identified by *Antconc* as four-word bundles.

3.3 Lexical bundle extraction

Three criteria were adopted to conduct automatic extraction of lexical bundles in MFC and NBC: (1) minimum frequency of lexical bundles, (2) the minimal number of texts in which lexical bundles occur, and (3) the length of lexical bundles.

The present study chose the length of four-word bundles because, as Biber et al. (1999) pointed out, three-word clusters are too common and five-word or longer clusters are 'more phrasal' and less encountered in the corpus. In the same vein, Stubbs and Barth (2003:76) stressed that longer bundles are easily limited to specific texts.

In terms of the minimum frequency of lexical bundles and the number of texts for MFC, the researcher set a minimum frequency at 100, and each bundle should occur more than 40 films. This is in order to avoid bringing up lexical bundles likely to be more specific to particular films or film genres. For NBC, the researcher set a minimum frequency at 6, and each bundle should occur more than one film (range>1) to avoid obtaining a lexical bundle that occurs repeatedly in one film only.

3.4 Data Analysis

After the software generated the list, I categorized the lexical bundles on the list into functional types on the basis of Biber et al. (2004)'s and Biber (2006)'s frameworks.

The data analyzed in this study are two corpora both composed of fictional film scripts, the meanings of bundles may varied depend on their film plots. Therefore, for a more contextualised view of the data, the analysis on linguistic patterns (e.g., concordance lines, collocation) and analysis on single excerpt are both used to

interpret the corpora. For the purpose of overview, the research looks at the common patterns in concordance lines of the bundles, in order formulate interpretations of how the bundles and their collocate used in corpora; for a more specific detail, the single scene analysis highlights how the bundle functions scripted language in real examples.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

CHAPTER 4 FINDINGS AND DISCUSSIONS

This chapter presents the results of functional analyses performed on both corpora. Two major sections report and analyze the findings. The functional types of lexical bundles from MFC and NBC are classified in sections 4.1 and 4.2, followed by a discussion of the similarities and differences between lexical bundles from the two corpora.

4.1 Lexical Bundles in MFC

This section address the first research question, reporting on lexical bundles in MFC and discussing their functions. Based on the criteria and threshold spelled out in 3.3, a total of 77 lexical bundle types were derived. Following Biber et al. (2004)'s framework spelled out in 2.3, the derived lexical bundles were categorized according to their functions. Note that several novel subcategories were added by the researcher to the group of Special Functions to accommodate occurrences of some lexical bundles newly found in this study. Table 2 below lists all 77 lexical bundle types identified in MFC according to their functional categories. Each lexical bundle is accompanied by its raw frequency shown in parenthesis.

Table 2: Lexical bundles in MFC and their functional categories

Functional category (Token No.)	Lexical bundle (Frequency)	No. of Types	Percentage of Types (No.)	Number of Tokens (per cent)
	A. Epistemic stance			
	I don't know (689) , don't know what (184), I don't think (175) , don't know I (74), you don't know(87), I didn't know (80), I can't believe (82), don't know how (75), but I don't (71), I think it's (63), and I don't (64), I'm not sure (59)	12		1,703 (43.73)
	B. Desire			
STANCE EXPRESSIONS (3894)	*I don't want (201), don't want to (191), I'd like to (100), do you want to (66), you don't want (60), I don't need (60), I don't care (75)*	7		762 (19.57)
	B2. Obligation			
	you don't have (98), I want you to (89), don't have to (86), you want me to (83), I need you to (70)	5	45.45% (35)	426 (10.94)
	B3.Prediction/intention			
	I'm going to (233), you're going to (168), we're going to(114), I'm not going (70)	4		585 (15.02)
	B4. Ability/effort	4		295 (7.6%)

		I don't have（107）, can't help but(69), I cant do(63), we don't have (56)			
	B5. Imperatives			
	get out of here (62) get out of the (61)	2		123 (3.2%)

DISCOURSE ORGANIZERS (230)	A. **Topic introduction/focus** *that's what I (82), you know what I (71), that's why I (55)*	3	5.19% (4)	137 (59.57%)
	B. Conditions *if you don't (93)*	1		93 (40.43 %)

REFERENTIAL BUNDLES (2895)	**A. Identification/focus** *it's not a (56)*	1		56 (1.93)
	B. Quantity Specification *the rest of the (169),*	1		169 (5.84%)
	C1. Place reference *in front of the (196), the edge of the (177), in the middle of (130), the end of the (165),the top of (142), the side of the (144), the middle of the(129), on the other side(119), out of the way (123), the other side of (116), at the end of (108),the center of the (101), in front of him (100), the back of the* (201), *other side of the(96), in front of a (87), at the top of (88),*	21	31.17% (24)	2,514 (86.84%)

	out of the car (82), out of the room (77), the front of the (70), in the back of(63)				
	C2. Time reference *for the first time (156)*	1			156 (5.9 %)
	A. Apology *I'm sorry I (131), i'm so sorry (87)*	2			218 (14.93)
SPECIAL FUNCTIONS (1460)	**B. Inquiries** *what are you doing (221) , what's going on (130), why don't you (87), are you talking about (74), what are you talking (63), what do you think (85), how do you know (61), what do you mean (79)*	8	8%	19.4 (15)	800 (54.79)
	C. Actions *look at each other (140), takes a deep breath (90), turns back to the(63), looks up at the (93), looks down at the (56)*	5			442 (30.27)
Total		77	100%		8,479

Table 2 presents the lexical bundles and their functional categories derived from MFC. As can be seen, stance bundles are the most dominant group in MFC, having the most lexical bundle types and the highest frequency of tokens. This is followed by Referential, Special Functions and Discourse Organizing bundles, respectively. Overall, these proportions reflect the nature of film scripts as a hybrid discourse which is heavily spoken and descriptive. Specifically, mainstream film scripts are predominantly dialogic, as manifested through a large density of spoken formulaic

expressions in the Stance, Special-function, and Discourse-organizing categories. This points to the dialogue-oriented storytelling technique employed for textual development. The text of film script, however, is also largely contributed by descriptive narration, as illustrated by the second-most frequent Referential lexical bundles in MFC. Major patterns and functions of each lexical bundle category are discussed in more detail in the following subsections. It should be noted here that functional analysis of the lexical bundles involves examination of their expanded context in relevant scripts and concordance lines to describe why and how they are used in the mainstream film scripts.

An excerpt from the film will be presented first, followed by contextual information regarding plots and further explanation regarding the functions of bundles, as demonstrated in the example provided below:

Excerpt: **(1)**

CHARLOTTE

The whole system cost eleven

dollars! It even smells expensive. ***You're going to*** look beautiful on your date tonight.

(2011 *The Help*)

Contextual information:

The excerpt takes place in a 1960s American southern family, Charlotte bought a hair-curler for her daughter.

Interpretation:

The bundle *You're going to* reveal Charlotte's wish and prediction that her daughter would look better by this new machine, the expression entails another underlying wish that she wants her daughter to marry a decent guy as soon as possible. This in

accordance with that social circumstance in 1960s. that a girl could only live better by a successful marriage.

4.1.1 Stance Expressions in MFC

Stance bundles make up the largest functional category in MFC, with 35 lexical bundle types in stance expressions, equaling 45.45% of all derived bundles. This highlights the prevalence of spoken discourse in film scripts as suggested by the predominance of stance bundles with the first-person pronoun "I" and such contracted forms as "don't", "I'm" and "you're". These bundles express characters' thoughts, opinions, attitudes, and evaluations which in turn serve to initiate and advance the story. While this does not seem very surprising (given the genre of texts in MFC), the corpus-based approach to film scripts here brings to attention a structural pattern shared by stance bundles across sub-categories, which has not been noted elsewhere, to the best of my knowledge: stance lexical bundles in American mainstream film scripts are mainly in negative forms, with 24 out of 35 bundle types (68.57%). This may be because, on the one hand, negative lexical bundles are characteristic of spoken language, as Biber et al. (1999) have found, and hence film dialogues are full of negative bundles. However, upon qualitative textual analysis of film scripts in MFC, the predominance of negated stance bundles cannot be attributed to a property of spoken registers alone. It is found that the prevalence of negative stance bundles can be linked to the creation of conflicts, both external and internal, e.g., characters' disagreement, unwillingness, or (in)difference, which constitutes the problem-solving convention in storytelling, as illustrated in the following subsections.

4.1.1.1 Epistemic bundles

Most bundles are realized in the epistemic stance sub-categories through cognition verbs, such as *know* and *think*, and mental state expressions, e.g. *sure*. Eleven out of

12 epistemic stances indicate uncertainty. In many cases, negative epistemic attitude bundles are often used when characters are uncertain about future events/status. For instance:

Uncertainty about the future

(1) ANDY

I need your help. ***I don't know*** what to

do. It's like I'm completely beneath her contempt.

(2006 *The Devil Wears Prada*)

(2) YOUNG CHARITY

I don't know what my future will be. Father says I am to be a lady, but... It's all so dull when you're not there.

(2017 *The Greatest Showman*)

The two extracts above illustrate how the negative stance bundle *I don't know,* the most frequent lexical bundles in MFC, plays a role in expressing characters' uncertainty that accounts for major conflicts underpinning the film stories. In extract 1), the character Andy is trying to gain approval from her boss, and the lexical bundle helps to express her problem that keeps the story going. In extract 2), the lexical bundle betrays the character's inner conflict, i.e., her uncertainty and anxiety.

Uncertainty about the past

When epistemic bundles are used to signify a previous occurrence (e.g., *I don't know what just happened*), they primarily serve to exonerate the speaker of an allegation or suspicion.

When anything unfavorable occurs, characters usually deny the allegations by pointing out that they were unaware of certain facts. Thus, in excerpt (1) below, Harry states that he has no idea why the goblet of fire (a magical artifact) chose him as a contestant for an upcoming tournament that he is not eligible for since he is under the age of seventeen. Because the goblet's choice perplexed the other students, many of them suspected Harry of cheating. Therefore, he has to protest against the other students' suspicion. The repeated lexical bundle, *I don't know*, in the excerpt helps reinforce his claim.

(1) HARRY

I didn't put my name in that cup. I

don't WANT eternal glory I just wanna

Be. Look, ***I don't know what*** happened

tonight and ***I don't know why***, it just

did ok.

(2005 *Harry Potter and the Goblet of Fire*)

At the same time, the bundles *I don't know* and *don't know what* were found to form a core part of a longer and more semantically loaded formulaic expression *I don't know what you're talking about,* which is used for a similar pragmatic purpose, i.e. denying the other character's suspicion. This is illustrated in examples (2) and (3) below:

(2) OBI-WAN

I know how he feels about you.

PADME

(nervous)

What did he say?

OBI-WAN

Nothing. He didn't have to.

PADME is a little flustered. She stands and Obi-Wan follows.

She walks to the balcony.

PADME

I don't know what you're talking

about.

(2005 *Star Wars Episode III – Revenge of Sith*)

(3)

TRUPO

Man walks around in a fifty thousand

dollar chinchilla coat and he never even

bought me a cup of coffee? Something

wrong there.

FRANK

I don't know what you're talking about.

(2007 *American Gangster*)

Finally, apart from negative stance lexical bundles, the only affirmative stance lexical bundle *I think it's* is also a highly functional bundle utilized to advance the storyline. It can be used to discourage someone from doing something (e.g., *I think it's unwise for you to*) or present a fact or opinion (e.g., *I think it's hot*). In MFC, it is noticeable that the bundle is often followed by the noun *time* (15 out of 62 occasions), as exemplified below:

we're not heroes! SUSAN We're from Finchley! PETER	*I think it's*	time we were going. LUCY But
out there, the Endeavour is coming up hard starboard, and	*I think it's*	time we embraced that oldest, noblest
bounds up the stairs. Fox follows, out of breath LUCIUS	*I think it's*	time to talk about my year-
your son. Dad is taken aback. PO'S DAD Po,	*I think it's*	time I told you something I
HICCUP STOICK I've decided I don't want to	*I think it's*	time you learn fight dragons. to
keep blasting the hearing, it's not gonna come back.	*I think it's*	time we reconsider the inner monitors.
A man only has his pride to stand on and	*I think it's*	time for you to get a
CONTINUED: (6) PROFESSOR McGONAGALL (CONT'D)	*I think it's*	time I ask a few of
bounds up the stairs. Fox follows, out of breath LUCIUS	*I think it's*	time to talk about my year-
great Mike Wazowski! You'll come up with something. MIKE	*I think it's*	time I leave the greatness to
for me to make a spectacular return to evil! Doctor,	*I think it's*	time we showed Gru what we'
up a lot of TV hours talking about me. Tonight?	*I think it's*	time we talk about my wife.
Until he runs right into -- Mrs. Starrett. 28. MRS. STARRETT	*I think it's*	time you leave. Ben. Catching his
What? AARON I'm late for class. Go home, mom.	*I think it's*	time for your nap. Aaron leaves.
down to her level, and tries to sound	*I think -it's*	time for you take a nappy-

Figure 4: Concordance lines for 'I think it's' in MFC

Upon looking at the phrase in its various contexts in MFC, the bundle *I think it's time* is potentially face-threatening and hence can lead to external and internal conflicts

between characters. For example, it is used as a challenge to others' opinions, e.g., "I think it's time we reconsider the inner monitors".

In some cases, it provides the impression to the listener that any suggestion made afterwards will be challengeable. The bundle here could be used to encourage the others to take on a challengeable task, e.g., '*I think it's time* to learn to fight dragons'.

Moreover, it proposes that the listener enter a new topic that the speaker wishes to focus on, e.g., 'It's time we talk about my wife', or expresses the desire to leave someone alone or dismiss someone, e.g., 'it's time you leave'.

4.1.1.2 Attitudinal and modality stance bundles

The next category is *attitudinal/modality stance*, the most frequent type in this category is *Desire.*

Desire

A closer look at desire bundles reveal that the negative form is also predominant (5 out of 7 types) in this subcategory. Its prevalence comes from the fact that characters in MFC tend to express their unwillingness (e.g., 'I don't want') or indifferences (e.g., 'I don't care'), by using these bundles frequently. Again, the negation of stance bundles seems to be a major linguistic strategy in creating tension and conflict in film stories. This observation gains support from comparing uses of desire bundles in MFC and those in previous studies on lexical bundles in other genres, e.g., business emails or academic writing (Biber & Barbieri, 2007; Siricharoen, 2016). While emails and academic writing tend to state the want or intention to do something (e.g., 'I'd like to'), film scripts contain a lot of bundles that express refusal to have or get something, such as *I don't need* or *don't want to*. This is also coherent with King's earlier

observation (2005) that a typical mainstream film character is not shy or embarrassed to uncover his/her heart. For example, in extract (1) below, the most frequent Desire bundle *I don't want* is used to exclaim the character's despise on the rich family she is facing:

(1) RACHEL

I don't want any part of your

family.

(2018 *Crazy Rich Asians*)

In another case, *I don't want* is also used to avoid further conflicts, as in example 2 below:

(2) LINGUINI (CONT'D)

Look. ***I don't want*** to fight. I've

been under a lot of- you know, pressure. A lot has changed in not very much time, you know. I'm suddenly a Gusteau and I gotta be a Gusteau or you know, people will be disappointed. It's weird...

(2007 *Ratatouille*)

This monologue was spoken by Linguini (a cook's assistant) right after he catches a rat, and it sounds hilarious to an audience. People would reckon it is entirely neutral for a man who works in the kitchen to chase the rats out. Followed by a more bumbling discussion, the whole monologue illustrates his timid and confused manner.

I'd like to is the only bundle type (100 tokens) in this subgroup that expresses the character's willingness instead of unwillingness. The bundle is a contracted form of *I would like to*, which is common in emails and academic writing while the full form only occurs 10 times in MFC. This mainly comes from the informality of film language, which abounds in dialogues and hence colloquial speech styles are preferred. Interestingly, unlike those used in emails or lectures, in which *I'd like to* expresses the speaker's desire through politeness strategies, several cases of *I'd like to* in MFC occur in contexts of hatred or face-threatening acts, such as refusal. In extract (1) below, Eduardo wants to freeze the bank account as revenge for his friend Mark, and in (2) Hazel wants the addressee to leave him alone.

(1) EDUARDO

I'd like to freeze this bank account and

cancel all existing checks and lines of

credit.

(2010 *Social Network*)

(2) HAZEL

No, I'd... *I'd like to* be alone for

a while.

(2015 *Creed*)

Obligation

For the bundle *You don't have*, 54 out of its 98 instances are the fragments of the five-word bundle *you don't have to*, the rest of the instances are related to Ability/effort

subcategory (e.g., *you don't have a choice*). In most scenarios, *you don't have to* is an expression of courtesy that can free the listener from some obligations, as illustrated below. In (1), for example, Enrico uses it to save Pizarro from answering a question; in (2) Stock tries to save the hearer's energy from going up somewhere, and in (3) Amy tries to free the hearer from a financial obligation for paying the fee.

(1) ENRICO

Mumble, We don't got the pebbles!

(To Pizarro)

Don't listen to him, senor. **You don't have** to answer that.

(2006 *Happy Feet*)

(2) STOCK

You don't have to go up there.

(2010 *How to Train Your Dragon*)

(3) AMY

I thought you'd be in a great mood.

You **don't have to** pay for dad's nursing home anymore.

(2015 *Trainwreck*)

In some cases, however, *you don't have to* in MFC occurs in a context opposite the above common use. It can perform a highly face-threatening function in utterances. In excerpt (4) below, for instance, the bundle is used explicitly to confront the other speaker:

(4) ERICA

--I have to study.

MARK

You ***don't have to*** study. You ***don't have***

to study. Let's just talk.

ERICA

I have to go study.

MARK

You ***don't have to*** study.

ERICA

Why do you keep saying I ***don't have to***

study?!

MARK

Because you go to B.U.!

(2010 *Social Network*)

Boston University (B.U.) is a well-known university near Harvard University. Mark despises Boston University since he attends Harvard, which has a higher rating, so he continued telling Erica that she didn't need to study because a lower-rated university doesn't need much effort.

Prediction/intention

I'm going to is the most frequent bundle in this subcategory. It can be used to describe one's prediction of the short-term future, as in extracts (1) and (2) or illustrate a long-term vision towards the future, as in extract (3):

(1) LANDY

I'm going to the office now.

I'll make my own introductions. Off VOSEN...

(2007 *The Bourne Ultimatum*)

(2) GRU

You are going to suffer the wrath

of Gru! Seriously I'm going to count to three, and you had better be in this car!

(2010 *Despicable Me*)

(3) SKEETER

I'm going to be a serious writer,

Mr. Blackly. I applied for a job, but Miss Stein just thought-

(2011 *The Help*)

However, a less common use of the phrase is found in MFC, again, as part of impolite acts. There are several cases in which characters use this bundle to threaten or intimidate hearers. The following excerpts display these cases:

(4) MURDOCK

(TO BUD)

Get me a beer, old man. And if it's not

good, ***I'm going to*** smash it across your

face.

(2007 *Wild Hog*)

(5) CHINESE GANGSTER

Nay tiey ching chaw yut bok mon.

[Make sure it's real. It's a million dollars.]. ALTERNATE/ALSO (insult): Ne dei yeeche ling ngnaw dong maw gok nay chun. [Next time you make me wait, ***I'm going to*** cut your dick off.

(2006 *The Departed*)

Ability/effort

All of the bundles in this subcategory are negative, implying the character's incompetence (e.g., 'I can't do this without your help') or insufficiency (e.g., 'I don't have a license'). A case of interest is *we don't have*, which is a highly functional bundle for screenwriters, especially in terms of accelerating the film's plot or creating a suspenseful atmosphere. As exemplified by the extract below, the character, Alan, uses *we don't have time* to push his/her partner to take an action. It also excites the audience for an upcoming rush moment.

(1) ALAN

We don't have time for this! We

gotta find Doug!

(2009 *The Hangover*)

4.1.2 Referential lexical bundles

As mentioned earlier, referential bundles constitute the second largest functional category in MFC. Its large proportion, however, comes mainly from lexical bundles that designate places (see Table 4.1), with 21 out of 24 referential bundle types (87.5%). While place and time are essential elements in narrative text world (cf. Gavins, 2007), it is interesting that in mainstream film scripts references to place far outnumber temporal expressions. Upon examination, it is found that the large density of place-referential bundles relates to characters' actions and movements, often described in action lines, where information about characters' physical actions is given (see 3.1). In other words, place references serve to construct spatial context of characters' actions, which in turn becomes contextual information for storyboard design and film shooting. The fact that place referential bundles, like stance bundles, largely constitute mainstream film scripts reflects another aspect of the nature of filmscripts, i.e., it involves descriptions of settings and actions, apart from the dialogue, to fulfill their communicative functions. Below are examples of place-referential lexical bundles in the context of mainstream filmscripts. As can be seen from the examples, the bundles occur not only in action films and they help discourse participants, be actors or directors, to build up a scenario in their minds before film shooting.

(1) Adelie climbs up anyway. Everyone budges along Mumble who has nowhere to move. With a despairing second Adelie falls off *the back of the* berg.

(2006 *Mission Impossible 3*)

(2) The weapons pile up ***in the middle of the*** stage. Everyone else cheers and starts to mingle, introducing themselves and shaking hands.

(2007 *Shrek 3*)

(3) Mary sits in a chair eating dinner from a tray. Miranda sits on ***the edge of the*** bed near her.

(2008 *Sex and the City*)

It must be noted that while place referential bundles are associated with characters' physical actions, the only temporal lexical bundle in MFC *for the first time* relates to characters' perceptions and emotions. Of its 156 entries, 36 (23.07 %) cases of the bundle are used in characters' dialogues and 120 (76.93%) in action lines. This suggest that like place referential bundles, it is more characteristic of the description part than the dialogue. Analysis of its concordance lines reveals that the bundle often co-occurs with words that describe characters' vision and visual acts, such as *see, seeing, noticing, gaze* and *eyes.* Moreover, these vision-related words often co-occur with those related to feelings and emotions, including *sympathetic gaze, deep emotion,* and *painful looking* (See Figure 4.2 below). Based on these patterns, it can be observed that a primary function of the lexical bundle *for the first time* is to highlight critical emotional moments in the stories, which are reflected through their visual perception. Given that readers of filmscripts are primarily people involved in film production, namely directors, actors, cinematographers, the lexical bundle also serves as a linguistic clue for them to interpret, direct and act in terms of emotional expression in films.

1. <u>Her eyes</u> return to Caesar -- who   *for the first time*   softens, offering a <u>sympathetic gaze</u>.

2. her little face inches from his...	*for the first time*	we see deep emotion in her
3. to the Colonel, and noticing	*for the first time*	painful looking BRAND SCARS on
4. here anyway. Mike's eye narrows.	*For the first time*	he looks determined. The light turns
5. on a White House tour.	*For the first time*	he looks irritated at work. INT.
6. that Ludendorff loosens his grip.	*for the first time*	we see that he is afraid.
7. alright? Ray looks up, and	*for the first time*	we see sadness. Tears trickle down
8. Goodbye, Bella. And here,	*for the first time*	we SEE how truly agonizing this
9. looks away and Dumbledore,	*for the first time*	notices his hands. They are splayed,
10. MIRANDA appears. They	*for the first time*	MIRANDA nods, almost
11. the tube. Ted sits on the sofa, and	*for the first time*	we see Ted in his present

Figure 5: Sample concordance lines for 'for the first time' in MFC

In dialogues, the bundle also serves to constitute a character's emotional expression in a highlighted moment, with a similar co-occurrence pattern, as illustrated by the highlighted parts in the extracts below.

(1)

LORRIE

For the first time, just now, I

realized that there were 155 people on that plane. And you were one of them.

More tears. Tears of relief and joy.

(2)
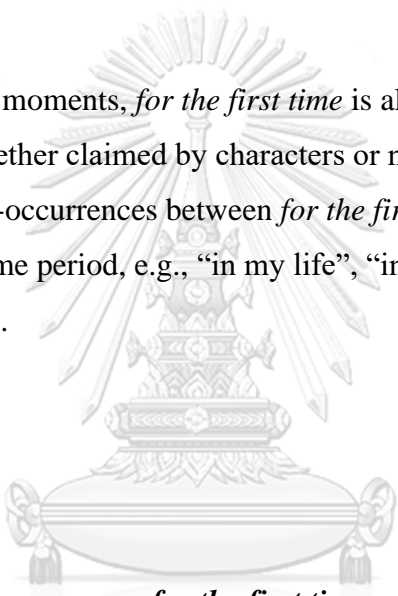
EGO (CONT'D)

It was with Meredith that I

<u>experienced love</u> **for the first time**. I called her my river lily. And from that love,
Peter, you.

(2017 *The Guardians of Galaxy vol.2*)

In relation to emotional moments, *for the first time* is also used to describe milestones
in a character's life, whether claimed by characters or narrated by screenwriters. This
is suggested through co-occurrences between *for the first time* and prepositional
phrases referring to a time period, e.g., "in my life", "in our history" and "in years", as
illustrated by Figure 4.3.

1. theft. They're suing me because   *for the first time*   <u>in their lives</u>, things didn't

2. So, what now? MIKE You know,   *for the first time*   <u>in my life</u>, I don't

3. Gloria's eyes are clear and   *for the first time*   <u>in a long time</u> she looks

4. man in the room is surprised   *for the first time*   <u>in a very long time</u> by

5. , all weakness gone. She smiles   *for the first time*   <u>since Stefan took her wings</u>. But

6. beat, husband and wife reunited   *for the first time*   <u>in years.</u> In the smoke and

7. and reading a newspaper   *for the first time*   <u>in 12 years</u>, plus I just met

8.	staring at the numbers. Stuck,	*for the first time*	in her life. Stafford stops at
9.	her voice to meet Hippolyta's	*for the first time*	in three thousand years.
10.	of the Nation of Wakanda and	*for the first time*	in our history we will be

Figure 6: Sample concordance lines for' for the first time' in MFC

4.1.3 Special-function lexical bundles

Not only through the high frequency of the above functional categories, the significance of the combination of dialogue and action in mainstream filmscripts can be realized qualitatively through Special-function lexical bundles. While the sub-categories of *Apology* and *Inquiry* bundles realize the importance of these speech acts in MFC, the *Actions* sub-catgory illustrate actions common in mainstream filmscripts, e.g., *looks at each other* and *turns back to the* (see Table 4.1). It should be noted that while the spoken sub-categories are also found in previous studies on lexical bundles in other types of discourse, e.g. business emails (Siricharoen and Wijitsopon, 2020) and textbooks (Biber, 2004), the *Actions* group is a new subset of Special-function lexical bundles added to the category in the present study. This is to accommodate a number of related lexical bundles that emerge in MFC. Furthermore, this suggests an important characteristic of mainstream film scripts, i.e., hybrid of common formulaic spoken expressions and descriptive expressions of actions, which in turn is linked to its function in cinematic discourse that features conversation and acting.

Given the *Actions* lexical bundles, it can be seen that the act of looking is the most common in the writing of film scripts, with 289 out of 442 tokens (65.38%) of *Actions* bundles containing the verb "look". The bundle *look at each other*, the most frequent

bundle in this group (see Table 4.1), is often followed by other actions, especially those that register feelings, e.g. *laugh, shrug, smile* and *hug*, as illustrated by the Figure 4.4 below:

1.	ENGAGEMENT RING... They	***look at each other***	ALLY (laughing) Are you kidding
2.	starts up his bike. Doug and Bob	***look at each other***	and shrug. Dudley hurries out of
3.	the boxers to shore. The guys all	***look at each other***	and smile. It's a nice
4.	me, understand?! Josh and Kitty	***look at each other***	Little awkward. GABE (Josh and
5.	right? Eddie, Richie and Stanley	***look at each other***	The mood has changed. Back to
6.	earthquake. Ajax and Deadpool	***look at each other***	Shrug. DEADPOOL (CONT'D)
7.	What was that? They turn back,	***look at each other***	Ginny smiles. Then: NEW
8.	New Year! Carrie and Miranda	***look at each other***	They hug. 168L INT. LOWER
9.	Carrie, Samantha and Miranda	***look at each other***	Carrie BURSTS out laughing.
10.	MIRANDA appears. They	***look at each other***	MIRANDA nods, almost

Figure 7: Concordance lines for look at each other in MFC

This pattern points to the significance of the act of looking in mainstream films as a common preliminary action, through which two characters communicate quietly before more explicit expression of feelings or attitudes is revealed. Directors and actors have to rely on this description for film shooting so that acting can lead the audience to appropriate interpretation.

4.1.4 Discourse-organizing lexical bundles

According to Table 4.1, Discourse-organizing bundles are the least frequent and varied functional category in MFC. Despite the limited occurrences, some patterns can be observed from this category. They are dialogic expressions where all contain either first- or second-personal pronouns, thereby contributing to the dialogic nature of film scripts. In the *Topic introduction/focus* subcategories, the three bundles (*that's what I, you know what I, that's why I*) contain the personal pronoun *I* and are said by characters to give explanations about themselves. In fact, many of these bundles are part of longer conventionalized expressions, such as *you know what I mean/think, that's what I'm talking about,* etc. This not only creates a style close to real-life conversation in a film script but also provides contextual information about characters for audiences of the movies. For example:

BELLA

You all know what Edward wants. And

you know what I want. But I won't

force myself on you...

(2009 *The Twilight Saga: New Moon*)

In this excerpt, Bella emphasizes knowledge commonly shared between her and the Cullen family: she desires to be a vampire like them, by using the bundle *you know what I want*. This in turn can prompt the audience to infer what that common knowledge the characters have in order to understand the meaning between the lines about the conflicts between both parties, i.e., that although both understand each other, they cannot fulfill the other's wants. Bella wants to be a vampire, but Edward does not want her to be.

4.2 Lexical bundles in NBC

This section addresses research question 2. Based on the threshold and criteria indicated in 3.3, a total of 100 lexical bundles in NBC were derived and categorized into four types, following Biber et al (2004)'s functional framework. Table 3 demonstrates the distribution of four primary functional categories in NBC.

Table 3: Four primary functional categories of lexical bundles in NBC

Functional category (Token No.)	Lexical bundle (Frequency)	Percentage. of Types (No)	Percentage of Tokens (No)
	A. Epistemic stance		
	I don't know (152), I don't think (33), don't know I (21), don't know what(17), I think it's (16), and I don't (15), but I don't(15), **don't know why(14),** *I didn't know(14), I can't believe(12),* **it doesn't matter(11), I don't see (10), don't know if(9),** *I'm not sure(8),* **'t know why I(8), don't think I(7), don't think you(7), I think he's (7), I think I'm (7), we don't know(7), didn't know you(6), don't know that(6), don't know where(6), I didn't realize(6), I don't remember(6), I just don't (6), I think that's(6),I thought it was (6), I thought you were(6),** *you don't know (6)*	30 (30%)	450 (47.22%)
	B. Desire		
STANCE EXPRESSIONS (953)	*I don't want (57), don't want to (55), , you don't want(16), do you want to (15),* **I don't like(13),** *I don't care (12), i want you to (11),* **'t want to do(11), didn't want to (8),** *I'd love to (7),* **I don't need(7), I want to be (7), 't want to be (7),** *I'd like to (6),* **want to be a(6), want to do it (6),** *you want me to (6)*	17 (17%)	244 (25.60 %)
	B2. Obligation		
	don't have to(9), **you don't have(8), do I have to (6),**	3	23

		(3%)	(2.41%)

	B3. Prediction/intention		
	*I'm going to(63), you're going to (23), we're going to(21), I'm not going (15) , **I was going to(10), going to have to (9), I'll see you (7) 'm going to go (7), re not going to (7), they're going to(7), he's going to (6), you're not going (6)*	13 (13 %)	194 (20.36%)
	B4. Ability/effort *I don't have (12), **I didn't get(9), I'm trying to(8)**, we don't have (7) **don't have a (6)***	5 (5%)	42 (5.41%)

DISCOURSE ORGANIZERS (50)	A. **Topic introduction/focus**		
	***i'm telling you (12),** you know what I (10)**, what I'm doing (9) I have to say(6), it's hard to(6),*	5 (5%)	43 (86 %)
	B. Conditions *if you don't (7)*	1 (1%)	7 (14 %)

REFERENTIAL BUNDLES (108)	**A. Identification/focus**		
	's not what I(12), that's not what (12)**, it's not a (7), **but it's not(6),	4 (4%)	37 (34.26%)
	B1 Place reference *in front of him (10), out of the car(9), the back of the(8), other side of the(7), **the other side of(7), a few feet away(6)**, at the end of(6), **in the passenger seat(6), stands in the doorway(6),***	9 (9%)	65 (60.18%)
	B2. Time reference		

	in the middle of (6),	1 (1%)	6 (5.56%)
SPECIAL FUNCTIONS (115)	**A. Politeness** *I'm sorry I (22),* **nice to meet you (9)**	2 (2%)	31 (26.96 %)
	B. Inquiries *what do you mean (17), what are you doing (14), why don't you (9), what's going on (6),* **what you want to (6), why didn't you (6),**	6 (6%)	58 (50.43 %)
	C. Actions **kisses him on the (7),** *takes a deep breath (7)*	2 (2%)	14 (12.17%)
	D. Deny **no I don't (6)**	1 (1%)	6 (5.22%)
	E. Script Jargon **v.o. she's (6)**	1(1%)	6(5.22 %)
Total		**100**	**1,226**

(The emboldened lexical bundles in the table are those that do not occur in MFC.)

CHULALONGKORN UNIVERSITY

As shown in Table 4.2, the most common category in NBC is stance bundles (68%), followed by referential bundles (14%), special functional bundles (13%) and discourse organizer (6%), respectively. This distribution pattern is similar to that in MFC (*see* 4.1). The similarity can be attributed to the fact that MFC and NBC are of the same genre, i.e. film scripts, in which dialogues and descriptive details are featured and proceed the texts. Nevertheless, upon a closer look, percentile proportions of some categories in NBC differ greatly from those in MFC. Such quantitative disparity is discussed in Section 4.3. In this section, each category of lexical bundles in NBC and some of its representative bundles are examined and discussed in terms of their contributions to Noah Baumbach's script writing.

4.2.1. Stance bundles in NBC

A total of 68 stance lexical bundle types turn up in NBC. It is found that stance lexical bundles from different subcategories also share co-occurrence patterns with the words pauses and hesitates, which suggest discontinuity in conversations. This in turn points to problematic interactions of some kind as a result of a character's uncertainty, anxiety or lack of understanding as the characters in NBC like to pause between or before their statements. This is reflected through the frequent use of the words pause and hesitate in the dialogues illustrated in the concordance lines below (most instances of both words occur in parenthesis).

1.	*I don't believe* he's read it.	**(pause)**	You both should talk to him. Bernard
2.	Good. *I don't know*. I do too.	**(pause)**	I can teach you to swim if
3.	*I can't believe* she did this to me!	**(pause)**	I didn't tell you because... I
4.	*You don't want to* be a pro.	**(pause)**	I'm sure I lost my parking
5.	*I don't know* why I said that.	**(pause)**	What's this cane? DANNY Oh...it'
6.	What's wrong? MARGOT *I don't know*.	**(pause)**	Before you gave me your sweater I
7.	the Galapagos? JOAN *I don't know*, Pickle.	**(pause)**	Ivan and I could take you to
8.	It sounds great. Who's in it? WALT	**(pause)**	Orson Welles? *I don't know*, I
9.	I come? (CONTINUED}: BERNARD Yau have tennis.	**(pause)**	*You're going to* he doing that
10.	No, Margot, *I don't want to* count.	**(pause)**	What was it about Dad that had
11.	I don't know why I said that.	**(pause)**	What's this cane? DANNY Oh...it'
12.	What's wrong? MARGOT *I don't know*.	**(pause)**	Before you gave me your sweater I
13.	Nah, *I don't have* the right shoes.	**(pause)**	Hey, I was thinking...you know how
14.	the Galapagos? JOAN *I don't know*, Pickle.	**(pause)**	Ivan and I could take you to

15. ***You don't want to***? She slowly shakes her head.   **(pause)**   Josh notices something. JOSH Is that the shaman's

16. here? CHARLIE Yeah. But we LIVED in New York.   **(pause)**   Why? Is there a problem? TED ***We're going to***

17. FRANK Could I come? BERNARD You have tennis.   **(pause)**   ***You're going to*** be doing that which is

18. in it. JOSH Only because ***I thought it was*** real!   **(pause**   with genuine feeling) I loved you. JAMIE I like

Figure 8: Concordance lines of 'pause' in NBC that collocates with lexical bundles

1. ***You don't think*** it's bad, do you? CHARLIE (**(hesitates)**   I don't ever watch TV so, you know,

2. re in the same room all of the time. CORNELIA   **(hesitates)**   ***I don't want to*** take away your enthusiasm

3. you WON'T LISTEN. Harold opens the car door and   **hesitates**   HAROLD ***I don't know how*** I could be

4. NICOLE (pause) Do ***you want me to*** cut it? CHARLIE (**(hesitates)**   OK. NICOLE I'll get scissors. EXT. NICOLE'S

5. Silence. CORNELIA ***You don't want*** kids, right? (Josh   **hesitates)**   Because I don't. (they both hesitate) I'm

6. the art to these facilities as a write-off. MAUREEN (**hesitates)**   Oh...OK. ***I can't believe*** Danny is happy

Figure 9 Concordance lines of 'hesitates' in NBC that collocates with lexical bundles

The patterns above illustrate an important aspect of Baumbach's scriptwriting, i.e., contextual information provided before, during or after characters' speaking, such as their gazes, facial expressions, tones, body language, and state of mind. These details are usually put in parenthesis as highlighted markers.

Another shared pattern among stance lexical bundles in NBC is that, like MFC, most stance bundles are in the negative form, such as *I don't know*, *I don't think*, etc. Specifically, of 68 stance bundles, 42 are negative (61.76%). This may be because negated bundles are characteristic of spoken English and film dialogues, as suggested in 4.1. However, as will be shown in the analysis of sub-categories of stance bundles below, negative stance bundles in NBC are also highly functional to Noah Baumbach's film scripts. It should also be noted that the words *pauses* and *hesitates* mentioned above are found to co-occur more frequently with negative stance

expressions than affirmative ones. This highlights a strong association between the two words and negative stance bundles as well.

4.2.1.1 Epistemic Stance Bundles

In this subcategory, NBC and MFC shared seven bundles; they are the seven most common bundles in MFC, namely *I don't know*, *don't know what*, *I don't think*, *don't know I*, *you don't know*, *I didn't know*, and *I can't believe*. They have similar usages to indicate uncertainty or lack of knowledge about the future and the past, like the examples of *I didn't know* and *I can't believe* below.

(1) REITBART

You know, Josh told me... ***I didn't know*** you <u>miscarried</u>.

(2007 *Margot at the Wedding*)

(2) MAUREEN

(hesitates)

Oh...OK. ***I can't believe*** Danny is happy about that and Jean...well, who knows what Jean feels about anything.

(2017 *The Meyerowitz Stories (New and Selected)*)

As illustrated above, the speakers sought to avoid bringing up sensitive topics (such as miscarriage) during the conversations. In extract (1) Reitbart was concerned to not directly discuss miscarriage, which she learned through Josh. The ellipsis signaled a pause; a brief pause would demonstrate her empathy for another lady's misfortune. Meanwhile, *I can't believe* in extract (2) includes a hesitation before the bundle. Maureen used this lexical bundle to show her amazement at her stepson Danny's attitude about family heritage; she then paused before speaking about another stepdaughter Jean, concealing her true feelings toward the two stepchildren. 'They're truly terrible disappointments,' she acknowledged in a later scene.

It must be noted that the bundle *I can't believe* is often associated with the characters' emotional turbulences toward a sudden loss or anger, as shown in the examples below:

(1) PAULINE

Ingrid's really upset about it.

Fuck. ***I can't believe*** she did this

to me!

(2007 *Margot at the Wedding*)

(2)

Nicole stamps her feet and shakes her fists like a child having a tantrum.

NICOLE

I can't believe I have to know you FOREVER!

(2007 *Margot at the Wedding*)

There are 20 out of 30 epistemic bundles that only occur in NBC, the majority of them are in negative forms such as *I don't see* and *don't know why*. The lexical bundle *don't know why* (14 times) is the most frequently occurring bundle exclusive to NBC. This bundle is often associated with emotional expressions in NBC related to family memories, infidelity, and personal trauma. For example, in *Meyerowitz Stories (New and Selected)* the bundle *don't know why* occurs five times, of which the character Matt uses twice and both instances are preceded by an interruption (cough and tear). See the extracts below:

(1) MATT

……

You know what's awesome about middle age, you now know more than your parents. You can guide the--

(coughing a bit now)

I'm sorry, ***I don't know why*** I'm suddenly, I think it's all this plaster dust and this coffee—

(2017 *The Meyerowitz Stories (New and Selected)*)

(2) MATT

……

Of being very proud. Of wanting to be an artist like my dad. Of being included -- he was interested in me, he...loved me...

(he wipes tears from his face)

I'm sorry, I ***don't know why*** I'm...

I've been angry at him for so much of my adult life. I guess I was trying to outrun him.

(2017 *The Meyerowitz Stories (New and Selected)*)

The use of *don't know why* in conjunction with emotional expressions illustrates a pattern of Matt's discontinued talk, which points to the film's central theme: a group of dysfunctional adult siblings (including Matt) struggling to live through their father's shadow. As evidenced by the fact that both interruptions occur during his emotional confessions about his father, 'cough' or 'cry' is the reaction along with his emotional outbursts. In extract (1), Matt uses a disguised tone to discuss his glamorous post-adult life but implies bitter feelings about his father's bad influence, before the coughing interrupts his statement. In extract (2), he speaks of the value of his father's love during his childhood but is then interrupted by tears. In both scenes, Matt states that he does not know why he interrupts himself, but he is fully cognizant

of his desire to be cherished by his father and of his sadness due to his dysfunctional family.

Another instance regarding interruption followed by *I don't know why* is also found in *While We are Young*. In the extract below, Darby laughs while discussing the tragic death of her dog. In fact, the death is made up by Darby to gain sympathy while the story also seems funny to herself.

(3) DARBY

When I was a child my dog was killed in front of me by two rottweilers.

(she starts to laugh)

I don't know why I'm laughing. It's not funny.

(2015 *While We're Young*)

The use of *I don't know why* in this case betrays contrastive feelings. This is also in line with the other instances of *don't know why* in NBC. Most of the characters are talking about their/others' internal minds when they use this expression, such as not knowing why they are crying or laughing. They rarely mention an objective fact, such as not knowing why the car broke down; yet, there is an exception in *While We are Young*, in which a characters says "I don't know why the PowerPoint didn't work".

Importantly, from the point of view of concordance analysis, epistemic lexical bundles have co-occurrence patterns that point to the association between these bundles and negative feelings or interactions. As can be observed from figure 4.7, these bundles co-occur with negative words, such as, annoyed (line 4), wearily (line 1) and dismissive (lines 7 and 8). Moreover, they also tend to co-occur with expressions of body language that suggest uncertainty, dissatisfaction, and discontinuity in conversations, e.g., pause, silence (line 2), shrugs (lines 3 and 5) and hesitates. This also includes an orthographic symbol of pauses or ellipsis like "…" (lines 9 and 10). These patterns are illustrated in sample concordance lines below.

This pattern of epistemic bundles and its frequent collocations suggests that Baumbach's screenplays are concerned with characters' discontinuities or negative emotions along with their uncertainty.

1. . He says <u>wearily</u>: MALCOLM *I don't know* what we are. 9 INT. CLAUDE'S

2. you must be. <u>Silence</u>. PAULINE *I don't know* where to begin. What can I tell

3. (<u>shrugs</u>) Maybe I'll try to see him *I don't know.* Things. when he's here. Life. HAROLD

4. are you so dressed up? DANNY *I don't know.* (so <u>annoyed</u> at Dad) I don't fucking

5. doesn't ultimately mean much. *I don't know,* ? L.J. L.J. (<u>shrugs</u>) I don't know either.

6. my eyes getting hollows? TIM *I don't know.* JOSH (<u>trying to be positive</u>)

7. our heads. PAULINE(<u>dismissive</u>) *I don't think* that's it. I just think it was

8. 114. HEDGE FUND DAVE *I don't care* JOSH (<u>dismissive</u>) I didn't expect

9. (trying to act cool) I guess<u>…I…I</u> *don't know what* … why that happened. SOPHIE It's okay

10. You know, Josh told me<u>…</u> *I didn't know* you miscarriaged. CORNELLA

Figure 10: Concordance lines of co-occurrence patterns of negative epistemic bundles

4.2.1.2 Attitudinal and Modality Stance Bundles

Desire

There are 17 desire lexical bundles in NBC, nine of which are negative, including the two most frequent *Desire* bundles *I don't want* and *don't want to.*

I don't want is the most common desire bundle (57 times), which characters mostly employ to distance themselves from other characters. Characters in NBC reject five sorts of desires with it, according to a concordance line assessment:

(1). having a conversation with someone: e.g., *I don't want to have this conversation with you; I don't want to hear from you.*

(2) Seeing someone: e.g., *I don't want to see you.*

(3) doing something as another's suggestion: e.g., *I don't want to do that; No, Margot, I don't want to count*

(4) tangible items: e.g., *I don't want money or anything; I don't want a bandage*

(5) *abstract items*: e.g., *I don't want to feel pressure from you*

On the other hand, the bundle *you don't want* is primarily used by egocentric characters to dismiss or show contempt to others, with dismissals being related with the character's needs to retain his or her privileged/dominant identities. For example, in *The Squid and the Whale*, Bernard's identities as a writer and a dominant father, are both endangered by his successful ex-wife's rising writing career. In context of this, he tries to downplay his son Walt's desire to become a professional tennis player, claiming that this is a job for philistines. For persuading purposes, *you don't want* occurs twice in two different scenes, as illustrated by the extracts below:

(1) BERNARD

You don't want to be a pro.

(pause)

I'm sure I lost my parking space so we're gonna have to drive around.

(2) BERNARD

Come on, *you don't want to* be a tennis pro.

FRANK

Why not?

BERNARD

It's not serious. I mean, McEnroe or Borg is an artist, it's like dance.

Connors has a brutish brilliance. But at Ivan's level...Ivan is fine, but he's not a serious

guy. He's a <u>philistine</u>.

FRANK

What's a philistine?

BERNARD

A guy who doesn't care about books or interesting films or things. Your mother's

brother Ned is also a <u>philistine</u>

FRANK

Then I'm a philistine.

(2005 *The Squid and the Whale*)

In Baumbach's cinematic world, we can see that he prefers one character to use the same expression in the different scenes for creating a sense of cohesion, like Matt's repeated "I'm sorry, I don't know why" (discussed in 4.2.1.1). Likewise, there are several instances in which characters repeatedly used the same desire bundles in different scenes, such as "I don't like", the most frequent exclusive desire bundle in NBC, which occurs 13 times in all five screenplays. The repetitions of "I don't like"

emphasizes the character's dislikes. It could be used by the same character in different scenes, for example:

(3) HAROLD (to DANNY)

I tell her, ***I don't like*** you when you drink. <u>She becomes a</u>

<u>different person.</u>

(*Meyerowitz Stories*, page 11)

(4)HAROLD (to WALT)

I tell Maureen, ***I don't like*** you when you drink. <u>You become a different person</u>, I say.

(exasperated)

(*Meyerowitz Stories*, page 69)

In the above extracts, Harold tells his sons, Danny and Walt, that he tried to persuade their stepmother Maureen to stop drinking, as depicted in two scenes from the early and later stages of the screenplay, respectively. As the plot develops, the conflicts between him and Maureen intensify, as evidenced by the second direct quotation of his phrase "you become a different person" in an exasperated tone.

Also, the character can repeat their dislikes in the same scene, as shown below,

(5) MARGOT

I don't like the girl and ***I don't***

like the way Malcolm looks at her.

This scene occurs when Margot's fiancé Malcolm is playing with another girl in the pool, and she indirectly expresses her concern over his infidelity. These concerns are foregrounded by the repeated *I don't like* as they create insecurities throughout the

extract, highlighting how Margot worries about her marriage and the potential infidelity of Malcolm, and thus contributing to her characterization as an insecure woman.

Obligation

Don't have to is the most common obligation bundle in NBC, with seven out of nine instances relating to speaking (e.g., *don't have to* 'tell', 'answer', 'respond'). This phrase is thereby often used to imply that someone divulged a secret, as shown in the following passage.

(1) PAULINE
You did. You ***don't have to*** tell

him everything.

MARGOT

He wants to know. If I don't tell

him, he figures it out.

In the example above, Pauline was accusing Margot of telling her secret to Margot's son, the underlying message of *you don't have to* is 'you should not'. While in other examples characters used this kind of expression to avoid further communication:

(2) CLAUDE

I masturbated last night. While

everyone was asleep I went into the

bathroom and did it.

MARGOT

You don't have to tell me, sweety.

When the teenager Claude talks about an embarrassing topic (masturbation) with his aunt Margot, she does not want a further discussion. You don't have to expresses her unwillingness to continue this topic. The vocative 'Sweety' is used for Claude, who just brought up an adult topic.

Prediction/intention

Twelve out of 13 bundles in this category contain 'going to'. The only exception is *I'll see you* as a routine when parting. The functions of these bundles differ depending on the subject; for example, when the first-person pronouns 'I' or 'we' are contained within the lexical bundle, the expressions frequently indicate a positive future. For example:

MATT

I'll see you soon!

Matt hangs up. Randy is talking to a guy with an open notebook. He turns to Matt and says decisively.

(2017 *The Meyerowitz Stories (New and Selected)*)

As for *you're going to,* it performs various functions, including: (1) reassuring someone that everything will be alright in the future (see extract 1), (2) warning someone that they are making a mistake (extract 2), or cursing someone (extract 3).

(1)

MARGOT

You're going to be fine.

(2007 *Margot at the Wedding*)

(2)

NICOLE

And now *you're going to* put Henry through this horrible thing so you can yet again get what you want.

(2019 *Marriage Story*)

(3) DANNY

I'm telling you, *you're going to* feel like crap.

(2017 *The Meyerowitz Stories (New and Selected)*)

It must be noted that within this sub-category, there are eight out of 15 bundles that are exclusive to NBC. The most frequent one is 'I was going to'. Out of its 10 instances, six are followed by verbs indicating 'speaking', such as 'say' 'call' 'ask' and 'warn'. In *The squid and the whale*, Joan used the phrases 'I was going to say' twice in the same scene, shown in the extracts below:

(1) JOAN

Umm...there was something else *I was going to say*...Oh, I ran into Celia, Lance's mother, on the street and she was telling me how wonderful she thinks you are. How polite and funny you are...

WALT

Uh huh.

(2) JOAN

I said, I know all those things about him already. But it's nice to hear it.

(pause)

I remember what else *I was going to say*. I wanted to know if you'd be interested in coming to dinner on Saturday because I'm having the Dicksteins over--

WALT

I'm going to a party on Saturday and I'm sleeping at Jeffrey's.

(2005 *The Squid and the whale*)

In these extracts, the repetition of *I was going to say* shows Joan's attempt to mend her relationship with her son, Walt, by introducing new topics to extend the conversations with him.

4.2.2 Referential Lexical Bundles

Referential expressions feature a total of 14 lexical bundle types, accounting for 14% of NBC's lexical bundles. They are divided into three sub-categories: identification, place, and time references. Like in MFC, place references are the most common type, accounting for 9 of the 14 bundle types. The majority of these place-referencing bundles state where the scenes are located, with no additional information. Unlike those in MFC, however, those scenes are largely of people having conversations, not moving. See the examples below:

(1) Josh films Jamie as he gets ***out of the car***. Tipper films the house.

(2) Josh stands ***in the doorway,*** he wears his bike helmet and holds his hat. Breitbart smiles.

There are four place referential bundles that occur in NBC, they are), *the other side of* (7 times), *a few feet away* (6 times), *in the passenger seat* (6), and *stands in the doorway* (6). Except for 'the other side of', the rest three bundles illustrate three typical scenes of Baumbach's film:

1. Two people in a near distance e.g., *She stands a few feet away from Pauline*
2. In driving scenes e.g., *Jim drives with Margot in the passenger seat.*
3. A person stands by the door and talks with someone inside.

Nevertheless, six bundles appear in both MFC and NBC, as shown in Table 4.2.2 below. It displays the noun collocates of shared referential bundles between the MFC and NBC[1]. It suggests some common location, vehicle (e.g., truck, bus), or event (e.g., cooking, talking) in Baumbach's and mainstream films. A collocate in the study suggested it is more likely to co-occur with certain bundles than other words. It is noteworthy that a collocate below could occur before or after the bundle in a span of five words. Thus, an example of collocation between *in front of him* and 'table' could be:

"She sits on the <u>table</u> directly ***in front of him***"

Table 4 Overlapping referential bundles in the corpora MFC and NBC, and their noun collocates

The Bundle	Noun collocates in MFC	Noun collocates in NBC
The back of the	Truck, room, bus, neck, boat. head	House, door, theatre, car
In Front of him	Table	Pancake, car
Other side of the	Room, door	Room, island, park
At the end of	Hall, hallway, door	Table, driveway
Out of the car	/	/
In the middle of	Nowhere, road, sentence	Street, cooking, talking

Another interesting point about Referential lexical bundles is found in the sub-category Identification/focus, in which all the bundles in this sub-group are in the negative form. It is found that these four bundles are mainly used when characters refute the others. Given the bundle *It's not a*, which is the only identification/focus bundle idiosyncratic to NBC, characters use it to refute other people's definitions of something, and then provide their own later. See the example below:

---

1. The collocates with MFC referential bundles were extracted using the 'collocate' function of *Antconc* and sorted by likelihood. Despite the limited occurrences in NBC, I manually selected the noun collocates.

(3) FRANK

When does Mom's story come out in the magazine?

WALT

It's not a magazine, it's a literary journal. Next month, I think.

According to the narrative, the terms "Magazine" and "literary journal" can be used interchangeably. The publication that contains Joan's (Frank's mother) story is a magazine devoted to creative writing. Walt's correction reveals his literary self-identity and his desire to establish patriarchal dominance within the family (by correcting his son).

4.2.3 Special-function Lexical Bundles

A number of lexical bundles in this category occur only in NBC, some of which are discussed in detail in this section. First, the bundle *v.o. she's* refers to the frequent voiceovers in Baumbach's recent films, particularly *Marriage Story* (five out of six instances occur in this film). Charlie, Nicole's ex-husband, uses the bundle to highlight Nicole at the beginning of the film. As the concordance lines shown in Figure 4.9.

. She sweeps up her own hair. CHARLIE (	***V.O.) She's***	always inexplicably brewing a
reading, but still managing to read. CHARLIE (	***V.O.) She's***	dancer. Infectious. She makes
towel and tries it that way. No luck. CHARLIE (	***V.O.) She's***	amazing at opening jars because
PAUSE mid-lift before it comes off. CHARLIE (	***V.O.) She's***	brave. After that movie, All Over
she's on a stage, rehearsing for a play. CHARLIE (	***V.O.) She's***	my favorite actress. INT.

Figure 11: Concordances for v.o.she's in Marriage Story

All of his comments on Nicole are positive, which rewinds their precious marriage times. Those words mainly contrast with the later plots in the film, which contain lots of quarrels. After the ex-couple pursue a divorce lawsuit, they accuse each other for destroying the martial life. In these scenes, Charlie's comments on Nicole are totally contrastive from the ones in the beginning, see the examples below:

(1) 'All your best acting is behind you. You're back to being a **HACK**.'
(2) 'Life with you was JOYLESS.'
(3) 'Every day I wake up and I hope you're dead! Dead, like if I could guarantee Henry would be okay, I'd hope you'd get an illness, and then get hit by a car and die!'

Next, while the act of looking is common in MFC, none is found in NBC. Instead, the bundle *kisses him on the* emerges, five of which are used to depict an elderly lady kissing her younger family member. The frequency of these scenes suggested the mother's too-close bond with her son, stepson, and son-in-law. The relationship of three characters are illustrated in figure 4.9 below:
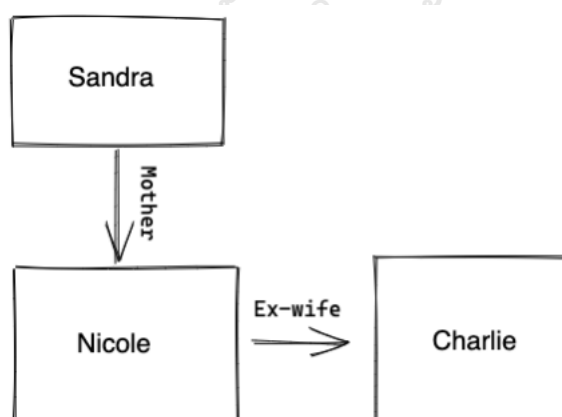

*Figure 12: **The relation of characters in** **Marriage Story***

(1) Sandra *kisses him on the* lips. He lifts her up. Nicole watches impatiently.

(Context: Charlie comes to visit his ex-wife Nicole and her Mother Sandra, then Sandra kissed Charlie)

The man Sandra was kissing in excerpt 1 is her ex-son-in-law, and it's worth noting that Baumbach did not shoot the kissing sequence (perhaps because he did not want to challenge the audience's moral acceptance). Sandra had just questioned her daughter Nicole, in the previous scene if she and Claude were still sharing the same bed. It ties along with the infantile characteristics discussed in Chapter 2.

4.2.4 Discourse Organizers

Discourse organizing bundles are the least common functional category in NBC, accounting for only 6% of all lexical bundles. Five were assigned to the Topic introduction/focus function sub-category, while the other one was assigned to the Conditions function.

I'm telling you is the most common bundle in this category, and it is worth noting that it is the catchphrase of the character Danny Meyerowitz (from *Meyerowitz Stories*, in which he said it eight times). Danny uses it to offer advice ('I'm telling you, I've got some nice recommendations.') and encouragement ('I'm telling you, you're going to meet a lot of wonderful, interesting new people.') in seven out of eight cases. Through these expressions, Baumbach attempted to create a warm-hearted character, which contrasts with other nonchalant, egocentric male characters in NBC. The other four examples of this bundle in NBC, on the other hand, tends to be have a face-threatening pragmatic force. For example, first, it is used to conclude that a lady's relationship with her boyfriend will not last long:

BERNARD

Ivan's not a serious possibility for your mother.

FRANK

I think he is.

BERNARD

I'm telling you he isn't, Frank.

You'll see. He won't last.

(2005 *The Squid and the Whale*)

In other cases, the bundle is used to insist someone's intention (e.g., '*I'm telling you* what I really feel'), emphasize omeone's attitude (e.g., 'No, don't. *I'm telling you*) and threaten someone (e.g., '*I'm telling you*, you're going to feel like a crap').

4.3 Comparison between lexical bundles in MFC and NBC

This section aims to answer research question 3 regarding differences between MFC and NBC. Overall, the distributional and functional analysis display similarities between NBC and MFC. To begin with, the distribution patterns of functional categories in both corpora are similar, with stance bundles topping the list, followed by referential bundles, special functional bundles, and discourse organizers, respectively. Second, despite their differences in size, both corpora contain 42 identical lexical bundle types. This shows that both Baumbach's and mainstream films share some core lexical bundles. The similarities can reflect and be attributed to the register of film scripts. At the same time, there are a number of lexical bundles that occur only in one of the corpora, listed below:

Table 5 The bundles that only occur in MFC

1. Don't know how
2. I think it's
3. I need you to
4. Can't help but
5. I can't do
6. Get out of here
7. Get out of the
8. That's what I

9. *You know what I*
10. *That's why I*
11. *The rest of the*
12. *In front of the*
13. *The edge of the*
14. *The end of the*
15. *The top of the*
16. *The side of the*
17. *On the other side*
18. *Out of the way*
19. *The center of the*
20. *Other side of the*
21. *In front of a*
22. *At the top of*
23. *Out of the room*
24. *The front of the*
25. *In the back of*
26. *For the first time*
27. *I'm so sorry*
28. *Are you talking about*
29. *What are you talking*
30. *What do you think*
31. *How do you know*
32. *Look at each other*
33. *Turns back to the*
34. *Looks up at the*
35. *Looks down at the*
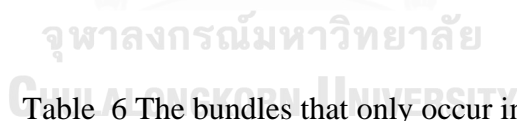
Table 6 The bundles that only occur in NBC

1. *don't know why*
2. *it doesn't matter*
3. *I don't see*
4. *Don't know if*
5. *'t know why I*
6. *Don't think I*
7. *Don't think you*
8. *I think he's*
9. *I think I'm*
10. *We don't know*
11. *Didn't know you*
12. *Don't know that*
13. *Don't know where*

14. I didn't realize
15. I don't remember
16. I just don't
17. I think that's
18. I thought it was
19. I thought you were
20. I don't like
21. 't want to do
22. Didn't want to
23. I'd love to
24. I want to be
25. 't want to be
26. Want to be a
27. Want to do it
28. Do I have to
29. I was going to
30. Going to have to
31. I'll see you
32. 'm not going to
33. 're not going to
34. They're going to
35. He's going to
36. 'm going to go
37. You're not going
38. Don't have a
39. I didn't get
40. I'm trying to
41. I'm telling you
42. You know what I
43. What I'm doing
44. I have to say
45. It's hard to
46. 's not what I
47. That's not what
48. But it's not
49. The other side of
50. A few feet away
51. In the passenger seat
52. Stands in the doorway
53. Nice to meet you
54. What you want to
55. Why didn't you
56. Kisses him on the
57. No I don't
58. V.o. she's

Here are differences between MFC and NBC. First, taking a quantitative perspective (as shown in table 4.3 below), it is found that there are more tokens of special function bundles than referential bundles in NBC. This suggests that, unlike MFC, Baumbach's scripts rely less on location and character movement descriptions. Second, NBC has noticeably more stance bundles than other bundle categories, most of these bundles occurred in conversations rather than narrations, which suggests that scripts in NBC rely heavily on characters' dialogues. Furthermore, 82.61 percent of NBC's special functional bundles are spoken bundles, such as inquiry ('what are you doing?') and deny ('no, I don't'); compared to MFC, in its special functional bundles, 69.73 percent are spoken bundles, while narration bundles, such as those in action subgroups, continue to play the importation roles.

Table 7 Frequencies of lexical bundles in MFC and NBC

	MFC	NBC
Stance Bundles	3,894 (45.93%)	953 (77.73%)
Referential Bundles	2,895 (34.14%)	108 (8.81%)
Special Functions	1,460 (17.22%)	115 (9.38%)
Discourse Organizers	230 (2.71%)	50 (4.08%)
Total	**8,479**	**1,226**

The stance bundles are predominant in both corpora and mainly express character's uncertainty, such as insufficient information, confusion, refusals, etc. Despite this, NBC has a greater variety of lexical bundles than MFC, particularly for stance bundles (68 to 35 types), even though is it much smaller in size. This suggests that Noah Baumbach's film scripts are heavily verbal and the plots are especially tied to conversations, rather than actions, as observed by film critics. Moreover, the dominance of the stance bundles and its great variety in NBC may be due to its

conversational styles. First, the characters' long and meandering dialogues create plenty of opportunities for stance bundles. Second, the characters like detailing their thoughts, featuring interruptions, avoidances, and hesitations (which bundles like *I don't know*, *I don't think* occur regularly).

In terms of desire and prediction/intention subcategories, while NBC is more constrained, generally focusing on the interpersonal interactions between the wants and intentions, MFC is more focused on the action-oriented aspects. In the collocates, it is observed that some insights like the frequent collocation of 'I don't want to+hurt' or 'I'm going to kill'.

In addition, since mainstream films rely heavily on scene transitions, MFC has a greater variety of place reference types. 86.84 per cent of MFC referential bundles pertain to place referential bundles, highlighting common vehicles such as 'bus,' 'boat,' and 'truck', and common settings such as 'desert,' 'room'. On the other hand, NBC's identification/focus bundles account for a greater proportion than MFC's (1.96 percent versus 34.26 percent), and bundles such as "it's not a" are typically used when characters argue with others.

Special functional bundles comprise 17.22 percent of MFC occurrences, with the subgroup 'action' comprising the majority. Those bundles associated with 'actions' are associated with their function in mainstream cinematic discourse, which includes both dialogue and action. Unlike MFC, NBC has more special-function bundles and the majority of them are spoken language, thereby reflecting NBC's heavily dialogic nature.

Discourse organizers have a low frequency in both corpora and perform similar functions in each corpus. Despite the fact that some catchphrases in certain films, such as Danny's frequent use of "I'm telling you" in Meyerowitz Stories, contribute to his characterization as a warm-hearted person. In general, discourse organizers primarily connect and project the conversational structures, while saving time by omitting background information.

CHAPTER 5 CONCLUSION

This chapter concludes the study with summary of findings, contributions and recommendations for future studies.

5.1 Summary of the Study

The present study investigated the lexical bundles in Noah Baumbach scripts and those in mainstream American films from a corpus linguistic perspective. It categorized the lexical bundles and identified their functions through Biber et al. (2004) and Biber (2006) frameworks. The researcher analyzed the lists of lexical bundles extracted from the two corpora using the criteria stated in chapter 3 section 3.3 to identify their functions. These results met the first two objective of the study, which were to look at the different functional types of lexical bundles identified in MFC and NBC. The study's third objective was to examine and contrast the functional forms of lexical bundles found in each corpus. The answers to each research question are summarized below.

5.1.1 Answers to Research Question 1

While mainstream film scripts have been described at the macro level in terms of its structural components, the findings of this study have demonstrated that American mainstream film scripts are mainly characterized by common spoken lexical bundles. This may account for previous observations of mainstream films as being formulaic (e.g., King, 2005) and also for their popularity since they contain a large number of dialogic expressions relatable to real-life spoken language which, however, are used to create a fictional text world by highlighting conflicts in stories, expressing characters' inner feelings and advancing the plots.

At the same time, American mainstream film scripts are also essentially made up of formulaic expressions that describe actions, locations, and movements. These groups of lexical bundles help enrich the texts with visualized contextual information and create cohesive meanings in the texts. If film scripts are considered as a discourse, looking at communicative acts and individual participants in the communication, the emerging lexical bundles serve to inform and facilitate film production teams, including actors, directors and administrators, allowing them to interpret cinematic texts and prepare for the production. Overall, the bundles in MFC primarily focus on dialogues and scene-building.

5.1.2 Answers to Research Question 2

The second research question addresses the functional types of lexical bundles in Baumbach's film scripts. It is found that the majority of functional categories lie in the stance bundle category, which focuses on 'problematic interactions' such as confrontations, discontinuity in conversations, accusations, etc. This, again, points to the characterization of indecisive characters in Baumbach's film. The negative stance expressions came with their uncertainty, anxiety or lack of understanding. Negative identification/focus bundles such as *it's not a* can indicate a character's attitude and contribute to their characterization, whereas the majority of referential bundles indicate the common location or physical distances between actors. In addition to their primary functions, certain discourse organizers and special functional bundles, such as "I'm telling you" or "v.o.she's," can also reflect character emotion/personality (e.g., topic introductions and script jargon for voiceover). Overall, NBC's primary focus is on its characters; the lexical bundles serve to highlight dialogues and their functions in characterization and character development.

5.1.3 Answers to Research Question 3

The results suggest that NBC and MFC share a similar distributional pattern in functional categories. Also, there are 37 overlapping bundles between the two

corpora. Those bundles suggest similar patterns contained in common film scripts, such as common spoken phrases like *I don't know*, *I don't want*; and place references in narration parts like *in the middle of*, *in front of him,* etc.

According to the occurrences of these bundles, NBC is more unbalanced than MFC, which heavily relies on bundles of stances that occur in dialogues. MFC has multiple functions that are served by referential bundles and special functional bundles. Many of those components contribute to the narration aspect. It can be observed that in NBC, the dialogue parts are more prominent than the narration parts whereas in MFC scripted narration requires extensive attention as well. This is probably because mainstream film production requires more investment, teamwork, and pre- and post-production efforts, so everything must be meticulously planned at the screenplay stage.

5.2 Contributions and implications of the Study

With a corpus linguistics method, the study shed light on cinema and telecinematic discourse by investigating the film scripts with a lexical-bundle approach. Second, it contributes to the growing body of research about American independent filmmaking and how it compares to mainstream films. The research likewise adds to the linguistic study of film scripts by taking a closer look at the entire script rather than just the spoken parts.

Another contribution of the present study is its approach in combining the fields of telecinematic discourse and telecinematic stylistics. The focus on American mainstream film scripts brings up common linguistic patterns that may be seen as a linguistic norm for the register, with which film scripts of a particular writer like Noah Baumbach can be compared. This in turn throws light on his distinctive style, i.e., his deviation from the norm. The study therefore responds to O'Meara's (2018a) suggestion on the conduct of research focusing on the screenwriters' scripted verbal style. In addition, it adds to the growing corpus-assisted telecinematic discourse

literature (Bednarek, 2011; Jautz & Minow, 2020; Pavesi, 2020; Reichelt, 2020) by examining both narration and dialogue in the scripts.

Finally, because there are so many common spoken lexical bundles, the study supports the widespread use of movies as a motivational tool in language learning. Teachers could use the corpus-assisted method to discuss the movie dialogue because of the film's popularity among pupils. By exposing the high-frequency bundles with related filmic speeches, the phraseological application in language instruction could lead students to realize the usage in real life.

5.3 Limitations and suggestions for further Study

In spite of this study's contributions, the manually compiled corpora also had some limitations. Due to NBC's small size and limited range of detected lexical bundles, bundles that occur in only two films may not be indicative of the writer's style.

While the study sheds light on general linguistic patterns shared across mainstream film scripts and those in Noah Baumbach's, a number of points are worth further study to gain a thorough understanding of the film discourse. First, it would be useful to pursue a contrastive study on spoken lexical bundles in film scripts and those in real conversations. Although the present study has shown that everyday stance bundles predominate in film scripts, which are fictional, it would be interesting to examine further to what extent they are used differently, both quantitatively and qualitatively, from those in real conversations. Likewise, while it has been shown that descriptive bundles in narration, especially place-referential and action lexical bundles, have quantitative significance in film scripts, the study only have slightly touched on their qualitative aspects due to space limits. It would be interesting to further investigate their patterns and functions as well as their relationship with spoken lexical bundles in film scripts.

At the same time, a focus of the study on Noah Baumbach's film scripts can be applied to the verbal styles of other writer-directors, such as Hong Sang-soo, Philip

Garrel, Eric Rohmer, and Whit Stillman, among others, who integrate dialogue and images with care. Baumbach's other screenplays, particularly those he co-wrote with Greta Gerwig, such as *Frances Ha*, could be examined in depth to be compared with findings of the present study. This is so that we can see whether the collaboration could affect this screenplay style. Also, the scripts influenced by Greta's style are also the legacy of the 'mumblecore' cinematic movement, which contrasts optimistic spirits with emotional dysfunctions.

Although there are still several other points that need to be explored to describe the discourse of film scripts, the study is hoped to illustrate the value of a multidisciplinary approach, which brings together corpus linguistics, discourse studies, and film studies, to a vital text in the film industry, a form of entertainment dear to people across the globe.

REFERENCES

Abrams, N. (2012). The new Jew in film: Exploring Jewishness and Judaism in contemporary cinema. 1-258.

Altenberg, B. (1998). Connectors and sentence openings in English and. *Corpora and cross-linguistic research: Theory, method, and case studies*(24), 115.

Androutsopoulos, J. (2012). Introduction: Language and society in cinematic discourse. *Multilingua*, *31*, 139-154. https://doi.org/10.1515/multi-2012-0007

Bacon, H. Synthesizing approaches in film theory. *available at www*.

Baker, A. (2013, 2013). *Frances Ha: The Green Girl*.

Balázs, B. (2010). *Béla Balázs: Early film theory: Visible man and the spirit of film* (Vol. 10). Berghahn Books.

Bateman, J. A., & Schmidt, K.-H. (2012). *Multimodal film analysis: how films mean*. Routledge.

Baumbach, N. (1995). *Kicking and Screaming* [Comedy, Drama, Romance]. Trimark Pictures, Castleberg Productions, Sandollar Productions.

Baumbach, N. (2005). *The Squid and the Whale* [Comedy, Drama]. Samuel Goldwyn Films, Sony Pictures International, Destination Films.

Baumbach, N. (2007). *Margot at the Wedding* [Comedy, Drama]. Scott Rudin Productions.

Baumbach, N. (2015). *While We're Young* [Comedy, Drama, Mystery]. IAC Films, Scott Rudin Productions.

Baumbach, N. (2017). *The Meyerowitz Stories (New and Selected)* [Comedy, Drama]. Netflix, IAC Films, Gilded Halfwing.

Baumbach, N. (2019). *Marriage Story* [Comedy, Drama, Romance]. Heyday Films, Netflix.

Bednarek, M. (2010). *The Language of Fictional Television: Drama and Identity*. Bloomsbury Publishing.

Bednarek, M. (2011). The stability of the televisual character. In *Telecinematic Discourse* (pp. 20).

Bednarek, M. (2012). "Get us the hell out of here": Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics*, *17*(1), 35–63.

Bednarek, M. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge University Press.

Bednarek, M., Veirano Pinto, M., & Werner, V. (2021). Corpus approaches to telecinematic language. *International Journal of Corpus Linguistics*, *26*(1), 1-9. https://doi.org/10.1075/ijcl.00034.int

Bellow, S. (1964). *Herzog. 1964*. New York, Viking Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, *25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.

Bihlmeyer, J. (2005). The (Un)Speakable FEMININITY in Mainstream Movies: Jane Campion's "The Piano". *Cinema Journal*, *44*(2), 68-88.

Bousfield, D. (2010). Researching impoliteness and rudeness: Issues and definitions. In *Interpersonal Pragmatics* (Vol. 6, pp. 101-136).

Bousfield, D., & McIntyre, D. (2018). Creative linguistic impoliteness as aggression in Stanley Kubrick's Full Metal Jacket. *Journal of Literary Semantics*, *47*. https://doi.org/10.1515/jls-2018-0003

Brigham, A. (2021, 2021). *A Tribute to Noah Baumbach: A Modern Auteur - Hollywood Insider*.

Brody, R. (2015, 2015). *Noah Baumbach's New Movie "While We're Young" | The New Yorker*.

Brueggemann, T., & Brueggemann, T. (2019). How Much Would 'Roma' Have Grossed With a Traditional Release? We Have An Answer. *IndieWire*.

Bucholtz, M., & Hall, K. (2005). Identity and Interaction: A Sociocultural Linguistic Approach. *Discourse Studies*, *7*, 585-614. https://doi.org/10.1177/1461445605054407

Calvo, C. (1990, 1990). *Power relations and fool-master discourse in Shakespeare: a discourse stylistics approach to dramatic dialogue*.

Casey Benyahia, S., & Mortimer, C. (2012). *Doing Film Studies* (0 ed.). Routledge.

Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning and Technology*, *14*.

Chew, K. (2019). On war and cuteness: the utopian politics of Disney's Zootopia. *Screen*, *60*(4), 567-586. https://doi.org/10.1093/screen/hjz043

Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*.

Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, *14*(1), 29-59. https://doi.org/10.1075/ijcl.14.1.03cul

Duralde, A. (2019, 2019). *'Marriage Story' Film Review: Scarlett Johansson and Adam Driver Break Apart in Noah Baumbach's Devastating Drama*.

Erbland, K., & Erbland, K. (2017). 'The Meyerowitz Stories (New and Selected)': Dustin Hoffman's Bad Dad Anchors Sprawling Family Saga. *IndieWire*.

Feller, M. (2020, 2020-02-08 04:29:00). *'Marriage Story' Is a Heartbreaking Film About Love and Divorce. Is It Also Based on Real Life?*

Film Studies. (2012, 2012-03-19).

Forchini, P. (2020). Marvel and DC studios compared. *The Routledge Handbook of Corpus Approaches to Discourse Analysis*.

Foundas, S. (2010, 2010). *Review: Greenberg*.

Freddi, M. (2011). A phraseological approach to film dialogue: Film stylistics revisited. *Yearbook of Phraseology*, *2*(1), 137-162. https://doi.org/10.1515/9783110236200.137

Gemser, G., Van Oostrum, M., & Leenders, M. A. A. M. (2007). The impact of film reviews on the box office performance of art house versus mainstream motion pictures. *Journal of Cultural Economics*, *31*(1), 43-63. https://doi.org/10.1007/s10824-006-9025-4

Genzlinger, N. (2019, 2019-04-05). Jonathan Baumbach, Novelist With an Experimental Bent, Dies at 85. *The New York Times*.

Gregori Signes, C. (2020). Victim-naming in the murder mystery TV series Twin Peaks: a corpus-stylistic study. *Series - International Journal of TV Serial Narratives*, Vol. 6 No. 2 (2020): Dialogues with Technology. https://doi.org/10.6092/ISSN.2421-454X/11218

Gregory, M. (1967). Aspects of varieties differentiation. *Journal of linguistics*, *3*(2), 177–198.

Grice, H. P. (1975). Logic and conversation. In (pp. 41 - 58). Brill.

Grobar, M., & Grobar, M. (2017, 2017-11-21T22:12:58+00:00). *'The Meyerowitz Stories' Director Noah Baumbach On His Personally Rooted Penchant For New York Stories*.

Handler, R. (2019, 2019-12-10T15:29:30.354-05:00). *Every Noah Baumbach Couple, Ranked by Sheer Misery*.

Hellerman, J. (2019, 2019-12-20T08:45:00-08:00). *How Does Noah Baumbach Create Realistic Dialogue?*

Hill, J., & Gibson, P. C. (2000). *Film studies: critical approaches*. Oxford University Press.

Hill, R. (2008). The New Wave Meets the Tradition of Quality: Jacques Demy's *The Umbrellas of Cherbourg*. *Cinema Journal*, *48*(1), 27-50. https://doi.org/10.1353/cj.0.0062

Hoffmann, C., & Kirner-Ludwig, M. (2020). *Telecinematic Stylistics*. Bloomsbury Academic.

Horn, J. (2010, 2010). *Noah Baumbach on Why Greenberg Isn't As Depressing As You'd Think*.

Hornaday, A. (2015, 2015-08-20T10:20-500). 'Mistress America' marks a return to comedy's screwball golden age(s). *Washington Post*.

Hsu, J. (2008). The Secrets of Storytelling: Why We Love a Good Yarn. 6.

Introduction: problems of definition and the discourse of American independent cinema. (2017). In Y. Tzioumakis (Ed.), *American Independent Cinema: Second Edition* (2 ed., pp. 1-14). Edinburgh University Press.

Jautz, S., & Minow, V. (2020). 'Drucilla, we : The formulaic nature of problem-oriented talk in soap operas. In C. Hoffmann & M. Kirner-Ludwig (Eds.), *Telecinematic Stylistics* (1 ed., pp. 63-86). Bloomsbury Academic.

Jenkins, H. (2006). *Fans, bloggers, and gamers: Exploring participatory culture*. nyu Press.

Jones, K. (2016, 2016). *The Squid and the Whale: 4 Way Street*.

Kemp, P. (2015, 2015). *Film of the week: While We're Young | Sight & Sound*.

King, G. (2005). *American independent cinema*. I B Tauris.

Kohn, E. (2013, 2013). *Charming 'Frances Ha' Reteams Greta Gerwig and Noah Baumbach For a Deceptively Simple Portrait of Post-College Woe | IndieWire*.

Lambrou, M., & Stockwell, P. (2010). *Contemporary Stylistics*. A&C Black.

Lane, A. (2017, 2017). *"The Meyerowitz Stories (New and Selected)" and "Wonderstruck"*.

Laplante, J. (2013). Cruel & Tender: An Appreciation of Noah Baumbach. *Miami Film Festival*.

Lee, K. K. H. (2018). "Watashi-tachi wa ningen da!": A Corpus-Assisted Analysis of a Non-Human Character in the Anime Series 'From the New World' – New Voices in Japanese Studies.

Margot at the Wedding (2007).

Marie, M. (2003). *The French New Wave* (1 ed.). John Wiley & Sons, Ltd.

Mast, G. (1990). Woody Allen: The Neurotic Jew as American Clown. In S. B. Cohen (Ed.), *Jewish wry : essays on Jewish humor* (pp. 125-141). Wayne State

University Press.

Mayer, S. (2017). Girl Power: Back to the Future of Feminist Science Fiction with Into the Forest and Arrival. *Film Quarterly*, *70*(3), 32-42. https://doi.org/10.1525/fq.2017.70.3.32

McIntyre, D. (2012). Prototypical Characteristics of Blockbuster Movie Dialogue: A Corpus Stylistic Analysis. *Texas Studies in Literature and Language*, *54*(3), 402-425. https://doi.org/10.7560/TSLL54307

Meyer, J., Song, R., & Ha, K. (2016). The effect of product placements on the evaluation of movies. *European Journal of Marketing*, *50*(3/4), 530-549. https://doi.org/10.1108/EJM-12-2014-0758

Movie blockbusters. (2013). (J. Stringer, Ed.). Taylor & Francis.

Mulvey, L., & En, P. W. (2008). From cinephilia to film studies. In *Inventing film studies* (pp. 217–232). Duke University Press.

Myers, E. (2015, 2015). Review: While We're Young. *Film Comment*.

Nerdwriter1. (2017, 2017-12-01). *What Realistic Film Dialogue Sounds Like*.

Neupert, R. (2007). A history of the French new wave cinema.

Newman, M. Z. (2011). *Indie: An American Film Culture*. Columbia University Press.

Noah Baumbach's Scores. (2021, 2021).

Noah Baumbach Awards.

Noah Baumbach shows 'Greenberg' how he sees it. (2010, 2010-03-21T07:00:00).

Noah Baumbach: Independent Film's Dramedy Mastro - Indie Film Hustle. (2019, 2019-12-30T01:00:12-08:00).

Nursanti, R. Y. (2015). *A Pragmatic Analysis of Maxim Flouting in Hunger Games Movie*

Nuryani, E. (2016). A Pragmatic Analysis of Politeness Features of Criticism in Joseph Mcginty's This Means War.

O'Meara, J. (2018). *Engaging Dialogue: Cinematic Verbalism in American Independent Cinema*. Edinburgh University Press.

O'Meara, J. (2014). A Shared Approach to Familial Dysfunction and Sound Design: Wes Anderson's Influence on the Films of Noah Baumbach. In P. C. Kunze (Ed.), *The Films of Wes Anderson: Critical Essays on an Indiewood Icon* (pp. 109-124). Palgrave Macmillan US.

O'Meara, J. (2018). Neurotic and Going Nowhere: Comedy and the Contemporary Jewish American Male. In S. Dibeltulo & C. Barrett (Eds.), *Rethinking Genre in Contemporary Global Cinema* (pp. 49-61). Springer International Publishing.

Parker, I. (2013, 2013). *Happiness*.

Pavesi, M. (2020). 'I shouldn't have let this happen' : Demonstratives in film dialogue and film representation. In C. Hoffmann & M. Kirner-Ludwig (Eds.), *Telecinematic Stylistics* (1 ed., pp. 19-38). Bloomsbury Academic.

Pepperell, R., & Punt, M. (2006). *Screen Consciousness: Cinema, mind and world*.

Pfefferman, N. (2005, 2005-10-13T12:00:00+00:00). *Prickly Fathers, Rebellious Sons*.

Piazza, R., Bednarek, M., & Rossi, F. (2011). *Telecinematic Discourse: Approaches to the Language of Films and Television Series*. John Benjamins Publishing Company.

'A Quiet Place Part II' Makes Serious Memorial Day Noise With A $48.4 Million Three-Day Bow; 'Cruella' Is Solid In Second With $21.3 Million. (2021, may 30 2021).

Reichelt, S. (2020). Innovation on screen: Marked affixation as characterization cue in Buffy the Vampire Slayer. https://doi.org/10.1075/ijcl.00038.rei

Roberts, G., & Wallis, H. (2001). *Introducing film*. Arnold.

Rosch, E., & Lloyd, B. B. (1978). Cognition and categorization.

Sarris, A. (1963). The Auteur Theory and the Perils of Pauline. *Film Quarterly*, *16*(4), 26-33. https://doi.org/10.2307/3185951

Scott, A. O. (2013, 2013-05-16). If 27 Is Old, How Old Is Grown Up? *The New York Times*.

Scott, M., & Tribble, C. (2006). *Textual patterns: key words and corpus analysis in language education / Mike Scott and Christopher Tribble*. John Benjamins Pub.

Sharf, Z., & Sharf, Z. (2021). Noah Baumbach Signs Exclusive Deal with Netflix, Next Film 'White Noise' to Shoot in 2021. *IndieWire*.

Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*, *49*, 100946. https://doi.org/10.1016/j.jeap.2020.100946

Siricharoen, A., & Wijitsopon, R. (2020). A Corpus-based Comparative Study of Lexical Bundles in Authentic and Textbook English Business Emails. *LEARN Journal: Language Education and Acquisition Research Network*, *13*(2), 41-63.

Statham, S. (2015). 'A guy in my position is a government target ... You got to be extra, extra careful: Participation and strategies in crime talk in The Sopranos. *24*(4), 322-337. https://doi.org/10.1177/0963947015605442

Sutton, G. (2015). An Ongoing Conversation: Overlapping Dialogue as Narration Strategy.

Takala, H. R. (1994). *The stylistics of resolution in virginia woolf's mrs dalloway and to the lighthouse* Theses (Dept. of English)/Simon Fraser University].

Taubin, A. (2015, 2015). *Amy Taubin on Noah Baumbach's Mistress America*.

Teo, S. (2019). The Chinese film market and the Wolf Warrior 2 phenomenon. *Screen*, *60*(2), 322-331. https://doi.org/10.1093/screen/hjz017

Thompson, K. (2003). *The Classical Hollywood Cinema : Film Style and Mode of Production to 1960*. Routledge.

Truffaut, F. (1957). Vous êtes tous témoins dans ce procès: le cinéma français crève sous les fausses légendes. *Article paru en*, 322–339.

Tzioumakis, Y. (2006). *American Independent Cinema*. Edinburgh University Press.

Vivarelli, N., & Vivarelli, N. (2019). 'The Irishman' Is Netflix's Biggest Theatrical Release at Home and Abroad. *Variety*.

Wilkins, K. (2021). Creative resistance: Noah Baumbach's literary filmography. *Textual Practice*, *35*(10), 1629-1646.

Williams, L., & Gledhill, C. (2000). *Reinventing film studies*.

Zago, R. (2020). Film discourse. In *The Routledge Handbook of Corpus Approaches to Discourse Analysis* (pp. 168–182). Routledge.

Zuckerman, E. W., & Kim, T. Y. (2003). The critical trade-off: identity assignment and box-office success in the feature film industry. *Industrial and Corporate Change*, *12*(1), 27-67. https://doi.org/10.1093/icc/12.1.27

APPENDICES

Appendix A The Films Selected in Noah Baumbach Corpus

The Squid and the Whale (2005) Comedy Drama

Margot at the Wedding (2007) Comedy Drama

While We're Young (2015) Comedy Drama Mystery

The Meyerowitz Stories (New and Selected) (2017) Comedy Drama

Marriage Story (2019) Comedy Drama Romance

Appendix B. The Films Selected in Mainstream Film Corpus

Table 8 The Films Selected in Mainstream Film Corpus

Film Title	Rank of the Year	Year	Genre
Star Wars Episode III	1	2005	Action Adventure Fantasy Sci-Fi
Harry Potter and Goblet of Fire	2	2005	Adventure Family Fantasy Mystery
War of the worlds	3	2005	Adventure Sci-Fi Thriller
The Chronicles of Narnia	4	2005	Adventure Family Fantasy
Wedding Crashers	5	2005	Comedy Romance
The Pacifier	17	2005	Action Comedy Drama Family
Flight Plan	21	2006	Action Comedy Drama Family
Pirates of the Caribbean: Dead Man's Chest	1	2006	Action Adventure Fantasy
Happy Feet	7	2006	Adventure Animation Comedy Family Music Musical Romance
MI3	12	2006	Action Adventure Thriller
The Devil Wears Prada	14	2006	Comedy Drama
The Departed	15	2006	Crime Drama Thriller
Shrek 3	2	2007	Adventure Animation Comedy Family Fantasy

Pirates of the Caribbean: at World's End	4	2007	Action Adventure Fantasy
The Bourne Ultimatum	6	2007	Action Mystery Thriller
Ratatouille	8	2007	Adventure Animation Comedy Family Fantasy
Wild Hogs	11	2007	Action Adventure Comedy
American Gangster	19	2007	Biography Crime Drama
Disturbia	34	2007	Crime Drama Mystery Thriller
Dark Knight	1	2008	Action Crime Drama Thriller
Hancock	4	2008	Action Drama Fantasy
Wall.E	5	2008	Adventure Animation Family Romance Sci-Fi
Kung Fu Panda	6	2008	Action Adventure Animation Comedy Family
Twilight	8	2008	Drama Fantasy Romance
Sex and the City	11	2008	Comedy Drama Romance
Harry Potter and the Half-Blood Prince	2	2009	Action Adventure Family Fantasy Mystery
Up	3	2009	Adventure Animation Comedy Family
Twilight: New Moon	4	2009	Adventure Drama Fantasy Romance

Avatar	5	2009	Action Adventure Fantasy Sci-Fi
The Hangover	6	2009	Comedy
Blind Side	10	2009	Biography Drama Sport
Toy Story 3	2	2010	Adventure Animation Comedy Family Fantasy
Inception	6	2010	Action Adventure Sci-Fi Thriller
Harry Potter and the Deathly Hallows: Part 1	7	2010	Adventure Family Fantasy Mystery
Despicable Me	8	2010	Animation Comedy Crime Family Fantasy
How to Train Your Dragon	10	2010	Action Adventure Animation Family Fantasy
Social Network	31	2010	Biography Drama
The Town	32	2010	Crime Drama Thriller
Harry Potter and the Deathly Hallows: Part 2	1	2011	Adventure Drama Fantasy Mystery
Pirates of the Caribbean: On Stranger Tides	5	2011	Action Adventure Fantasy
Fast Five	6	2011	Action Adventure Crime Thriller
Cars 2	7	2011	Adventure Animation Comedy Family Sci-Fi Sport
Thor	8	2011	Action Adventure Fantasy
The Help	11	2011	Drama
The Avengers	1	2012	Action Adventure

			Sci-Fi
The Dark Knight Rises	2	2012	Action Adventure
The Hunger Games	3	2012	Action Adventure Sci-Fi Thriller
Skyfall	4	2012	Action Adventure Thriller
Ted	9	2012	Comedy
Lincoln	19	2012	Biography Drama History War
Despicable Me 2	3	2013	Adventure Animation Comedy Crime Family Fantasy Sci-Fi
Monster University	5	2013	Adventure Animation Comedy Family Fantasy
Frozen	6	2013	Adventure Animation Comedy Family Fantasy Musical
Gravity	7	2013	Drama Sci-Fi Thriller
The Croods	14	2013	Adventure Animation Comedy Family Fantasy
The Butler	25	2013	Adventure Biography Crime Drama Thriller
Captain Philips	28	2013	Biography Drama
Maleficent	6	2014	Action Adventure Family Fantasy Romance
X-Men: Days Of Future Past	7	2014	Action Adventure Sci-Fi Thriller
Dawn Of The Planet Of The Apes	8	2014	Action Adventure

			Sci-Fi Thriller
Godzilla	11	2014	Adventure Drama Sci-Fi
Interstellar	15	2014	Action Biography Drama War
Gone Girl	17	2014	Drama Mystery Thriller
Lone Survivor	24	2014	Action Adventure Family Fantasy Romance
The Fault in Our Stars	25	2014	Drama Romance
Star Wars: The Force Awakens	2	2015	Action Adventure Sci-Fi
Inside Out	4	2015	Adventure Animation Comedy Drama Family Fantasy
American Sniper	6	2015	Action Biography Drama War
The Martian	9	2015	Adventure Drama Sci-Fi
Pitch Perfect 2	13	2015	Comedy Music
Straight Outta Compton	19	2015	Biography Drama History Music
Trainwreck	27	2015	Comedy Drama Romance
Creed	28	2015	Drama Sport
The Imitation Game	39	2015	Biography Drama Thriller War
Deadpool	6	2016	Action Adventure Comedy Sci-Fi
The Revenant	14	2016	Action Adventure Drama History Western

Sing	18	2016	Animation Comedy Family Musical
Sully	24	2016	Biography Drama
Bad Mom	25	2016	Comedy
Beauty And The Beast	2	2017	Family Fantasy Musical Romance
Wonder Woman	3	2017	Action Adventure Fantasy Sci-Fi War
The Guardians Of Galaxy Vol.2	4	2017	Action Adventure Comedy Sci-Fi
It	6	2017	Horror
Thor: Ragnarok	7	2017	Action Adventure Comedy Fantasy Sci-Fi
Hidden Figure	20	2017	Biography Drama History
War For the Planet of the Apes	22	2017	Action Adventure Drama Sci-Fi Thriller
Black Panther	1	2018	Action Adventure Sci-Fi
A Star Is Born	12	2018	Drama Music Romance
Bohemian Rhapsody	14	2018	Biography Drama Music
A Quiet Place	15	2018	Drama Horror Sci-Fi
Crazy Rich Asians	17	2018	Biography Drama Musical
The Greatest Showman	24	2018	Drama Sport
Creed 2	28	2018	Action Adventure Sci-Fi
The Avengers IV: Endgame	1	2019	Action Adventure Drama Sci-Fi

Toy Story 4	3	2019	Adventure Animation Comedy Family Fantasy
Joker	9	2019	Crime Drama Thriller
Us	12	2019	Horror Mystery Thriller
Knives Out	21	2019	Comedy Crime Drama Mystery Thriller
Ford VS Ferrari	27	2019	Action Biography Drama Sport
Hustlers	29	2019	Comedy Crime Drama

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

VITA

NAME	Runze Xu
DATE OF BIRTH	12 May 1994
PLACE OF BIRTH	China
INSTITUTIONS ATTENDED	Assumption University Chulalongkorn University
HOME ADDRESS	Hefei, Anhui, China