Original Article

Exploring agreement on social media: a small-scale corpus study with proposals for technical and analytical frameworks

Explorando o ato de concordar nas mídias sociais: um estudo de corpus de pequena escala com propostas de estruturas técnica e analítica

Elnaz Gholami (iD)

Universida de Valencia, Valencia, Spain
Email: elg@alumni.uv.es

Abstract

The article focuses on the methodology of creating a small-sized corpus from social media platforms like LinkedIn and Instagram to study speech acts, particularly the act of agreement, within the context of Spanish social networks. It aims to demonstrate the contributions of such a corpus to the understanding of speech acts while addressing the complexities and challenges associated with collecting conversational interactions and user data from these platforms, including factors such as representativeness, privacy, and platform limitations. The article proposes a technical sheet for capturing contextual information and users' sociolinguistic data, along with a pragmatic analysis sheet for the act of agreement. The latter delineates the variables and variants studied at the interactive and illocutionary levels, covering aspects such as the types of communicative turns seeking agreement, direct resources for expressing agreement, their functions, and complexities. Ultimately, the article seeks to enrich

the field of corpus pragmatics by providing a resource for researchers interested in annotating pragmatic corpora for agreement research.

Keywords

Corpus pragmatics; Social media; Agreement; Speech acts.

Resumo

O artigo foca na metodologia de criação de um corpus de pequenas dimensões a partir de plataformas de mídia social como LinkedIn e Instagram, com o objetivo de estudar atos de fala, particularmente o ato de concordar no contexto das redes sociais espanholas. O objetivo é demonstrar as contribuições desse corpus para a compreensão dos atos de fala, abordando as complexidades e os desafios associados à coleta de interações conversacionais e dados dos usuários dessas plataformas, incluindo questões como representatividade, privacidade e limitações da plataforma. O artigo propõe uma ficha técnica para capturar informações contextuais e dados sociolinguísticos dos usuários, juntamente com uma ficha de análise pragmática para o ato de concordar. Esta última delineia as variáveis e variantes estudadas nos níveis interativo e ilocucionário, cobrindo aspectos como os tipos de turnos comunicativos que buscam concordância, os recursos diretos para expressar concordância, suas funções e complexidades. Em última análise, o artigo visa enriquecer o campo da pragmática de corpus, fornecendo um recurso para pesquisadores interessados na anotação de corpora pragmáticos para o estudo do ato de concordar.

Palavras-chaves

Pragmática de corpus; Mídia social; Concordar; Atos de fala.

Introduction

The study of social media has evolved into an integral and significant component of linguistics, given its pervasive role as a tool for communication, social interaction, and information dissemination in today's digital age. There is growing interest in creating corpora from social media, as understanding the dynamics of language within these digital realms is crucial for comprehending the complexities of online interactions. Among various linguistic phenomena, the study of speech acts holds particular significance in deciphering the intricacies of social media discourse. By delving into the realm of speech acts on social media, researchers can gain valuable insights into how individuals negotiate meaning, express themselves, and establish connections within online communities.

This article aims to demonstrate how a small-sized corpus derived from social media platforms can enhance the study of speech acts, while also shedding light on the challenges associated with compiling a corpus in this domain. Our focus is on one of the most prevalent speech acts in everyday conversations: agreement, often regarded as the apex of discourse (cf. Briz, 2001), and investigated by several scholars, each probing different aspects of this phenomenon (cf. Baym, 1996; Guiller; Durndell, 2007; Bolander, 2012; Cantamutto; Fainstein, 2019; Vigara; Hernández, 2011).

This manuscript is part of the methodology adopted for a multilevel analysis of agreement on Spanish social networks. Here, we will elucidate the methodological approaches and corpus collection processes that guided us in developing an analysis sheet for exploring the mentioned act. In doing so, we aim to contribute to the burgeoning field of corpus pragmatics, offering insights that will facilitate the work of other scholars endeavoring to construct and annotate pragmatic corpora for agreement research.

The importance of a pragmatic corpus

A pragmatic analysis that is unbiased and valid must be carried out in the situated context and within the sociocultural system of a language. This has led many scholars to create various corpora of real speaker interactions (e.g., Corpus Val.Es.Co., COSER, PRESEEA) to study contextualized utterances and speaker goals. The significance of corpora in pragmatic analysis is highlighted by the emergence of Corpus Pragmatics, a field combining pragmatics and corpus linguistics, two approaches that were long considered mutually exclusive (cf. Rühlemann; Clancy, 2018).

Aijmer and Rühlemann (2015) formally articulated Corpus Pragmatics, combining qualitative and quantitative methods to study pragmatic concepts, form, function, and variation across social, cultural, and regional levels. This approach provides a "methodological framework that allows for the interpretation of spoken or written meaning, with an emphasis on providing empirical evidence for this interpretation" through the analysis of naturally occurring, uncensored data, aided by concordancing software (Clancy; O'Keeffe, 2015, p.235). As a result, it enriches pragmatic research by incorporating contextual elements such as language users, social situations, and activity types, thereby enhancing our understanding of how communication unfolds in real-life contexts.

Echoing Clancy and O'Keeffe (2015), in this subdiscipline "[i]n addition to the richness of the results, and the surprises they hold, the research creates a really useful and reusable resource through tagging and coding" (Clancy; O'Keeffe, 2015, p.251). To date, various studies have been conducted in the field of corpus pragmatics and the obtained results, such as Milà-Garcia's studies (2016, 2018) on agreement and

disagreement, confirm the viability and advantages of creating pragmatic corpora for studying speech acts. However, pragmatic annotation is time-consuming, labor-intensive, and subject to interpretation, requiring inter-rater validation (O'Keeffe; Clancy; Adolphs, 2019, p.60). Despite these challenges, the field aims to develop tools for automating the annotation process, moving closer to more efficient and inclusive methodologies (O'Keeffe; Clancy; Adolphs, 2019, p.67).

Development of small-sized corpus

There are numerous large corpora compiled from different discursive genres, which can be utilized for pragmatic studies. However, the vast size of these corpora can itself pose a problem for pragmatists for various reasons (cf. Milà-Garcia, 2018; Moreno-Ortiz; García-Gámez, 2023). Firstly, since they were not originally intended for pragmatic analyses, they may not include the necessary information for such studies. Additionally, they may suffer from either an insufficient or excessive presence of the trait being examined, and their pre-existing grammatical annotations may not align with the requirements of researchers, necessitating additional annotation (Milà-Garcia, 2018, p.2). As a result, small-sized corpora have begun to emerge with the following characteristics: they are usually designed for specific research purposes, often compiled by the researchers themselves (which facilitates familiarity with the samples) and tend to focus on the analysis of phenomena where context is particularly important (García Ramón, 2018, p.113, based on Walsh, 2013).

This type of corpus can easily offer a wealth of extralinguistic and discursive information, and as Albelda (2022) emphasizes, greater contextual detail leads to better access to discourse. Consequently, researchers gain deeper insights into the values that speakers assign to their usage preferences (Albelda; Álvarez, 2010, p.81), the patterns imposed by discursive genres, the internal structure of the speech acts studied, the type of relationship of greater distance or proximity between interlocutors, and other socio-situational factors that may affect communication (Albelda, 2022, p.229).

The other aspect we would like to mention is that small-sized corpora allow for the adoption of both a function-to-form approach (common in pragmatics) and a form-to-function approach (common in corpus linguistics) when studying speech acts, as well as the combination of these two approaches. The blending of these two perspectives offers a balance that improves both the breadth and frequency of form, while adequately capturing and documenting the richness of the contextual conditions in which the forms initially emerged (cf. O'Keeffe; Clancy; Adolphs, 2019).

An example of a small-sized corpus

To pursue our study of the agreement speech act, we found it necessary to develop our own corpus, as existing ones did not align with our specific research aims. Given the considerable time, resources and effort typically involved in constructing an extensive corpus, we opted for a small-scale approach. This allowed us to maintain a high level of informational richness while efficiently meeting the objectives of our analysis. Below, we provide an overview of the corpus collection process, from its conception to the challenges we faced in carrying it out.

Corpus description

The compiled corpus consists of digital conversations in Spanish that integrate the act of direct agreement on LinkedIn and Instagram. The conversations were gathered from public pages or accounts, covering five specific thematic areas: sports, television series or movies, lifestyle (both personal and professional), news, and health. These topics were chosen to avoid controversial and contentious subjects, aiming to obtain more expressions of agreement and to cover various areas to provide a general perspective on how agreement is treated on these social networks. The corpus comprises a total of 70,000 words and 3,258 comments. These conversations, all transcribed *verbatim*, were manually collected and are divided into two categories: long conversations (CCL) and short conversations (CCC). Long conversations (CCL) refer to responses to comments made on posts on the initial pages of selected accounts, comprising between 120 and 400 words. The set of long conversations consists of 168 conversations, including a total of 2,068 comments, collected from Instagram and LinkedIn, covering 50,000 words in total (25,000 from each platform and 5,000 for each topic). Short conversations (CCC) refer to interactions consisting of a maximum of three interventions and containing fewer than 120 words. These involve only two speakers and may arise as a combination of the main post and a comment, or between a comment and its response in the comment section. The short conversations form a 20,000-word corpus (10,000 words from each platform and 2,000 for each topic).

The creation of the corpus took place in two phases: in the first phase, the interactions were collected, and in the second phase, the sociolinguistic data of the users were gathered, considering factors such as gender, age, and educational level. The preparation of both phases was carried out by the researcher, and methodological issues related to internet data extraction were considered throughout, such as representativeness (ensuring a similar number of words for each platform), the amount of necessary contextual information, and privacy (guaranteeing the anonymity of users on these platforms) (cf. Androutsoupoulos; Beiβwenger, 2008; Mancera, 2014). Additionally, efforts were

made to ensure that the conversations were recent, with a publication date no earlier than the last four years (from October 2020 to January 2021).

The following tables schematically present all the information provided. Table 1 offers a summary of the quantity and length of both long and short comments, as well as the quantity and length of comments containing direct agreement on each platform and category. Table 2 includes the thematic line with the corresponding proportion of comments included in the corpus.

Table 1 – Quantity and length of comments in corpus

	CCL	CCC		CCL	CCC
N. of comments ING	1456	840	N. of comments ING – agreement	334	540
N. of comments LKDN	612	350	N. of comments LKDN – agreement	192	211
Total	2068	1190		526	751
N. of words ING	25.000	10.000	N. of words in comments ING – agreement	4411	5140
N. of words LKDN	25.000	10.000	N. of words in comments LKDN – agreement	7472	5997
Total	50.000	20.000		11.883	11.137

Table 2 – Total number of comments for each topic in CCL and CCC

CCL	ING	LKDN	CCC	ING	LKDN
N. of comments – HEALTH	223 15.31%	122 19.93%	N. of comments with agreement – HEALTH	149 17.73%	52 14.85%
N. of comments – SPORTS	390 26.78%	141 23.03%	N. of comments with agreement – SPORTS	230 27.38%	84 24%
N. of comments SERIES/MOVIES	301 20.67%	124 20.26%	N. of comments with agreement – SERIES/MOVIES	134 15.95%	78 22.28%
N. of comments – NEWS	227 15.59%	101 16.50%	N. of comments with agreement – NEWS	189 22.50%	79 22.57%
N. of comments – LIFESTYLE	315 21.63%	124 20.26%	N. of comments with agreement – LIFESTYLE	138 16.42%	57 16.28%
Total	1456	612		840	350

User description

When gathering user data, it's crucial to note that accessing or uncovering users' real identities can often prove challenging due to the ever-changing nature of information on the internet. Consequently, many studies on social media platforms are restricted to analyzing variables such as origin, gender, and age, often neglecting factors related to educational level or social class. In our research, we categorized users based on variables such as origin (Spanish or long-term residents in the country), gender (male and female), age (≤18, 19-25, 26-55, ≥56), and educational level (with or without university education).

The number of users participating in each interaction ranges from a minimum of two to a maximum of 40, considering the word limit established for each conversation. The CCL involves 1,400 users (1,010 on Instagram and 390 on LinkedIn), along with their respective contributions, while the CCC comprises 1,193 users (878 on Instagram and 315 on LinkedIn). However, due to the aforementioned limitations, which we will detail in the next section, we could not obtain sociolinguistic data from all 2,593 users. Table 3 represents only the categorization of users who will be the focus of our study, namely those who have engaged in direct agreement interactions. The distribution of users according to social factors such as gender, age, and educational level is depicted in the subsequent table.

Table 3 – Sociolinguistic data of the studied users

Social factor	Variants				
Gender	**Woman**		**Man**		**NO ID**
	646		609		8
Percentage	51.14		47.98		0.63
Academic background	**Without university studies**		**With university studies**		**NO ID**
	417		773		73
Percentage	33.01		61.20		5.77
Age group	**≤ 18**	**19-25**	**26-55**	**≥ 56**	**NO ID**
	86	144	755	246	32
Percentage	6.80	11.40	59.77	19.47	2.53
Total number			1263		

Reasons for choosing Instagram and LinkedIn

The COVID-19 pandemic resulted in a significant increase in the use of social media during this period and in the following years. Over the course of the first year since the beginning of the pandemic in Spain, more than 37 million people, equivalent to 80% of the population, turned to social media, according to the *Digital 2021* report compiled by *Hootsuite* and *We are Social*[1]. This surge in social media usage sparked our interest in investigating agreement in this domain, and we decided to study two of the most popular platforms during this time: Instagram and LinkedIn. We specifically chose these two due to their distinctive differences, which would provide more meaningful and comparative results. The most notable characteristics of these social media platforms are as follows:

a) Each platform caters to specific user needs: Instagram promotes socialization, while LinkedIn serves a transactional purpose.

b) Socialization dynamics differ on both: relationships on Instagram can be either symmetrical or asymmetrical, whereas on LinkedIn, they are mainly symmetrical.

c) Different types of identities can be constructed on these platforms: Instagram allows for anonymous profiles, whereas LinkedIn typically uses authentic identities.

d) Immediacy varies on these networks: conversations on Instagram tend to be more synchronous than on LinkedIn.

e) Although both can be accessed via mobile devices and computers, Instagram is primarily designed for mobile devices, while LinkedIn is for computers.

f) The platforms belong to different colloquial registers according to Briz's (2010) situational variation scale: conversations are more colloquial on Instagram than on LinkedIn.

g) Each network has its own usage norms and courtesy protocols.

Challenges in internet data collection and adopted solutions

We conclude this section by addressing the challenges we encountered when creating the corpus, as constructing a corpus of communicative interactions on the internet proves to be more complex than it may seem at first glance. Following the

---

[1] Source: https://www.expansion.com/economia-digital/innovacion/2021/02/10/6022c89de5fdea-59448b459b.html (Consulted in July 2021)

insightful observation of Pano and Moya (2015), it is easy to assume that access to online interactions is straightforward, as they are already digitized and, in principle, do not require transcription for subsequent analysis. However, several aspects need to be considered, such as the corpus's extent or representativeness, as well as the data generated by the system —for example, the input and output lines of a chat channel or the date a comment was sent— and the data generated by the user, including text, links, emoticons, and images (Pano; Moya, 2015, p.118). Additional factors complicating the collection of a digital corpus include the geographical location of the speakers, the validity of some interactions for linguistic or sociolinguistic analysis, and even the peculiar spelling observed in certain types of Computer-Mediated Communication (Pano; Moya, 2015, p.118).

Next, we will present the challenges we faced during the process of collecting our empirical material, which we divided into two categories: problems related to obtaining conversations and issues linked to user data collection.

Difficulties in conversation collection

Regarding conversations, after conducting a preliminary study on various topics, we observed that the selected thematic areas generated the most comments with expressions of agreement. Despite this, we encountered difficulties in collecting the LinkedIn corpus. At the time, LinkedIn was gaining popularity due to the employment situation created by COVID-19 but had not yet become widespread. This limited the availability of posts, particularly in the thematic areas of sports and series/movies, making it difficult to gather a sufficient number of conversations to build the long conversation corpus (CCL).

Additionally, we faced the idiosyncrasy of topics on LinkedIn; on this platform, interactions and posts are expected to focus on work-related matters. Therefore, there is no acceptance or interest from users to engage in common or ordinary topics. Consequently, we had to visit more pages and collect more conversations from LinkedIn compared to Instagram. This is reflected in the number of long conversations collected: 90 from LinkedIn and 78 from Instagram. Specifically, for the two challenging topics, we obtained 22 sports conversations on LinkedIn and 17 on Instagram, and 22 series/movies conversations on LinkedIn and 14 on Instagram.

Difficulties in user data collection

We also encountered several challenges in extracting sociolinguistic data from users. The volatility of information on the internet, which allows for identity changes and the creation of private accounts, complicated access to account holder data. To distinguish between Spaniards and other Spanish speakers (such as Latin Americans or Spanish learners), we used linguistic markers, including specific verb forms and

pronouns from Spain, as well as colloquial markers and expressions. We also evaluated various aspects of users' profiles, such as shared photos, post locations, and activity on other social media platforms.

To gather information about users' age, gender and level of education, we examined their posts and profiles on various social media platforms. In extreme cases, we directly contacted users to obtain the required information; however, some did not respond, leading to their categorization as *unidentified.*

An additional problem, particularly on Instagram, was that users could change their identification or deactivate their accounts. This made it complicated to track users between the first and second phases of data collection and forced us to discard their comments. Finally, we faced issues with suspicion of our activities by the platforms, which led to restrictions on certain actions and extended the data collection phase beyond our initial expectations.

Case study: agreement speech act

In this section, we'll outline how the corpus guided us in developing a technical sheet for conversations and an analysis sheet for the speech act of agreement. To develop these sheets, we initially consulted existing taxonomies (Pomerantz, 1984; Schneider, 1988; Gille, 2001; Santamaría, 2003; Henning, 2015; Milà-Garcia, 2016; Al-Khanaifsawy, 2016), which we then refined through corpus observation and contextualized analysis. We believe that these sheets will serve other researches interested in designing and annotating a pragmatic corpus for agreement.

Technical sheet for conversations

Each conversation in the corpus is accompanied by a technical sheet that provides relevant data for research purposes. This sheet aims to offer adequate contextual information to facilitate the analysis of the speech act. The structure of the technical sheet is based on the model used in the Corpus Val.Es.Co. (Briz; Val.Es.Co. Group, 2002, p.36-38):

a. **Conversation identifiers:**

Publication date:
Collection place: LinkedIn/Instagram
Conversation length: Word count

b. **Communicative situation:**

Topic: sports/news/health/lifestyle/TV series or movies
Dominant tenor: interpersonal/transactional

Gholami | Exploring agreement on social media: a small-scale corpus study with proposals for technical …
Revista Diadorim | v.26, n.2, e63939, 2024

10

Tone: formal/informal

Mode: since we work with written colloquial conversations, it is written.

Degree of colloquial prototypicality: Prototypical colloquial/Peripheral colloquial

c. Description of participants:

Number of participants: from 2 to 40

Participant role: administrator or account owner/follower or contacts (these last two terms are established by the platforms, the first is used in Instagram, the second in LinkedIn)

Type of relationship: follower-follower/follower-administrator or account owner

Sociolinguistic features chart: (Table 4)

Table 4 – Table of sociolinguistic data of users

Speaker key	Gender			Age					Educational level		
	M	W	NO ID	≤18	19-25	26-55	≥56	NO ID	With	Without	NO ID
1											
2											
3											

Proposal of an analysis sheet for agreement

Different levels have been recognized for a variational pragmatic analysis (cf. Barron; Schneider, 2009; Félix-Brasdefer, 2019). Our research focused on two of these levels: interactive and illocutionary. Below, we present the analysis sheet we developed to examine the speech act of agreement at these levels.

Interactive and illocutionary levels, a scheme

The interactive level focuses on acts in sequence, that is, on the sequential structure of adjacent pairs and follow-up moves. On the other hand, the illocutionary level relates to the communicative function of an act, exploring the various ways to perform such an act and the conventional pragmatic-linguistic resources used to express it (Félix-Brasdefer, 2019, p.189). Therefore, the interactive level scrutinizes the first part of adjacent pairs, while the illocutionary level is concerned with the second part of these pairs. Prior to detailing the variables and their corresponding variants for analysis at these levels, we present them schematically in Table 5 to provide a comprehensive view.

Gholami | Exploring agreement on social media: a small-scale corpus study with proposals for technical ...

Revista Diadorim | v.26, n.2, e63939, 2024　　　　**11**

Table 5 – Analysis sheet of illocutionary-interactive level

Interactive level: First part of the adjacent pair		
Initiation activity type		1. Information 2. Opinion 3. Evaluation 4. Elicitation 5. Proposal
Act type		1. Action 2. Statement
Illocutionary level: Second part of the adjacent pair		
Head act	Agreement resourcee	1. Emphatic backchannelers 2. Neutral backchannelers 3. Acknowledgment constructions 4. Evaluative constructions 5. Identification constructions 6. Performative constructions
	Agreement function	1. Confirmation 2. Agreement 3. Positive evaluation 4. Acknowledgment 5. Identification 6. Acceptance
	Totality of agreement	1. Partial 2. Total
	Intensification	1. Grammatical 2. Lexical 3. Orthotypographical 4. Non-verbal

Table 5 – Cont.

Interactive level: First part of the adjacent pair			
Support	Appellatives	1. Common 2. Proper noun 3. Affectionate term 4. Pejorative	
	Support moves	1. Repetition 2. Reformulation 3. Elaboration 4. Irony 5. Laughter 6. Narration	7. Assessment 8. Compliment 9. Clarification 10. Evidence 11. Rhetorical questioning
	Position of support move	1. Before the head act 2. After the head act	
Complexity of agreement		1. Simple 2. Complex with two resources 3. Complex with three resources 4. Complex with more than 4 resources	

Interactive level: first part of the adjacent pair

Initiation activity type. In this stage, the type of communicative first turn seeking agreement will be identified. The literature has recognized different sequences in which agreement can occur (cf. Tsui, 1994; Santamaria, 2003; Henning, 2015; Milà-Garcia, 2016). After observing the corpus, we have distinguished 5 general types of preceding turns that may precede agreement: information, opinion, evaluation, elicitation, and proposal, characterized as follows:

Information: Informational statements provide objective data without including opinions, judgments, or personal feelings. They focus on informing about events, states of affairs, or recounting personal experiences (cf. Tsui, 1994, p.135).

Opinion: Opinion involves an assessment of a referent in terms of certainty/uncertainty or the assignment of characteristics/qualities, without necessarily including positive or negative value judgments.

Evaluation: Evaluation involves the speaker's judgment and reflects an internal evaluation, whether positive or negative, of people, objects, events, situations, etc.

Evaluations are "comparative, subjective, and value-laden" and the essential criterion for identifying them focuses on the spectrum of good-bad (cf. Thompson; Hunston, 2000, p.13 & 25).

Elicitation: Elicitation refers to any statement that functions to obligatorily elicit verbal or non-verbal responses (cf. Tsui, 1994, p.81).

Proposal: A proposal is a "proposition or idea that is expressed and offered to someone for a purpose" (RAE).

Act type. Secondly, we verify whether the content of the first turn refers to an action or something previously said, i.e., if it reports on an event or if it comments on or expresses something about it. Comments may include statements of information, evaluations, opinions, requests, and proposals, while actions are communicated exclusively through statements of information. As an example, in sequence (1), the first turn informs about an action, specifically the postponement of the Barcelona football club elections, and user 1142 expresses agreement with this action.

(1)

1141: "🚨🚨APLAZADAS LAS ELECCIONES del BARÇA del día 24 de enero ✖️📭."

("🚨🚨BARÇA ELECTIONS POSTPONED from January 24th ✖️📭.")

1142: Bien hecho no es serio con lo que estamos pasando

(Well done, it's not serious considering what we're going through)

(Instagram, Sports, 2021)

However, in the following sequence, user 31 expresses his opinion about the work of healthcare staff during COVID-19, which user 32 agrees with:

(2)

31:     Toda la sociedad debemos tanto a nuestro personal sanitario que demuestra cada día su gran valor y profesionalidad, que toda nuestra de elogio se queda corta. GRACIAS!!!

(All of society owes so much to our healthcare staff who demonstrate their great value and every day, that all our praise falls short. THANK YOU!!!)

32:     Gran verdad!

(So true!)

(LinkedIn, Health, 2021)

Illocutionary level: second part of the adjacent pair

Agreement resources. At this level, we will first identify the direct resources used to express agreement (referred to as DREA henceforth), as outlined in the literature and evidenced in the corpus. These resources encompass emphatic backchannelers, neutral backchannelers, acknowledgment constructions, identification constructions, performative verbs or constructions and evaluative constructions (cf. Pomerantz, 1984; Schneider, 1988; Gille, 2001; Henning, 2015; Milà-Garcia, 2016; Al-Khanaifsawy, 2016). Table 6 summarizes all the mentioned strategies.

Table 6 – Direct resources of expressing agreement (DREA)

Resource	Example
Emphatic backchannelers	*Exacto (exactly)*
Neutral backchannelers	*Sí (yes), no*
Acknowledgment constructions	*Verdad (true), tienes razón (you're right)*
Evaluative constructions	*Perfecto (perfect), bien (fine)*
Identification constructions	*Pienso lo mismo (I think the same), también (too)*
Performative constructions	*Estoy de acuerdo (I agree)*

Agreement function. Next, we examine the function of the agreement produced. Several definitions have been offered for agreement, according to which we can assume that it is an overarching term encompassing multiple functions. As Schneider (1984) points out, agreement is primarily considered as a positive reaction to statements, usually to opinion statements. However, this concept can extend to other types of positive reactions, such as confirmation, recognition, and even backchanneling behavior, as essentially the same elements are used in all these cases (Schneider, 1984, p.160). Building on this idea, we aim to specify the functions that DREA can perform. After reviewing the literature and observing the DREA in the corpus and their relationship to the preceding turn, we have distinguished six general functions. DREA are used to:

- show agreement or adherence to a certain statement or action: *estoy de acuerdo* (*I agree*)
- confirm or approve a statement or action: *exacto (*exactly*)*
- accept something said previously: *ok (okay), vale (alright)*
- positively evaluate or assess a statement or action: *me parece bien (sounds good to me)*

- acknowledge the truth, certainty, or rationality of a statement: *tienes razón (you're right), es verdad (it's true)*

- and demonstrate that speakers have similar thoughts and ideas: *pienso igual (I think the same)*

Totality of agreement. Later, we determine whether the agreement is total or partial; the latter case is usually indicated by an adversative conjunction like '*pero*' (*but*) or conditional expressions like '*si*' (*if*).

Intensification. Once we have defined the resources, their functions, and totality, we assess whether intensifications are used. Drawing from Vivas's framework (2014) regarding intensification on the internet, we have classified intensifications into four categories: grammatical, lexical, orthotypographical and non-verbal, such as the use of emoticons. Each category will further be subdivided into subcategories (to read more, refer to Gholami; Llopis, 2023).

Appellatives and support moves. In the next step, we examine the presence of other elements, such as appellatives and support moves, in the intervention where the agreement occurs. This idea is derived from the study of Blum-Kulka et al. (1989) on the act of request and applied to the study of (dis)agreement by López Sako (2008) and Milà-Garcia (2016). First, we assess whether appellatives —labeled as proper names (*Ana*), common nouns (*Mr.*), affectionate terms (*dear*) or pejorative terms (*you, idiot*)— are used, based on Vidán (2004). Next, we analyze the presence of support moves, including both resources and strategies identified as indirect agreement by previous studies, as well as additional ones identified by us. We posit that when DREA is employed to convey the illocutionary force of the agreement act, the other elements present in the act primarily serve to support the head act. The differentiation between direct and indirect agreement resources was established by considering the alignment between explicit linguistic forms and the illocutionary force or intention of the act, following Searle (1975).

Thus, support moves encompass strategies for the indirect expression of agreement, such as repetitions, paraphrastic reformulations, elaborations, irony, laughter, narratives, assessments, compliments, clarifications, evidences and rhetorical questions. *Repetitions* occur when the interlocutor repeats either literally or partially what the first speaker said, while *reformulation* involves a paraphrastic repetition of what was said in the previous turn. *Elaborations* allow the speaker to develop their reasoning and introduce new information to support another speaker's opinion (cf. Gille, 2001; Henning, 2015). Through *irony*, the speaker makes a statement that humorously implies agreement with the interlocutor's opinion or stance. *Laughter*, in turn, serves as a communicative response indicating comprehension, appreciation of the irony, and social acceptance of humor. Similarly, a listener's reaction to a previous statement may include a *narrative* —often of personal experiences— that can either reflect or contrast with the interlocutor's stance (cf. PILLET-SHORE, 2017).

On the other hand, *positive assessments* consist of statements that reflect favorable evaluations of previously expressed ideas or actions, while *compliments* are directed specifically at the interlocutors themselves (cf. Tsui, 1994). *Clarifications* may be needed to resolve misunderstandings and demonstrate alignment with the other speaker's viewpoint. In other instances, a speaker might provide an example or concrete *evidence* in support of both the other speaker's statement and their own opinion. Finally, *rhetorical questions* can also accompany the central act of agreement to reinforce either their own or another's idea. These rhetorical questions are not intended to be answered literally and are semantically equivalent to statements that represent the obvious response (Dumitrescu, 2016, p.769).

In the next two sequences, we observe the use of appellatives and three different support moves (repetition in sequence 3; narration and reformulation in sequence 4).

(3)

70:    <71> Desde luego..y volver la mirada a Dios también!! Y a la Virgen, España es tierra de María. Es tiempo de trabajo y oración. Ora et labora. Ninguna sin la otra.

       (Of course... and also turn our gaze to God!! And to the Virgin, Spain is the land of Mary. It is a time for work and prayer. Ora et labora. Neither without the other.)

71:    <70> **Bien** *Luis*, Ora et Labora, **muy bueno**

       (**Well said**, *Luis*, Ora et Labora, **very good**)

                                                                    (LinkedIn, News, 2021)

(4)

1000: Dicen que pareja que entrena unida, se mantiene unida... pero NO es cierto. Pareja que se comunica, que se respeta, que se esfuerza por comprenderse, que da un paso atrás si es necesario, que perdona, que reconoce, que lucha como equipo, que ama... esa si se mantiene unida 🫶 💪

       (They say that a couple that trains together, stays together... but it is NOT true. A couple that communicates, respects each other, strives to understand each other, takes a step back if necessary, forgives, recognizes, fights as a team, loves... that couple does stay together 🫶 💪)

1004: **Así es** *mi @vikikacosta* yo tenía una pareja que entrenábamos juntos pero no había casi nada de lo que mencionas... *Y una relación es todo o nada por que a medias no funciona...*

       (**That's right**, *my @vikikacosta*, I had a partner with whom we trained together but almost none of what you mentioned was there... *And a relationship is all or nothing because halfway does not work...*)

                                                                    (Instagram, Style, 2021)

In sequence 3, User 71 shows agreement using a complex structure that includes an evaluative construction ("bien" [*well said*]), supported by an appellative (proper name: *Luis*) and a support move (literal repetition of part of User 70's opinion: "Ora et Labora"). At the end, User 71 uses another evaluative construction ("muy bueno" [*very good*]) to further intensify the agreement.

In sequence 4, a similar pattern occurs. The direct agreement is accompanied by support elements such as an appellative, affectionate term ("mi @vikikacosta"), a narrative of the user's own life experiences (underlined in the intervention), and a reformulation of user 1000's opinion at the end (underlined and italicized in the intervention).

Position of support move. Subsequently, we investigate the position of the support moves relative to the central act of agreement —whether they occur before or after it. This analysis aims to identify established patterns in the preferred sequencing of acts, such as agreement (cf. Levinson, 1983; Pomerantz, 1984; Schegloff, 2007; Félix-Brasdefer, 2019).

Complexity of agreement. Finally, we delve into the complexity of agreement, considering that literature suggests that preferred acts are often short and simple. However, we acknowledge that agreements may stand alone without any supporting moves, or may manifest in a more intricate manner by combining two or more direct/indirect resources.

Analysis of a conversation, a sample

In this section, we will examine a conversation selected from the CCL corpus of LinkedIn (Figure 1), using the previously proposed technical and pragmatic analysis sheets. This particular sequence was chosen because it features three distinct types of agreement. The conversation involves four male users, all with academic backgrounds: 316 (the account owner), 315, 317, and 318. Their ages range from 26 to 55, except for 318, who is over 56. The discussion focuses on the importance of honesty in the business world.
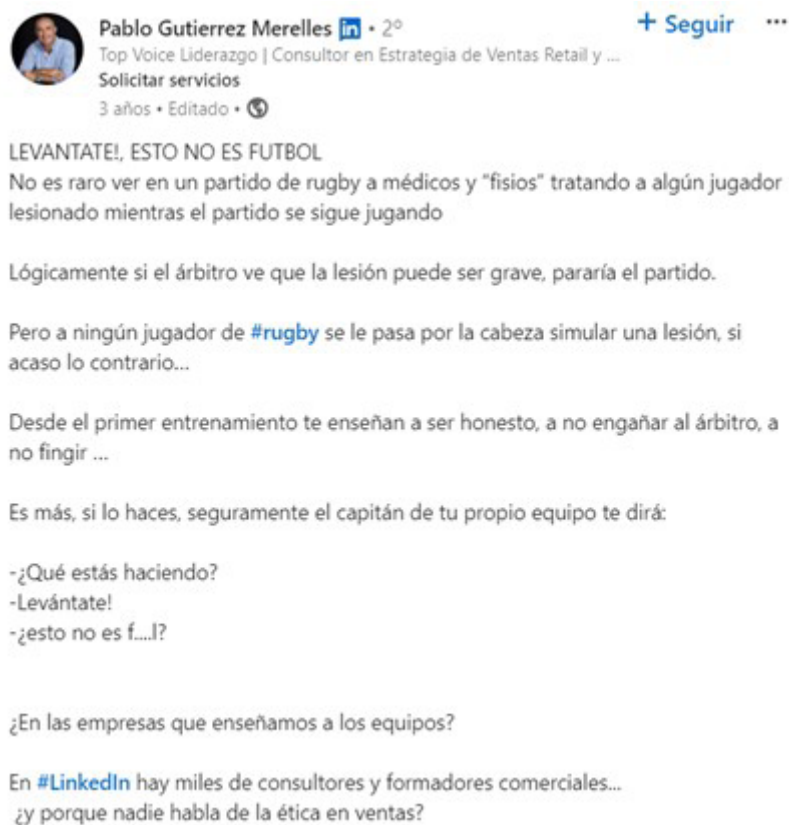
Figure 1 – A LinkedIn Post
Source: LinkedIn, 2020

(5)

1   315:   Y por lo que a mi me toca en las compras. Hay que ser digno no solo con tu equipo y empresa si no con tus proveedores. Igual que en el rugby no hay enemigos sino compañeros contra los que jugar, en la empresa los proveedores tampoco son alguien a quien enfrentarse pues al final son parte de tu negocio.

(As for my role in purchasing, it's important to be respectful not only to your team and company but also to your suppliers. Just as in rugby there are no enemies, only teammates to play against, in business, suppliers are not adversaries to confront, as they ultimately become part of your business.)

2   316:   **Sin duda** *Iván*, son parte del negocio no empresas a las que exprimir hasta que den pérdidas... Gracias

(**without a doubt**, *Iván*, they are part of the business, not companies to be squeezed until they incur losses... Thank you.)

3   317:   **No puedo estar más de acuerdo contigo** *Ivan*.

(**I couldn't agree more with you**, *Iván*.)

4   318: Es difícil hacer ver esa teoría, pero **sin duda es así**. La fortaleza de los proveedores se convierte en ayuda para que los clientes alcancen el éxito, sin embargo, cuantas veces el cliente ahoga a sus proveedores haciendo que se enfrenten entre ellos para despues recoger las nueces?. Es importante aplicar la regla win to win, para que el resultado sea bueno y duradero para ambos.

(It's difficult to convey this concept, but **it's undoubtedly true**. The strength of suppliers becomes a resource to help customers succeed. However, how often does the customer stifle its suppliers, forcing them to compete against each other, only to pick up the benefits afterward? It's crucial to apply the win-win rule so that the outcome is positive and sustainable for both parties.)

(LinkedIn, Sports, 2020)

The completed technical sheet for this conversation includes the following details. Table 7 provides general information about the conversation, including the publication date, platform, and conversation length. It also covers the communicative context and details about the users participating in the conversation. Table 8 focuses on the sociolinguistic features of these participants.

Table 7 – Technical sheet of Conversation 5

CONVERSATION IDENTIFIERS			
Publication date	2020	**Collection place**	LinkedIn
Conversation length	156 words		
COMMUNICATIVE SITUATION			
Topic	sports	**Dominant tenor**	interpersonal
Tone	informal	**Mode**	written
Degree of colloquial prototypicality		Peripheral colloquial	
DESCRIPTION OF PARTICIPANTS			
Number of participants	4	**Participant role**	owner/contact
Type of relationship	Owner >> contact / contact >> contact / contact >> contact		

Table 8 – Sociolinguistic features chart of participants in conversation 5

Speaker key	Gender			Age					Educational level		
	M	**W**	**NO ID**	**≤18**	**19-25**	**26-55**	**≥56**	**NO ID**	**With**	**Without**	**NO ID**
315 (contact)	M					26-55			With		
316 (owner)	M					26-55			With		
317 (contact)	M					26-55			With		
318 (contact)	M						≥56		With		

Now, we proceed to analyze this conversation using the analysis sheet proposed earlier. At the interactive level, the initial communicative turn (Intervention 1) by user 315 expresses an opinion on the importance of suppliers in the business world; therefore, it is a statement rather than an action.

At the illocutionary level, we observe three interventions that include an act of agreement. In the first one (Intervention 2), user 316, the account owner, expresses agreement using an emphatic backchannel response ("sin duda" [*without a doubt*]), which serves to confirm user 315's opinion. This DREA is not intensified internally but is reinforced by an appellative (the proper name "Iván") and a supportive move through paraphrastic reformulation (underlined in the intervention). This results in a total and complex agreement.

The second agreement (Intervention 3) is a simple one, consisting of a single DREA, a performative construction intensified through a grammatical procedure (emphatic construction: "No puedo estar más de acuerdo contigo" [*I couldn't agree more with you*]). This DREA serves to express total agreement or adherence to user 315's opinion. The intervention also includes an appellative, again using a proper name, which is the most common form on LinkedIn.

The last intervention (N.4) is slightly different. It begins with an assessment of user 315's opinion, which shifts the DREA from the initial position in the response. Although it may initially appear to be a disagreement, reading the entire intervention makes it clear that it is actually a total agreement. User 318 employs an acknowledgment construction ("así es" [*that's right*]), which is further intensified by an adverb ("sin duda" [*without a doubt*]). This agreement acknowledges the truth or certainty of user 315's statement and is followed by a supportive move that elaborates on 315's opinion (underlined in the intervention).

Gholami | Exploring agreement on social media: a small-scale corpus study with proposals for technical ...

Revista Diadorim | v.26, n.2, e63939, 2024　　　　　　　　　　　　　　　　　　21

Conclusion

In conclusion, this manuscript has delved into the nuanced realm of speech acts within social media discourse, with a focus on the pivotal act of agreement. By emphasizing the importance of pragmatic corpora in comprehending speech acts within their contextualized online environments, it underscores the valuable insights derived from the synergy of qualitative and quantitative approaches.

To facilitate a comprehensive study of agreement, we created a small-sized corpus of Spanish-language content from LinkedIn and Instagram. This corpus enabled us to conduct a pragmatically multilevel analysis of the phenomenon of agreement. Specifically, we developed a technical sheet for each conversation, which includes vital contextual data such as the communicative situation, the tenor of discourse, the participants and their sociolinguistic features. This sheet ensures that each conversation is analyzed within its specific social and linguistic context, thereby providing a robust framework for subsequent analysis.

In addition to the technical sheet, we designed a pragmatic analysis sheet tailored to the speech act of agreement. This sheet allowed us to systematically explore both the interactive and illocutionary levels of agreement. It categorizes the types of communicative turns that precede agreement, the resources used to express agreement (e.g., emphatic and neutral backchannelers, evaluative constructions), and the various functions that agreement can serve (e.g., confirmation, acknowledgment, positive evaluation). Moreover, the analysis sheet accounts for the totality and intensity of agreement, as well as the presence of support moves —including appellatives and indirect agreement resources— that accompany the expression of agreement.

These tools not only enriched our own study but also offer tangible resources for researchers interested in constructing and annotating pragmatic corpora for speech act analysis. As the field of corpus pragmatics continues to advance, this manuscript aims to contribute to this discipline by navigating the intricacies of online data collection, corpus development, technical sheet formulation, and speech act analysis in social media interactions.

References

AIJMER, K.; RÜHLEMANN, C. (ed.). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, 2015.

ALBELDA, M. Los corpus del español hablado y los estudios pragmáticos. *Lingüística de corpus en español*, n. 1, p. 223-238, 2022.

ALBELDA, M.; ÁLVAREZ, A. Los corpus discursivos en el estudio pragmático de la atenuación y de la intensificación. *Revista internacional de lingüística iberoamericana*, n. 16, p. 79-100, 2010.

AL-KHANAIFSAWY, A. N. Investigating Iraqi EFL learners' use of the speech act of agreement. *Adab AlKufa*, v. 1, n. 27, p. 11-30, 2016.

ANDROUTSOUPOULOS, J.; BEIβWENGER, M. Introduction. Data and Methods in Computer-Mediated Discourse Analysis. *Language@Internet*, v. 5, n. 9, p. 1-7, 2008.

BARRON, A.; SCHNEIDER, K. P. Variational pragmatics: Studying the impact of social factors on language use in interaction. *Intercultural Pragmatics*, v. 6, n. 4, p. 425-442, 2009.

BAYM, N. K. Agreements and Disagreements in a Computer-Mediated Group. *Research on Language and Social Interaction*, n. 29, p. 315-346, 1996.

BLUM-KULKA, S.; HOUSE, J.; KASPER, G. *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex Publishing Corporation, 1989.

BOLANDER, B. Disagreements and agreement in personal/diary blogs: A closer look at responsiveness. *Journal of Pragmatics*, n. 44, p. 1607-1622, 2012.

BRIZ, A. *El español coloquial en la conversación: esbozo de pragmagramática.* Barcelona: Ariel, 2001.

BRIZ, A. El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas. *In:* FONTE, I.; RODRÍGUEZ ALFANO, L. (ed.). *Perspectivas dialógicas en estudios del lenguaje*. Nuevo León: Universidad Autónoma de Nuevo León, 2010. p. 21-56.

BRIZ, A.; VAL.ES.CO. GROUP. *Corpus de conversaciones coloquiales. Anejo de la revista Oralia*. Madrid: Arco/Libros, 2002.

CANTAMUTTO, L.; FAINSTEIN, P. Marcadores conversacionales de acuerdo en la enseñanza de Español como Lengua Extranjera. *Verba Hispánica*, n. 27, p. 153-172, 2019.

CLANCY, B.; O'KEEFFE, A. Pragmatics. *In:* Biber, D.; Reppen, R. (ed.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, 2015. p. 235-251.

DUMITRESCU, D. Oraciones interrogativas indirectas y otras estructuras. *In:* GUTIÉRREZ-REXACH, J. (ed.). *Enciclopedia de Lingüística Hispánica*. London: Routledge, 2016. p. 761-772.

FÉLIX-BRASDEFER, J. C. *Pragmática del español. contexto, uso y variación*. New York: Routledge, 2019.

GARCÍA RAMÓN, A. *Epistemicidad en interacción: (a)simetrías epistémicas en secuencias de acuerdo y su relación con la construcción de roles funcionales en conversaciones y entrevistas.* Dissertação (Doutorado em Aplicação Linguística) – PhD em Programa em Estudos Hispânicos Avançados, Universidade de València, València, 2018.

GHOLAMI, E.; LLOPIS, A. El uso de recursos de intensificación en la expresión del acuerdo en Instagram y LinkedIn. *SOPRAG*, v. 11, n. 2, p. 1-26, 2023.

Gholami | Exploring agreement on social media: a small-scale corpus study with proposals for technical ...

Revista Diadorim | v.26, n.2, e63939, 2024   **23**

GILLE, J. *Pautas argumentativas en el diálogo espontáneo. Un estudio de conversaciones intra e interculturales*. Dissertação (Doutorado) – Departamento de Espanhól e Português, Universidade de Stockholm, Stockholm, 2001.

GUILLER, J.; DURNDELL, A. Students' linguistic behaviour in online discussion groups: Does gender matter? *Computers in Human Behavior*, v. 23, n. 5, p. 2240-2255, 2007.

HENNING, S. *La construcción de la imagen social en dos pares adyacentes: Opinión-acuerdo/ desacuerdo y ofrecimiento-aceptación/rechazo. Un estudio de la conversación familiar sueca y española*. Dissertação (Doutorado em Espanhól) – Departamento de Estudos Românticos e Clássicos, Universidade de Stockholm, Stockholm, 2015.

LEVINSON, S. C. *Pragmatics*. Cambridge: Cambridge University Press, 1983.

LÓPEZ SAKO, N. I. *Politeness in American English, Spanish and Japanese: The Case of (Dis) Agreements in Conversation*. Dissertação (Doutorado) – Departamento de Filologia Inglesa e Germânica, Universidade de Granada, Granada, 2008.

MANCERA, A. Cortesía en 140 caracteres: Interacciones en Twitter entre periodistas y prosumidores. *Revista de Filología de la Universidad de la Laguna*, n. 32, p. 163-181, 2014.

MILÀ-GARCIA, A. *L'acord i el desacord en català en tres gèneres discursius: anàlisi sociopragmàtica*. Dissertação (Doutorado Catalan) – PhD em Programa de Tradução e Ciências da Linguagem, Pompeu Fabra University, Barcelona, 2016.

MILÀ-GARCIA, A. Pragmatic annotation for a multi-layered analysis of speech acts: A methodological proposal. *Corpus Pragmatics*, v. 2, n. 3, p. 263-287, 2018.

MORENO-ORTIZ, A.; GARCÍA-GÁMEZ, M. Strategies for the analysis of large social media corpora: Sampling and keyword extraction methods. *Corpus Pragmatics*, n. 7, p. 241-265, 2023.

O'KEEFFE, A., CLANCY, B. Y ADOLPHS, S. *Introducing Pragmatics in Use.* 2. ed. London: Routledge, 2019.

PANO, A.; MOYA, P. CorpusRedEs. Proyecto de creación y anotación de un corpus de comunicación mediada por ordenador en español. *CHIMERA:* Revista De Corpus De Lenguas Romances Y Estudios Lingüísticos, n. 2, p. 117-129, 2015.

POMERANTZ, A. Agreeing and disagreeing with assessments: Some features of preferred/ dispreferred turn shapes. *In:* DINS ATKINSON, J. M.; HERITAGE, J. (ed.). *Structures of Social Action*. New York: Cambridge University Press, 1984. p. 57-101.

RÜHLEMANN, C; CLANCY, B. *Corpus Linguistics for Pragmatics:* A guide for research. Abingdon, Oxon; New York, NY: Routledge, 2018.

SANTAMARÍA, C. *La negociación de acuerdo en la conversación coloquial. Estudio Contrastivo Español-Inglés*. Dissertação (Doutorado em Espanhíl) – Departamento de Filologia Inglesa, Complutense Universidade de Madrid, Madrid, 2003.

SCHEGLOFF, E. A. *Sequence Organization in Interaction:* A Primer in Conversation Analysis. Cambridge: Cambridge University Press, 2007.

SCHNEIDER, K. *Small Talk:* Analyzing Phatic Discourse. Marburg: Hitzeroth, 1988.

Gholami | Exploring agreement on social media: a small-scale corpus study with proposals for technical ...

Revista Diadorim | v.26, n.2, e63939, 2024   **24**

SEARLE, J. R. Speech acts and recent linguistics. *Annals of the New York Academy of Sciences*, n. 263, p. 27-38, 1975.

THOMPSON, G.; HUNSTON, S. Evaluation: An Introduction. In: HUNSTON, S.; THOMPSON G. (ed.). *Evaluation in Text:* Authorial Stance and the Construction of Discourse. Oxford: Oxford University Press, 2000. p. 1-27.

TSUI, A. B. *English Conversation*. Oxford: Oxford University Press, 1994.

VIDÁN, D. La interculturalidad en el aula de ELE. *Biblioteca virtual redELE*, v. 2, 2004.

VIGARA, A. M.; HERNÁNDEZ, I. Ciber(des)cortesía en los foros de opinión de la prensa escrita: un ejemplo. *ELUA*, n. 25, p. 353-379, 2011.

VIVAS, J. La cortesía valorizadora en las redes sociales. Análisis de un corpus de publicaciones en Facebook. *Pragmalingüística*, n. 22, p. 154-172, 2014.