

# THE ENGLISH OF SOCIAL MEDIA

## Investigating the linguistic features of Instagram posts

ELENA MANCA  
UNIVERSITÀ DEL SALENTO

**Abstract** - In the contemporary digital environment, adolescents and young adults are increasingly exposed to English through various forms of digital media, with social media playing a particularly prominent role. While this exposure often occurs incidentally during leisure activities, it contributes to the unconscious acquisition of linguistic structures, vocabulary, and communicative norms. This study investigates the nature of the English language encountered on social media by analysing three corpora of Instagram posts focused on cinema, music, and travel, topics identified as particularly engaging by Italian university language students (Manca 2024). The corpora were compiled from posts published between February 2024 and February 2025 on six high-activity Instagram pages selected for each theme. Adopting a quantitative approach grounded in Biber's (1995; Biber and Conrad 2009) multidimensional analysis, the study explores the distribution of linguistic features characteristic of spoken versus written English. The corpora are also compared against the British National Corpus (BNC) to identify statistically significant deviations in part-of-speech (POS) frequencies. Analysis was conducted using SketchEngine (Kilgariff *et al.* 2004) and WMatrix7 (Rayson 2008), employing keyword lists, wordlists, and concordances. The findings provide insights into the hybrid linguistic nature of English in digital spaces and may provide the basis for future research on second language acquisition through social media.

**Keywords:** Instagram language; written and spoken English; quantitative analysis; social media.

## 1. Introduction

In today's digital landscape, adolescents and young adults are exposed to the English language through a wide range of digital media, with social media platforms playing a particularly significant role (Manca 2024; Pavesi *et al.*, 2023; Pavesi, Bianchi 2024; Pavesi *et al.* forthcoming). Whether engaging with short-form videos, memes, tweets, or online discussions, young users interact with English daily, often without the explicit intention of learning the language. However, this frequent exposure for entertainment purposes contributes to an incidental acquisition of English, as users absorb linguistic structures, vocabulary, and communicative norms by interacting with them.

Despite the growing presence of English in digital spaces, important questions remain about the nature of the language used in these contexts. In particular, it is worth exploring what distinctive linguistic features emerge in the use of English on platforms such as social media, and to what extent the English found in these digital environments more closely resembles spoken or written varieties. As Benamara *et al.* (2018, p. 663) describe, social media posts are short online conversational posts or comments sharing properties with both spoken and written language. Their interactional nature may make them more similar to some spoken genres, such as conversation, while their written and mainly asynchronous nature may contribute to make them closer to written genres. However, as Biber (1995, p. 37) suggests, there are no linguistic or situational features of speech and writing that can be applied to all spoken and written genres. Spoken public speeches and written expositions are very similar, while conversations and spoken public

speeches are quite different. If we consider the most frequent types of speech and writing (face-to-face conversations and informational expositions) some distinctive features can be identified. Based on its situational characteristics, typical speech is interactive, dependent on shared space, time, and background knowledge; from a linguistic perspective, it is structurally simple, fragmented, concrete, and dependent on exophoric (situation-dependent) reference; typical writing has opposite features.

This study seeks to explore the linguistic and genre distinctive features of social media posts through the analysis of three corpora consisting of Instagram posts published on six pages focused on three main themes: music, travel, and cinema (see Section 3). These pages were chosen according to follower count and posting frequency and all the posts published on these pages between February 2024 and February 2025 were collected to construct the three corpora under investigation.

The study adopts a quantitative approach grounded in Biber's ([1988], 1995) and Biber and Conrad (2009) multidimensional analysis, which examines the frequency of specific linguistic markers and associates their distribution with features typical of either spoken or written genres. Additionally, the three corpora are compared with the Spoken and Written Samplers of the British National Corpus (BNC) in order to identify part-of-speech (POS) tags with significantly unusual frequencies in relation to written and spoken genres of English. The quantitative analysis was carried out using the software tools SketchEngine (Kilgarriff *et al.* 2004) and WMatrix 7 (Rayson 2008), employing wordlists, keyword lists, and concordances as primary analytical instruments.

The remainder of this paper is organised as follows: Section 2 provides an overview of Biber's framework of multidimensional analysis and, drawing on previous studies, outlines the linguistic features of digital and social media communication. Section 3 describes the data and the methodology adopted. Section 4 presents the analysis of the three corpora along with an interpretation of the findings, while Section 5 offers some concluding remarks.

## 2. Written and spoken varieties through multidimensional analysis

Biber (1995, p. 56ff) compares speech and writing based on six dimensions of linguistic variation grouping a series of linguistic features sharing an underlying communicative function.

Dimension 1 is labelled 'Informational vs. Involved Production' and its poles represent highly interactive, affective discourse produced under real-time constraints versus highly informational discourse, which is carefully crafted and highly edited (1995, p. 115, 135). The features with positive weight on this Dimension (1995, p. 105-107) are private verbs (e.g., 'think', 'love'), present tense, first- and second-person pronouns, *wh*-questions, emphatics and amplifiers ('just', 'really', 'very', 'absolutely', etc.), reduced surface form (e.g., *that*-deletions, contractions, pro-verb 'do', demonstrative pronouns, and indefinite pronouns which substitute for fuller noun phrases, and final stranded prepositions), generalised or uncertain presentation of information (e.g., hedges and possibility modals), and fragmented production of text (e.g., 'be' as main verb, non-phrasal coordination, analytic negation). Furthermore, there are four highly occurring subordination features, namely causative subordination ('because'), sentence relatives, *wh*-clauses, and conditional subordination (e.g., 'if', 'unless'). The features with negative weight on this Dimension are nouns and prepositions, careful selection of vocabulary, high type/token ratio. High scores of the features with positive weight of Dimension 1 can be

observed in face-to-face and telephone conversations while moderately high scores are found in personal letters, spontaneous speeches, and interviews (1995, p. 129).

Dimension 2 is labelled as ‘Narrative vs. Non-narrative Concerns’ and the features with positive weight are past tense and perfect aspect verbs, third person pronouns, public verbs (‘tell’, ‘speak’, ‘say’, ‘explain’), present participial clauses, and synthetic negation (1995, pp. 135-136). Fiction genres have by far the highest mean scores on this dimension, while intermediate values include prepared and spontaneous speeches, biographies, personal letters, face-to-face conversation, and press reportage.

Dimension 3 is defined ‘Explicit versus Situation-Dependent Reference’ and the features with positive weight are *wh*-relative clauses, pied-piping constructions (e.g., ‘a financial base on which to work’), phrasal coordination (e.g., ‘such classes and such examinations’), and nominalisations. The highest scores of features with positive weight on this Dimension can be observed in genres such as official documents and professional letters, while moderately high scores are typical of press reviews, academic prose, and religious genres (1995, p. 142).

Dimension 4 is labelled ‘Overt Expression of Persuasion’ and is defined by linguistic features with positive weight such as prediction modals, necessity modals, possibility modals, conditional clauses, suasive verbs (e.g., ‘command’, ‘instruct’, ‘demand’), infinitives used as adjective and verb complements, and split auxiliaries. Genres with high scores on this Dimension are professional letters and editorials (1995, p. 111).

Dimension 5 (1995, pp. 151ff) has been named as ‘Abstract versus Non-Abstract Information’. Features with positive weight are frequent use of conjuncts, agentless and *by* passives, adverbial past participial clauses, past participial WHIZ deletions (e.g., ‘cold moulds tested in’) and certain types of adverbial subordination. Genres such as academic prose and official documents have the highest score on this dimension.

Dimension 6, labelled as ‘On-line Informational Elaboration’ includes features with positive weight such as *that* complements to verbs, *that* complements to adjectives, *that* relatives in object positions, and demonstratives. Other elements, occurring slightly less frequently, are final prepositions, existential *there*, *wh*-relative clauses in object position (1995, pp. 113-114). Three genres score high on this Dimension, namely prepared speeches, interviews, and spontaneous speeches. In these genres, the focus is informational, but the speaker must deal with the time-related limitations of real-time production.

## 2.1. Linguistic features in electronic communication

Biber and Conrad (2009) analysed the language used in three registers relying on electronic means to convey a message, namely e-mails, e-forum postings and text messages, and compared the findings obtained to prototypical forms of speech and writing, that is to say conversation and academic prose.

Similarly to conversation, e-mail messages are interactive (Biber and Conrad 2009, p. 178), as e-mail senders expect the receivers to reply or at least to acknowledge their messages; furthermore, both e-mail messages and conversation communicate feelings and personal stance. Differences between the two registers can be found in the mode (spoken vs. written) and medium (the channel). Conversation is spoken while e-mail messages are written and, for this reason, they can be planned and revised. Furthermore, while in conversation sender and addressee share time and space, in e-mail messages this is less so. From a linguistic perspective (Biber and Conrad 2009, p. 182), e-mail messages share similarities with conversation as regards the frequency of lexical verbs. Biber and Conrad (2009, p. 182) argue that this is because both e-mail messages and conversation share an

interpersonal or task-focused focus and, for this reason, they tend to adopt a more clausal style. As for pronoun usage, they show a frequency falling between spoken dialogue and academic writing, incorporating significantly more pronouns than academic prose but fewer than face-to-face interactions. The first-person pronouns ('I', 'we') and second-person pronouns ('you') show the same frequency in e-mail messages and conversation. Conversely, third person pronouns are rare in e-mail and academic prose but common in conversation. In terms of noun frequency, e-mails closely resemble academic writing, where nouns are considerably more prevalent than in casual conversation.

E-forum posts are similar to e-mails in that they are written and they can be planned and revised prior to their sending (Biber and Conrad 2009, p. 191). Among the linguistic features of e-forum postings, the absence or only occasional use of capitalisation and ellipses can be observed. These postings exhibit grammatical features similar to those of emails, with a slightly higher frequency of verbs, a significantly higher frequency of adverbs, slightly fewer nouns and adjectives, and an almost identical frequency of pronouns (2009, p. 195). They are also characterised by short clauses and by a more frequent use of third person pronouns with respect to e-mails.

Text messages are a highly interactive form of communication and, for this reason, they are closer to conversation than to emails or forum postings. However, unlike conversation, text messages are written and can be planned and edited. Biber and Conrad (2009, p. 205) observe that two distinguishing features of text messages are abbreviations and nonstandard punctuation and capitalisation. From a linguistic perspective, text messages are characterised by a high number of verbs (higher than e-forum postings and conversation) as well as a high number of pronouns (higher than conversation). As for nouns, their occurrence is similar to emails and e-forum postings, while adverbs and adjectives have a very low frequency of occurrence.

## **2.2. Linguistic features of social media posts**

A substantial body of research has examined the broader topic of language use on social media. Scholars have explored a range of themes within this domain, including the prevalence and dynamics of hate speech (Castaño-Pulgarín *et al.* 2021; Guiora and Park 2017; Schmid *et al.* 2024), the construction of identity, ideology, and gender (Chałupnik and Brookes 2022; Esposito and KhosraviNik 2025; Featherman 2015; Georgalou 2015; Zappavigna 2012), the role of social media in language learning (Aloraini and Cardoso 2022; Alazemi *et al.* 2023; Prichard 2013; Reinhardt 2019), the dissemination of disinformation (Aïmeur *et al.* 2023; Colliander 2019; Moravec *et al.* 2019), and the nature of political discourse (Ceron and D'Adda 2016; Coesemans and De Cock 2017; Kreis 2017; McDonnell 2020). In addition, considerable attention has been given to the use of social media as marketing tools, investigating what strategies are employed to engage audiences, shape brand identity, and influence consumer behavior (Benthaus *et al.* 2016; Cade 2018; Voramontri *et al.* 2019).

The studies on social media described below are more language-oriented, in that they analyse a series of Facebook posts, aiming to identify the linguistic promotional strategies used by travel operators and winery owners in English-speaking countries.

Bianchi (2017) analysed a corpus that included the Facebook posts of three large international travel operators based in English-speaking countries, and compared this corpus against parts of the BNC Sampler to examine how the language of the posts under analysis relates to everyday language and where it falls on the written-spoken/formal-informal continuum. The keyword comparison with the BNC Written Sampler generated a list primarily consisting of items such as 'you', 'n't', 'what', 'why', 's', 've', 'yours',

'want', 'get', 'here', 'do', 're', 'go on', 'miss out', which Biber (1995; see Section 2) identifies as linguistic features of involved production (Bianchi 2017, p. 16). Interestingly, the comparison with the BNC Spoken Sampler does not provide as keywords pronouns, contractions, adverbs, or phrasal verbs, but rather verbs and adjectives. The presence of positive keywords characteristic of informal communication in the comparison to the BNC Written Sampler, and their absence in the comparison to the BNC Spoken Sampler, suggest that, despite being written, the posts under analysis resemble spoken communication more than written informal communication (Bianchi 2017, p. 16).

Manca (2021) examined two corpora of Facebook posts published by a selection of Australian and US wineries. The analysis was carried out according to Hyland's (2005) framework, which includes interactive and interactional strategies. The most frequent interactive feature used in both corpora (Australian and US) was represented by Transition Markers, in particular additive conjunctions ('and', 'or') occurring more frequently than other types of conjunctions, such as the subordinating ones. Text stages or sequences, and indications of discourse goals or topic shifts were not used probably because their presence could add complexity to a post that should be read quickly by Facebook users when scrolling down the newsfeed page. Furthermore, the absence of Code Glosses, namely explanations, clarifications, reformulations, suggests that post authors and post readers possess a high degree of shared knowledge about the topics dealt with and no additional information is needed to make contents clearer and more accessible to readers (Manca 2021, p. 125). As for interactional features, first- and second-person pronouns showed a high percentage of usage followed by attitude adjectives and attitude verbs (most of them being private verbs). Furthermore, posts in both corpora often included direct invitations to take advantage of what is described and proposed, and rhetorical and non-rhetorical questions which act as a way to convince readers to take action or to make them feel as an important part of the community.

Bianchi and Manca (2023) examined the Facebook posts written by wine producers from New Zealand, South Africa and the UK. The findings revealed a frequent use of additive conjunctions, first- and second-person pronouns, and attitude adjectives. In all posts, attitude was conveyed primarily through adjectives, while sentence adverbs were less frequently used, and verbs were rarely used (Bianchi and Manca 2023, p. 260). Hedges and boosters occurred very infrequently. This was ascribed to the promotional nature of the posts which directly address Facebook users by inviting them to take advantage of what is offered and proposed through reader pronouns, directives and questions.

All the case studies described in this section exhibit several common linguistic features, such as the use of first- and second-person personal pronouns, a reduced use of subordination, the presence of finite verbs, and the frequent occurrence of questions, particularly *wh*-questions. These linguistic characteristics, grouped by Biber (1995) under Dimension 1, define the involved mode of production, which is typical of spoken language and conversation. It can, therefore, be hypothesised that written/texted, interactive, computer-mediated communication tends to adopt features that bring it closer to spoken rather than written traditional genres. To identify the linguistic features emerging in the use of English on platforms such as social media and to ascertain the extent to which the English found in these digital environments more closely resembles spoken or written varieties, the following sections will analyse six Instagram pages focused on music, travel, and cinema. This analysis may also help us understand what kind of English social media users are exposed to when navigating and interacting on these platforms.

### 3. Materials and methods

The present analysis of social media linguistic features is conducted by examining three corpora, which were compiled between February 2024 and February 2025. These consist of all posts (excluding followers' comments) published during that period on six Instagram pages managed by prominent influencers specialising in music, travel, and cinema. These topics were selected based on a study conducted by Manca (2024), who investigated the preferences, perceptions, and attitudes of Italian university language students, particularly those enrolled at the University of Salento, towards the role of social media in learning English. Using a questionnaire inspired by the *IECoL* tool, the Informal English Contact and Learning questionnaire (Pavesi *et al.*, 2023) and administered to 213 language students, Manca's (2024) study explored students' motivations for using social media and their perceptions of how such use may influence their language proficiency. The results indicated that students engage with social media not only for entertainment and cross-cultural interaction, but also as a means to practice English and access information in the language. Among the thematic interests most commonly cited by the participants were cinema, music, and travel. Based on these findings, the present study selects two Instagram pages for each of the three topics. Influencers were selected based on their number of followers, posting frequency on Instagram, and the inclusion of text alongside pictures and reels in their posts.

The Music corpus (henceforth InstaMusic) collects 402 posts from the Instagram accounts of *Somewhere Soul* (@somewhere\_soul) and *Pigeons and Planes* (@pigsandplans) and includes 26,741 words. *Somewhere Soul* was founded by Josh Mason, an artist manager and digital marketer who presents the best new music from around the world and advertises music events to his 912,000 Instagram followers. *Pigeons and Planes* currently has 905,000 followers and was founded by Jacob Moore. It is now part of the entertainment company *Complex Network* and provides a comprehensive overview of new music across various genres and geographic regions, curates and promotes high-quality music created by both established artists and emerging talents, and offers a diverse and inclusive representation of contemporary musical developments.

The Travel corpus (henceforth InstaTravel) includes 294 posts from two Instagram pages, namely *Renee Roaming* (@reneeroaming) and *Nomadic Matt* (@nomadicmatt), and currently has 38,547 words. *Renee Roaming* is Renee Hahnel's Instagram page, an Australian-born travel photographer, blogger, and author. She has journeyed to nearly 40 countries across five continents, sharing her experiences and inspiring others to explore the world. At the time of writing, she has 944,000 followers. *Nomadic Matt* is Matthew Kepnes, an American travel blogger who shares his travel experiences and provides tips on how to travel on a budget and which destinations to visit around the world. His Instagram page has 153,000 followers.

The Cinema corpus (henceforth InstaCinema) includes 167 posts and 26,088 words coming from two Instagram pages, namely *Cinéma* (@lynncinema) and *Cinematheque* (@cinematheque). *Cinéma* is Lynn Sfeir's Instagram account, a Lebanese filmmaker known for her work as a director, writer, and editor. On her Instagram page, she reviews films and provides her own interpretation of film scenes, with a particular focus on those films that have made history. She currently has 175,000 followers. *Cinematheque* is Dakota Joaquin's Instagram page focusing on film reviews and analysis and it is followed by 132,000 people.

Table 1 summarises the features of the corpora considered for analysis.

| Corpus      | Number of posts | Number of words |
|-------------|-----------------|-----------------|
| InstaMusic  | 402             | 26,741          |
| InstaTravel | 294             | 38,547          |
| InstaCinema | 167             | 26,088          |

Table 1  
Details of corpora used.

The analysis was based on Biber's (1995) six dimensions and was carried out using a quantitative approach. Two different tools were used in the quantitative analysis, namely SketchEngine, and the BNC Sampler Written and Spoken available on WMatrix7. Using SketchEngine, the frequency distribution values of the linguistic markers indicated by Biber were identified, namely conjunctions, pronouns, verbs, nouns, adjectives, adverbs, emphatics, amplifiers, hedges, downtoners, demonstratives, existential *there*, synthetic and analytic negation, phrasal verbs, contractions, *that* and *wh*-relative clauses, *wh*-questions, final stranded prepositions, pied-piping in relative clauses, infinitives and participial clauses. Each of the three corpora has been examined separately. Wordlists and concordances generated using SketchEngine were analysed to determine the usage and the frequency of occurrence of the linguistic features under analysis. Percentage values were calculated based on the total number of words in each corpus. Using WMatrix 7, the three corpora were compared with the BNC Sampler Written and the BNC Sampler Spoken in order to identify the key differences and similarities in the frequency of parts of speech (POS) tags. Key POS tags indicate statistically unusual frequencies of parts of speech and are meaningful in that they reveal when certain POS tags occur significantly more often in a target corpus (in this analysis, the Instagram corpora) than in a reference corpus (the BNC reference samples). The 20 most frequent key POS tags resulting from the comparison between each of the two BNC Samplers and the three Instagram corpora were selected for analysis, along with their corresponding log-likelihood (LL) values, a statistical measure used to determine keyness (Rayson, Garside, 2000). This test evaluates whether differences in relative frequencies between two corpora are statistically significant, taking into account their respective sizes. A log-likelihood value above 6.63 is generally accepted as the threshold for statistical significance, corresponding to a 99% confidence level (Rayson 2008; Gabrielatos, Marchi 2012). The greater the log-likelihood value, the more statistically significant the POS tag is considered to be.

## 4. Findings

### 4.1. The InstaMusic corpus

In the InstaMusic corpus there are 18 different conjunctions, most of them being coordinating conjunctions: 'and' is at the top of the list with 448 occurrences (1.7%), followed by 'but' (60 times – 0.22%), 'so' (40 times – 0.15%), and 'or' (20 times – 0.07%). The conjunction 'if' is the most frequent subordinating conjunction occurring 58 times (0.21%) and its presence can be considered as a marker of stance, a condition for actions or beliefs (Biber 1995, p. 107). Temporal conjunctions are less frequent and are 'when' (16 entries – 0.05%), 'after' (14 entries – 0.05%), 'since' (9 entries – 0.03%), 'while' (8 entries – 0.03%), 'until' (4 entries – 0.01%), 'before' (3 entries – 0.01%). The causative conjunction 'because' occurs 9 times (0.03%), while the concessive 'although' occurs only three times (0.01%).

The first-, second-, and third-person singular pronouns have very similar frequency distributions. First- and third-person plural pronouns are also present, although less frequently. Frequency distribution and percentages are reported in Table 2.

| Pronoun | Frequency | %    | Pronoun | Frequency | %     |
|---------|-----------|------|---------|-----------|-------|
| I       | 304       | 1.1  | their   | 74        | 0.3   |
| you     | 271       | 1    | she     | 55        | 0.2   |
| it      | 268       | 1    | they    | 53        | 0.2   |
| my      | 132       | 0.5  | our     | 40        | 0.14  |
| he      | 132       | 0.5  | them    | 30        | 0.1   |
| his     | 114       | 0.4  | us      | 24        | 0.1   |
| her     | 104       | 0.38 | him     | 17        | 0.06  |
| we      | 96        | 0.35 | its     | 15        | 0.05  |
| me      | 88        | 0.32 | yours   | 2         | 0.007 |
| your    | 78        | 0.3  | theirs  | 1         | 0.003 |

Table 2  
Personal pronouns and possessive adjectives and pronouns in the InstaMusic corpus.

There are 623 verbs with a total frequency of 4,200 tokens (15.7%). Most of them (97%) occur only twice or once. The 30 most frequent verbs are listed in Table 3 (the frequencies of the verbs ‘to be’, ‘to have’, ‘can’, ‘would’ and ‘will’ also include the contracted and negative contracted forms).

| Verb | Entries | %    | Verb   | Entries | %    | Verb      | Entries | %    |
|------|---------|------|--------|---------|------|-----------|---------|------|
| is   | 436     | 1.6  | watch  | 47      | 0.17 | make      | 30      | 0.11 |
| are  | 134     | 0.5  | can    | 45      | 0.16 | dropped   | 27      | 0.10 |
| was  | 137     | 0.5  | had    | 42      | 0.15 | making    | 24      | 0.09 |
| have | 115     | 0.4  | hope   | 40      | 0.15 | think     | 24      | 0.09 |
| been | 106     | 0.4  | get    | 37      | 0.13 | would     | 23      | 0.09 |
| be   | 103     | 0.38 | do     | 33      | 0.12 | were      | 22      | 0.08 |
| has  | 103     | 0.38 | let    | 33      | 0.12 | hear      | 20      | 0.07 |
| will | 59      | 0.2  | got    | 31      | 0.11 | going     | 19      | 0.07 |
| see  | 53      | 0.2  | listen | 31      | 0.11 | listening | 19      | 0.07 |
| know | 51      | 0.2  | shared | 31      | 0.11 | said      | 18      | 0.06 |

Table 3  
The 30 most frequent verbs in the InstaMusic corpus.

The verb wordlist suggests the presence of four groups of verbs: auxiliaries (‘be’, ‘have’, ‘will’, ‘do’, ‘would’), generic verbs whose meaning strongly depends on the context in which they are used (‘do’, ‘get’, ‘make’, ‘take’, ‘let’), private verbs (verbs that are used for the overt expression of private attitudes, thoughts, and emotions, such as ‘hope’, ‘think’), and verbs specifically related to the topic or to the nature of posts (‘see’, ‘watch’, ‘share’, ‘come’, ‘listen’, ‘drop’, ‘play’, ‘hear’).

‘Be’ used as main verb, particularly when employed in predicative expressions (e.g., ‘the house is big’ instead of ‘a big house’), is associated with a fragmented presentation of information implying a low informational density (Biber 1995, p. 106), that is to say with a presentation having many loosely joined clauses and few content words or novel semantic elements. In this corpus, it is indeed mainly used in predicative expressions (80%), but also as an auxiliary (14%). Similarly, the verb ‘do’, in this corpus, is mainly used as a main verb with generic meaning (52%) or as an auxiliary (39%), but it is also used as a pro verb (9%), as in ‘I hope it nourishes your soul in the way it did mine’. Features of informality can be observed in the frequency of ‘get’, ‘make’, ‘take’, and ‘let’,

generic verbs whose meaning depends on the context in which they are used, whereas involved production is suggested by the patterns of ‘see’ and ‘know’. The former is frequently used in expressions ‘I hope to see you there’ referring to music events and concerts, while ‘know’ is mainly used to engage readers and ask for their opinion as in ‘Let me know what you think’ or ‘Let me know if you are a fan’. Furthermore, in the concordance of ‘think’, several examples of reduced surface form marked by *that* deletions can be noticed, as in ‘So I thought I’d start off with a little spotlight of some of my favourite albums’.

Among the 70 most frequent verbs (with a range of occurrence going from 436 to 10), 38 are in the base form (present simple, imperative or infinitive), 20 are in the past, 12 are in the ‘-ing’ form.

There are about 2,400 nouns in the InstaMusic corpus with a total frequency of 7,800 (29%). This high percentage value suggests that posts have an informational communicative function. The nouns in this corpus can be loosely grouped into three main semantic fields: 1. Music (‘album’, ‘music’, ‘artist’, ‘record’, ‘song’, and so on), 2. Time (‘year’, ‘time’, ‘Friday’, ‘day’, ‘night’), 3. Instagram (‘link’, ‘bio’, ‘video’). Table 4 lists the first 20 most frequent nouns in this corpus.

| Noun    | Frequency | %    | Noun    | Frequency | %    |
|---------|-----------|------|---------|-----------|------|
| music   | 217       | 0.8  | albums  | 45        | 0.16 |
| album   | 187       | 0.7  | video   | 42        | 0.15 |
| year    | 117       | 0.4  | tour    | 41        | 0.15 |
| time    | 75        | 0.28 | show    | 37        | 0.13 |
| artists | 63        | 0.23 | release | 38        | 0.14 |
| record  | 63        | 0.23 | night   | 36        | 0.13 |
| link    | 55        | 0.2  | project | 35        | 0.13 |
| bio     | 50        | 0.18 | artist  | 34        | 0.12 |
| song    | 50        | 0.18 | day     | 34        | 0.12 |
| Friday  | 48        | 0.17 | track   | 34        | 0.12 |
| years   | 48        | 0.17 | songs   | 31        | 0.11 |

Table 4  
The 20 most frequent nouns in the InstaMusic corpus.

Adjectives are another frequent feature of the posts included in the InstaMusic corpus: there are 569 different adjectives with a total frequency of 1,961 tokens (7.3%). They are mainly evaluative/descriptive adjectives and are used to emphasise the quality and the validity of the albums and records proposed. Their presence suggests that the InstaMusic posts have an informational aim because as Biber (1995, p. 105) explains “they are a more integrated form of nominal elaboration than predicative adjectives or relative clauses, since they pack information into relatively few words and structures”. Table 5 lists the 20 most frequent evaluative/descriptive adjectives in the InstaMusic corpus.

| Adjective | Entries | %    | Adjective  | Entries | %    |
|-----------|---------|------|------------|---------|------|
| new       | 222     | 0.8  | jazz       | 29      | 0.1  |
| first     | 81      | 0.3  | good       | 27      | 0.1  |
| full      | 60      | 0.2  | big        | 25      | 0.09 |
| last      | 43      | 0.16 | beautiful  | 23      | 0.08 |
| happy     | 40      | 0.15 | few        | 23      | 0.08 |
| favourite | 40      | 0.15 | special    | 21      | 0.07 |
| best      | 38      | 0.14 | great      | 19      | 0.07 |
| next      | 36      | 0.13 | wonderful  | 18      | 0.06 |
| top       | 35      | 0.13 | incredible | 15      | 0.05 |
| two       | 33      | 0.12 | three      | 15      | 0.05 |

Table 5  
The 20 most frequent adjectives in the InstaMusic corpus.

There are 191 different adverbs in this corpus with a total frequency of 1,300 (4.8%). They can be grouped into seven sets: adverbs of time, locative adverbs, amplifiers, emphatics, hedges, downtoners, and adverbs of addition.

Among the adverbs of time, the most frequent are ‘now’ (67 entries – 0.25%), ‘just’ (28 times – 0.1%), ‘ever’ (23 entries – 0.08%), ‘always’ (22 entries – 0.08%), ‘still’ (20 entries – 0.74%), ‘then’ (20 entries – 0.74%), ‘recently’ (19 entries – 0.7%), ‘never’ (19 entries – 0.7%), ‘today’ (17 entries – 0.06%), ‘soon’ (13 entries – 0.04%), ‘so far’ (13 entries – 0.04%), ‘again’ (11 entries – 0.04%), ‘yet’ (11 entries – 0.04%), ‘already’ (9 entries – 0.03%), ‘once’ (8 entries – 0.03%), ‘currently’ (7 entries – 0.02%), ‘finally’ (7 entries – 0.02%). The most frequent locative adverbs are ‘here’ (72 entries – 0.26%), ‘there’ (18 entries – 0.7%), and ‘away’ (17 entries – 0.06%). Amplifiers are ‘even’ (17 entries – 0.06%), ‘absolutely’ (14 entries – 0.04%), ‘very’ (14 entries – 0.04%), and ‘highly’ (5 entries – 0.2%). Emphatics include ‘so’ (37 entries – 0.14%), ‘most’ (29 entries – 0.1%), ‘really’ (19 entries – 0.7%), ‘truly’ (9 entries – 0.03%), ‘more’ (8 entries – 0.03%), ‘actually’ (7 entries – 0.02%). Hedges comprise ‘almost’ (7 entries – 0.02%) ‘probably’ (7 entries – 0.02%), and ‘maybe’ (6 entries – 0.02%). The adverb ‘only’ (23 entries – 0.08%) is the only downtoner appearing in the corpus. Adverbs of addition include ‘also’ (41 entries – 0.15%), and ‘too’ (18 entries – 0.07%).

The demonstrative ‘this’ occurs frequently with 260 entries (1%), 223 times as a demonstrative adjective and 37 times as a demonstrative pronoun. ‘That’ occurs 83 times (0.3%), 71 times as an adjective and 12 times as a pronoun. ‘These’ occurs 27 times (0.1%), 22 times as an adjective and 5 times as a pronoun, while ‘those’ occurs 15 times (0.05%), 6 times as an adjective and 9 times as a pronoun.

Existential *there* occurs with 23 entries (0.08%), thus suggesting that it is not a preferred structure in the Instagram posts considered for analysis. Analytic negation can be found 139 times (0.5%) and, for this reason, it is preferred to synthetic negation occurring 74 times (0.27%). The total frequency of phrasal verbs is 264 (1%), and the most frequent phrasal verbs are those combined with ‘up’ (89 entries) and ‘out’ (77 entries).

Relative clauses, especially those with ‘who’ and ‘which’ have 155 examples (0.57%), while those encoded in participial structures are less frequent, with 47 examples (0.17%); there are only 11 cases (0.4%) of final stranded prepositions and none exemplifies pied-piping. *Wh*-questions can be observed in 40 cases (0.15%).

Figure 1 summarises the results of the analysis described.

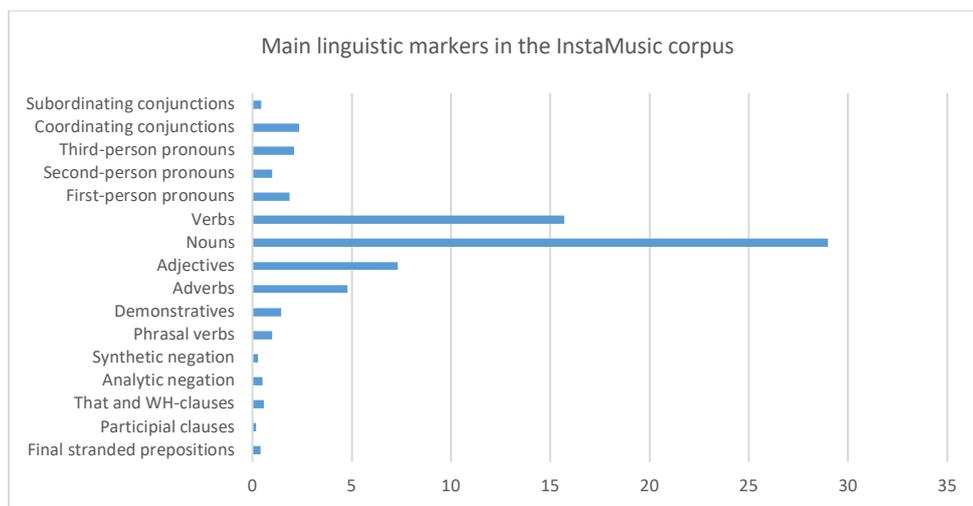


Figure 1  
Main linguistic markers in the InstaMusic corpus.

#### 4.1.1. Discussion of findings in the InstaMusic corpus

The analysis reveals that the InstaMusic corpus displays many of the features identified by Biber (1995) under Dimension 1, particularly those associated with involved production. The frequently occurring coordinating conjunctions noted above function as markers of fragmentation, while the frequent use of conditional subordination with ‘if’ serves to express stance, that is, a condition for actions or beliefs (Biber 1995, p. 107).

First- and second-person pronouns directly reference the addressor and addressee and, for this reason, they can be considered indicators of highly interactive discourse. Along with contractions, these pronouns contribute to a colloquial and informal tone, characteristic of interactive contexts involving rapid language production (Biber 1995, pp. 19, 105).

In this corpus, ‘be’ as a main verb is primarily found in predicative constructions. Together with non-phrasal coordination and analytic negation – the latter being preferred over synthetic negations – it contributes to a fragmented presentation of information (Biber, 1995, p. 106). Reduced surface forms can be observed in the frequent use of demonstrative pronouns, *that* deletions, and ‘do’ used as a pro-verb. Features of informality are also evident in the frequent use of ‘get’, ‘make’, ‘take’, and ‘let’, generic verbs whose meaning depends on their context of usage, and in the high frequency of phrasal verbs. Involved production can be further observed in the patterns of the verbs ‘see’ and ‘know’. The most frequent verb tense in the InstaMusic corpus is the base form (present simple, imperative, or infinitive), while past tenses occur much less frequently. Furthermore, the preference for time adverbs suggests a tendency towards the description of present or recent events. The adverb ‘now’, which is the most frequent time adverbial in the corpus, also performs the function of discourse particle with the aim of maintaining conversational coherence and monitoring the information flow in involved discourse (Biber 1995, p. 241; Schiffrin 1981; Chafe 1982). Emphatics are also commonly used and serve as markers of informal, colloquial discourse, signalling involvement with the topic (Biber 1995, p. 241; Chafe 1982).

Although all these features suggest a type of involved production, the high frequency of occurrence of nouns and adjectives reveal an informational focus. As Biber observes (1995, p. 104), “nouns are the primary bearers of referential meaning in a text, and a high frequency of nouns thus indicates great density of information”. Furthermore, the frequent use of adjectives reinforces the informational aim of the texts as they act as an

integrated form of nominal elaboration (Biber 1995, p. 105). Reduced grammatical structure combined with frequent nominalisation is typical of broadcasts, a specialized genre, which is spoken and produced in real-time, but has the characteristics of informational production (Biber 1995, p. 134). The frequent use of demonstratives in this corpus suggests immediate reference to entities, ideas, or prior discourse, interactive and real-time processing of information, and direct engagement with the subject matter. This is reinforced by the frequent use of the adverb ‘here’ used as a deictic with the function of text internal referent. On-line informational elaboration (Dimension 6) relies heavily on demonstratives (Biber 1995, p. 114) which are used for both text-internal deixis and for exophoric, text-external reference.

The frequent occurrence of third-person pronouns, features associated with Dimension 2, is not accompanied by past verb tenses. For this reason, their presence does not indicate a narrative function; rather, it relates to the content of the posts, which often involve reviews of artists and their music. Overall, the findings suggest that the Instagram posts in the InstaMusic corpus are characterised by features typical of involved production, while also displaying informational concerns.

#### 4.1.2. The InstaMusic corpus and the BNC Samplers

To further check differences and similarities of the Instagram posts in the InstaMusic corpus with spoken and written discourse, the analysis continues with a comparison with the BNC Samplers Written and Spoken. The comparison is carried out by investigating key POS tags, which are presented in Table 6 in order of log-likelihood value.

| Key POS with BNC Written                          |          | Key POS with BNC Spoken            |          |
|---|----------|------------------------------------|----------|
| POS   | LL value | POS                                | LL value |
| lexical verbs in the base form                    | 426.86   | singular proper nouns              | 1718.57  |
| singular temporal nouns                           | 236.19   | singular common nouns              | 1421.61  |
| cardinal numbers                                  | 172.22   | general adjectives                 | 886.91   |
| possessive adjectives and pronouns                | 158.87   | general prepositions               | 519.57   |
| second-person pronouns                            | 132.01   | possessive adjectives and pronouns | 482.22   |
| demonstratives                                    | 99.66    | plural common nouns                | 430.46   |
| singular weekday nouns                            | 95.37    | singular temporal nouns            | 285.73   |
| ‘-ing’ participle of lexical verb                 | 74.96    | ‘of’ as a preposition              | 268.03   |
| singular proper nouns                             | 71.94    | cardinal numbers                   | 242.36   |
| singular indefinite pronouns                      | 67.34    | articles                           | 174.55   |
| first-person singular objective personal pronoun  | 63.60    | ‘-ing’ participle of lexical verb  | 165.62   |
| ordinal numbers                                   | 63.03    | singular article                   | 155.42   |
| singular cardinal numbers                         | 51.12    | ordinal numbers                    | 147.96   |
| prepositional adverbs and particles               | 31.13    | past tense of lexical verb         | 100.22   |
| comparative general adverbs                       | 30.23    | ‘for’ as a preposition             | 94.23    |
| singular common nouns                             | 28.25    | past participle of ‘to be’         | 74.75    |
| past participle of ‘to be’                        | 28.17    | singular weekday nouns             | 73.29    |
| first-person singular subjective personal pronoun | 27.99    | base form of lexical verbs         | 68.06    |
| plural proper nouns                               | 26.75    | singular month nouns               | 63.10    |
| locative adverbs                                  | 26.70    | plural temporal nouns              | 62.65    |

Table 6

The first 20 key POS tags in the InstaMusic corpus compared to the BNC Samplers.

Results emerging from the two comparisons reflect both linguistic tendencies and content-specific features. For example, key parts of speech referring to temporal nouns (e.g.,

‘year’, ‘time’, ‘day’), cardinal and ordinal numbers, weekday nouns, proper nouns, and locative adverbs (e.g., ‘there’) are linked to the content of the posts, which present dates of music events and record releases.

The comparison with the BNC Written Sampler aligns with the results of the analysis described in Section 4.1.1. Key POS features include verbs in their base form (simple present or imperative), personal pronouns (particularly second- and first-person), possessive adjectives (e.g., ‘my’, ‘its’, ‘his’, ‘your’), demonstratives, ‘-ing’ participles of lexical verbs, prepositional adverbs and particles, comparative general adverbs, past participle of ‘to be’, and the locative adverb ‘here’. This suggests that although the Instagram posts selected for analysis are written, they differ from other written genres.

The comparison with the BNC Spoken Sampler reveals that the language of the InstaMusic corpus contains an excess of nouns, adjectives, and prepositions to be considered similar to spoken English.

Interestingly, several key POS features that emerged in the comparison with the Written Sampler and that were not linked to post content (e.g., possessive adjectives, ‘-ing’ participles of lexical verbs, past participle of ‘to be’, and the base form of lexical verbs) are also prominent in the comparison with the Spoken Sampler. It may, therefore, be hypothesised that these are distinctive features of Instagram posts. However, this hypothesis needs to be further validated through the comparisons of the InstaTravel and InstaCinema corpora with the BNC Samplers.

## 4.2. The InstaTravel corpus

The InstaTravel corpus has 38,547 running words, 4,757 word types and 294 posts. As with the InstaMusic corpus, the analysis proceeds with the identification of the frequency of entries and percentages of occurrence of conjunctions, pronouns, verbs, nouns, adjectives, adverbs, demonstratives, existential there, synthetic and analytic negation, phrasal verbs, contractions, relative clauses, *wh*-questions, final stranded prepositions, pied-piping in relative clauses, and participial clauses.

There are 17 different conjunctions with a total occurrence of 1,905 (4.8%). The most frequently occurring are coordinating conjunctions: ‘and’ (938 times – 2.4%), ‘but’ (176 times – 0.4%), ‘or’ (159 times – 0.4%), and ‘plus’ (20 times – 0.05%). The subordinating conjunctions are ‘if’ (165 times – 0.4%), ‘so’ (118 times – 0.3%), ‘when’ (110 times – 0.28%), ‘because’ (75 times – 0.2%), ‘as’ (39 times – 0.1%), ‘while’ (32 times – 0.08%), and ‘before’ (25 times – 0.06%).

First-, second-, and third-person singular pronouns have very similar frequency distributions. First- and third-person plural pronouns are also present, although less frequently. Frequency distribution and percentage values are reported in Table 7.

As already noted above, in highly interactive discourse first- and second-person pronouns directly reference the sender and the addressee, while third-person pronouns mark reference to persons outside of the immediate interaction, in this case the description of the whole travel experience.

There are 644 verbs with a total frequency of 6,500 tokens (16.8%). Only 330 verbs occur more than twice. In Table 8, the 30 most frequent verbs are listed (the frequencies of the verbs ‘to be’, ‘to have’, ‘can’, ‘would’ and ‘will’ also include the contracted and negative contracted forms).

| Pronoun | Frequency | %    | Pronoun | Frequency | %     |
|---------|-----------|------|---------|-----------|-------|
| I       | 813       | 2.1  | their   | 39        | 0.1   |
| you     | 769       | 2    | us      | 23        | 0.06  |
| it      | 435       | 1.1  | its     | 17        | 0.04  |
| my      | 333       | 0.8  | he      | 16        | 0.04  |
| your    | 196       | 0.5  | she     | 16        | 0.04  |
| we      | 160       | 0.4  | her     | 9         | 0.02  |
| they    | 103       | 0.26 | his     | 8         | 0.02  |
| me      | 95        | 0.24 | ours    | 3         | 0.007 |
| our     | 57        | 0.14 | him     | 3         | 0.007 |
| them    | 56        | 0.14 | yours   | 3         | 0.007 |

Table 7  
Personal pronouns and possessive adjectives and pronouns in the InstaTravel corpus.

| Verb | Entries | %    | Verb   | Entries | %    | Verb   | Entries | %    |
|------|---------|------|--------|---------|------|--------|---------|------|
| is   | 701     | 1.8  | go     | 111     | 0.28 | find   | 52      | 0.13 |
| are  | 317     | 0.8  | travel | 95      | 0.24 | make   | 51      | 0.13 |
| have | 273     | 0.7  | see    | 90      | 0.23 | stay   | 50      | 0.12 |
| can  | 231     | 0.6  | want   | 80      | 0.2  | know   | 48      | 0.12 |
| get  | 184     | 0.47 | been   | 72      | 0.18 | save   | 47      | 0.12 |
| was  | 176     | 0.45 | has    | 70      | 0.18 | love   | 46      | 0.11 |
| be   | 160     | 0.4  | would  | 70      | 0.18 | think  | 46      | 0.11 |
| will | 154     | 0.4  | take   | 67      | 0.17 | hiking | 45      | 0.11 |
| do   | 130     | 0.3  | going  | 63      | 0.16 | use    | 41      | 0.10 |
| am   | 125     | 0.3  | had    | 59      | 0.15 | eat    | 41      | 0.10 |

Table 8  
The 30 most frequent verbs in the InstaTravel corpus.

Among the 30 most frequent verbs, three groups of verbs can be identified: auxiliaries ('be', 'have', 'can', 'will', 'do', 'would'), generic high-frequency verbs that derive their meanings from their co-text ('do', 'get', 'make', 'take'), private verbs ('want', 'hope', 'think', 'love'), and verbs specifically related to the topic of posts ('go', 'travel', 'see', 'find', 'stay', 'save', 'use', 'eat').

As for the verb 'be', when in the simple present, in most cases (70%), it is used in predicative expressions, which are associated with fragmented presentation of information (Biber 1995, p. 106). The verb 'do' is mainly used as a main verb with generic meaning and less frequently as an auxiliary. Involved production is suggested by the presence of private verbs such as 'want', 'hope', 'think', and 'love', used by the two travel influencers to express their feelings and attitudes towards what is described. As for verb tenses, among the 70 most frequent verbs (with a range of occurrence going from 701 to 15), 48 are in the base form (mainly present simple or imperative), 12 are in the past, 10 are in the '-ing' form.

There are 2,821 nouns with a total frequency of 10,700 tokens (28%). About 1,300 nouns occur more than twice. The 30 most frequent nouns prominently include two types of nouns: temporal nouns and nouns generically referring to travelling.

| Noun   | Frequency | %    | Noun   | Frequency | %    |
|--------|-----------|------|--------|-----------|------|
| travel | 132       | 0.34 | year   | 58        | 0.15 |
| trip   | 123       | 0.3  | food   | 54        | 0.14 |
| time   | 117       | 0.3  | places | 48        | 0.12 |
| day    | 111       | 0.28 | tips   | 45        | 0.11 |
| people | 94        | 0.24 | area   | 44        | 0.11 |
| park   | 85        | 0.2  | life   | 44        | 0.11 |
| city   | 84        | 0.2  | lake   | 42        | 0.1  |
| tour   | 68        | 0.17 | home   | 41        | 0.1  |
| trail  | 66        | 0.17 | road   | 41        | 0.1  |
| world  | 61        | 0.15 | summer | 41        | 0.1  |

Table 9  
The 20 most frequent nouns in the InstaTravel corpus.

There are 689 adjectives with a total frequency of 2,700 (7%). They are qualifying adjectives used to describe destinations, temporal adjectives referring to the trip, and adjectives indicating dimensions and measures. As noticed by Biber (1995, p. 105), adjectives are frequently used in texts whose aim is informational, as they contribute to nominal elaboration. Table 10 includes the 20 most frequent qualifying adjectives in the InstaTravel corpus.

| Adjective | Entries | %    | Adjective  | Entries | %    |
|-----------|---------|------|------------|---------|------|
| good      | 160     | 0.4  | perfect    | 31      | 0.08 |
| favorite  | 65      | 0.16 | easy       | 31      | 0.08 |
| new       | 57      | 0.14 | local      | 30      | 0.08 |
| great     | 52      | 0.13 | small      | 28      | 0.07 |
| free      | 40      | 0.1  | incredible | 26      | 0.06 |
| last      | 38      | 0.09 | first      | 25      | 0.06 |
| next      | 38      | 0.09 | super      | 23      | 0.06 |
| beautiful | 36      | 0.09 | cool       | 23      | 0.06 |
| cheap     | 32      | 0.08 | amazing    | 22      | 0.05 |
| long      | 31      | 0.08 | full       | 22      | 0.05 |

Table 10  
The 20 most frequent adjectives in the InstaTravel corpus.

There are 233 different adverbs with a total frequency of 2,100 (5.44%). The same categories identified in the InstaMusic corpus are found here, namely adverbs of time, locative adverbs, amplifiers, emphatics, hedges, downtoners, and adverbs of addition and extent. The most frequent adverbs of time are ‘always’ (66 entries – 0.17%), ‘never’ (46 entries – 0.11%), ‘still’ (30 entries – 0.08%), ‘early’ (21 entries – 0.05%), ‘often’ (19 entries – 0.05%), ‘ever’ (17 entries – 0.04%), ‘soon’ (11 entries – 0.02%), ‘already’ (10 entries – 0.02%), and ‘recently’ (10 entries – 0.02%). There are four locative adverbs, namely ‘here’ (78 entries – 0.2%), ‘there’ (58 entries – 0.15%), ‘away’ (17 entries – 0.04%), and ‘home’ (15 entries – 0.03%). The most frequent amplifiers are ‘even’ (33 entries – 0.08%), ‘very’ (27 entries – 0.07%), ‘definitely’ (22 entries – 0.05%), ‘absolutely’ (13 entries – 0.03%), ‘highly’ (16 entries – 0.04%), and ‘totally’ (6 entries – 0.01%), while emphatics are ‘so’ (74 entries – 0.2%), ‘really’ (67 entries – 0.17%), ‘most’ (66 entries – 0.17%), ‘more’ (14 entries – 0.03%), and ‘just’ (6 entries – 0.01%). Hedges include ‘maybe’ (13 entries – 0.03%) and ‘probably’ (10 entries – 0.02%), while ‘only’ (40 entries – 0.01%) is the only downtoner identified. ‘Enough’ (17 entries – 0.04%) is

used to describe extent, while adverbs of addition comprise ‘too’ (55 entries – 0.14%) and ‘also’ (49 entries – 0.12%).

The demonstrative ‘this’ occurs frequently, that is, with 321 entries (0.8%), 241 times as a demonstrative adjective and 80 times as a demonstrative pronoun. ‘That’ occurs 71 times (0.18%), 30 times as an adjective and 41 times as a pronoun. ‘These’ occurs 51 times (0.13%), 41 times as an adjective and 10 times as a pronoun, while ‘those’ occurs 22 times (0.05%), 14 times as an adjective and 8 times as a pronoun. As mentioned above, the frequent use of demonstratives is connected with interactive and real-time processing of information.

Existential *there*, used to introduce a new entity, occurs with 81 entries (0.2%), and this implies a significant presence of constructions that establish or highlight the existence of something in the discourse, emphasising their existence within the narrative or description. This pattern may indicate a preference for setting the scene or adding new elements to the discussion, reinforcing the idea that the focus is on presenting information rather than just describing existing situations.

Analytic negation can be found 296 times (0.76%) and, for this reason, it is preferred to synthetic negation occurring 146 times (0.37%). The total frequency of phrasal verbs is 296 (0.76%), and the most frequent phrasal verbs are those combined with ‘up’ (93 entries) and ‘out’ (101 entries). As for relative clauses, *that* clauses and *wh*-clauses with ‘who’, ‘which’, and ‘what’ occur with 277 entries (0.71%), while participial clauses are absent. There are 20 cases (0.05%) of final stranded prepositions, while pied-piping in relative clauses can be observed in 6 cases (0.01%). *Wh*-questions have 40 occurrences (0.1%).

Figure 2 shows the most representative linguistic markers in the InstaTravel corpus, along with the percentage values of their occurrence.

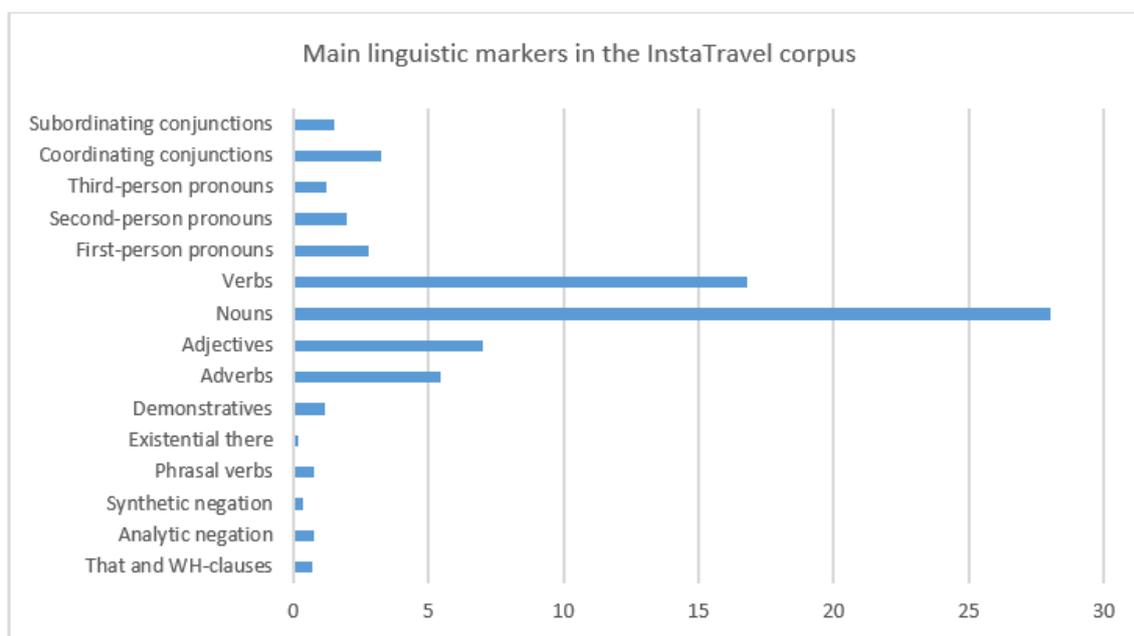


Figure 2  
Main linguistic markers in the InstaTravel corpus.

#### 4.2.1. Discussion of findings in the InstaTravel corpus

The linguistic markers associated with involved production exhibit the highest frequency of occurrence. Coordinating conjunctions, which serve as markers of fragmentation, are used more frequently than subordinating conjunctions. Among the latter, the conditional subordinating conjunction ‘if’, often used to indicate stance, has the highest frequency. First-person pronouns are preferred over second- and third-person pronouns, even though all three function as markers of involved and interactive discourse. As previously observed in the InstaMusic corpus, personal pronouns and contractions are characteristic of interactive discourse (Biber 1995, pp. 19, 105).

The main verb ‘be’ is the most frequently occurring verb, primarily used in predicative expressions, thereby contributing to the fragmented presentation of information (Biber 1995, p. 106). Involved production is also evident in the frequent use of private verbs such as ‘want’, ‘hope’, ‘think’, and ‘love’, which the two travel influencers use to express emotions and attitudes towards the described content.

Demonstratives, generic verbs like ‘do’, ‘get’, ‘make’, and ‘take’, as well as phrasal verbs, appear frequently in the corpus. Their high frequency suggests reduced surface forms and a general informality in the posts. The most common verb tense in the InstaTravel corpus is the base form, mainly present simple or imperative, while past tenses are used less often.

Among adverbs, emphatics and locative adverbs are the most frequent, with ‘so’, ‘here’, ‘really’, and ‘there’ showing particularly high percentages. The temporal adverb ‘always’ is also common. This combination of emphatics and temporal adverbs reflects an intention to convey strong emotions, highlight experiences, and create a vivid sense of place, while also signalling informality and engagement with the topic (Biber 1995, p. 241; Chafe 1982).

Additionally, the high frequencies of *that* clauses and *wh*-clauses support Biber’s (1995, p. 230) observation that these subordinating constructions frequently co-occur with interpersonal and reduced-content features such as first- and second-person pronouns, questions, contractions, hedges, and emphatics. These subordinators are commonly found in spoken genres, both interactional (e.g., conversation) and informational (e.g., speeches), but occur only rarely in informational written genres. Finally, the high frequency of nouns and adjectives in the selected Instagram posts indicates a high informational load; however, the posts’ reduced grammatical structure sets them apart from other written genres.

#### 4.2.2. The InstaTravel corpus and the BNC Samplers

Table 11 summarises the results of the comparison with the BNC Samplers Written and Spoken.

Some key POS tags appear in both comparisons, as they are connected to the content-specific features of the posts. These include singular and plural proper nouns, temporal and locative nouns, singular and plural units of measurement, and singular cardinal numbers. The unusually high frequency of these elements can be attributed to the nature of the posts, which describe travel experiences and therefore frequently refer to place names, distances, times, days, months, and numbered lists of tips or suggestions.

The comparison with the BNC Written Sampler reveals an unusually high frequency of lexical verbs in their base form (simple present or imperative), second-person pronouns, the verb ‘do’ (in its base form, ‘-ing’ form, and infinitive), superlative general adjectives, comparative general adverbs, subordinating conjunctions, and the base form of

the verb ‘have’. These findings align with the observations discussed in Section 4.2.1: although the Instagram posts analysed are written texts, they differ from traditional written genres. This distinction is evident in the choice of verb tenses, which reflect a present-oriented mode of communication, in the use of generic verbs, which suggests a fragmented and informal presentation of information, and in the frequent use of the second-person pronoun, which implies a high level of interactivity.

| Key POS with BNC Written                              |          | Key POS with BNC Spoken            |          |
|---|----------|------------------------------------|----------|
| POS   | LL value | POS                                | LL value |
| lexical verbs in the base form                        | 12632.93 | singular proper nouns              | 52194.25 |
| singular proper nouns                                 | 11977.89 | singular common nouns              | 9398.55  |
| 2nd person personal pronoun (‘you’)                   | 2104.98  | lexical verbs in the base form     | 4922.36  |
| plural proper nouns                                   | 1079.61  | general adjectives                 | 3393.66  |
| ‘do’, base form (finite)                              | 404.62   | singular units of measurement      | 2111.16  |
| singular temporal nouns                               | 361.03   | general prepositions               | 1955.68  |
| ‘wh’- general adverbs (‘where’, ‘when’, ‘why’, ‘how’) | 351.00   | plural proper nouns                | 1403.77  |
| cardinal numbers                                      | 276.84   | plural common nouns                | 1369.66  |
| ‘do’, infinitive                                      | 196.20   | superlative general adjectives     | 714.97   |
| superlative general adjectives                        | 187.12   | singular cardinal numbers          | 629.94   |
| singular locative nouns                               | 155.18   | singular locative nouns            | 590.67   |
| comparative general adverbs                           | 154.19   | singular temporal nouns            | 582.77   |
| base form of ‘be’                                     | 136.61   | possessive pronouns and adjectives | 433.12   |
| singular cardinal numbers                             | 95.48    | comparative general adverbs        | 281.21   |
| plural units of measurement                           | 78.17    | singular month nouns               | 254.18   |
| locative adverbs                                      | 73.63    | ‘for’ as a preposition             | 213.60   |
| singular units of measurement                         | 61.64    | plural units of measurement        | 196.92   |
| subordinating conjunctions                            | 59.66    | ‘be’ infinitive                    | 191.81   |
| ‘doing’   | 52.46    | superlative degree adverbs         | 155.15   |
| ‘have’, base form                                     | 41.35    | plural temporal nouns              | 136.63   |

Table 11

The first 20 key POS tags in the InstaTravel corpus compared to the BNC Samplers.

The comparison with the BNC Spoken Sampler suggests that the InstaTravel corpus displays an unusually high occurrence of singular and plural common nouns, lexical verbs in the base form, prepositions, superlative and comparative degree adverbs, and the verb ‘be’ in the infinitive form. As previously noted in the InstaMusic corpus, lexical verbs in the base form emerge as key POS features in both comparisons, which may lead us to hypothesise that this is a characteristic feature of Instagram posts. A comparison between the InstaCinema corpus and the BNC samplers may further confirm this trend and strengthen this assumption.

### 4.3. The InstaCinema corpus

The InstaCinema corpus has 26,088 running words, 4,597 word types and 167 posts.

There are 14 different conjunctions, and the most frequently occurring are coordinating conjunctions: ‘and’ (678 times – 2.6%), ‘but’ (185 times – 0.7%), ‘or’ (87 times – 0.3%). The subordinating conjunctions are ‘if’ (74 times – 0.3%), ‘while’ (60 times – 0.2%), ‘when’ (37 times – 0.14%), ‘because’ (27 times – 0.1%), ‘as’ (16 times – 0.06%), ‘although’ (14 times – 0.05%), ‘though’ (10 times – 0.03%), ‘whether’ (9 times – 0.03%), ‘after’ (4 times – 0.01%), and ‘before’ (2 times – 0.007%).

The third-person singular pronoun ‘it’ has the highest frequency among pronouns (422

entries – 1.6%), probably due to the fact that posts mainly review films. The second most frequent pronoun is ‘I’ (355 entries – 1.36%) followed by ‘you’ (147 entries – 0.56%), and ‘they’ (93 entries – 0.35%). Table 12 sums up all the frequencies of occurrence in raw figures and percentage values of personal subject and object pronouns, and possessive adjectives and pronouns.

In the InstaCinema corpus, there are 775 verbs with a total frequency of 4,400 tokens (17%). The verb ‘be’, in the present simple, past simple, and infinitive, is the preferred choice in the posts, with a total percentage value of 3.6%. It is followed by ‘have’, ‘had’, and ‘has’ (0.9%). Several modal verbs (‘would’, ‘can’, ‘could’, ‘will’, ‘might’, and ‘should’) account for 0.8% of total usage. The generic verbs ‘do’ and ‘make’ together account for 0.6%, while specific verbs related to the content of the posts, such as ‘see’, ‘say’, ‘offer’, ‘watch’, and ‘add’, make up 0.5%. The private verbs ‘think’ and ‘feel’ (the latter used in the sense of ‘think’ or ‘believe’) also have a frequency of 0.5%. The verb ‘be’, in most cases (70%), is used in predicative expressions. As for tenses, among the 70 most frequent verbs (with a range of occurrence going from 512 to 8), 36 are in the base form (present simple, imperative or infinitive), 20 are in the past, 9 are in the ‘-ing’ form.

Table 13 summarises the 30 most frequent verbs appearing in the wordlist.

| Pronoun | Frequency | %    | Pronoun | Frequency | %    |
|---------|-----------|------|---------|-----------|------|
| it      | 422       | 1.60 | her     | 35        | 0.13 |
| I       | 355       | 1.36 | we      | 34        | 0.13 |
| its     | 159       | 0.60 | your    | 34        | 0.13 |
| you     | 147       | 0.56 | me      | 33        | 0.12 |
| his     | 115       | 0.40 | she     | 24        | 0.09 |
| my      | 94        | 0.36 | them    | 20        | 0.07 |
| they    | 93        | 0.36 | our     | 10        | 0.03 |
| he      | 74        | 0.30 | us      | 9         | 0.03 |
| their   | 68        | 0.26 | him     | 8         | 0.03 |

Table 12  
Personal pronouns and possessive adjectives and pronouns in the InstaMusic corpus.

| Verb  | Entries | %    | Verb  | Entries | %    | Verb     | Entries | Entries |
|-------|---------|------|-------|---------|------|----------|---------|---------|
| is    | 512     | 2    | does  | 50      | 0.2  | will     | 27      | 0.1     |
| are   | 206     | 0.8  | were  | 47      | 0.18 | might    | 24      | 0.09    |
| was   | 201     | 0.8  | can   | 46      | 0.18 | say      | 24      | 0.09    |
| have  | 124     | 0.5  | seen  | 37      | 0.14 | felt     | 23      | 0.09    |
| be    | 104     | 0.4  | feel  | 36      | 0.14 | made     | 22      | 0.08    |
| would | 69      | 0.26 | could | 36      | 0.14 | offer    | 22      | 0.08    |
| had   | 58      | 0.2  | been  | 30      | 0.11 | watching | 19      | 0.07    |
| think | 55      | 0.2  | did   | 30      | 0.11 | see      | 17      | 0.06    |
| has   | 51      | 0.2  | being | 28      | 0.1  | should   | 17      | 0.06    |
| do    | 50      | 0.2  | feels | 28      | 0.1  | add      | 16      | 0.06    |

Table 13  
The 30 most frequent verbs in the InstaMusic corpus.

Nouns include 2,000 types, with a total frequency of 6,200 (23.7%), and 847 of them occur more than twice in this corpus (Table 14). All the nouns included in the list refer to the topic that posts focus on, namely films. The only exception is ‘time’, that is used in expressions such as ‘every time’, ‘the first time’, ‘spend/worth your time’, ‘a very long time’, and ‘at the same time’.

| Noun         | Frequency | %    | Noun        | Frequency | %    |
|--------------|-----------|------|-------------|-----------|------|
| film         | 237       | 0.90 | plot        | 34        | 0.13 |
| movie        | 116       | 0.40 | life        | 33        | 0.13 |
| story        | 98        | 0.37 | movies      | 30        | 0.11 |
| film         | 66        | 0.25 | show        | 28        | 0.10 |
| time         | 60        | 0.22 | action      | 26        | 0.09 |
| character    | 45        | 0.17 | performance | 26        | 0.09 |
| performances | 45        | 0.17 | sense       | 26        | 0.09 |
| characters   | 44        | 0.17 | people      | 25        | 0.09 |
| audience     | 40        | 0.15 | dialogue    | 24        | 0.09 |
| season       | 40        | 0.15 | scene       | 24        | 0.09 |

Table 14  
The 20 most frequent nouns in the InstaMusic corpus.

There are 860 adjective types, totaling 2,600 tokens (10%). Table 15 includes the 20 most frequent qualifying adjectives in this corpus.

| Adjective | Entries | %    | Adjective    | Entries | %    |
|-----------|---------|------|--------------|---------|------|
| narrative | 52      | 0.2  | entire       | 19      | 0.07 |
| first     | 48      | 0.18 | favorite     | 18      | 0.06 |
| great     | 36      | 0.13 | profound     | 18      | 0.06 |
| best      | 30      | 0.11 | short        | 18      | 0.06 |
| original  | 30      | 0.11 | cinematic    | 17      | 0.06 |
| emotional | 24      | 0.09 | interesting  | 17      | 0.06 |
| thematic  | 23      | 0.09 | bad          | 16      | 0.06 |
| real      | 22      | 0.08 | strong       | 16      | 0.06 |
| new       | 21      | 0.08 | incredible   | 15      | 0.05 |
| better    | 20      | 0.07 | entertaining | 14      | 0.05 |

Table 15  
The 20 most frequent adjectives in the InstaMusic corpus.

Films and plots are evaluated from various perspectives, ranging from emotional depth and thematic richness to overall quality and cinematic impact. Adjectives like ‘great’, ‘best’, ‘strong’, ‘incredible’, and ‘favorite’ are used to evaluate the overall quality of a film, often reflecting personal preferences or broader critical reception. Adjectives such as ‘emotional’, ‘profound’, ‘interesting’, and ‘entertaining’ describe how a film engages the audience, while ‘thematic’ and ‘narrative’ describe the film’s deeper meaning and storytelling structure. Finally, adjectives such as ‘original’, ‘new’, and ‘cinematic’ point to a film’s innovation.

There are 330 different adverbs, which reach a total frequency of 2,000 (7.6%). The most frequent categories are amplifiers, emphatics, hedges, and downtoners. The most frequent amplifiers are ‘even’ (41 entries – 0.15%), ‘absolutely’ (34 entries – 0.13%), ‘completely’ (30 entries – 0.11%), ‘very’ (37 entries – 0.14%), ‘truly’ (19 items – 0.07%), ‘fully’ (16 entries – 0.06%), ‘highly’ (13 entries – 0.05%), ‘incredibly’ (13 entries – 0.05%), ‘particularly’ (13 entries – 0.05%), and ‘entirely’ (6 entries – 0.01%). Emphatics include ‘most’ (55 entries – 0.14%), ‘more’ (41 entries – 0.15%), ‘really’ (29 entries – 0.11%), and ‘so’ (23 entries – 0.08%). Hedges are made up of ‘perhaps’ (25 entries – 0.09%), ‘almost’ (18 entries – 0.06%), and ‘maybe’ (12 entries – 0.04%). Downtoners comprise ‘only’ (45 entries – 0.17%), ‘nearly’ (4 entries – 0.01%), and ‘somewhat’ (4 entries – 0.01%). Adverbs of time include ‘still’ (28 entries – 0.1%), ‘never’ (18 entries – 0.06%), ‘always’ (16 entries – 0.06%), and ‘ever’ (15 entries – 0.05%), while locative

adverbs consist of ‘here’ (38 entries – 0.14%), and ‘away’ (23 entries – 0.08%). A frequent adverb of addition is ‘also’ (44 entries – 0.16).

As for demonstratives, ‘this’ has 342 occurrences (1.3%), 219 times as a demonstrative adjective and 123 times as a demonstrative pronoun, whereas ‘that’ is used 73 times (0.27%), 27 times as an adjective and 46 times as a pronoun. ‘These’ occurs 31 times (0.11%), 29 times as an adjective and twice as a pronoun, while ‘those’ occurs 23 times (0.08%), 10 times as an adjective and 13 times as a pronoun. Existential *there* has 45 occurrences (0.17%), while analytic negation can be found 227 times (0.87%), slightly preferred to synthetic negation occurring 163 times (0.6%).

Phrasal verbs occur 133 times (0.5%), and the most frequent phrasal verbs are those combined with ‘out’ (37 entries – 0.1 %) and ‘up’ (25 entries – 0.09%). There are 427 occurrences (1.6%) of *that* clauses and *wh*-clauses with ‘who’, ‘which’, and ‘what’, while participial clauses have 26 occurrences (0.09%). Final stranded prepositions are rare (only 5 cases are observed – 0.01%), while pied-piping in relative clauses can be observed in 40 cases (0.15%). *Wh*-questions can be observed in 33 cases (0.12%).

Figure 3 summarises the most frequent linguistic markers identified in the analysis of the InstaCinema corpus.

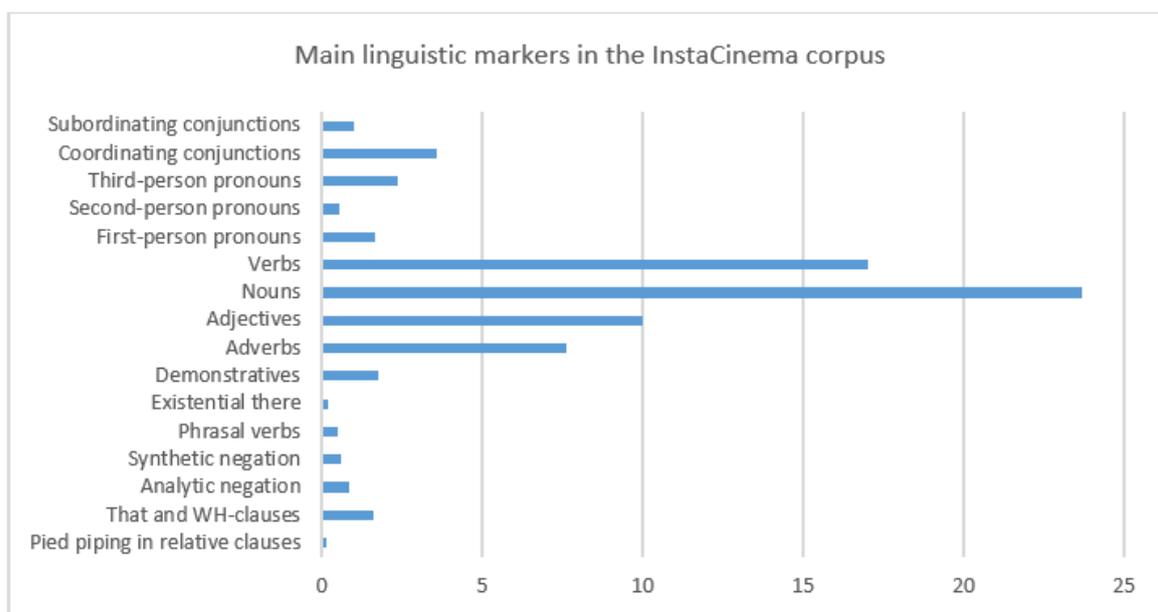


Figure 3  
Main linguistic markers in the InstaCinema corpus.

#### 4.3.1. Discussion of findings in the InstaCinema corpus

As already observed in the other two corpora, features of fragmentation/integration and involvement/detachment are also evident in Figure 3. Fragmentation is marked by clauses occurring in succession without connectives or joined by coordinating conjunctions, which, in this corpus, are used more frequently than subordinating conjunctions. Involvement is reflected in linguistic features that highlight interaction between participants in the communicative context, such as personal pronouns, and in the expression of personal thoughts and feelings, evidenced by the high frequency of first-person pronouns, amplifiers, emphatics, and private verbs such as ‘think’ and ‘feel’ (the latter used in the sense of ‘think’ or ‘believe’).

Generic verbs like ‘do’ and ‘make’, along with phrasal verbs and demonstratives, also occur frequently in the corpus. Their prominence contributes to reduced surface forms and general informality. Furthermore, the co-occurrence of *that* clauses and *wh*-clauses with these reduced-content features serves as further evidence of fragmentation. However, this fragmented and imprecise structure is counterbalanced by the frequent use of nominalisation, which reflects the informational intent of the posts and the tendency to create detachment. Integration is also evident in the high frequency of adjectives, linguistic features that contribute to packaging information within the text (Biber 1995, p. 21).

#### 4.3.2. The InstaMusic corpus and the BNC Samplers

The comparison of the InstaCinema corpus with the two BNC Samplers reveals its key POS tags. The first 20 tags are reported in Table 16 in order of log-likelihood value.

| Key POS with BNC Written  |          | Key POS with BNC Spoken             |          |
|---|----------|-------------------------------------|----------|
| POS   | LL value | POS                                 | LL value |
| general adverbs   | 933.32   | general adjectives                  | 2787.41  |
| ‘-s’ form of lexical verb   | 369.83   | singular common nouns               | 2293.38  |
| general adjectives  | 350.01   | article                             | 845.43   |
| singular determiner (e.g., ‘this’, ‘that’, ‘another’)                         | 298.39   | plural common nouns                 | 825.73   |
| possessive pronouns   | 247.59   | possessive pronouns                 | 716.90   |
| indefinite pronouns, singular (e.g., ‘anyone’, ‘everything’, ‘nobody’, ‘one’) | 155.52   | ‘of’ (as preposition)               | 700.18   |
| ‘-ing’ participle of lexical verb   | 139.38   | general prepositions                | 649.18   |
| base form of lexical verbs  | 135.22   | ‘-s’ form of lexical verbs          | 620.29   |
| singular article  | 128.19   | general adverbs                     | 600.51   |
| 1st person sing. subjective personal pronoun                                  | 115.54   | singular proper nouns               | 451.75   |
| singular common nouns   | 90.86    | singular article                    | 441.49   |
| ‘be’, base form   | 88.02    | ‘-ing’ participle of lexical verbs  | 282.84   |
| ‘does’  | 75.11    | superlative general adjectives      | 187.32   |
| superlative general adjectives  | 74.47    | comparative degree adverbs          | 178.90   |
| ‘is’  | 65.67    | superlative degree adverbs          | 160.11   |
| adversative coordinating conjunction (‘but’)                                  | 59.93    | ‘with’, ‘without’ (as prepositions) | 152.76   |
| 3rd person sing. neuter personal pronoun (‘it’)                               | 52.91    | coordinating conjunctions           | 130.90   |
| degree adverbs  | 51.06    | infinitive marker (‘to’)            | 104.41   |
| comparative degree adverb   | 44.78    | past participle of lexical verbs    | 66.38    |
| infinitive marker (‘to’)  | 42.52    | singular reflexive personal pronoun | 57.68    |

Table 16

The first 20 key POS tags in the InstaMusic corpus compared to the BNC samplers.

As shown in Table 16, many key POS tags are common to both comparisons, although they differ in their log-likelihood values. Linguistic markers that show unusually high frequency in both spoken and written genres include general adverbs, the ‘-s’ form of lexical verbs, general adjectives, possessive pronouns, the ‘-ing’ participle of lexical verbs, the singular article, singular common nouns, superlative general adjectives, comparative degree adverbs, and the infinitive marker ‘to’.

Some of these tags can be better understood when considered in the context in

which they are used. The posts in this corpus describe and review films, and this accounts for the unusual presence of subjectivity markers such as adjectives, possessive pronouns, and adverbs, particularly amplifiers and emphasisers. Descriptions and reviews typically involve a high level of nominalisation, which explains the notably high frequency of common nouns.

The comparison with the BNC Written Sampler reveals an unusually high frequency of first- and third-person singular pronouns. Consequently, there is also a higher frequency of lexical verbs conjugated in the third-person singular, as well as the verbs ‘do’ and ‘be’ in the same form. Lexical verbs in their base form (simple present or imperative) also emerge as key tags in this comparison, a trend already noted in the two previously analysed corpora. Additionally, the base form of ‘be’ appears to be a remarkable feature, with its high frequency (4.4%) already discussed in Section 4.3.

The unusual frequency of the adversative coordinating conjunction ‘but’ can probably be attributed to the argumentative nature of the posts, in which pros and cons of films are discussed.

These features distinguish the posts from the written genres included in the BNC Written Sampler, particularly due to the frequent use of ‘be’ and ‘do’, which contributes to a fragmented and informal presentation of information. Moreover, the prevalence of base-form lexical verbs suggests a lack of narrative structuring, further differentiating these posts from conventional written texts.

The comparison with the BNC Spoken Sampler reveals an unusually high frequency of prepositions such as ‘of’, ‘with’, and ‘without’, which typically serve to condense information and are, therefore, not commonly associated with spoken English. Furthermore, the presence of superlative adverbs and possessive pronouns as key POS tags suggests that the level of subjectivity conveyed in these posts is unusually high—even for spoken genres, where involvement and personal opinion are expected.

## 5. Concluding remarks

This paper set out to identify the distinctive lexical and grammatical features of Instagram posts and to determine whether such texts more closely resemble typical spoken or written varieties. Before moving to the discussion of the findings, it is important to acknowledge the limitations of this study. Notably, the relatively small size of the corpora restricts the generalisability of the results, and larger and more varied datasets would be required to draw more robust conclusions. Furthermore, the reference corpus used to detect key POS tags may not reflect the specific varieties of English used by Instagram users, which could pose potential constraints. However, the findings offer insights into the hybrid nature of language use on social media platforms, suggesting meaningful linguistic tendencies that warrant further investigation.

Regarding the first research question, the analysis of the three corpora reveals several shared linguistic features, suggesting that Instagram posts are characterised by both involved production and informational orientation. All three corpora highlight the frequent use of markers indicative of interactive and real-time language, such as first- and second-person pronouns and contractions, which contribute to an informal and engaged tone. They also emphasise fragmentation due to the prevalent use of coordinating conjunctions, base-form lexical verbs (i.e., simple present, imperative, or infinitive), and reduced surface forms marked by generic verbs, phrasal verbs, demonstratives, and *that* clauses or *wh*-clauses. At the same time, an informational focus is evident in each corpus through the high frequency of nouns, adjectives, and nominalisations, elements that pack

dense referential content into the text. This combination of involved production markers (e.g., interactive personal pronouns and subjectivity indicators) with features that serve to condense and clarify information (e.g., nominalisations and integrated adjectives) suggests that, despite variations in genre specifics, the Instagram posts included in the three corpora consistently embody a blend of real-time, interactive communication and an efficient, information-dense style.

The comparisons of the three corpora with the spoken and written components of the British National Corpus identified shared and distinct linguistic traits differentiating Instagram posts from traditional genres.

The InstaMusic corpus displays shared POS features with the two Samplers, which include temporal nouns, cardinal/ordinal numbers, weekday and proper nouns, and locative adverbs, reflecting the content of posts (e.g., event dates). Compared to the BNC Written Sampler, the posts show frequent use of base-form verbs, personal/possessive pronouns, demonstratives, and ‘-ing’ participles, pointing to an informal, interactive written style. When compared to spoken data, InstaMusic shows an overuse of nouns, adjectives, and prepositions, making it less similar to spoken English. Some POS tags (e.g., possessive adjectives, ‘-ing’ participles, base-form verbs) are common to both comparisons, suggesting they may be specific to Instagram language.

The InstaTravel corpus compared to the BNC Samplers shows content-driven tags including proper nouns, locative/temporal nouns, units of measurement, and cardinal numbers, mirroring the descriptive and informative nature of travel posts. The written comparison highlighted frequent base-form verbs, second-person pronouns, verbs like ‘do’ and ‘have’, superlative adjectives, and subordinating conjunctions, indicating fragmented, present-focused discourse. The comparison with the Spoken Sampler again revealed frequent common nouns, base-form verbs, prepositions, and superlative adverbs. Like InstaMusic, base-form lexical verbs emerge as a recurring feature in both comparisons, reinforcing their potential status as a distinct marker of Instagram discourse.

The InstaCinema corpus compared to the BNC Samplers presented POS tags common to both written and spoken genres, namely adverbs, ‘-s’ verb forms, adjectives, possessive pronouns, ‘-ing’ participles, superlatives, and common nouns, indicating subjectivity and evaluative language, typical of film reviews. The comparison with the Written Sampler showed high use of first-/third-person pronouns, lexical verbs in third-person, and verbs like ‘do’ and ‘be’, again pointing to an informal and fragmented register. The comparison with the Spoken Sampler registers a marked use of prepositions and superlative adverbs, marking the language as more complex and subjective than everyday speech.

Regarding the second research question, the present findings illustrate that, while Instagram posts are written texts, they do not conform to traditional written genres. Instead, they blend features of spoken language (personal pronouns, interactivity, and informality) with content-specific lexical patterns (e.g., time/place nouns in travel and music, adjectives in cinema). Across all the three corpora, lexical verbs in the base form stand out as a distinctive linguistic feature, appearing consistently in the comparison with the BNC Written Sampler and the comparison with the BNC Spoken Sampler and signalling a present-tense style that suits the platform’s engaging nature. Moreover, Instagram language shows an informal, often fragmented structure, relying on generic verbs and subjective markers, especially in evaluative contexts like cinema reviews. Interestingly, some features usually not typical of spoken English (e.g., nominalisation and prepositions for information compression) appear frequently, suggesting that Instagram may be considered a hybrid register, characterised by a concise, expressive, and informational style, which is functional for digital communication.

These findings may also serve as a foundation for more targeted investigations into

second language acquisition through social media. The language encountered in this context is practical, emotionally expressive, and topic-specific, blending features of everyday conversation with written promotional and descriptive discourse. It may therefore be hypothesised that exposure to Instagram content supports the development of familiarity with instructional, descriptive, evaluative, and promotional language, as well as conversational expressions, both general and domain-specific vocabulary, and varied ways of articulating opinions, preferences, and intensities. Nevertheless, the relationship between social media language and second language learning constitutes a complex domain that necessitates sustained and targeted investigation; however, it remains a field that is undoubtedly worthy of scholarly inquiry.

**Bionote:** Elena Manca is an Associate Professor of English Linguistics, Language and Translation at the University of Salento (Lecce) and teaches ‘English Language and Linguistics’ and ‘Translation into/from English and Italian’ at the BA in Linguistic Mediation of the University of Salento. Her areas of research are Corpus Linguistics, specialised languages, cross-cultural studies, translation studies and language teaching. In particular, she has extensively investigated the linguistic, visual and cultural features of tourism discourse by cross-comparing several multimodal tourist texts in English and in Italian produced by a variety of cultures. She has proposed a four-step methodology for the identification of translation equivalents across cultures and she has elaborated the Cultural Communication Grammar framework for the classification of cultural features through the analysis of language and visuals. She has also developed language-teaching materials for familiarising students with Corpus Linguistics and with the notion of extended units of meaning and functionally complete units of meaning which are at the basis of the interplay between language and meaning.

**Author’s address:** [elena.manca@unisalento.it](mailto:elena.manca@unisalento.it)

**Acknowledgements:** This work has been supported by the PRIN project *The informalisation of English language learning through the media: Language input, learning outcomes and sociolinguistic attitudes from an Italian perspective* funded by the Italian Ministry for University and Research (MUR) – Bando 2020 – grant 2020NNJTW3\_001.

## References

- Aïmeur E., Amri S. and Brassard G. 2023, *Fake news, disinformation and misinformation in social media: a review*, in “Social Network Analysis and Mining” 13 [1], pp. 1-36.
- Alazemi A.F.T., Gheisari A. and Patra I. 2023, *The consequences of task-supported language teaching via social media on academic engagement, emotion regulation, willingness to communicate, and academic well-being from the lens of positive psychology*, in “Asian-Pacific Journal of Second and Foreign Language Education” 8, pp. 1-26.
- Aloraini N. and Cardoso W. 2022, *Social media in language learning: a mixed-methods investigation of students’ perceptions*, in “Computer Assisted Language Learning” 35 [8], pp.1707-1730.
- Benthous J., Risius M. and Beck R. 2016, *Social media management strategies for organizational impression management and their effect on public perception*, in “The Journal of Strategic Information Systems” 25 [2], pp. 127-139.
- Biber D. [1988] 1995, *Variation across speech and writing*, Cambridge University Press, Cambridge.
- Biber D. and Conrad S. 2009, *Register, genre, and style*, Cambridge University Press, Cambridge.
- Bianchi F. 2017, *The social tricks of advertising. Discourse strategies of English-speaking tour operators on Facebook*, in “Iperstoria” 10, pp. 3-32.
- Bianchi F. and Manca E. 2023, *Wine promotion on Facebook. A linguistic comparison of posts by producers from English-speaking countries*, in Cacchiani S. and Bondi M. *Positioning, point of view and ideology in the digital transformation*, in “Lingue e Linguaggi” 58 (Special Issue), pp. 247-269.
- Cade N.L. 2018, *Corporate social media: How two-way disclosure channels influence investors*, in “Accounting, Organizations and Society” 68, pp 63-79.
- Castañó-Pulgarín S.A., Suárez-Betancur N., Vega L.M.T. and López H.M.H. 2021, *Internet, social media and online hate speech. Systematic review*, in “Aggression and violent behaviour” 58, 101608.
- Ceron A., D’Adda G. 2016, *E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election*, in “New Media & Society” 18 [9], pp. 1935-1955.
- Chafe W.L. 1982, *Integration and involvement in speaking, writing, and oral literature*, in Tannen D. (ed.), *Spoken and written language: exploring orality and literacy*, Ablex, Norwood, N.J., pp. 35-54.
- Chalupnik M. and Brookes G. 2022, *Discursive acts of resistance: a multimodal critical discourse analysis of All-Poland Women’s Strike’s social media*, in “Gender and Language” 16 [3], pp. 308-333.
- Coesemans R. and De Cock B. 2017, *Self-reference by politicians on Twitter: Strategies to adapt to 140 characters*, in “Journal of Pragmatics” 116, pp. 37-50.
- Colliander J. 2019, *“This is fake news”. Investigating the role of conformity to other users’ views when commenting on and spreading disinformation in social media*, in “Computers in Human Behavior” 97, pp. 202-215.
- Esposito E. and KhosraviNik M. 2025, *Discourse in the digital age. Social media, power, and society*. Routledge, New York.
- Featherman C. 2018, *Discourses of ideology and identity. Social Media and the Iranian election protests*. Routledge, New York.
- Gabrielatos C. and Marchi A. 2012, September 13-14, *Keyness: Appropriate metrics and practical issues* [Conference presentation], CADS International Conference, *Corpus-assisted Discourse Studies: More than the sum of Discourse Analysis and computing?*, Bologna, Italy.
- Georgalou M. 2015, *Beyond the timeline: Constructing time and age identities on Facebook*, in “Discourse Context & Media” 9, pp. 24-33.
- Guiora A. and Park E.A. 2017, *Hate speech on social media*, in “Philosophia” 45, pp. 957-971.
- Hyland K. 2005, *Metadiscourse. Exploring Interaction in Writing*, Continuum, London and New York.
- Kilgarriff A., Rychlý P., Smrž P. and Tugwell D. 2004, *The Sketch Engine*, Proceedings of the 11th EURALEX International Congress, pp. 105-116. [www.sketchengine.eu](http://www.sketchengine.eu).
- Kreis R. 2017, *The “Tweet Politics” of President Trump*, in “Journal of Language and Politics” 16 [4], pp. 607-618.
- Manca E. 2021, *Australian and US wineries on Facebook: Analyzing interactive and interactional resources*, in D’Angelo L. and Consonni S. (eds.) *New Explorations in Digital Metadiscourse*, CERLIS Series, vol. 10, pp. 115-142.
- Manca E. 2024, *Social media and online informal learning of English. University language students’ preferences, attitudes, and perceptions*, in Pavesi M. and Bianchi F. (eds) *‘In the Presence of English’. Informal English Language Learning in Italy*, in “Lingue e Linguaggi” 66 (Special Issue), pp. 219-239.
- McDonnell A. 2020, *Clinton stated, Trump exclaimed! Gendered language on Twitter during the 2016 presidential debates*, in “Journal of Language and Politics” 19 [1], pp. 71-88.

- Moravec P.L., Minas R.K. and Dennis A.R. 2019, *Fake News on Social Media*, in "MIS quarterly" 43 [4], pp. A1-A13.
- Pavesi M., Ghia E. and Ciabattoni T. 2023, *Profiling informal experience with English. An instrument for data collection*, in "Lingue e Linguaggi" 59, pp. 303-332.
- Pavesi M. and Bianchi F. 2024, 'In the Presence of English'. *Informal English Language Learning in Italy*, in "Lingue e Linguaggi" 66 (Special Issue).
- Pavesi M., Bianchi F., Bruti S. and Zago R. forthcoming, *In contact with English, informal media-driven learning among Italian university students*, Carocci, Roma.
- Prichard C. 2013, *Training L2 learners to use Facebook appropriately and effectively*, in "CALICO Journal" 30 [2], pp. 204-225.
- Rayson P. 2008, *From key words to key semantic domains*, in "International Journal of Corpus Linguistics" 13 [4], pp. 519-549. <https://ucrel.lancs.ac.uk/wmatrix/>
- Rayson P. and Garside R. 2000, *Comparing corpora using frequency profiling*, in Kilgariff A. and Berber Sardinha T. (eds.), *Proceedings of the workshop on Comparing Corpora*, held in conjunction with The 38th annual meeting of the Association for Computational Linguistics, pp. 1-6, Association for Computational Linguistics. <https://aclanthology.org/volumes/W00-09/>
- Reinhardt J. 2019, *Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking*, in "Language Teaching" 52 [1], pp. 1-39.
- Schiffirin D. 1981, *Tense variation in narrative*, in "Language" 57, pp. 45-62.
- Schmid U.K., Kümpel A.S. and Rieger D. 2024, *How social media users perceive different forms of online hate speech: A qualitative multi-method study*, in "New media & society" 26 [5], pp. 2614-2632.
- Voramontri D. and Klieb L. 2019, *Impact of social media on consumer behaviour*, in "International Journal of Information and Decision Sciences" 11 [3], pp. 209-233.
- Zappavigna M. 2012, *Discourse of Twitter and social media: How we use language to create affiliation on the web*, Bloomsbury Publishing, London.