International Journal of Translation, Interpretation, and Applied Linguistics
Volume 6 • Issue 1

Pronoun Processing:
A Bibliometric Analysis of Research Trends and Topics

Qingxu Zhai, Beijing Foreign Studies University, China*
Lin Fan, Beijing Foreign Studies University, China

ABSTRACT

Using bibliometric analysis, this paper provides an overview of pronoun research from 2012 to 2022. It collected and analyzed 2,774 articles on pronouns from Web of Science (WoS) categories related to linguistics or language using CiteSpace, a citation analysis tool. This paper examined the intellectual framework and patterns of pronoun research through co-citation analysis and identified the most productive journals, influential articles, intellectual base, and trending topics in pronoun studies. The main intellectual base includes anaphora resolution, referring expression, grammatical category, eye movement, subject expression, and thematic analysis. Trending topics comprise studies on English pronouns, acquisition of pronouns, and information cues such as syntax and discourse that affect the comprehension of pronouns. The results reveal the complexity and diversity of pronoun processing. This research contributes to the ongoing discussion and debate on pronoun studies, helps researchers understand the state of pronoun research, and offers suggestions for relevant future research.

KEYWORDS

Anaphora Resolution, Bibliometric Analysis, CiteSpace, Pronoun, Pronoun Comprehension, Pronoun Processing, Pronoun Production

INTRODUCTION

Personal pronouns are important linguistic components that have attracted considerable research interest. They relate to the speech act roles of the speaker, the addressee, and others. Numerous factors might affect the determination of the antecedent of a pronoun, which include syntactic factors such as the grammatical number and gender agreement, semantic factors such as animacy and specificity, and pragmatic factors such as discourse coherence and salience. Resolving pronouns appears to be complex and prone to ambiguity when these factors either conflict with each other or are underspecified. For instance, in the sentence "Mary told Jane that she was pregnant", it is unclear whether "she" refers to Mary or Jane. Understanding pronoun functionality can aid in comprehending the inherent Universal Grammar that facilitates language acquisition and usage, as well as illuminate the role of syntactic constraint and other factors influencing language processing. Hence, it provides a useful example

DOI: 10.4018/IJTIAL.334702 *Corresponding Author

to compare different approaches that seek to understand the processing mechanisms of pronouns at various levels of analysis.

Recent trends in pronoun studies have led to a focus on several areas. One major area of investigation in pronoun studies is the role of various factors that affect pronoun processing. It is important because pronouns are frequently used in natural language and can convey different meanings depending on the context. Researchers have examined the influence of discourse and syntax on pronoun comprehension (Felser, 2019; Grüter et al., 2017; Kush & Dillon, 2021; Puebla & Felser, 2022). For instance, Kush and Dillon (2021) found that during the processing of cataphoric pronouns, native English speakers were capable of utilizing Principle B to guide their antecedent search. Some studies have examined the variability of pronoun processing across individuals, depending on factors like working memory capacity or age differences (Cunnings, 2017; Eilers et al., 2018). According to Cunnings (2017), working memory is essential for L2 learners to link anaphors to their antecedents. At the same time, a few studies have employed techniques such as EEG and ERP to study pronoun processing (Pablos et al., 2015; Xu et al., 2019; Zhao et al., 2020). These technologies can unveil different brain responses elicited by various types of stimuli and illuminate the neural mechanisms underlying pronoun comprehension. Moreover, some recent research has contrasted the use of pronouns in different languages, such as those allow null subjects and those require overt ones. Contemori et al. (2022) investigated bilingual Spanish-English speakers, implying that bilingual individuals' abilities to make inferences about coherence relations and probability distributions are influenced by their native and second languages. These studies demonstrated, to a certain extent, the complexity and diversity of pronoun processing.

Although much research has focused on pronoun processing, the general state of pronoun studies still remains unclear. The present study uses CiteSpace, a bibliometric analysis software, to visually explore the intellectual framework and changing research trends in pronoun studies, aiming to achieve a more comprehensive understanding of the research field. Maps of co-cited documents and clusters, and co-occurring keywords were generated from articles obtained from the Web of Science (WoS) Core Collection using CiteSpace, to reveal the most influential articles, intellectual base, and research topics in pronoun studies and examine their intellectual links. The primary objective of this study is to contribute to the scholarly discourse in the field of pronoun studies by providing critical insights into its intricate nature, analyzing and offering a comprehensive overview of the current state of knowledge, and suggesting directions for future research.

METHOD

Data Collection

Bibliometric analysis is a widely used and methodologically rigorous technique for systematically analyzing and evaluating extensive amounts of scientific data (Donthu et al., 2021). It entails systematic statistical analyses of various literary works, such as books, articles, and other publications, aiming to monitor and evaluate the productivity and influence of authors or researchers. It allows us to analyze in detail the intricate evolutionary nuances of a particular field and illuminates its nascent domains. For this study, articles on pronouns were collected from the WoS database. The WoS database was selected because it provides access to the world's top scholarly literature and is known for its comprehensiveness and reliability. The selection includes various citation indexes, i.e., Social Sciences Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Science Citation Index Expanded, Book Citation Index-Social Science & Humanities, Book Citation Index-Science, Conference Proceedings Citation Index-Social Science & Humanities. The following search method was used to refine the dataset:

```
Topic = "pronoun"
Time span= "2012.01.01-2022.12.31"
Language = "English"
Document type = "article" or "early access" or "review article"
```

Altogether, 3,841 articles were gathered from 1,491 journals comprising 159 Wos Categories. The categories include various fields, such as "linguistics", "psychology multidisciplinary", "communication", "philosophy", and "literature". However, our focus is on linguistic studies, we restricted our search to categories that are relevant to linguistics and language. These categories consist of "language linguistics", "linguistics", "psychology experimental", "psychology developmental", "education educational research", and "psychology multidisciplinary". After screening the selected categories, we identified 2,774 articles for bibliometric analysis.

Instrument

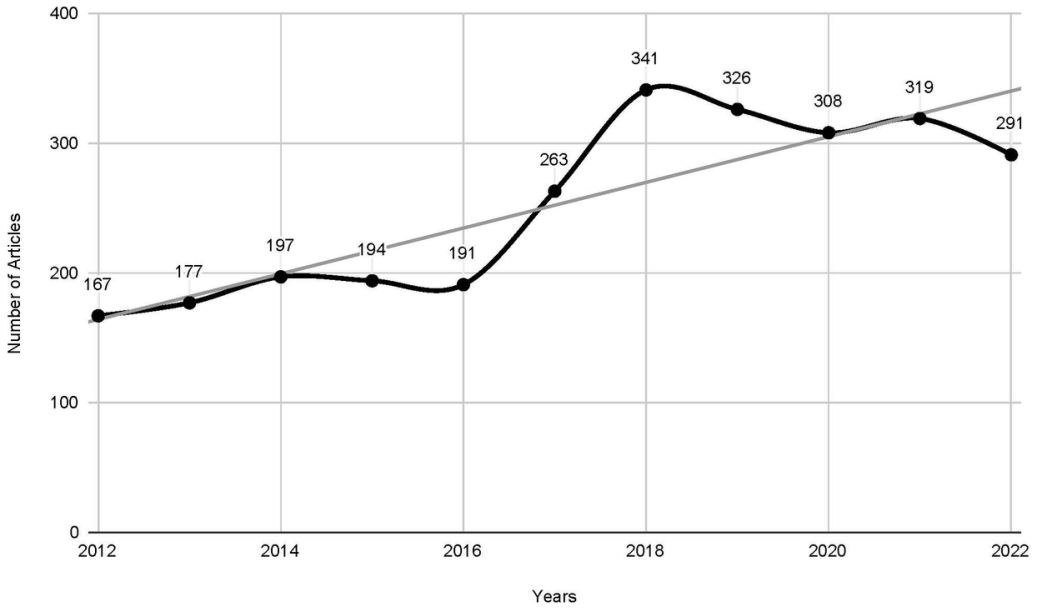
To gain insights into the current state of research on pronouns, the present study utilized CiteSpace (version 6.1.R4 Basic). This specialized citation analysis tool offers a holistic and historical view of evolving specialties (Chen, 2017). Through co-citation analysis, the study identified influential articles, intellectual base, and trending topics in the field of pronoun research.

RESULTS

Trends in Publications Over the Past Decade

Figure 1 shows the yearly distribution of articles on pronoun research throughout the last decade and their corresponding publication counts. A conspicuous observation from this figure is the upward trajectory in publications. The number of published papers exhibited a steady rise from 2012 to 2016, followed by a more substantial surge in 2017 and culminating in its zenith in 2018. Over the past

Figure 1. The annual publication of pronoun research in WoS



five years, there has been a consistent annual publication of roughly 300 studies, demonstrating an unwavering and persistent interest in pronoun research.

The current study examined 2,774 articles on pronoun processing published in 766 journals. Among these, 140 journals were found to have published at least five articles on pronouns, offering insight into the multifaceted research aspects pertaining to this topic. Table 1 presents the top ten productive journals, ranked by the number of articles published on pronouns. The results indicate that *Frontiers in Psychology* had the highest number of publications, with 114 articles on pronouns, followed by *Glossa: A Journal of General Linguistics*, which published 84 articles, and *Journal of Pragmatics* and *Lingua*, each of which published 75 articles. Other productive journals in the top ten include *Natural Language Linguistic Theory*, and *Linguistic Inquiry*, among others. By analyzing the top productive journals, researchers can better understand the research trends and priorities in this area, as well as the various approaches and methodologies used by scholars. This knowledge can also assist in identifying potential collaborators or target journals for future publications.

Table 1. Most productive journals in pronoun studies

Rank	Journal	Number
1	*Frontiers in Psychology*	114
2	*Glossa: A Journal of General Linguistics*	84
3	*Journal of Pragmatics*	75
4	*Lingua*	75
5	*Natural Language Linguistic Theory*	44
6	*Linguistic Inquiry*	41
7	*Journal of Child Language*	39
8	*Linguistics*	39
9	*Journal of Psycholinguistic Research*	38
10	*Journal of Language and Social Psychology*	32

Document Co-Citation Analysis on Pronouns

Bibliographic records were analyzed and visualized to gain a comprehensive understanding of academic research pertaining to pronoun studies from 2012 to 2022. Using a one-year time slice, the top fifty most cited or frequently occurring articles were identified and selected for further analysis. These results were visualized in Figure 2, which displays a network map of co-cited documents. The map includes 651 individual nodes and 2,271 links. The results were highly reliable, as indicated by the modularity Q value of 0.7455. In addition to the visualization map, Table 2 lists the most influential papers in pronoun research during the past decade. The findings uncovered the most influential and frequently referenced articles in the field, providing valuable sources for researchers.

The article produced by Dillon et al. (2013) stands out as the most cited one. In this study, the researchers explored the association between linguistic representation, memory access, and the operation of subject-verb agreement and reflexive anaphors. They examined how syntactically inappropriate noun phrases affected the calculation of these dependencies. During comprehension, agreement resolution exhibited prominent interference effects, whereas reflexives did not demonstrate any similar interference effect originating from illegitimate antecedents. The researchers also presented evidence indicating that the lack of interference for reflexives was best explained by a retrieval process that primarily used syntax-based guidance for anaphor resolution, as opposed to the combined

Figure 2. Critical articles in pronoun studies

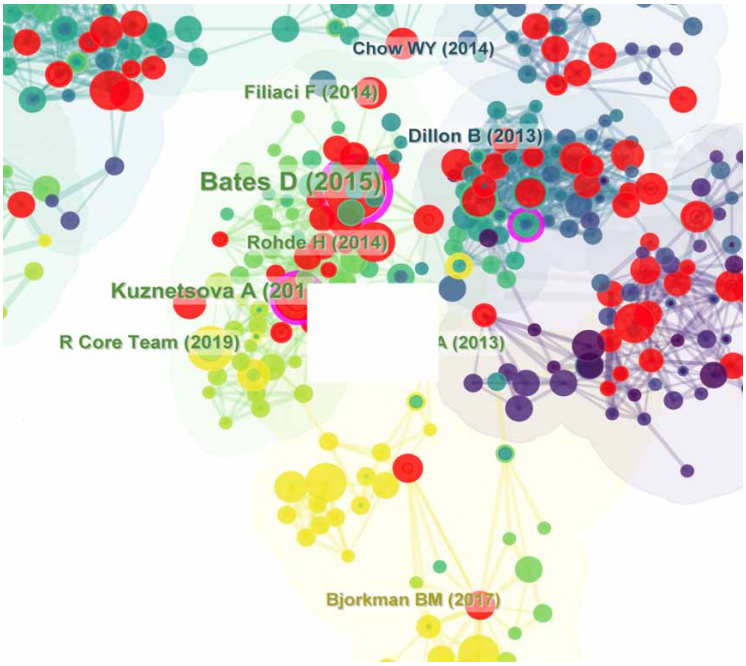


Table 2. The most influential papers in pronoun research

Citation Counts	Author (Year)	Title	Journal
25	Dillon et al. (2013)	Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence	*Journal of Memory and Language*
24	Rohde and Kehler (2014)	Grammatical and information-structural influences on pronoun production	*Language, Cognition and Neuroscience*
21	Filiaci et al. (2014)	Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison	*Language, Cognition and Neuroscience*
19	Chow et al. (2014)	Immediate sensitivity to structural constraints in pronoun resolution	*Frontiers in Psychology*
19	Kehler and Rohde (2013)	A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation	*Theoretical Linguistics*

morphosyntactic cues used to address subject-verb agreement dependencies. The findings suggested that although participants were bound by a similar morphological agreement restriction, they employed different approaches to enforce this constraint for the two dependencies during online processing.

According to the citation count, the article by Rohde and Kehler (2014) ranks second. They examined the impact of grammatical and information-structural factors on pronoun production in English. Sentence completion tasks were used to elicit participants' pronoun production. The authors predicted that both factors would influence participants' production of pronouns. The findings showed that semantic bias did not affect pronoun production, while only the grammatical role of antecedents did. The subject bias in production was found to be motivated by topicality. Furthermore, the empirical evidence substantiated the Bayesian model of pronominal selection posited by Kehler et al. (2008).

The article by Filiaci et al. (2014) is the third most-cited publication. They investigated Spanish and Italian speakers' interpretation of null subjects and overt pronouns in sentences with potential antecedents. A null-subject language refers to a language that permits sentences lacking an explicit subject. Overt pronouns are pronouns that are explicitly expressed in a sentence. The authors hypothesized that null subjects and overt pronouns have different anaphoric biases based on the language and the syntactic placement of the antecedents. They tested the hypothesis with Spanish and Italian participants and found that null subjects were interpreted similarly in both languages, following Accessibility Theory, which predicts that null subjects prefer more salient antecedents. However, overt pronouns showed cross-linguistic differences: overt pronouns in Spanish were easier to process when they matched a prominent antecedent than in Italian but harder to process when they switched the reference from the previous subject comparable to Italian. They suggested that minor differences limited to the pronoun could cause consistent differences between the two languages.

The article by Chow et al. (2014) takes the fourth spot regarding citations. They investigated the relationship between pronoun interpretation and memory retrieval processes that rely on agreement features and structural constraints. Through five reading comprehension experiments utilizing self-paced reading and eye-tracking techniques, they found that agreement characteristics and syntactic limitations were used instantly when finding a suitable antecedent for a pronoun, suggesting that antecedents that violate grammar rules might be considered due to restoration processes activated when finding an antecedent failed at first. The authors provided evidence for the Simultaneous Constraint hypothesis by endorsing cue-based retrieval models of sentence processing that utilized both feature-matching and structure-sensitive cues. This study has important implications for pronoun processing and comprehension, which suggests that readers can rapidly integrate agreement features and structural constraints to make sense of complex sentences.

The article by Kehler and Rohde (2013) ranks fifth in citations. The paper introduced a probabilistic model that integrates two established theories of pronoun interpretation. While the coherence-driven theory emphasizes contextual coherence as the key factor in determining the referent of a pronoun, the centering-driven theory highlights the salience and accessibility of referents in discourse. By leveraging a probabilistic framework, the proposed model combined both theories to weigh their contributions toward determining the referent of a pronoun. The authors provided evidence for its efficacy by comparing its predictions with data from previous studies on pronoun interpretation, demonstrating better performance over both coherence-driven and centering-driven theories across several measures of accuracy and efficiency. One possible implication of their paper is that it provided a unified framework for modeling pronoun interpretation, suggesting that pronoun interpretation is a dynamic and probabilistic process that involves multiple sources of information and expectations.

Cluster Interpretations

In the study of pronouns, understanding the relationships between research papers and the impact of their citations is crucial to identifying key areas of interest and advancing our understanding of the research domain. The use of co-citation analysis identified 15 distinct clusters within this field. Among these clusters, the largest six have been highlighted as particularly significant. Analysis with CiteSpace revealed that three clusters - cluster #0, #1, and #2 - experienced the most prominent citation bursts from 2012 to 2022, suggesting that these clusters represented the most dynamic and influential research areas within the field during this period.

To offer a more general understanding of the clusters, Figure 3 visually represents some major clusters identified within the network. Additionally, Table 3 has been included to summarize their most important aspects. By examining these clusters in detail, researchers can better grasp the patterns and trends within the field of pronoun research and identify potential fields for further exploration.

The largest cluster #0 is highly cohesive, with a silhouette value of 0.861. It is classified as anaphora resolution by both LLR and LSI. Anaphora resolution is interpreting the link between a pronoun and its antecedent (the previous mention of the entity) in a discourse (Garrod, 2001). This

Figure 3. Cluster view of pronoun studies

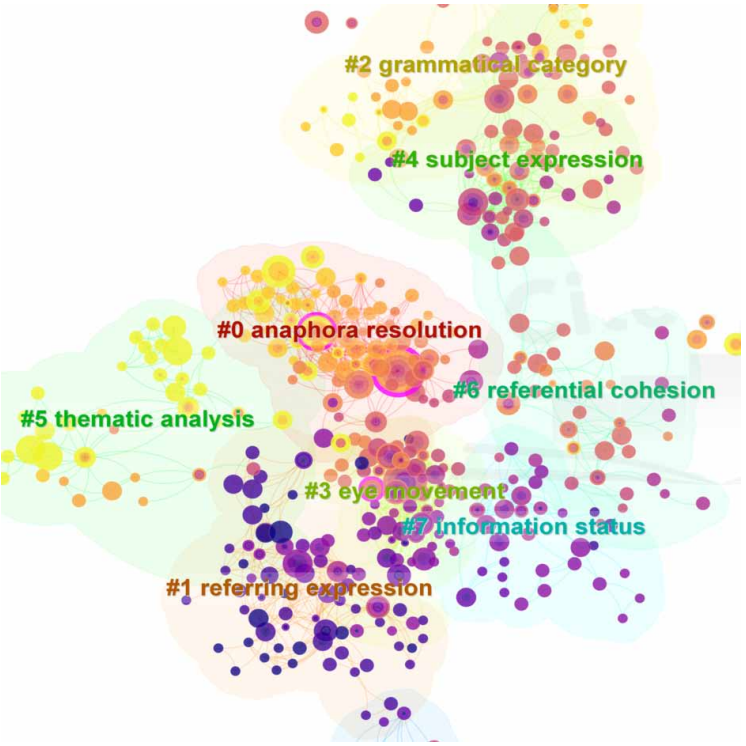


Table 3. Top 6 clusters of pronoun studies

Cluster ID	Size	Silhouette	Label (LLR)	Average Year
0	74	0.861	anaphora resolution	2016
1	68	0.925	referring expression	2009
2	60	0.947	grammatical category	2015
3	58	0.843	eye movement	2012
4	47	0.962	subject expression	2013
5	46	0.947	thematic analysis	2018

process involves using contextual information to determine the referent of a pronoun, which helps to establish coherence and meaning in discourse. Rohde and Kehler (2014), Filiaci et al. (2014), and Kehler and Rohde (2013) are the three articles with the most citations, as reviewed earlier (see section 3.2). Using a sentence completion task, Rohde and Kehler (2014) conducted a study on the effects of factors related to grammar and information structure on English pronoun production. They found that only the grammatical role of the antecedent was significant in determining pronoun production, supporting their Bayesian model. Meanwhile, Filiaci et al. (2014) examined the comprehension of null subjects and overt pronouns in sentences with two potential antecedents among speakers of Spanish and Italian. They found that null subjects were interpreted similarly in both languages, but overt pronouns showed cross-linguistic differences. Finally, the probabilistic model of Kehler and

Rohde (2013) integrated coherence-driven theory with centering-driven theory in order to determine the referent of a pronoun more accurately.

With 68 members and a silhouette value of 0.925, cluster #1 is the second largest. It is classified as referring expression by both LLR and LSI. Referring expression is defined as specific and non-predicative noun phrases (Jørgensen, 1998). According to his description, "referring expression" cannot be classified as a mere syntactic category. Rather, it encompasses various forms of proper names, pronouns, definite and indefinite descriptions. Fukumura and van Gompel (2011), Arnold (2010), and Kehler et al. (2008) are the three articles that have received the most citations. Fukumura and van Gompel (2011) investigated the influence of animacy on reference selection (pronoun or noun phrase) in reference production. Their results showed that native speakers of British English preferred to use pronouns for animates, suggesting that the reference strategy varies with the animacy of the referent. Arnold (2010) examined the influence of accessibility on speakers' use of language to refer to objects and concepts during production. Factors that impact accessibility, such as givenness, and syntactic prominence, were discussed, and examples were provided. Kehler et al. (2008) examined people's interpretation of pronouns, specifically looking at the role of coherence in guiding interpretation. The article reported three experimental studies that supported a coherence-driven analysis of pronoun interpretation and proposed that this approach could explain previously suggested interpretation biases. The results showed that probabilistic expectations about coherence relations and entities to be mentioned play a significant role in pronoun interpretation.

The third largest cluster #2 has a high silhouette value of 0.947 and comprises 60 members. It is labeled as grammatical category by both LLR and LSI. Grammatical category focuses on the morphological and syntactic properties of pronouns. Grammatical categories are a class of linguistic elements that possess grammatical significance, as opposed to lexical categories (Matthews, 2014). They are characteristics of linguistic items within a language's grammar that can take on different values. According to citation records, Harbour (2016), Ackerman et al. (2018), and Moskal (2015) are the three most frequently cited articles on this topic. Harbour (2016) investigated the ways in which languages employ to express the concept of person. He addressed the interaction of pronouns with number features and their ability to create intricate structures when combined with other elements. Ackerman et al. (2018) investigated the employment of resumptive pronouns to resolve illicit island extractions in language. They conducted experiments and found that the use of resumptive pronouns could facilitate sentence processing with these extractions. Moskal (2015) explored the difference in case-driven suppletion between lexical nouns and pronouns. They proposed that a minimal locality principle based on syntactic hierarchy could explain the reason that suppletion rarely affects lexical nouns but often occurs in pronouns.

With 58 members and a silhouette value of 0.843, cluster #3 is the fourth largest cluster. It is classified as eye movement by LLR. Online methodologies such as eye-tracking, which operate in real-time and concurrently, are a type of data collection method that can provide insights into a participant's language processing, whether receptive or productive, as it unfolds (Godfroid, 2019). According to the citation data, the most frequently cited article is Dillon et al. (2013), followed by Chow et al. (2014), and Cunnings and Felser (2013). Dillon et al. (2013) (see section 3.2) examined the processing strategies of agreement and reflexives of the English language using eye-tracking experiments. Their study found that structurally illicit noun phrases affected subject-verb agreement, while reflexives were not, and this was attributed to differences in memory retrieval mechanisms. Chow et al. (2014) (see section 3.2) conducted experiments on pronoun interpretation and memory retrieval processes. They found that agreement features and structural restrictions were used during initial retrieval, and that readers might consider grammatically illicit antecedents if initial retrieval failed. Cunnings and Felser (2013) presented two eye-movement experiments examining the impact of working memory capacity on the comprehension of sentences with reflexive pronouns that adhered to Principle A. The authors investigated whether readers applied Principle A consistently or if it depended on their working memory capacity, and the results showed that English readers with

limited working memory capacity were more likely to consider alternative referents for the reflexive pronouns, even if it violated Principle A.

It is worth noting that other significant clusters, such as cluster #4 (subject expression) and cluster #5 (thematic analysis), deserve mention. Subject expression refers to a subject pronoun which typically refers to the person or thing that performs an action and is usually located at the beginning of a sentence, preceding the verb. In cluster #4, Carvalho and Bessett (2015) examined the impact of various factors on the usage of subject pronouns in Uruguayan border Spanish. The investigation considered both linguistic and extra-linguistic factors. The findings revealed that bilingual speakers employed subject pronouns at comparable rates to monolingual speakers, indicating minimal influence from Portuguese. Travis and Torres Cacoullos (2012) examined instances where speakers of Spanish, a "pro-drop" language that allows for the omission of subjects, opt to use the first-person singular pronoun "yo". They employed multivariate analysis on conversational data to evaluate various hypotheses regarding factors that affect the usage of "yo". The findings reveal that cognitive, mechanical, and constructional elements drive the employment of "yo" more than a contrastive purpose. In cluster #5, Bjorkman (2017) discussed the use of the singular "they" in English, which is usually used to refer to a person whose gender is unknown or specified. The study provided explanations for the use of "they" as a pronoun that can refer to a person of any gender and as a pronoun that refers to a specific person whose gender is unknown or irrelevant. Meanwhile, Arnold et al. (2018) explored the effect of language experience on the comprehension of ambiguous pronouns and found that individuals with greater written language exposure were more proficient at assigning pronouns to their grammatical subject.

Trending Topics for Pronoun Studies

The study of pronouns has become increasingly relevant in recent years, with researchers exploring various aspects of their usage and processing mechanisms. One method of analyzing intellectual structure is by examining frequently used keywords. The current study identified the top eight most used keywords (counts > 100): pronoun, language, English, acquisition, discourse, comprehension, syntax, and information. These keywords provide insight into the most pressing concerns within the field. As expected, "pronoun" is the most frequently mentioned keyword, emphasizing its significance to this study area. Meanwhile, the predominant focus of studies is on the English language, indicating that research in this field primarily centers around English pronouns. The acquisition of pronouns is also a notable theme, with studies investigating the process by which individuals acquire pronouns in their native language (mostly English) and second language. Additionally, comprehension of pronouns requires attention to discourse and syntactic cues and the retrieval of other relevant information. Table 4 displays the top-mentioned keywords and their corresponding frequency counts, highlighting the

Table 4. Top 8 keywords in research on pronouns

Rank	Keyword	Counts
1	pronoun	323
2	language	242
3	English	163
4	acquisition	159
5	discourse	149
6	comprehension	140
7	syntax	130
8	information	115

most prominent research topics on pronouns. These findings can inform future studies and guide researchers in identifying important areas for further investigation.

Firstly, research on pronouns has been widely conducted in the English language, with abundant theoretical and empirical studies. Theoretical studies of English pronouns have mainly focused on two aspects. One area of research has centered on syntactic constraints and their application in sentence structure. Most of the work follows theories such as the Government and Binding Theory and the Minimalist Program proposed by Chomsky (1981, 1992). Other studies have followed and focused on the role of pronouns in establishing coherence and cohesion, including how they contribute to the overall structure (Halliday & Hasan, 1976). Empirical studies of English pronouns have also been abundant and have shown, to a certain extent, the influence of the above-mentioned theoretical studies. Researchers have conducted extensive investigations into a diverse range of topics concerning pronouns, encompassing but not limited to the processing and comprehension of pronouns in discourse (Grüter et al., 2017; Puebla & Felser, 2022; Trnavac & Taboada, 2016), and syntactic constraints on pronoun resolution (Drummer & Felser, 2018; Kazanina et al., 2007; Kush & Dillon, 2021; Patterson & Felser, 2019; Sutton, 2015). Although compared to studies on pronouns in other languages, such as Spanish and Portuguese, investigations on English pronouns account for a substantial portion, it does not mean that pronouns in other languages are less interesting and important. In fact, these languages have some unique features and variations in their pronoun systems that deserve attention. For example, Spanish and Portuguese have variable subject pronoun expressions, meaning that verbs may or may not be accompanied by a phonetically realized subject. Meanwhile, Chinese does not have a direct equivalent for the English pronoun "it". It uses alternative structures to convey a similar meaning. For instance, instead of saying, "It's raining", a Chinese speaker might say, "The rain is falling".

Secondly, researchers have been particularly interested in the study of pronoun acquisition. Language comprehension refers to an individual's capacity to comprehend spoken or written languages. This skill involves the capacity to understand the meanings of words and to comprehend their organization in order to construct coherent sentences. Successful language comprehension also requires the ability to supervise one's own understanding and use strategies to repair breakdowns in communication. Native language speakers can typically complete this task quickly and subconsciously. Even native child speakers show impressive proficiency in this area. For example, Sutton (2015) explored the acquisition of Principle C by children and found that children had adult-like competence by 30 months of age. However, the task of processing appropriate meaning representations becomes even more impressive when considering individuals learning a language as a second or third language later in life. Research on L2 learners' pronoun processing has shown that L1 and L2 speakers have different processing patterns. Some have attempted to identify the underlying causes of these performance discrepancies, with some suggesting insufficient or inadequate knowledge or experience of the second language (Contemori et al., 2022), while others have attributed the differences to L1 influence (Murakami & Alexopoulou, 2016) or age-of-acquisition effects (Wang et al., 2023). Moreover, studies have also documented that L2 pronoun processing may be subject to limited mental capacity or resources, less awareness or attention to grammar, or increased vulnerability to confusion or distraction in the memory when using the second language (Cunnings, 2017). Nevertheless, despite the obstacles encountered while acquiring a second language, L2 learners' ability to attain proficiency is truly remarkable.

Thirdly, various types of information influence referential decisions in language comprehension. These include structure-sensitive and morphological restrictions, discourse-level and contextual cues, and the degree of referents' elaboration, etc. These factors interact during language processing and guide the decision-making process for language users. However, the weighting of these information sources can sometimes be improper or delayed, causing short-term confusion, distraction, or retrieval failure of candidates that match some features but are not the intended target. The effects of interference provide valuable insights into how information sources interact during the construction of dependencies. Comparing the interference patterns between L1 and L2 speakers can help enlighten how these populations utilize the different information sources that aid in reference resolution and to what extent

they use them. Many studies in this field have focused on identifying cues, such as the focus of verbs, discourse prominence, and syntactic cues, that influence pronoun interpretation (Patterson et al., 2017; Puebla & Felser, 2022; Sportiche, 2013). Earlier research on non-native anaphor resolution, in particular, has shown that L2 speakers are significantly more affected by discourse cues than L1 speakers (Puebla & Felser, 2022). This suggests that L2 speakers may be more dependent on higher-level information, such as discourse context, to guide their referential decisions. In contrast, L1 speakers may rely more on lower-level information, such as morphological and syntactic cues. Overall, understanding L1 and L2 speakers' processing and resolving references in language can provide valuable insights into language processing, which can facilitate individuals' learning of a second language.

DISCUSSION

This study utilized CiteSpace software to conduct a document co-citation analysis of 2,774 articles on pronouns published between 2012 and 2022. Through the exploration of influential articles, main clusters, and trending topics, this study uncovered a reliable historiographic and systematic science mapping of pronouns. This approach has enabled us to understand the state of the field better and identify potential future research directions.

Multiple prominent clusters are revealed by the results of our document co-citation analysis, including anaphora resolution, referring expression, grammatical category, eye movement, subject expression, and thematic analysis. Trending topics mainly focus on the pronouns of the English language, the acquisition of pronouns, and discourse and syntactic information that affect pronoun comprehension. These major clusters and trending topics represent a diverse range of research domains in pronoun studies. Although there are various kinds of studies on pronouns, most studies can be generalized into two important aspects of pronouns, i.e., comprehension and production.

Investigations into pronoun resolution commonly analyze the instantaneous deployment of structure-dependent restrictions in first language and second language sentence comprehension and the interplay between diverse sources of knowledge during online linguistic analysis. Some investigations have demonstrated that the likelihood of a constraint being violated can vary depending on the type of constraint being studied, differing in languages, techniques, etc. For instance, in their study of structural restrictions on coreference, Kazanina et al. (2007) found that Principle C of the Government and Binding Theory had an early impact on the search process of native English speakers, resulting in interference from gender incongruent referents in structurally licit antecedent locations but not in syntactically illicit antecedent positions. However, Patterson & Felser (2019) and Drummer & Felser (2018) investigated German pronouns and found that Principle C effects showed a delayed processing effect during online reading in their study of German pronouns. Some other research investigates whether speakers of two typologically related languages comprehend pronouns using the same mechanism, whether the performance of constraints is the same across languages, and whether the interactions of various information, such as syntactic, discourse, and pragmatic factors, share the same parameter settings. For instance, Contemori et al. (2022) investigated bilingual Spanish-English speakers' representations of pronominal expressions and found that bilingual individuals' ability to make inferences about coherence relations and probability distributions was influenced by both their native and second language. Meanwhile, Puebla & Felser (2022) found that L2 participants exhibit a high level of sensitivity to discourse factors, i.e., topic shifts, suggesting that discourse-level cues carry greater weight in L2 processing than in L1 processing.

In addition to the comprehension of pronouns, there is also significant research interest in the production of pronouns. Research has shown that pronoun production is influenced by structural biases, such as syntactic roles. Buckle et al. (2017) investigated if the match between animacy and semantic function in prior dative sentences and following scenes influenced the syntactic constructions and noun order based on animacy chosen by children and adults. The research found that animacy information could impact speakers' word order separately from syntax, as well as through interactions with syntax. However, these mechanisms varied between children and adults. Similar to pronoun comprehension,

discourse-level factors have also been found to play a crucial role in pronoun production. Vogels et al. (2015) demonstrated that speakers who experienced cognitive load display a greater preference for pronouns, particularly when referring to less salient referents, showing that speakers experiencing cognitive load encountered challenges in considering the salience of discourse, leading them to choose expressions that were more efficient for themselves. Pronoun production has also been documented to exhibit cross-linguistic variation, showing different preferences by immigrants for pronoun use. Shin and Otheguy (2013) explored the influence of social rank and gender on a linguistic adaptation observed in Spanish spoken in New York City, specifically the increased use of Spanish subject pronouns. The research indicated that exposure to a different life in New York City and the use of the English language was correlated with the change in pronoun use.

Upon reviewing these studies, it becomes evident that many of them have sought to determine or pinpoint the variables that influence the comprehension and production of pronouns, regardless of whether the focus is on pronouns in English or other languages, the acquisition of pronouns in a native or second language, or typological interferences between pronouns. However, as Felser (2019) emphasized, there is a need for a comprehensive examination of the interplay between various constraints and information sources, focusing on identifying the linguistic cues that must be extracted from or re-examined in the input to apply a specific constraint effectively.

CONCLUSION

Through analysis and elaboration, our interpretation not only elucidated the intellectual underpinnings and structural components of the research domain on pronouns but also effectively characterized its major topics. Over the past decade, there has been a remarkable surge of publications in pronoun studies. Our results from co-cited documents analysis, clusters, and co-occurring keywords analysis demonstrated that the research on pronouns is closely linked with various fields such as anaphora resolution, grammatical category, and eye movement. By examining and interpreting scientific knowledge related to pronoun studies, this research clarifies the intricate research surrounding pronouns. Despite significant achievements, our understanding of the nature of pronoun processing still requires further theoretical and empirical studies. More cross-linguistic and typological studies are needed to uncover the encoding and processing of pronouns in different languages. Such studies could enlighten the different parameter settings of pronouns across languages, which may assist in language acquisition or related studies. Additionally, empirical studies are still needed to explore the impact of syntactic, discourse cues, and other factors on pronoun comprehension and production. An integrated and dynamic model of pronoun processing would be beneficial. While eye-tracking technology has been widely used in prior studies, future studies could benefit from the employment of neuroimaging techniques such as fMRI, ERP, or EEG. With the rapid development of these technologies, we may be able to understand the complex processing mechanism of pronouns better. Finally, more interdisciplinary studies are needed to explore how pronouns interact with other cues besides syntactic and discourse factors, such as prosody and gesture, in natural communication. These directions might lead to a deeper insight into the function and processing mechanisms of pronouns in human language.

COMPETING INTERESTS

The authors of this publication declare there are no competing interests.

FUNDING AGENCY

This study was supported by the Fundamental Research Funds for the Central Universities to Qingxu Zhai, grant number 2023JX097, and the Double First-Class Discipline Major Project of Beijing Foreign Studies University to Lin Fan, grant number 2022SYLZD008.

REFERENCES

Ackerman, L., Frazier, M., & Yoshida, M. (2018). Resumptive pronouns can ameliorate illicit island extractions. *Linguistic Inquiry*, *49*(4), 847–859. doi:10.1162/ling_a_00291

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203. doi:10.1111/j.1749-818X.2010.00193.x

Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, *102*, 41–54. doi:10.1016/j.jml.2018.05.002

Bjorkman, B. M. (2017). Singular they and the syntactic representation of gender in English. *Glossa (London)*, *2*(1), 1–13. doi:10.5334/gjgl.374

Buckle, L., Lieven, E., & Theakston, A. L. (2017). The effects of animacy and syntax on priming: A developmental study. *Frontiers in Psychology*, *8*, 2246. Advance online publication. doi:10.3389/fpsyg.2017.02246 PMID:29326639

Carvalho, A. M., & Bessett, R. M. (2015). Subject pronoun expression in Spanish in contact with Portuguese. In A. M. Carvalho, R. Orozco, & N. L. Shin (Eds.), Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective (pp. 143–166). Academic Press.

Chen, C. (2017). Science mapping: A systematic review of the literature. *Journal of Data and Information Science*, *2*(2), 1–40. doi:10.1515/jdis-2017-0006

Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. De Gruyter Mouton.

Chomsky, N. (1992). *The Minimalist Program*. MIT Press.

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, *5*, 1–16. doi:10.3389/fpsyg.2014.00630 PMID:25018739

Contemori, C., Mossman, S., & Ramos, A. K. (2022). Acquiring L2 pronoun interpretation biases: Implicit learning at the discourse level in L2 English. *Studies in Second Language Acquisition*, *44*(1), 210–227. doi:10.1017/S0272263121000152

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678. doi:10.1017/S1366728916000675

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1-2), 188–219. doi:10.1080/01690965.2010.548391

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. doi:10.1016/j.jml.2013.04.003

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. doi:10.1016/j.jbusres.2021.04.070

Drummer, J.-D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language*, *101*, 97–113. doi:10.1016/j.jml.2018.04.001

Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018). Individual differences in children's pronoun processing during reading: Detection of incongruence is associated with higher reading fluency and more regressions. *Journal of Experimental Child Psychology*, *173*, 250–267. doi:10.1016/j.jecp.2018.04.005 PMID:29753908

Felser, C. (2019). Structure-sensitive constraints in non-native sentence processing. *Journal of the European Second Language Association*, *3*(1), 12–22. doi:10.22599/jesla.52

Filiaci, F., Sorace, A., & Carreiras, M. (2014). Anaphoric biases of null and overt subjects in Italian and Spanish: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, *29*(7), 825–843. doi:10.1080/01690965.2013.801502

Fukumura, K., & van Gompel, R. P. G. (2011). The effect of animacy on the choice of referring expression. *Language and Cognitive Processes*, *26*(10), 1472–1504. doi:10.1080/01690965.2010.506444

Garrod, S. (2001). Anaphora Resolution. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 490–494). doi:10.1016/B0-08-043076-7/01527-8

Godfroid, A. (2019). *Eye Tracking in Second Language Acquisition and Bilingualism*. Routledge. doi:10.4324/9781315775616

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, *7*(2), 199–229. doi:10.1075/lab.15011.gru

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge.

Harbour, D. (2016). *Impossible Persons*. MIT Press. doi:10.7551/mitpress/9780262034739.001.0001

Jørgensen, S. (1998). What are referring expressions? How do we recognize them? *Pragmatics & Cognition*, *6*(1-2), 99–120. doi:10.1075/pc.6.1-2.07jor

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384–409. doi:10.1016/j.jml.2006.09.003

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44. doi:10.1093/jos/ffm018 PMID:22923856

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1-2), 1–37. doi:10.1515/tl-2013-0001

Kush, D., & Dillon, B. (2021). Principle B constrains the processing of cataphora: Evidence for syntactic and discourse predictions. *Journal of Memory and Language*, *120*, 305–324. doi:10.1016/j.jml.2021.104254

Matthews, P. H. (2014). *The Concise Oxford Dictionary of Linguistics* (2nd ed.). Oxford University Press.

Moskal, B. (2015). Limits on allomorphy: A case study in nominal suppletion. *Linguistic Inquiry*, *46*(2), 363–376. doi:10.1162/LING_a_00185

Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition*, *38*(3), 365–401. doi:10.1017/S0272263115000352

Pablos, L., Doetjes, J., Ruijgrok, B., & Cheng, L. L. S. (2015). Active search for antecedents in cataphoric pronoun resolution. *Frontiers in Psychology*, *6*, 1638–1638. doi:10.3389/fpsyg.2015.01638 PMID:26579023

Patterson, C., Esaulova, Y., & Felser, C. (2017). The impact of focus on pronoun resolution in native and non-native sentence comprehension. *Second Language Research*, *33*(4), 403–429. doi:10.1177/0267658317697786

Patterson, C., & Felser, C. (2019). Delayed application of binding Condition C during cataphoric pronoun resolution. *Journal of Psycholinguistic Research*, *48*(2), 453–475. doi:10.1007/s10936-018-9613-4 PMID:30443890

Puebla, C., & Felser, C. (2022). Discourse prominence and antecedent mis-retrieval during native and non-native pronoun resolution. *Discours*, *29*(29), 3–31. doi:10.4000/discours.11720

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, *29*(8), 912–927. doi:10.1080/01690965.2013.854918

Shin, N. L., & Otheguy, R. (2013). Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society*, *42*(4), 429–452. doi:10.1017/S0047404513000468

Sportiche, D. (2013). Binding theory—Structure sensitivity of referential dependencies. *Lingua*, *130*, 187–208. doi:10.1016/j.lingua.2012.11.005

Sutton, M. (2015). *Competence and performance in the development of Principle C* [Unpublished doctoral dissertation]. University of Maryland.

Travis, C. E., & Torres Cacoullos, R. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics*, *23*(4), 711–748. doi:10.1515/cog-2012-0022

Trnavac, R., & Taboada, M. (2016). Cataphora, backgrounding and accessibility in discourse. *Journal of Pragmatics*, *93*, 68–84. doi:10.1016/j.pragma.2015.12.008

Vogels, J., Krahmer, E., & Maes, A. (2015). How cognitive load influences speakers' choice of referring expressions. *Cognitive Science*, *39*(6), 1396–1418. doi:10.1111/cogs.12205 PMID:25471259

Wang, J., Liang, L., & Chen, B. (2023). The age of acquisition effect in processing second language words and its relationship with the age of acquisition of the first language. *Language and Cognition*, *15*(2), 1–22. doi:10.1017/langcog.2022.40

Xu, X., Pan, M., Dai, H., Zhang, H., & Lu, Y. (2019). How referential uncertainty is modulated by conjunctions: ERP evidence from advanced Chinese–English L2 learners and English L1 speakers. *Second Language Research*, *35*(2), 195–224. doi:10.1177/0267658318756948

Zhao, M., Yin, Z., & Lu, Z. (2020). EEG evidence for Game-Theoretic Model to ambiguous pronoun resolution. In B. Liao & Y. N. Wáng (Eds.), *Context, Conflict and Reasoning* (pp. 47–62). Springer Singapore. doi:10.1007/978-981-15-7134-3_4

Qingxu Zhai is a PhD candidate in Research Institute of Foreign Languages at Beijing Foreign Studies University. His research interests mainly focus on second language acquisition.

Lin Fan is a professor in the Artificial Intelligence and Human Languages Lab at Beijing Foreign Studies University. Her research interests concentrate on the interplay between language and emotion, the inferential processing in discourse comprehension, as well as code-switching.