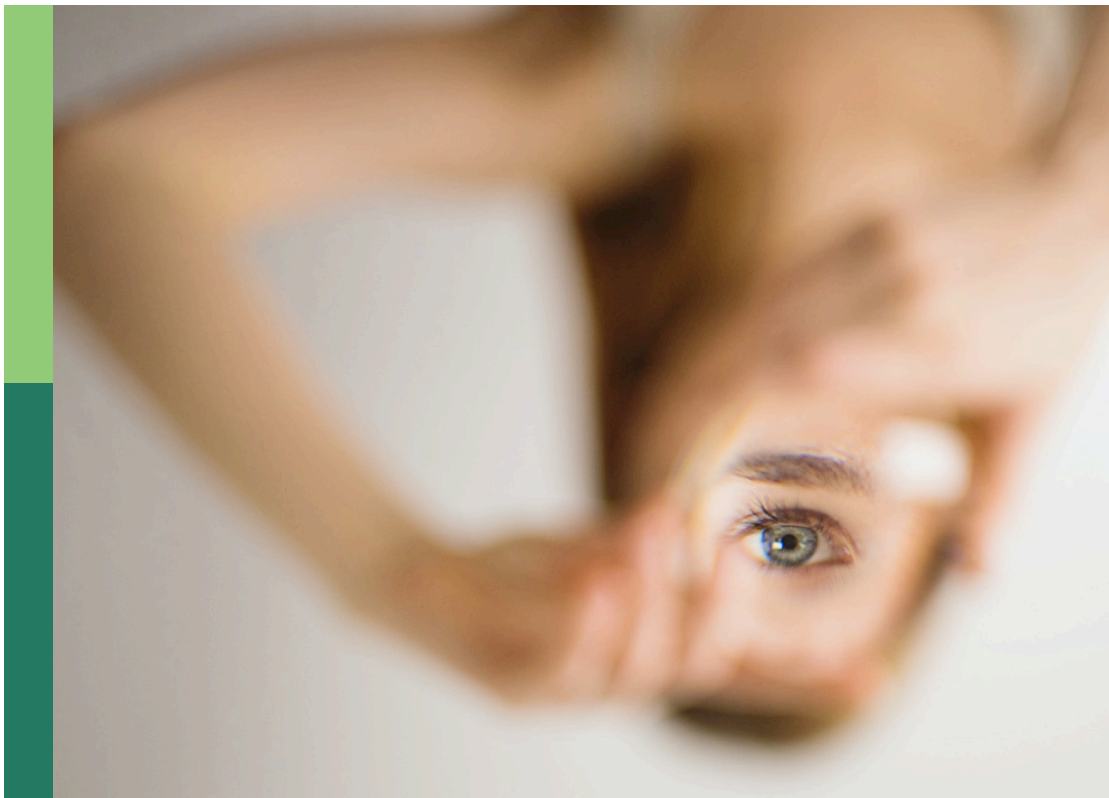
frontiers | Research Topics

Reviews in psychology of language

Edited by
Antonio Bova, Thomas L. Spalding and
Antonio Benítez-Burraco

Published in
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-6094-5
DOI 10.3389/978-2-8325-6094-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Reviews in psychology of language

Topic editors

Antonio Bova — Catholic University of the Sacred Heart, Italy

Thomas L. Spalding — University of Alberta, Canada

Antonio Benítez-Burraco — University of Seville, Spain

Citation

Bova, A., Spalding, T. L., Benítez-Burraco, A., eds. (2025). *Reviews in psychology of language*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6094-5

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 25 February 2025
DOI 10.3389/fpsyg.2025.1569614

OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE
Antonio Benítez-Burraco
✉ abenitez8@us.es

RECEIVED 01 February 2025
ACCEPTED 11 February 2025
PUBLISHED 25 February 2025

CITATION
Benítez-Burraco A, Bova A and Spalding TL (2025) Editorial: Reviews in psychology of language. *Front. Psychol.* 16:1569614. doi: 10.3389/fpsyg.2025.1569614

Editorial: Reviews in psychology of language

Antonio Benítez-Burraco [1]*, Antonio Bova[2] and Thomas L. Spalding[3]

[1]Department of Spanish, Linguistics, and Theory of Literature, Faculty of Philology, University of Seville, Seville, Spain, [2]Department of Psychology, Faculty of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy, [3]Department of Psychology, Faculty of Arts, University of Alberta, Edmonton, AB, Canada

KEYWORDS

psycholinguistics, cognition, language acquisition, second language learning, language disorders, language evolution

Editorial on the Research Topic
Reviews in psychology of language

Traditionally, research on language facts has focused on overt linguistic behaviors. Linguists have examined human languages to learn about their fundamental components and how these basic pieces are arranged into more complex sets, from syllables to words to sentences to discourses. Typologists have described hundreds of languages and found that they share a core set of components and structural principles, supporting the view that all human languages are similarly designed and fulfill similar roles equally well. Dialectologists and sociolinguists have characterized different varieties of each single language and show that intralinguistic variation follows similar paths and results from similar triggering factors as interlinguistic diversity. And the same is true for language change as characterized by historical linguistics. Nonetheless, for a long time, languages were regarded as cultural artifacts mostly, like food practices, religions, or types of costumes. In the second half of the twentieth century, insights on how languages are acquired by children started to change this traditional conceptualization of languages (and of language as a human distinctive trait). Nowadays, language is generally construed as a key component of the human phenotype, particularly, of our mind/brain. Nativist views of language gained preeminence during the last decades, to the extent that language was even thought of as an organ that grows in our brain under genetic guidance. This view has been toned down, so that both our genome and our environment are thought to contribute to our distinctive linguisticality. In any case, it is generally acknowledged that if we want to understand the ultimate nature of language, it is necessary to delve into the brain black box in order to know which aspects of our mind/brain support language and in particular, if they are specific to language or domain-general by nature.

Two related disciplines have led such a crucial line of inquiry: psycholinguistics and neurolinguistics. The former aims to know about the mental processes that allow us to understand and produce language, but also to acquire our mother tongue and to learn other languages. The latter tries to identify and characterize the brain circuits that support language processing and language acquisition/learning. In other words, psycholinguistics is mostly concerned with the software of language, whereas neurolinguistics is mostly interested in its hardware. Over the years, as with linguistics more generally, psycholinguistics has evolved to be more and more methodologically complex and theoretically diverse. Many technological advances (e.g. eye-tracking)

allow psycholinguists to conduct truly sophisticated experiments to address questions previously impossible to answer. Similarly, different theories about language processing and acquisition have emerged with time, and research has become increasingly diverse, as non-European languages and non-standard varieties of languages have been examined by psycholinguists. Finally, research has also evolved to be more multidisciplinary, as contacts with other subfields of linguistics (particularly, neurolinguistics), and other disciplines (like computational science, or biology) are helping psycholinguists to construct more robust hypotheses about the nature of language and to explore new avenues of research.

The aim of this Research Topic is to gather comprehensive and up to date review articles on key aspects of psycholinguistics. Because psycholinguistics is a notably dynamic and increasingly complex field, as noted, it is difficult (and urgent) for researchers to be up to date. In this Research Topic, we have brought together 10 contributions from 30 scholars.

Starting with articles addressing basic aspects of psycholinguistic research, the article authored by Sun and Lin provides a state-of-the-art review of past research on metonymy, a core cognitive operation in language processing, acquisition, and change. The bibliometric analysis performed by these authors reveals that theoretical and cognitive issues are still at the forefront of research on this topic, but also that some underexplored aspects are gaining attention, particularly, the intersection between metonymy and other diverse domains, including the emotional sphere, selected social and cultural dimensions, and other communicative modalities, particularly, vision. In turn, Renström reviews the potential impact of pronoun usage on gender conceptualization, with some attention to attitudinal issues, to understand both the resistance and the promotion of gender-inclusive language and linguistic gender reforms, more generally.

Reflecting ample interest, both past and present, in the psycholinguistics of language change during the lifespan, but also of language disorders, we have included 3 contributions on these issues in the Research Topic. Ansari et al. have conducted a meta-analysis of recent literature about word learning by children with Developmental Language Disorder (DLD), which contributes to clarify key features of the process, but also to suggest ways of improving the interventions aimed to facilitate vocabulary learning by these children. Janssen et al. discuss the role of executive function in the (dis)abilities of children with DLD for storytelling, also with the ultimate objective of achieving better strategies to ameliorate their reduced narrative capacities. Finally, Lesecq et al., which also focus on non-typical children, have found that gifted children exhibit more heterogeneous reading abilities compared to their peers.

For a long time, the field of psycholinguistics has been highly interested in the mechanisms involved in second language learning. Two of the articles comprising the Research Topic address this issue. Zhang et al. examine the impact of parents' investment behavior on the learning success of a sample of Chinese students of English as a second language (L2). They found that both participation and investment by the parents have a positive effect on the motivation and learning behavior of children. In turn, Hui and Chen provide insights on the effect of socioeconomic status on pragmatic awareness, using a sample of Chinese students learning English as an L2.

Two additional contributions to the Research Topic examine aspects of the psycholinguistics of language acquisition/learning, but with a focus on methodological issues. Harrag et al. have authored a article evaluating the utility of semantic indices as diagnostic and assessment tools of language change. Specifically, they focus on the role of propositional density, i.e. the amount of information conveyed per language segment, as a reliable tool for tracking language change during aging. With regards to the article by Li and Zhong, it reviews the usage of eye-tracking techniques for gaining insights on the psycholinguistics of translation. Their review article highlights several domains to which eye-tracking has been successfully applied, mostly related to human-machine interaction, as well as ongoing trends in the psycholinguistics of translation, which seems to be evolving to be more empirical and multidisciplinary, but less theoretically-motivated with time.

Finally, the article by Benítez-Burraco provides a general framework for language evolution studies in the human species. According to the author, modern languages resulted from selected changes in our cognition, including the emergence of improved ways of processing grammar rules. These changes might have resulted in part from our trend toward more prosocial behaviors, which also fostered the cultural evolution of languages. Ultimately, the article argues for a multidisciplinary approach to language evolution in which psycholinguistic research would still play a key role.

Author contributions

AB-B: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. AB: Conceptualization, Investigation, Writing – review & editing. TS: Conceptualization, Investigation, Writing – review & editing.

Acknowledgments

We would like to thank all the contributors to this volume (both authors and reviewers).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 May 2024
DOI 10.3389/fpsyg.2024.1380346

OPEN ACCESS

EDITED BY
Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Yuan Yao,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Yang Frank Gong,
University of Macau, China

*CORRESPONDENCE
Xiaobing Lu
✉ luxiaobing@nju.edu.cn
Honggang Liu
✉ liuhonggang@suda.edu.cn

RECEIVED 01 February 2024
ACCEPTED 02 May 2024
PUBLISHED 20 May 2024

CITATION
Zhang Y, Lu X and Liu H (2024) Relationship between learners' L2 Motivational Self System and parental investment behavior in learners' English learning.
Front. Psychol. 15:1380346.
doi: 10.3389/fpsyg.2024.1380346

Relationship between learners' L2 Motivational Self System and parental investment behavior in learners' English learning

Yu Zhang[1], Xiaobing Lu[2]* and Honggang Liu[3]*

[1]Gongping Middle School of Changchun, Changchun, China, [2]Department of Applied Foreign Language Studies, Nanjing University, Nanjing, China, [3]School of Foreign Languages, Soochow University, Suzhou, China

Students' motivation and learning behavior are significantly impacted by parents' participation and investment. It has been demonstrated that parental investment behavior could exert a direct effect on students' L2 Motivational Self System (L2MSS) mediated by parental investment belief. Nevertheless, the relationship between components of parental investment behavior and students' language learning motivation remains a topic necessitating further scholarly investigation. In response to this gap, we conducted a quantitative study involving a survey of 900 high school students to explore the relationship between students' English learning motivation, as conceptualized by the L2 Motivational Self System and parental investment behavior based on a four-component model. The findings in this study indicated that high school students exhibited moderate levels of L2MSS and relatively low levels of parental investment behavior. Moreover, students' L2MSS was found to have a significant positive correlation with the global parental investment behavior, with parental emotional investment behavior emerging as a positive predictor of high school students' L2MSS. These results underscore the importance of parental provision of sufficient economic, relationship, knowledge, and emotional support in cultivating a nurturing and supportive familial context conducive to the development of students' positive future selves.

KEYWORDS

correlation, L2 Motivational Self System, parental investment behavior, senior high school students, English learning

1 Introduction

Parents' active involvement and substantial investment within the family context serve as crucial factors that can significantly affect students' academic engagement (Fan and Williams, 2010; Castro et al., 2015; Hosseinpour et al., 2015), enhance students' self-regulation capabilities (Gonzalez-DeHass et al., 2005), support students' goal-setting processes (He et al., 2015), and improve students' academic outcomes (Fan and Chen, 2001). In the Chinese EFL (English as a Foreign Language) educational framework, English is a required component of the senior high school curriculum. Parents demonstrate a commitment to investing financial and material resources in their children's language education to enhance future career prospects. Parents' perceptions of the importance and necessity of English learning significantly shape students' motivational levels.

Parental investment behavior was found to directly influence students' English learning motivation (Liu, 2024). Research on the interplay between parental investment behavior and students' English learning motivation warrants further exploration, as it aligns with the burgeoning focus on learner-internal factors within external contexts. This study aims to examine the relationship between parental investment behavior and students' L2MSS to achieve a more nuanced understanding of how various forms of parental investment behavior—including economic, knowledge-based, relational, and emotional—impact students' motivation to learn English.

2 Literature review

2.1 L2 Motivational Self System

Motivation has been a prominent research topic due to its key role in supporting long-term acquisition (Dörnyei and Ryan, 2015; Csizér, 2019; Al-Hoorie and Szabó, 2022; Yousefi and Mahmoodi, 2022). Dörnyei (2005) introduced the L2 Motivational Self System consisting of three components: the ideal L2 self, the ought-to L2 self and the L2 learning experience. The ideal L2 self represents the aspirational self-concept that second language learners aim to embody in the future. The ought-to L2 self encompasses the perceived obligation to meet external expectations and avert negative outcomes. Finally, the L2 learning experience pertains to learners' educational experiences, influenced by pedagogical contexts such as instructors, peers, curriculum, and personal achievements (Dörnyei, 2005, 2009).

Subsequently, the L2MSS won a popularity in researching L2 motivation across diverse social-cultural contexts, owing to its adaptability and its capacity to integrate a multiplicity of theoretical perspectives (Al-Hoorie, 2018). A growing interest in motivation has emerged with two major topics. Some studies specifically validated the L2MSS in different contexts, such as China (You and Dörnyei, 2016) and Iran (Papi, 2010) and confirmed the validity of the anticipated construct. Lee and Lee (2020) investigated the L2MSS of 105 college students and 112 high school students in South Korea and found that the level of the ought-to L2 self of high school students was higher than that of university students because compared with university students, the English learning of high school students was mainly motivated to meet the expectations and recognition of others, such as family, teachers and peers.

Several scholarly investigations have examined the association between the L2MSS and other learner-internal and external factors, such as anxiety (Papi, 2010), willingness to communicate (Lee and Lee, 2020), engagement (Zhu et al., 2022), L2 achievement (Moskovsky et al., 2016) and learning style (Kim and Kim, 2014). Research by Magid (2009) revealed distinctions in the conceptualization of the ideal L2 self between middle school and college students. College students' conceptualization of the ideal L2 self was found to be more intricate and well-developed than that of middle school students, who appeared to struggle with formulating a cohesive and mature vision. College students not only experienced intrinsic satisfaction in learning English but also placed significant emphasis on the practical application of the language for future career advancement.

According to Papi (2010), there is a tendency for students to experience increased levels of anxiety when their learning behavior is heavily influenced by their ought-to L2 self. The external factors that influence students' L2 learning motivation included students' family background and learning contexts. Ryan (2009) explored the English learning motivation of 2,397 Japanese middle school students and college students through a mixed-methods study and found that students could gain a higher social status and positive learning experience through proficient English. Khany and Amiri (2018) demonstrated that the ideal L2 self and the L2 learning experience of Iranian high school students were the main motivational factors to stimulate learners' learning behavior. Li (2019) explored the English learning motivation of 128 Chinese high school students within the framework of L2MSS. The results showed that high school students' English learning was more inclined to be influenced by language learning experience, and the correlation between English learning experience and expected effort was the highest compared with the ideal L2 self and the ought-to L2 self.

English education constitutes a comprehensive and multifaceted endeavor that encompasses a wide array of elements, including the development of educational policies and curriculum standards, the design of teaching materials, and the establishment of examination and evaluation criteria at the macro level (Liu, 2017). At the micro level, it involves the support of teachers (Liu and Li, 2023) and the financial and emotional investment of parents (Zhang and Liu, 2022). Given this complex structure, it is essential to consider the impact of parental influence within the family environment on students' motivation to learn a language. Understanding the role of parents in shaping students' language learning experiences can contribute to the development of more effective strategies to enhance students' motivation and engagement.

2.2 Parental investment behavior

Parents' involvement in education and educational psychology research has attracted scholarly attention for its significant impact on their children's academic and emotional development (Boonk et al., 2018). Parental involvement is an essential factor in enhancing students' academic development (Wigfield et al., 2006; Asgari and Mustapha, 2011; Huang, 2013), promoting students' self-regulation ability (Gonzalez-DeHass et al., 2005), and improving students' academic performance (Fan and Chen, 2001). In the language education research, Liu (2017) developed the concept of "parental investment behavior" underpinned by Bourdieu (1986)'s Capital Theory and the concept of "investment" (Peirce, 1995). The parental investment behavior refers to "forms of investment in their children's English learning that are commensurate with the parents' capital in multiple ways and to diverse extents" (Liu, 2024, p.2). The parental investment behavior is classified into economic, relationship, knowledge, and emotional investment behaviors (Liu, 2017, 2024).

According to Liu (2017, 2024), the economic investment behavior pertains to the materials and resources parents provide for their children based on their economic capital, such as enrolling them in English extracurricular activities. The relationship investment behavior involves parents utilizing their social capital to help their children gain access to a key school with higher quality teaching. The knowledge investment behavior is linked to parents' symbolic and cultural capital, including tutoring their children in language knowledge and discussing foreign language cultural backgrounds at

home. The emotional investment behavior relates to parents leveraging their emotional capital, such as offering encouragement and support when their children struggle with academic performance. Parental investment behavior is primarily influenced by factors such as parental investment beliefs, social class, and educational level.

The concept of parental investment behavior aligns with the notion of parental involvement in the broader field of general education; however, parental investment behavior is more specifically grounded in Bourdieu's (1986) Capital Theory. This framework has been adapted and applied in the context of foreign language education, offering a more targeted approach to understanding the nuances of parental engagement in this specialized domain. Moreover, it covers wider range of parental investment behavior, which enriched and extended the concept of parental involvement. Parental investment behavior encompasses not only specific parental involvement practices such as tutoring children at home (Epstein et al., 2002), participating in school activities (Grolnick and Slowiaczek, 1994), and providing emotional support (Allatt, 1993), but also extends to relational investment behaviors at the social level. Given these considerations, the present study has chosen to adopt the concept of parental investment behavior as the focus of investigation.

2.3 L2 learning motivation and parental involvement

While limited research has been conducted on parental investment behavior in language education, existing studies in general education and educational psychology provide valuable insights into the role of parental involvement. The present study reviewed the relevant studies on parental involvement in general education and educational psychology, from which the research in language education also gained affordance and insights. Positive parental support and investment have been shown to enhance students' learning motivation, attitudes, and behaviors (e.g., Epstein et al., 2002; Pomerantz et al., 2007; Fan and Williams, 2010; Kong and Wang, 2021; Zhu et al., 2022; Wang and Liu, 2024). In a review of prior literature, Gonzalez-DeHass et al. (2005) concluded that parental engagement and participation positively impact students' motivational constructs, such as self-regulation, intrinsic and extrinsic motivation, and mastery goals.

Fan and Williams (2010) found that four parental involvement variables (i.e., parental guidance, home-school contact, aspirations for tenth-grade students' postsecondary education, and family rules concerning television watching) positively affected students' intrinsic English motivation. The research results of Villiger et al. (2014) showed that parents' emotional support predicted fourth-grade students' reading enjoyment and curiosity, and parental expectations also impacted students' reading curiosity. Kong and Wang (2021) demonstrated that parents' support and perception of usefulness are positively correlated with children's flow experience and intrinsic motivation in programming education. Boonk et al. (2022) concluded that parental involvement variables (parent–child discussions about education and parents' expectations) had a positive impact on children's motivation. Zhu et al. (2022) highlighted the importance of parental support in young L2 learners' learning, suggesting that parents should spend time reading with their children to increase the learners' motivation, learning engagement, and academic performance.

Based on the structural equation model, Liu (2017) identified a linear relation between parental investment belief, parental investment behavior, and middle school students' English learning motivation. Specifically, parents' investment behavior directly impacted students' English learning motivation. Additionally, students' motivation to study abroad was positively correlated with parental economic investment behavior, suggesting that more economic capital invested by parents would increase the opportunities for students to study abroad, as well as motivate them to do so. A further study conducted by Liu (2024) validated the internal structure of parental investment and found that parental investment belief had an indirect influence on high school students' L2MSS mediated by parental investment behavior.

In general education, there has been extensive research on the impact of parental investment and involvement on students' learning outcomes. However, there is limited research on the role of parental investment in learners' language learning motivation. The family environment, serving as an immediate environment for students, exerts a substantial impact on students' motivation to learn a foreign language (Gonzalez-DeHass et al., 2005; Schunk et al., 2014; Gong et al., 2023). Students' English learning is supported by parental investment, facilitated through close communication within the family environment. The students' motivation to learn English is influenced by both parental investment and interactions between students and their parents. This study aims to examine the relationship between different components of students' L2 Motivational Self System (L2MSS) and various aspects of parental investment behavior to further explore the critical role of parents in the family context during the process of students' English learning. Consequently, this study seeks to address three research questions.

> *RQ1*: What are the levels of high school students' L2MSS in terms of the ideal L2 self, the ought-to L2 self and the L2 learning experience?
>
> *RQ2*: What are the levels of the parental investment behavior in terms of the economic, relationship, knowledge and emotional investment behavior?
>
> *RQ3*: What is the relationship between students' L2MSS and the parental investment behavior?

3 Methodology

3.1 Research contexts and participants

In China, English is a mandatory subject in high schools and is a significant component of the college entrance examination (Gao, 2014; Zhang and Liu, 2022). Parents play a crucial role in their children's education, often actively engaging in their academic journey and aspiring for them to achieve high exam scores and improved career prospects (Liu, 2017, 2024). Given this context, this study selected high school students as its research participants. The convenience sampling method was utilized to survey Chinese high

school students due to the accessibility it offered (Rose et al., 2019). A total of 900 students from two northeastern cities in China were involved in this study, consisting of 403 male students (44.8%) and 497 female students (55.2%).

3.2 Research instrument

The study utilized a composite questionnaire including the demographic information, the L2MSS Scale, and the Parental Investment Behavior Scale. All items employed the Likert 5-point scale ranging from "1 (Strongly Disagree)" to "5 (Strongly Agree)." Demographic information collected the students' age, gender, and grade level as well as their parents' educational level and social status.

The L2MSS Scale was based on the L2MSS part of *English Learner Questionnaire* by You and Dörnyei (2016), including 16 items distributed in three dimensions, namely, ideal L2 self (e.g., Q2: I can imagine myself in the future giving an English speech successfully to the public in the future.), ought-to L2 self (e.g., Q7: Studying English is important to me in order to gain the approval of the society.), and L2 learning experience (e.g., Q15: I think time passes faster while studying English.). The scale demonstrated high reliability, with global and dimensional Cronbach's alpha coefficients of 0.885, 0.702, 0.684, and 0.760, respectively. The model's fit met the ideal criteria set by Hair et al. (2019): $\chi^2/df = 3.942$ (<5), CFI (Comparative Fit Index) = 0.979 (>0.90), TLI (Tucker-Lewis Index) = 0.967 (>0.90), RMSEA (Root Mean Square Error of Approximation) = 0.057 (<0.08), and SRMR (Standardized Root Mean Square Residual) = 0.033 (<0.10).

The Parental Investment Behavior Scale was taken from the *Family Education and English Learning Questionnaire* designed and validated by Liu (2024), consisting of 14 items classified into four sub-dimensions of parental investment behavior, namely, economic investment behavior (e.g., Q17: My parents pay for my English tutoring classes), relationship investment behavior (e.g., Q22: My parents made an effort to find a way to enroll me in the class I am currently attending), knowledge investment behavior (e.g., Q23: My parents converse with me in English as a means of improving my language skills), and emotional investment behavior (e.g., Q18: My parents encourage me when I make progress in learning English). This scale demonstrated higher reliability, with Cronbach's alpha coefficients for the four dimensions measuring 0.615, 0.803, 0.844, and 0.774, respectively. While the overall scale had a reliability of 0.847. The model fits meet the ideal criteria (Hair et al., 2019): $\chi^2/df = 4.428 < 5$; CFI = 0.948 > 0.90, TLI = 0.931 > 0.90; RMSEA = 0.062 < 0.008, and SRMR = 0.053 < 0.10.

3.3 Data collection and analysis

The questionnaire was administered in Chinese to facilitate accurate comprehension and appropriate responses from the students. Prior to commencing the survey, participants were provided with a thorough explanation of the survey's purpose, and informed consents were obtained from both students and their teachers. A total of 900 valid questionnaires were collected. Data were subsequently entered and analyzed using SPSS version 24.0. First, descriptive statistics were calculated for both the L2MSS and the parental investment behavior, including maximum, minimum, mean, and standard deviation values. Subsequently, the Pearson correlation coefficient was computed to examine the relationship between students' L2MSS and the parental investment behavior in response to the third research question. Lastly, a linear regression analysis was conducted to explore the relationship between the parental investment behavior and students' L2MSS. This analysis utilized the three dimensions of L2MSS as dependent variables and the four dimensions of the parental investment behavior as independent variables.

4 Results

4.1 Profiles of the L2MSS and parental investment behavior

As shown in Table 1, the global L2MSS and its three components received ratings that were marginally above the midpoint level of 3, with the L2 learning experiences domain achieving the highest mean score ($M = 3.33$, $SD = 0.84$). Repeated measures ANOVA indicated significant variations in mean scores across the dimensions of the L2MSS: $F(2, 1798) = 19.733$, $p < 0.001$, partial $\eta^2 = 0.022$. Specifically, pairwise comparisons with a Bonferroni adjusted alpha demonstrated that ideal L2 self ($M = 3.16$, $SD = 0.82$) achieved lower scores than L2 learning experience ($M = 3.33$, $SD = 0.84$) and ought-to L2 self ($M = 3.31$, $SD = 0.85$). No statistically significant difference was observed between the mean scores of L2 learning experiences and ought-to L2 self.

The level of parental investment behavior in high school students was 2.00 ($SD = 0.66$), below average. In terms of parental economic investment behavior, the level was 2.08 ($SD = 0.96$). Two related items (i.e., paying for English tutoring classes and arranging for one-on-one English tutoring) had the mean values of 2.42 and 1.75. This study found that the level of parental relationship investment behavior was 1.96 ($SD = 1.08$). The mean values of three related items (i.e., finding a way to enroll in the current class, in the current school, and a conductive English learning environment) were 1.96, 2.06, and 1.85. The mean of parental knowledge investment behavior in this study was 1.44 ($SD = 0.70$). The mean values of the four related items (i.e.,

TABLE 1 Descriptive analyses of variables.

Dimension	Max	Min	M	SD
Ideal L2 self	5.00	1.00	3.16	0.82
Ought-to L2 self	5.00	1.00	3.31	0.85
L2 learning experience	5.00	1.00	3.33	0.84
L2 Motivational Self System	5.00	1.19	3.26	0.66
Economic investment	5.00	1.00	2.08	0.96
Relationship investment	5.00	1.00	1.96	1.08
Knowledge investment	5.00	1.00	1.44	0.70
Emotional investment	5.00	1.00	2.52	0.90
Parental investment behavior	4.88	1.00	2.00	0.66

conversing in English, telling English stories, teaching English songs, teaching English knowledge) were 1.43, 1.44, 1.45, and 1.44. According to this study, the level of parental emotional investment behavior was 2.52 ($SD = 0.90$). As for the five related items (encouragement; helping analyze reasons about falling behind; setting an example by highlighting successful learners or by highlighting classmates; engagement in children's English learning), the mean values were 3.18, 2.51, 2.14, 2.45, and 2.32.

4.2 Relationship between the L2MSS and parental investment behavior

Table 2 presented the results of Pearson correlation between students' the L2MSS and parental investment behavior. There was a significant low correlation between the students' L2MSS and the parental investment behavior ($r = 0.170$, $p < 0.05$) (see in Table 2). The results showed no significant correlation between students' L2MSS and parental economic investment behavior. However, a significant low positive correlation existed between students' L2MSS and parental relationship investment behavior ($r = 0.262$, $p < 0.05$). Students' L2MSS was also positively correlated with parental knowledge investment behavior ($r = 0.132$, $p < 0.05$). Students' L2MSS and parental emotional investment behavior exhibited a significant low positive correlation ($r = 0.262$, $p < 0.05$).

Multiple linear regression was conducted to determine the best linear combination of the economic investment, the relationship investment, the knowledge investment and the emotional investment for predicting students' L2MSS. Statistical assumptions, such as the normal distribution of residuals and the non-linear correlation between predicted variables and residuals were all met in the analysis. The means, standard deviations, and correlation coefficients could be found in Tables 1, 2. The regression method of "enter" showed that the combination of the four independent variables significantly predicted students' L2MSS, F (4, 895) = 18.638, $p < 0.01$, with the economic investment and the emotional investment significantly contributing to the prediction ($p < 0.05$) except the relationship investment and the knowledge investment ($p > 0.05$) (see Table 3).

Models 2 to 4 in Table 3 examined the effects of the parental investment behavior on the three sub-dimensions of students' L2MSS. The results showed that the parental investment behavior had a positive effect on the three sub-dimensions of the L2MSS. Similarly, when the parental investment behavior was divided into four sub-dimensions for analysis on each sub-dimension of L2MSS, the economic investment behavior predicted students' ideal L2 self and ought-to L2 self. The emotional investment behavior significantly predicted the three dimensions of students' L2MSS. However, the relationship investment behavior only had a positive effect on the ideal L2 self, and the knowledge investment behavior had an effect on the ought-to L2 self.

5 Discussion

5.1 Profiles of the L2MSS and parental investment behavior

This study found that high school students had a higher level of the L2MSS, suggesting that the majority of students held a positive attitude towards learning English and maintained an ideal self-image related to language acquisition (You and Dörnyei, 2016; Lee and Lee, 2020; Thorsen et al., 2020). The levels of students' ideal L2 self, ought-to L2 self and L2 learning experience were slightly higher than the average. The level of students' L2 learning experience might relate to the learning atmosphere and teacher-student relationship (Liu, 2024). Students with a strong proficiency in English were more likely to receive recognition and support from peers and teachers, thereby achieving higher social standing in class and experiencing a more rewarding learning environment (Ryan, 2009; Zhang and Liu, 2022). Moreover, the high level of students' ought-to L2 self could be explained that parents emphasized English as more than just an academic subject, positioning it as a key factor in their children's future career prospects (Taguchi et al., 2009). High school students demonstrated motivation primarily driven by the desire to meet the expectations or gain recognition from parents, teachers, and peers (Lee and Lee, 2020; Zhang and Liu, 2022). Therefore, high school students' primary focus was on achieving higher marks rather than envisioning English in future career contexts, which might result in a less defined ideal L2 self-image (Magid, 2009).

This study found that the level of the parental investment behavior in high school students was below average, indicating that parents may not fully recognize the significance of English and consequently they may not invest sufficient capital in their children's English education. The parental investment and involvement could decrease as students grow older and the growing need of autonomy would reduce students' dependence on parents (Boonk et al., 2022). The level of the parental economic investment behavior was below average which was in line with the findings of Liu (2024). The parental economic investment behavior is influenced by their investment beliefs and the economic resources available to them (Liu, 2017, 2024).

The study also indicated that the level of parental relationship investment behavior was below average, paralleling the findings of Liu

TABLE 2 Correlation coefficients of variables.

	Glboal PIBeH	EcoInBe	EmoInBe	RelationInBe	KnowInBe
Global L2MSS	0.170**	0.004	0.262**	0.262**	0.132**
Ideal L2 Self	0.188**	0.033	0.244**	0.158**	0.106**
Ought-L2Self	0.102**	−0.027	0.184**	0.034	0.133**
L2 learning experience	0.116**	0.006	0.197**	0.066*	0.074*

$N = 900$, ** $p < 0.05$, * $p < 0.1$.
L2MSS, L2 Motivational Self System; PIBeH, parental investment behavior; EcoInBe, economic investment behavior; RelationInBe, relationship investment behavior; KnowInBe, knowledge investment behavior; EmoInBe, emotional investment behavior.

TABLE 3 Baseline regression analysis.

	Model (1)		Model (2)		Model (3)		Model (4)	
	L2MSS		IdealL2Self		OughtL2Self		L2LearnExp	
PIBeH	0.170***		0.188***		0.102**		0.116***	
EcoInBe		−0.106**		−0.078*		−0.105**		−0.069
EmoInBe		0.277***		0.236***		0.193***		0.219***
RelationInBe		0.021		0.098**		−0.037		0.008
KnowInBe		0.034		−0.010		0.092*		−0.004
_cons	2.923***	2.835***	2.691***	2.630***	3.047***	2.942***	3.029***	2.934***
N	900	900	900	900	900	900	900	900
R^2	0.029	0.078	0.035	0.069	0.010	0.049	0.014	0.043

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

L2MSS, L2 Motivational Self System; IdealL2Self, ideal L2 self; OughtL2Self, ought-to L2 self; L2LearnExp, L2 learning experience; PIBeH, parental investment behavior; EcoInBe, economic investment behavior; RelationInBe, relationship investment behavior; KnowInBe, knowledge investment behavior; EmoInBe, emotional investment behavior.

(2024). This may be attributed to the relatively stable nature of school and class assignments, which reduces the likelihood of school or class changes. It might be difficult for teachers who were faced with the management and guidance of a large number of students to sustain a productive relationship with each student's parents (Hill and Tyson, 2009).

The level of parental knowledge investment behavior was below average which might be influenced by parents' educational level (Alawawda and Razi, 2020; Liu, 2024). Parents with lower educational level may be unable to offer sufficient English knowledge tutoring to their children (Fan and Chen, 2001). In most cases, parents can only assist their children in checking the completion of homework as the English knowledge become more complicated in high school.

Moreover, the study noted that parental emotional investment behavior was also below average. It might relate to high school students' psychological characteristics. High school students had an increasing demand for independence from their parents and may not communicate with their parents frequently (Boonk et al., 2022). Additionally, some parents may not be able to dedicate time to accompanying their children in learning due to the challenge of balancing family economic responsibilities and their involvement in their children's education (Hall and Quinn, 2014).

5.2 Relationship between the L2MSS and parental investment behavior

Parental investment and support serve as one of the most important factors for students' English learning motivation (Gonzalez-DeHass et al., 2005; Schunk et al., 2014; Liu, 2024). The L2MSS is a habitus shaped within the sociolinguistic context and influenced by parents' cultural, relationship, knowledge, and emotional capital (Liu, 2024). The study aimed to investigate the potential predictive relationship between different dimensions of parental investment behavior—namely, economic, relationship, knowledge, and emotional investment—and students' L2MSS.

The results found in this study a significant low positive correlation between students' L2MSS and parental investment behavior. This relationship suggests that students who benefit from higher levels of parental investment behavior may exhibit more pronounced L2MSS, and that increased student motivation could prompt greater parental investment in their children's English education. The results were similar to the results reported by Liu (2024) who found that the parental investment behavior had a direct influence on students' L2 learning motivation. The study's findings underscore the critical role of parental support and investment in augmenting students' motivation to learn English (Fan and Williams, 2010; Schunk et al., 2014; Boonk et al., 2022). By acknowledging the influence of parental investment on students' language motivation, this research contributes to the broader understanding of the interplay between family dynamics and educational outcomes.

The results demonstrated no correlation between students' L2MSS and economic investment behavior which was in line with the research findings of Liu (2024). There was a significant low positive correlation between students' L2MSS and the relationship investment behavior, indicating that students whose parents offered more relationship investment might exhibit higher levels of the ideal L2 self, the ought-to L2 self, and the L2 learning experience. Parents' active participation in school activities, such as attending parent meetings, may foster students' academic participation to a certain extent and make students more motivated to engage in learning (Hill and Tyson, 2009). Parents' participation in educational activities was the cornerstone of students' success in school, attracting students' interest in engaging in school activities (He et al., 2015).

This study identified a significant low positive correlation between L2MSS and the knowledge investment behavior. It indicated that students who received sufficient knowledge investment from their parents might develop a more specific image of their ideal L2 self and ought-to L2 self or obtain a greater L2 learning experience. Parents' guidance in English knowledge learning could enhance students' motivation to learn English (Hosseinpour et al., 2015; Alawawda and Razi, 2020). The parental knowledge investment behavior was mainly related to the cultural capital (Bourdieu, 1986) invested by parents in educating their children affected by their education level and background. Parents with higher education levels could be more aware of the vital role of mastering a foreign language in their children's academic and future career development (Liu, 2024). The study identified a significant low positive relationship between students'

L2MSS and the emotional investment behavior, suggesting that with the continuous increase of parental emotional engagement, students might have higher levels of the three dimensions of L2MSS. Parents' emotional encouragement and support helped students to form a positive evaluation of themselves. Even in the face of difficulties, students with enough emotional support could be motivated and confident to devote themselves to learning (Gonzalez-DeHass et al., 2005; He et al., 2015).

Multiple regression analyses indicated that parental investment behavior explained 7.8% of students' L2MSS. It could be explained that students' motivation to learn English was affected by various factors, including not only parents' investment and involvement (Gonzalez-DeHass et al., 2005; Schunk et al., 2014; Liu, 2024), but also students' internal factors (e.g., anxiety, personality, learning strategies or styles) (Papi, 2010; Dörnyei and Ryan, 2015) and other social factors, such as peers and teachers in the school, and other important people in the community (Fan and Williams, 2010; Schunk et al., 2014). The parental economic investment behavior was found to significantly predict students' L2MSS, indicating that the more economic capital parents invest, the lower students' English learning motivation. One possible reason might be that high school students become more independent psychologically and emotionally from parents. They may have heavy schoolwork, so the excessive parental economic investment in extracurricular classes might be seen as controlling and negatively influence students' motivation (Fan and Williams, 2010). Parental emotional investment behavior emerged as a significant positive predictor of students' L2MSS, which indicated that students with a higher level of parental emotional investment behavior had a higher level of L2MSS. Students whose parents affirm their worth and provide emotional support are more likely to attain their ideal self-images and have more confidence to avoid negative outcomes. When parents invest in students' education such as setting an example and providing guidance, they can act as the capital which could help students encounter challenges and reinforce the value of learning (Gonzalez-DeHass et al., 2005).

6 Conclusion and implications

The present study investigated the relationship between high school students' L2MSS and parental investment behavior. The results showed that the level of high school students' L2MSS was slightly higher than the average and the level of parental investment behavior was below average. Students' L2MSS is greatly affected by students' psychological characteristics, parents' expectations, and school requirements. The level of the parental emotional investment behavior was the highest while the parental knowledge investment behavior was the lowest. The parental investment behavior was influenced by factors such as the socioeconomic status of the parents, their educational level, and the amount of time they have available between work and tutoring their children at home. The present study identified a significant low positive correlation between students' L2MSS and the parental investment behavior. The students with a higher level of parental investment behavior are likely more motivated to learn English. The increase in students' English learning motivation may also promote more parental investment.

The present study implies the significance of parental emotional investment behavior such as encouraging students when they make progress, communicating frequently with students about their English learning as well as setting English learning role models. When parents are invested, students reported more motivation and take responsibility in learning English. Moreover, as significant others in the family environment, parents should create a comfortable and caring learning atmosphere and provide emotional support and encouragement for their children, taking into account adolescents' psychological characteristics without focusing on peer comparisons. Parents might be more involved when they know their children are motivated. The findings provide insights for parents in which the four types of parental investment behavior do not play the same role in students' English learning motivation. Parents could compensate for the disadvantage in financial support through adequate emotional investment for their children.

7 Limitations and recommendations

The present study was limited by the following aspects. In regards to research methods, this study was limited to obtaining data through questionnaires. To provide a comprehensive analysis and understanding of how parental investment behavior impacts students' motivation to learn English, it would be beneficial to conduct further research using a variety of data collection methods such as observations, student journals, and interviews with parents and teachers (Rose et al., 2019). Furthermore, due to the availability of data, this study investigated only parental investment behavior and students' L2MSS through a cross-sectional design. Future studies are advised to measure students' motivation and the parental investment behavior over time and track changes in a longitudinal way. Additionally, the differences between various contexts, such as domestic or foreign educational contexts and urban or rural areas, should also be considered in the future research on parental investment.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: further inquiries can be directed to the first author. Requests to access these datasets should be directed to zhangy435@nenu.edu.cn.

Ethics statement

The studies involving humans were approved by Department of Applied Foreign Language Studies, Nanjing University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

YZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing, Funding acquisition. XL:

Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing, Investigation. HL: Conceptualization, Data curation, Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank Prof. Hongchen Xu at Beijing Language and Culture University; Mr. Yunkai Zou in the Institute of Education, Nanjing University and Mr. Xu Wang from School of Foreign Languages, Soochow University, for their support in processing research data. We also appreciated the participants in the current study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Alawawda, M., and Razi, O. (2020). Parental involvement in early second language learning: the role of the immediate environment. *Rev. Cercet. Interv. Soc.* 69, 23–48. doi: 10.33788/rcis.69.2

Al-Hoorie, A. H. (2018). The L2 motivational self system: a meta-analysis. *Stud. Sec. Lang. Learn. Teach.* 8, 721–754. doi: 10.14746/ssllt.2018.8.4.2

Al-Hoorie, A. H., and Szabó, F. (Eds.) (2022). *Researching language learning motivation: A concise guide*. New York: Bloomsbury.

Allatt, P. (1993). "Becoming privileged: the role of family processes" in *Youth and inequality*. eds. I. Bates and G. Iseborough (Buckingham: Open University Press).

Asgari, A., and Mustapha, G. B. (2011). The influence of informal language learning environment (parents and home environment) on the vocabulary learning strategies. *Eng. Lang. Lit. Stud.* 1, 7–13. doi: 10.5539/ells.v1n1p7

Boonk, L., Gijselaers, H. J., Ritzen, H., and Brand-Gruwel, S. (2018). A review of the relationship between parental involvement indicators and academic achievement. *Educ. Res. Rev.* 24, 10–30. doi: 10.1016/j.edurev.2018.02.001

Boonk, L. M., Gijselaers, H. J., Ritzen, H., and Brand-Gruwel, S. (2022). Student-perceived parental involvement as a predictor for academic motivation in vocational education and training (VET). *J. Vocat. Educ.* 74, 187–209. doi: 10.1080/136368 20.2020.1745260

Bourdieu, P. (1986). "The forms of capital" in *Handbook of theory and research for the sociology of education*. ed. J. Richardson (Westport: Greedwood), 241–258.

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., and Gaviria, J. L. (2015). Parental involvement on student academic achievement: a meta-analysis. *Educ. Res. Rev.* 14, 33–46. doi: 10.1016/j.edurev.2015.01.002

Csizér, K. (2019). "The L2 motivational self system" in *The Palgrave handbook of motivation for language learning*. eds. M. Lamb, K. Csizér, A. Henry and S. Ryan (London: Palgrave Macmillan).

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. London: Lawrence Erlbaum.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Dörnyei, Z., and Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.

Epstein, J. L., Sanders, M. G., Simon, K. C., Salinas, K. C., Jansorn, N. R., and Van Voorhis, F. L. (2002). *School, family and community partnerships: Your handbook for action. 2nd* Edn. Corwin: Thousand Oaks.

Fan, X., and Chen, M. (2001). Parental involvement and students' academic achievement: a meta-analysis. *Educ. Psychol. Rev.* 13, 1–22. doi: 10.1023/A:1009048817385

Fan, W. H., and Williams, C. M. (2010). The effects of parental involvement on students' academic self-efficacy, engagement and intrinsic motivation. *Educ. Psychol.* 30, 53–74. doi: 10.1080/01443410903353302

Gao, F. (2014). Social-class identity and English learning: studies of Chinese learners. *J. Lang. Identity Educ.* 13, 92–98. doi: 10.1080/15348458.2014.901820

Gong, Y., Zheng, Y., and Gao, A. X. (2023). Portuguese language learning and capital access: a study on university students in Macau. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras.* 8, 133–149. doi: 10.30827/portalin.viviii.29243

Gonzalez-DeHass, A. R., Willems, P. P., and Holbein, M. F. D. (2005). Examining the relationship between parental involvement and student motivation. *Educ. Psychol. Rev.* 17, 99–123. doi: 10.1007/s10648-005-3949-7

Grolnick, W. S., and Slowiaczek, M. L. (1994). Parents' involvement in children's schooling: a multidimensional conceptualization and motivational model. *Child Dev.* 65, 237–252. doi: 10.2307/1131378

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2019). *Multivariate data analysis. 8th* Edn. Boston, MA: Cengage.

Hall, N., and Quinn, R. (2014). J. Parental involvement at the high school level: parents' perspectives. *J. Ethn. Cult. Stud.* 1, 13–21. doi: 10.29333/ejecs/9

He, T., Gou, W. J., and Chang, S. (2015). Parental involvement and elementary school students' goals, maladaptive behaviors, and achievement in learning English as a foreign language. *Learn. Individ. Differ.* 39, 205–210. doi: 10.1016/j.lindif.2015.03.011

Hill, N. E., and Tyson, D. F. (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Dev. Psychol.* 45, 740–763. doi: 10.1037/a0015362

Hosseinpour, V., Sherkatolabbasi, M., and Yarahmadi, M. (2015). The impact of parents' involvement in and attitude toward their children's foreign language programs for learning English. *Int. J. Appl. Linguist. Eng. Lit.* 4, 175–185. doi: 10.7575/aiac.ijalel.v.4n.4p.175

Huang, S. (2013). The use of literacy bags promotes parental involvement in Chinese children's literacy learning in the English language. *Lang. Teach. Res.* 17, 251–268. doi: 10.1177/1362168813475950

Khany, R., and Amiri, M. (2018). Action control, L2 motivational self system, and motivated learning behavior in a foreign language learning context. *Eur. J. Psychol. Educ.* 33, 337–353. doi: 10.1007/s10212-016-0325-6

Kim, T. Y., and Kim, Y. K. (2014). A structural model for perceptual learning styles, the ideal L2 self, motivated behavior, and English proficiency. *System* 46, 14–27. doi: 10.1016/j.system.2014.07.007

Kong, S. C., and Wang, Y. Q. (2021). The influence of parental support and perceived usefulness on students' learning motivation and flow experience in visual programming: investigation from a parent perspective. *Br. J. Educ. Technol.* 52, 1749–1770. doi: 10.1111/bjet.13071

Lee, J. S., and Lee, K. (2020). Role of L2 motivational self system on willingness to communicate of Korean EFL university and secondary students. *J. Psycholinguist. Res.* 49, 147–161. doi: 10.1007/s10936-019-09675-6

Li, M. (2019). Research on Chinese high-school students' English learning motivation from the self system perspective. *J. Educ. Pract.* 10, 77–87. doi: 10.7176/JEP/10-26-07

Liu, H. (2017). *Empirical research on students' English learning motivation: A social class perspective*. Beijing: Higher Education Press.

Liu, H. (2024). Demystifying the relationship between parental investment in learners' English learning and learners' L2 motivational self system in the Chinese context: a Bourdieusian capital perspective. *Int. J. Educ. Dev.* 104, 102973–102910. doi: 10.1016/j.ijedudev.2023.102973

Liu, H., and Li, X. (2023). Unravelling students' perceived EFL teacher support. *System* 115, 103048–103012. doi: 10.1016/j.system.2023.103048

Magid, M. (2009). The L2 motivational self system from a Chinese perspective: a mixed methods study. *J. App. Ling.* 6, 69–90. doi: 10.1558/japl.v6i1.69

Moskovsky, C., Assulaimani, T., Racheva, S., and Harkins, J. (2016). The L2 motivational self system and L2 achievement: a study of Saudi EFL learners. *Mod. Lang. J.* 100, 641–654. doi: 10.1111/modl.12340

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: a structural equation modeling approach. *System* 38, 467–479. doi: 10.1016/j.system.2010.06.011

Peirce, N. B. (1995). Social identity, investment, and language learning. *TESOL Q.* 29, 9–31. doi: 10.2307/3587803

Pomerantz, E. M., Moorman, E. A., and Litwack, S. D. (2007). The how, whom, and why of parents' involvement in children's academic lives: more is not always better. *Rev. Educ. Res.* 77, 373–410. doi: 10.3102/003465430305567

Rose, H., Mckinley, J., and Baffoe-Djan, J. (2019). *Data collection research methods in applied linguistics*. London: Bloomsbury Academic.

Ryan, S. (2009). "Self and identity in L2 motivation in Japan: the ideal L2 self and Japanese learners of English" in *Motivation, language identity and the L2 self*. eds. Z. Dörnyei and E. Ushioda (Clevedon: Multilingual Matters), 120–143.

Schunk, D. H., Meece, J. L., and Pintrich, P. R. (2014). *Motivation in education: Theory, research and applications. 4th* Edn. Boston: Pearson.

Taguchi, T., Magid, M., and Papi, M. (2009). "The L2 motivational self system amongst Chinese, Japanese, and Iranian learners of English: a comparative study" in *Motivation, language identity and the L2 self*. eds. Z. Dörnyei and E. Ushioda (Clevedon: Multilingual Matters), 66–97.

Thorsen, C., Henry, A., and Cliffordson, C. (2020). The case of a missing person? The current L2 self and the L2 motivational self system. *Int. J. Biling. Educ. Biling.* 23, 584–600. doi: 10.1080/13670050.2017.1388356

Villiger, C., Wandeler, C., and Niggli, A. (2014). Explaining differences in reading motivation between immigrant and native students: the role of parental involvement. *Int. J. Educ. Res.* 64, 12–25. doi: 10.1016/j.ijer.2013.10.004

Wang, X., and Liu, H. (2024). Exploring the moderation roles of emotions, attitudes, environment, and teachers on the impact of motivation on learning behaviors in students' English learning. *Psychol. Rep.*, 1–27. doi: 10.1177/00332941241231714

Wigfield, A., Essoles, S. U., Schiefele, J., Roeser, R. W., and Davis-Kean, P. (2006). "Development of achievement motivation" in *Handbook of child psychology*. eds. W. Damon, R. M. Lerner and N. Eisenberg (Hoboken: Wiley), 933–1002.

You, C., and Dörnyei, Z. (2016). Language learning motivation in China: results of a large-scale stratified survey. *Appl. Linguis.* 37, 495–519. doi: 10.1093/applin/amu046

Yousefi, M., and Mahmoodi, M. H. (2022). The L2 motivational self-system: a meta-analysis approach. *Int. J. Appl. Linguist.* 32, 274–294. doi: 10.1111/ijal.12416

Zhang, Y., and Liu, H. (2022). Exploring Chinese senior high school students' L2 motivational self system. *Eur. J. Engl. Stud.* 2, 89–96. doi: 10.12973/ejels.2.2.89

Zhu, X., Chan, S. D., and Yao, Y. (2022). The associations of parental support with first-grade primary school L2 Chinese learners' ideal selves, motivation, engagement, and reading test performance in Hong Kong: a person-centered approach. *Early Educ. Dev.* 34, 1647–1664. doi: 10.1080/10409289.2022.2139544

frontiers | Frontiers in Psychology

TYPE Mini Review
PUBLISHED 29 August 2024
DOI 10.3389/fpsyg.2024.1434506

OPEN ACCESS

EDITED BY
Thomas L. Spalding,
University of Alberta, Canada

REVIEWED BY
Laura Giglio,
University of South Carolina, United States

*CORRESPONDENCE
Chaimae Harrag
✉ chaimaeharrag@gmail.com;
✉ charrag@nd.edu

RECEIVED 17 May 2024
ACCEPTED 16 August 2024
PUBLISHED 29 August 2024

CITATION
Harrag C, Sabil A, Conceição MC and Radvansky GA (2024) Propositional density: cognitive impairment and aging. *Front. Psychol.* 15:1434506. doi: 10.3389/fpsyg.2024.1434506

Propositional density: cognitive impairment and aging

Chaimae Harrag[1,2,3]*, Abdelkader Sabil[1], Manuel Célio Conceição[2] and Gabriel A. Radvansky[3]

[1]Arts, Languages, and Literature, Faculty of Languages, Arts, and Human Sciences, Hassan 1 University, Settat, Morocco, [2]Language Sciences, Faculty of Human Sciences, University of Algarve, Faro, Portugal, [3]Department of Psychology, University of Notre Dame, Notre Dame, IN, United States

It is important to understand the relationship between cognitive abilities and language processing. Here, we explore a burgeoning area of research that harnesses semantic indices to predict cognitive impairment and track cognitive decline. One such index, *propositional density*, quantifies the information conveyed per language segment. Despite some variation stemming from methodological, sampling, and measurement differences, we suggest that propositional density has diagnostic and assessment value. This paper surveys existing studies that have used propositional density in the context of cognitive aging and impairment and offers some insights into the use of this index to highlight differences in cognition. We also suggest further explorations of basic research involving this concept, and some applications for assessing cognitive health.

KEYWORDS

propositional density, cognitive aging, cognitive impairment, dementia, aphasia

1 Introduction

Exploring and understanding the interplay between language processing and cognitive abilities is a longstanding pursuit. Here we discuss how research has yielding insights into how cognitive processes shape, and are shaped by, language, with an emphasis on how the theoretical concept of propositional density can advance some research areas. We place a special emphasis on work on cognitive aging and cognitive impairment. Cognitive aging has been characterized by a decline in some mental abilities that are closely tied to language production (Craik and Salthouse, 2011; Mortensen et al., 2006) and comprehension (DeDe and Flax, 2016). Moreover, cognitive impairment, ranging from mild cognitive deficits to more severe neurodegenerative disorders, has an influence on language production (Bayles et al., 1993; Kempler and Van Lancker, 2002) and comprehension (Maseda et al., 2014). For instance, there is a decline in syntactic complexity of language produced by people with some cognitive impairment (Sand Aronsson et al., 2021). Also, impairment in some language processes is found in patients with Alzheimer Disease (Fraser et al., 2016). Our aim here is to overview work on propositional density as an index of linguistic complexity to show that it can then be used as an effective tool in research.

As detailed in the next section, *propositional density* quantifies the number of propositions (simple idea units) relative to words present (Brown et al., 2008). It has emerged as a diagnostic tool for studying cognitive aging (Kemper et al., 2001; Kemper and Sumner, 2001; Snowdon et al., 1996) and impairment (Snowdon et al., 1996, 2000) but has not gained the wide-spread use, even though it has the potential for assessing cognitive processes using both language production and comprehension. For language production, language samples can be assessed

for their complexity and content. Changes in the density of these production samples can provide insights into current cognitive health, serve as early indicators of impairment, and help track any declines over time. This is based on the idea that cognitive declines would be revealed in simpler, less dense output. In contrast, for language comprehension people can be given language samples (e.g., sentences, paragraphs, or stories) that vary in their propositional density. Differences in understanding and memory for those samples can provide an index of cognitive performance (Stine and Hindman, 1994). While there is potential for using propositional density as an index of cognitive processing, there has been some variation in its application (Spencer et al., 2012). Here, we suggest that the major source of this variation centers on the use of different methods and populations, along with differences in scoring methods (Chand et al., 2012).

In this paper, we first provide a definition of propositional density. Then we overview some major findings and variations in the literature within the context of how propositional knowledge is represented and processed in memory in the domains of aging and cognitive impairment. Finally, we suggest some future directions of exploration, some of which inform our own research.

2 Defining propositional density

In philosophy and psychology, propositions are defined as fundamental units of meaning. They served as the basis for early models of human memory (e.g., Anderson, 1983; Anderson and Bower, 1974). In language, propositions are the smallest components assessable for their truth value, consisting of a predicate and an argument, each with unique semantic roles (Kintsch, 1998; Van Dijk and Kintsch, 1983). For example, in "John loves Mary," the predicate is "loves" and the arguments are "John" and "Mary." Sentences may contain multiple propositions, such as "John loves Mary deeply," which includes "John loves Mary" and "John's love is deep."

Propositional density refers to the number of idea units within a linguistic segment, adjusted for the number of words. For example, "The old gray mare has a very large nose" contains five propositions: (1) the mare has a nose, (2) the mare is old, (3) the mare is gray, (4) the nose is large, and (5) the nose is very large. Propositional density is calculated as:

Propositional Density = Number of Propositions/Total Number of Words

For our example sentence, this would be:

Propositional Density = 5 propositions/9 total words = 0.56

It is important to note that propositional density is not simply a reflection of the number of words present. For instance, "The farmer in the field is growing wheat diligently, and he is wearing a blue hat" has a propositional density score of 0.375 (i.e., 6 propositions/16 words), while "The weary farmer, battling drought, planting wheat, seeking help, praying for rain, is wearing his hat" has a score of 0.563 (i.e., 9 propositions/16 words) despite both having the same word count.

It has been observed that propositional density scores are associated with differences in comprehension and memory. Increasing the number of propositions while holding word count constant leads to longer reading times (Kintsch and Keenan, 1973; Kintsch and Monk, 1972) and lower recall rates (Kintsch and Keenan, 1973). To wit, people with milder deficits would be expected to have difficulty with high density, but not low-density language. However, people with more severe deficits would be expected to have greater difficulty overall, and across a wider range of densities. Additionally, multiple propositions referring to a common concept increases retrieval time, as with the fan effect (Anderson, 1974). However, that said, memory is not solely guided by propositions. If people hear multiple sentences of overlapping content that refers to a common event, then they will use the degree to which memory probes match the integrated memory of all the propositions, rather than memory for the individual items that were heard (Bransford and Franks, 1971).

In the research discussed here, propositional density is used in two ways. First, it may be used to quantify the language produced by people. Specifically, how propositionally dense is their output? This is propositional density as a dependent variable. Second, it may be used to create different sets of materials, as was done by Kintsch, that vary in their densities. Specifically, this is done to assess language comprehension and memory. This is propositional density as an independent variable.

3 Cognitive aging

Cognitive aging brings about changes in mental function, such as slower processing speeds, diminished working memory capacity, and declining inhibitory function (Connelly et al., 1991; Park et al., 2002; Salthouse, 1990, 1996). These changes affect the ability to process and retain information, including some impact on language production and comprehension. For example, Kemper and Sumner (2001) found that older adults' oral language samples were less syntactically complex, less propositionally dense, and shorter compared to younger adults. These changes were associated with processing efficiency. Less efficient processors tend to have lower propositional density scores and longer utterances. Moreover, lower working memory span scores were associated with decreased lexical variety and syntactic complexity. Here we consider age-related changes in cognition, and how these are related to changes in propositional density during language production, and how propositional density can influence language comprehension.

3.1 Language production

In terms of language production, as noted above, older adults are more likely to produce less propositionally dense utterances. Some research has suggested that changes in output density are a result of changes to cognitive processes operating prior to language formulation, *per se* (Madden et al., 2019). For language production models, this phase is called conceptual preparation (Levelt, 1993, 2000). During this time people select and assemble ideas that are the basis of language production. This requires executive control, involving decisions about the structure and content of the forthcoming message (Harley, 2016; Levelt, 1993, 2000).

Research indicates that age-related cognitive declines can impact conceptual preparation, and subsequently this is reflected in reduced verbal output and idea density (Kemper and Sumner, 2001; Mortensen et al., 2006; Soares et al., 2014). This conceptual difficulty is also revealed in a higher dysfluency rate for older adults when they are

asked to talk about less familiar topics (Bortfeld et al., 2001; Mortensen et al., 2006). Because older adults have difficulty assembling the basic ideas that will underlie an utterance, they are more likely to have dysfluencies (um, er, uh, etc.) to give conceptual preparation more time to assemble the ideas that they are trying to convey.

More generally, word fluency reflects the ease with which people can generate multiple ideas and concepts. Such difficulties in idea generation then manifest themselves as declines in older adults' speech (Clark et al., 2009). An exploratory analysis by Barker et al. (2022) suggested that executive functioning may be responsible for the inability to conceptually select between competing ideas and concepts. This then results in diminished propositional density. More specifically, underlying executive function deficits, such as decreased inhibitory abilities, may allow irrelevant concepts and associations to intrude, rendering the propositional language smaller and less coherent than that found with younger adults (Barker et al., 2020; Barker et al., 2022; Hoffman et al., 2018; Pushkar et al., 2000).

While changes in propositional density could reflect changes in idea selection (Barker et al., 2022; Frederiksen et al., 1990; Levelt, 1993, 2000), other work has suggested that such changes are associated with diminished abstract reasoning and executive function (Barker et al., 2022; Hoffman, 2018). Idea generation relies on executive resources (Alexander, 2006) to select relevant information so that any language output aligns with communicative intentions (Alexander, 2006). Thus, lower propositional density would reflect executive decline, such as is seen in older adults (Arbuckle et al., 2000). Moreover, it has been reported that declines in planning and producing language varying together result in a dual task cost (Kemper et al., 2001, 2009, 2011).

Overall, there is an age-related change in language production, with older adults producing less dense utterances. This may reflect challenges with underlying cognitive processes associated with the creation and assembly of the ideas underlying the language output. This is crucial for real-world tasks like eyewitness reports, where older adults may need adjustments to express memories appropriately, such as speaking slower, and taking more time to put their thoughts together.

3.2 Language comprehension

Age-related changes in language processing are due, in part, to changes in working memory (Gilchrist et al., 2008; Salis, 2011; Wingfield and Stine-Morrow, 2000). Specifically, older adults typically do not maintain as much information as younger adults do, and so are more likely to have difficulty processing more complex language and, thus, are more likely to show a deficit. Although some problems may arise from more syntactically complicated language structures (Norman et al., 1991), another significant factor could be how these age-related changes could give rise to difficulties in language comprehension with increased propositional density.

As noted earlier, research with younger adults has shown that increases in propositional density leads to longer reading times and decreased memory. The changes that accompany the natural aging process leading to greater challenges for older adults for denser texts. Essentially, declines in working memory capacity make it harder to manage more propositions in a text. This was found in studies by Fraser et al. (2016), Kemper et al. (2001), and Mitzner and Kemper (2003). As a result of older adults' increased sensitivity to propositional density, they often need more time and resources to process denser sentences.

The impact of propositional density on comprehension for older adults was also seen in a study by Stine and Hindman (1994). This study compared reading and memory for younger and older adults for sentences with varying densities. Relative to younger adults, older adults spent more time reading denser sentences, and that this difference was correlated with working memory span scores. This was also associated with reduced memory. Stine and Hindman suggested that the slower reading time reflects a compensatory strategy with older adults taking more time to process the information because they are dissecting denser sentences into smaller, more manageable units. These smaller units could then be more easily handled with reduced working memory resources, allowing cognitive processes to be completed in a timely manner.

Overall, cognitive changes in older adults bring about changes in information processing. Materials that take into account these changes can enhance language comprehension and processing, supporting effective communication and lifelong learning. In essence, older adults may show a larger benefit to comprehension if presented with less dense materials.

4 Cognitive impairment

In this section, we consider two types of cognitive impairment, and how propositional density can be used as an index of the degree of disruption. These are the cognitive impairments that come with forms of dementia, such as Alzheimer's Disease, and language specific deficits, such as aphasia.

4.1 Dementia

With dementia, subtle linguistic changes may signal early cognitive dysfunction, because language processing involves various cognitive mechanisms that are compromised (Caplan, 1993). Language production issues would reflect the downstream problems that result from such deficits. As an illustration of the usefulness of propositional density for assessing such cognitive changes, a study by Medina et al. (2011) examined the relationship among familial Alzheimer's disease (FAD) mutation status, apolipoprotein E (APOE) genotype, and propositional density for non-demented people at risk for FAD. People provided biographical essays, which were analyzed for propositional density. Their results revealed no significant association between FAD mutation status and propositional density. However, the presence of the APOE E4 allele was strongly correlated with lower density.

Also, Engelman et al. (2010) found that cognitively intact people produce propositionally denser output than Alzheimer's patients, likely due to their intact cognitive resources (Stine and Hindman, 1994; Stine-Morrow et al., 2006). In contrast, impaired people have limitations that hinder the production of denser output (Snowdon et al., 1996, 2000). More specifically, the propositional density of the output of language production tasks can be used to predict Alzheimer's disease. As an example, Snowdon et al. (1996) analyzed autobiographies written by women around the age of 22 and compared them to subsequent outcomes. They found that lower density was

associated with lower cognitive test scores and higher occurrence of Alzheimer's disease later in life. Engelman et al. (2010) did a similar analysis using medical school admissions essays. Again, finding that lower density in the earlier writings of people who later developed Alzheimer's disease. Thus, early life language production can be used as a predictor of later cognitive health.

A longitudinal analysis of language samples from healthy and adults with dementia showed a progressive decline in grammatical complexity and propositional density with age (Kemper et al., 2001). Alzheimer disease was accompanied by accelerated deterioration. Moreover, grammatical complexity decline was linked to digit span, while propositional density decline was associated with vocabulary differences. More recently, Mueller et al. (2016) assessed whether people with Mild Cognitive Impairment (pMCI) and memory decline exhibit deficits in connected language measures. The people described a picture, and these productions were analyzed for semantic content (total semantic units, propositional density, and unique words), syntactic complexity, and speech fluency. The pMCI group had fewer unique words and semantic units than the controls, and importantly, differed in propositional density. No differences were found in speech fluency tasks or syntactic complexity. Thus, measures of propositional density can capture cognitive changes that might be missed with other measures.

Propositional density can be used to detect declines in semantic memory, as is found with Alzheimer's disease (Kirshner, 2012; Mascali et al., 2018; Zahn et al., 2004). During the early stages of the disease, people exhibit declines in semantic processing, as with lexical errors, delays in word finding, semantic paraphrases, and verbose language (Forbes-McKay et al., 2013), all of which reduce the density of their language productions. Venneri et al. (2016, 2018) suggested that tests of semantic processing, such as propositional density, can detect changes early on.

Supporting this, a study by Farias et al. (2012) examined whether density from oral language samples, obtained from cognitively intact, impaired, and demented groups, could predict subsequent trajectories of cognitive change. They found that density scores were more closely related to changes in overall cognitive function in the MCI and healthy groups than in the dementia group. This highlights the potential of density as a way to predict cognitive decline, particularly early on. Additionally, the study demonstrated that density was correlated with semantic memory, executive function, and spatial abilities, and minimally correlated with episodic memory.

It is important to note that all of the studies discussed in this section use measures of the density of language productions. To our knowledge, there are no studies of language comprehension and memory that use propositional density as a way of manipulating the materials presented and exploring language processing in this way. This is an open avenue for future research.

4.2 Aphasia

Propositional density has been used with nonfluent (Broca's) aphasia patients to assess the informativeness and communicative adequacy of their language production (Barker et al., 2020; Bryant et al., 2013). In one study, Bryant et al. (2013) investigated the extent to which propositional density scores differed between aphasic and non-aphasic discourse, and whether it could adequately index the severity of the aphasia. The study included people from the Goals in Aphasia Project with post-stroke aphasia following a cerebrovascular accident in their language-dominant hemisphere, and their family members, who served as controls. The language production samples were analyzed for propositional density, lexical diversity, complexity (measured as the mean length of an utterance and number of utterances), and overall verbal productivity.

The results revealed that propositional density scores differed in the two groups. There was a negative correlation between these scores and the severity of aphasia, indicating that more severe aphasia compromised language production to a degree that results in less dense and information-impoverished discourse. These results were validated by correlations with other language measures including Number of Different Words (NDW), Mean Length of Utterance (MLU), and Number of Utterances (NU). An unexpected increase in the Type-Token Ratio (TTR) was observed in cases of aphasia, possibly due to the large sample sizes of aphasic language data in the present research.

Similarly, a study by Fromm et al. (2016) evaluated how proposition density can differentiate between people with aphasia and controls, as well as among subtypes of aphasia, based on procedural discourse and personal narratives. There were six aphasia types assessed: Broca's, Wernicke's, anomic, conduction, transcortical motor, and people that had an Aphasia Quotient greater than 93.8. The controls scored higher than people with aphasia on both tasks. Additionally, density scores differed among the aphasia types. Density scores for the Broca group were lower than those for all the others. Moreover, everyone with aphasia scored lower on discourse tasks than on the narrative tasks. This shows that propositional density can be used to distinguish between different types of language deficits and task types.

Some studies on aphasia indicate that even with reduced propositional output and impaired executive functioning, skills such as comprehension, repetition, reading, and naming often remain intact (Crescentini et al., 2008; Robinson et al., 2006; Robinson, 2013). Webster et al. (2018) found that propositional density did not impact reading time or accuracy in people with aphasia, indicating intact comprehension despite production deficits. Fromm et al. (2017) also observed no differences in comprehension between healthy people and people with aphasia. The deficits studied so far appear to be confined to language production. The intact skills along with reduced propositional language during discourse generation may not be strictly language-based but could also be attributed to an inability to select novel thoughts (Barker et al., 2022; Robinson et al., 1998, 2010). Difficulties in the sequencing and selecting of thoughts may hinder the fluid connection of ideas during language production. This may be why patients often do well on word and sentence-level generation tasks, which require generating only a single idea and focusing attention on the current message. Another possible explanation is that people with aphasia during comprehension use context and redundancy in connected speech, which enables them to infer meaning and compensate for their linguistic impairments (Huber, 1990).

5 Variation in propositional density

Propositional density has been useful for assessing linguistic ability and cognitive health, but there are challenges. These include

variability in its reliability as a predictor of cognitive decline, effectiveness across modalities, study context, sample differences, and calculation methods. Each of these are considered in turn.

5.1 Predictability

While there have been studies showing that propositional density scores for language production can be used to help predict cognitive functioning in later life, there are also some inconsistencies. The predictability of propositional density scores from language output for later cognitive performance has relied heavily on rich production corpora. Ferguson et al. (2014) and Spencer et al. (2012) have both noted that longer writing samples tend to yield more consistent density measures, emphasizing the need for larger datasets to enhance reliability. Thus, it is expected that cases in which the language production output are smaller is likely to lead to less stable propositional density measures. Any predictions using such scores are likely to be less reliable.

5.2 Modalities

While there have been studies showing that propositional density scores for language production can be used as an index of cognitive functioning, there are also some inconsistencies depending on the modality of production. For example, Smolík et al. (2016) reported a decrease in density for amnestic mild cognitive impairment (aMCI) patients in spoken but not written language. This aligns with the Nun Study (Mitzner and Kemper, 2003), which found higher propositional density in written compared to oral samples. Written narratives tend to have higher density than spoken narratives. This difference may reflect the fact that the writing process serves to offload cognitive processes to some degree, which opens the door for denser language units.

5.3 Study context

As noted earlier, propositional density has been used in the context of both language production and language comprehension and memory studies. Each of these has its strengths and weaknesses. In terms of language production, it has clearly been shown that samples of early language production can, to some degree, predict cognitive deficits later on. This has not been demonstrated with language comprehension. That said, propositional density scores derived from language production require a large output sample to produce scores that are predictable. Moreover, in some cases, people may be less willing, or less able, to produce a great deal of linguistic output when asked, because, by its very nature, is cognitively demanding.

In comparison, for language comprehension tasks, researchers can use materials of various densities to assess performance. This can be done in a much shorter period and involves explicit experimental manipulation. This sort of assessment can involve both reading time measures, as well as memory measures. Such multimethod approaches are always preferred to single methods, such as just using memory. This also would allow for a better comparison between groups because the nature of the materials (the language input) would be the same in both cases. Any differences would be due to cognitive processes. That said, as noted earlier, comprehension measures are less likely to be useful in predicting later performance given the lack of background research on the topic, and the lack of test administration early in life. Still, overall, when faced with a choice, we would recommend going into any new evaluation to assess cognitive function using comprehension in situations where there is more task administration control.

5.4 Sample differences

Studies on conditions like MCI and aphasia report differences in propositional density due to the different cognitive profiles. For example, Mitzner and Kemper (2003) and Fraser et al. (2016) noted that Alzheimer's disease impacts different brain regions, causing heterogeneity in language samples. Distinct subgroups, such as amnestic and dysexecutive deficit groups, may demonstrate varying effects on density.

In the context of healthy aging, Véliz et al. (2013) found that there are no differences in syntactic complexity and propositional density between healthy younger and older adults in sentence production. Similarly, Ferguson et al. (2014) noted that propositional density during production remained consistent in healthy women from young to mid-adulthood but began to decline in older adulthood. That said it should be noted that these studies involved small sample sizes ($N = 20$). Thus, it is important to consider the nature of the samples that are being worked with to best understand how propositional density scores reflect what sorts of differences in cognitive processing that are known to be present.

5.5 Genre influences

Propositional density scores can also vary with different linguistic genres (Alyahya et al., 2020; Fromm et al., 2016). For example, storytelling narratives often have more content words and lexical diversity compared to composite picture description and expository discourse. This may reflect the greater ease that narrative language is processed relative to other kinds. This ease of processing may free up resources, making it easier to integrate more basic idea units into a given segment of language.

5.6 Calculation methods

Several methods have been used to calculate propositional density. These include the Language Across the Lifespan (LAL) coding manual, the Computerized Propositional Idea Density Rater (CPIDR) tool, and Analysis of Idea Density (AID). Each of these are considered in turn.

LAL (Kemper, 1993) is an adaptation of Turner and Greene's (1977) original manual defining propositional density. For this approach, propositions are grouped into three classes: predication (often verbs), modification (adjectives, adverbs), and connection (conjunctions, prepositions). The total number of propositions is divided by the total number of words.

The CPIDR tool (Brown et al., 2008) automates propositional density calculation by analyzing semantic content using parts of speech tagging, counting the derived propositions, and dividing them by the number of words. This is then followed by a refinement using post-analysis rules. These rules include flagging conjunctions, numerals, determiners, prepositions, adjectives, adverbs, possessives, verbs, relatives, or interrogatives as propositions. Additional rules condense complex verb phrases into single propositions; for instance, "may have been signing" would be condensed to a single proposition. Subject-auxiliary inversion is also used to correctly process questions. For example, "Has he come?" is converted to "he has come."

The AID has been used for oral language samples Chand et al. (2012). It is based on Kintsch's narrative analysis. This approach uses Turner and Greene's (1977) classification (predication, modification, and connectives) but diverges by emphasizing the extraction of meaningful content, differentiating between semantic significance and grammatical structure. In this approach, an idea is only counted if it contributes to the overall meaning. This removes verbal asides that may occur during oral output (e.g., "Oh, hi there!").

Thus, these guidelines also help distinguish between words and phrases that add new information versus those that fulfill grammatical necessities. For example, "And then, surprisingly, she left" counts "surprisingly" as it introduces a new idea, while "and then" is a narrative progression tool. Similarly, the repetition of "very" in "It was very, very cold" is evaluated based on whether it introduces new information or merely emphasizes the existing description. The manual also provides detailed guidelines on what constitutes a word, including lexical fillers, unfinished words, repeated words, utterance-initial conjunctions, and acronyms.

Overall, the LAL-based measure, CPIDR, and AID each have benefits and limitations. Incorporating propositional density into future research requires understanding the strengths and limitations of available tools. LAL provides a comprehensive analysis but is complex and less accessible without substantial linguistic training. CPIDR 3 segments text into propositional units using speech tags, offering an automated approach that simplifies the process but may overlook semantic richness and syntactic complexity. AID prioritizes semantic content, introducing rules to improve inter-rater reliability. However, it is labor-intensive and may allow for some element of subjectivity. This method relies on what the raters deem meaningful.

To illustrate differences among these measures, take the repetition of word very in "very, very cold." In approaches such as LAL and AID, this may be counted as meaningful because it is used for emphasis, but another rater may discard it given it does not introduce new information. Thus, it is necessary to define what constitutes a "word" for accurate calculations. CPIDR 3 counts "very, very exhausting" as three words, while AID counts one "very." Thus, clearer guidelines, such as standardizing the treatment of contractions, possessives, multi-word numbers, acronyms, and repeated words can ensure consistency in measurement, would be helpful.

6 Future research

Going beyond measurement issues, further work can be done looking at how this measure relates to various levels of language comprehension. There are three levels of representation that can be identified, namely the surface form, the textbase, and the mental model levels (Van Dijk and Kintsch, 1983). The surface form captures the verbatim wording of language, including the specific words and syntax that were used. The textbase captures the propositional idea units that are present in language, apart from the specific wording. For example, the sentences "The boy helped the girl" and "The girl was helped by the boy" have different wordings and syntax, but map onto the same underlying idea. Finally, the mental model level refers to what the language is about (Glenberg et al., 1987). It includes both information from the language itself, as well as inferences drawn from a person's long-term knowledge. For example, if you read that the firecracker exploded, you are likely to infer that someone lit the fuse, even if it was not explicitly stated.

Propositional density primarily assesses the textbase level of language comprehension and memory, although there is some involvement of surface form knowledge as well, especially in the calculation of a propositional density score (i.e., the number of words used). However, it is unclear how this measure relates to the mental model level. This is important because research indicates older adults may perform similarly to younger adults at the mental model level despite differences at the textbase level (Radvansky and Dijkstra, 2007). Do propositional density differences also manifest themselves as higher level comprehension and memory differences as well?

Another issue is how propositional density affects memory over time, such as forgetting rates for denser language. Research has shown that different types of information follow different patterns of retention and forgetting (e.g., Fisher and Radvansky, 2018). For example, surface form information is forgotten quite quickly, within a few seconds or minutes (Sachs, 1967), whereas the mental model level may endure for days, weeks, or even decades (e.g., Doolen and Radvansky, 2021). It may be that language at different levels of propositional density may exhibit different patterns of forgetting. For example, it may be that denser texts result in longer lasting memory because there is more supportive informational content. Alternatively, it may be that denser texts actually result in shorter lasting memory because there is more information to encode, which taxes cognitive resources, resulting in poorer initial encoding.

Future studies should explore this and consider practical issues like eyewitness report accuracy, memory veracity, and confidence. Is propositionally denser output linked to greater accuracy, truth-telling, and confidence? This could be if a strong, more detailed memory trace is present, then there is more underlying information to draw upon. Therefore, this would allow language production to include more idea units within a given utterance.

Expanding propositional density as a tool to other cognitive domains offers promising avenues for research, particularly in understanding autobiographical memory and its relation to memory report generalization. For instance, changes in language production density could indicate the presence of conditions such as depression, which is often associated with over general memories that lack detail and specificity (Williams et al., 2007).

There is also a lack of studies in clinical settings of how materials of different propositional density affect comprehension in people with dementia, aging adults, and those with cognitive impairments. Most studies of these populations focus on language production. Understanding this could reveal valuable insights into the cognitive processes underlying language use, aiding in the development of better communication strategies and interventions. Extending this

tool beyond its traditional applications could therefore provide additional insights into various conditions.

Author contributions

CH: Conceptualization, Writing – original draft, Writing – review & editing. AS: Supervision, Validation, Writing – review & editing. MC: Supervision, Validation, Writing – review & editing. GR: Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alexander, M. P. (2006). Impairments of procedures for implementing complex language are due to disruption of frontal attention processes. *J. Int. Neuropsychol. Soc.* 12, 236–247. doi: 10.1017/S1355617706060309

Alyahya, R. S. W., Halai, A. D., Conroy, P., and Lambon, M. A. (2020). A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain* 143, 1541–1554. doi: 10.1093/BRAIN/AWAA074

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cogn. Psychol.* 6, 451–474. doi: 10.1016/0010-0285(74)90021-8

Anderson, J. R. (1983). The architecture of cognition. Psychology Press: New York.

Anderson, J. R., and Bower, G. H. (1974). A propositional theory of recognition memory. *Mem. Cogn.* 2, 406–412. doi: 10.3758/BF03196896

Arbuckle, T. Y., Nohara-LeClair, M., and Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychol. Aging* 15, 65–77. doi: 10.1037/0882-7974.15.1.65

Barker, M. S., Ceslis, A., and Robinson, G. A. (2022). Idea selection for propositional language production. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 29, 260–283. doi: 10.1080/13825585.2020.1862044

Barker, M. S., Nelson, N. L., and Robinson, G. A. (2020). Idea formulation for spoken language production: the Interface of cognition and language. *J. Int. Neuropsychol. Soc.* 26, 226–240. doi: 10.1017/S1355617719001097

Bayles, K. A., Tomoeda, C. K., and Trosset, M. W. (1993). Alzheimer'S disease: effects on language. *Dev. Neuropsychol.* 9, 131–160. doi: 10.1080/87565649109540549

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Lang. Speech* 44, 123–147. doi: 10.1177/00238309010440020101

Bransford, J. D., and Franks, J. J. (1971). The abstraction of linguistic ideas. *Cogn. Psychol.* 2, 331–350. doi: 10.1016/0010-0285(71)90019-3

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., and Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behav. Res. Methods* 40:540. doi: 10.3758/BRM.40.2.540

Bryant, L., Spencer, E., Ferguson, A., Craig, H., Colyvas, K., and Worrall, L. (2013). Propositional idea density in aphasic discourse. *Aphasiology* 27, 992–1009. doi: 10.1080/02687038.2013.803514

Caplan, D. (1993). Toward a psycholinguistic approach to acquired neurogenic language disorders. *Am. J. Speech Lang. Pathol.* 2, 59–83. doi: 10.1044/1058-0360.0201.59

Chand, V., Baynes, K., Bonnici, L. M., and Farias, S. T. (2012). A rubric for extracting idea density from oral language samples. *Curr. Protoc. Neurosci.* Chapter 10:Unit10.5. doi: 10.1002/0471142301.NS1005S58

Clark, L. J., Gatz, M., Zheng, L., McCleary, C., MacK, W. J., and Chen, Y. L. (2009). Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. *Am. J. Alzheimers Dis. Other Demen.* 24, 461–468. doi: 10.1177/1533317509345154

Connelly, S. L., Hasher, L., and Zacks, R. T. (1991). Age and reading: the impact of distraction. *Psychol. Aging* 6, 533–541. doi: 10.1037/0882-7974.6.4.533

Craik, F. I. M., and Salthouse, T. A. (2011). The Handbook of aging and cognition: third edition. New York: Psychology Press, 1–657.

Crescentini, C., Lunardelli, A., Mussoni, A., Zadini, A., and Shallice, T. (2008). A left basal ganglia case of dynamic aphasia or impairment of extra-language cognitive processes? *Neurocase* 14, 184–203. doi: 10.1080/13554790802108380

DeDe, G., and Flax, J. K. (2016). "Language comprehension in aging" in Cognition, language and aging. ed. H. H. Wright (Amsterdam, The Netherlands: John Benjamins Publishing Company), 107–133.

Doolen, A. C., and Radvansky, G. A. (2021). A novel study: long-lasting event memory. *Memory* 29, 963–982. doi: 10.1080/09658211.2021.1953079

Engelman, M., Agree, E. M., Meoni, L. A., and Klag, M. J. (2010). Propositional density and cognitive function in later life: findings from the precursors study. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 65, 706–711. doi: 10.1093/GERONB/GBQ064

Farias, S. T., Chand, V., Bonnici, L., Baynes, K., Harvey, D., Mungas, D., et al. (2012). Idea density measured in late life predicts subsequent cognitive trajectories: implications for the measurement of cognitive reserve. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 67, 677–686. doi: 10.1093/GERONB/GBR162

Ferguson, A., Spencer, E., Craig, H., and Colyvas, K. (2014). Propositional idea density in women's written language over the lifespan: computerized analysis. *Cortex* 55, 107–121. doi: 10.1016/J.CORTEX.2013.05.012

Fisher, J. S., and Radvansky, G. A. (2018). Patterns of forgetting. *J. Mem. Lang.* 102, 130–141. doi: 10.1016/J.JML.2018.05.008

Forbes-McKay, K., Shanks, M. F., and Venneri, A. (2013). Profiling spontaneous speech decline in Alzheimer's disease: a longitudinal study. *Acta Neuropsychiatrica* 25, 320–327. doi: 10.1017/NEU.2013.16

Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *J. Alzheimers Dis.* 49, 407–422. doi: 10.3233/JAD-150520

Frederiksen, C. H., Bracewell, R. J., Breuleux, A., and Renaud, A. (1990). "The cognitive representation and processing of discourse: function and dysfunction" in Discourse ability and brain damage: theoretical and empirical perspectives. eds. Y. Joanette and H. Brownell (New York, NY: Springer Science & Business Media), 69–110.

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Richardson, J., and MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western aphasia battery cutoff. *Am. J. Speech Lang. Pathol.* 26, 762–768. doi: 10.1044/2016_AJSLP-16-0071

Fromm, D., Greenhouse, J., Hou, K., Russell, G. A., Cai, X., Forbes, M., et al. (2016). Automated proposition density analysis for discourse in aphasia. *J. Speech Lang. Hear. Res.* 59, 1123–1132. doi: 10.1044/2016_JSLHR-L-15-0401

Gilchrist, A. L., Cowan, N., and Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: recall of fewer but not smaller chunks in older adults. *Memory* 16, 773–787. doi: 10.1080/09658210802261124

Glenberg, A. M., Meyer, M., and Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *J. Mem. Lang.* 26, 69–83. doi: 10.1016/0749-596X(87)90063-5

Harley, T. A. (2016). The psychology of language: From data to theory. London, UK: Routledge. Available at: https://www.routledge.com/The-Psychology-of-Language-From-Data-to-Theory/Harley/p/book/9781848720893f

Hoffman, P. (2018). An individual differences approach to semantic cognition: divergent effects of age on representation, retrieval and selection. *Sci. Rep.* 8, 1–13. doi: 10.1038/s41598-018-26569-0

Hoffman, P., Loginova, E., and Russell, A. (2018). Poor coherence in older people's speech is explained by impaired semantic and executive processes. *Elife* 7:e38907. doi: 10.7554/ELIFE.38907

Huber, W. (1990). Text comprehension and production in aphasia: analysis in terms of micro-and macroprocessing. in *Discourse Ability and Brain Damage*. Eds. Joanette, Y., Brownell, H. H. Springer, New York, NY: Springer Series in Neuropsychology. doi: 10.1007/978-1-4612-3262-9_7

Kemper, S. J. (1993). "Language across the life span: coding manual" in Coding manual. ed. K. S. Lawrence. [*This may be an unpublished manuscript]

Kemper, S., Hoffman, L., Schmalzried, R., Herman, R., and Kieweg, D. (2011). Tracking talking: dual task costs of planning and producing speech for young versus older adults. *Aging Neuropsychol. Cognit.* 18, 257–279. doi: 10.1080/13825585.2010.527317

Kemper, S., Schmalzried, R., Herman, R., Leedahl, S., and Mohankumar, D. (2009). The effects of aging and dual task demands on language production. *Aging Neuropsychol. Cognit.* 16, 241–259. doi: 10.1080/13825580802438868

Kemper, S., and Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychol. Aging* 16, 312–322. doi: 10.1037/0882-7974.16.2.312

Kemper, S., Thompson, M., and Marquis, J. (2001). Longitudinal change in language production: effects of aging and dementia on grammatical complexity and propositional content. *Psychol. Aging* 16, 600–614. doi: 10.1037/0882-7974.16.4.600

Kempler, D., and Van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: a case study of Parkinson's disease. *Brain Lang.* 80, 449–464. doi: 10.1006/brln.2001.2602

Kintsch, W. (1998). Comprehension: a paradigm for cognition. Cambridge, UK: Cambridge University Press.

Kintsch, W., and Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cogn. Psychol.* 5, 257–274. doi: 10.1016/0010-0285(73)90036-4

Kintsch, W., and Monk, D. (1972). Storage of complex information in memory: some implications of the speed with which inferences can be made. *J. Exp. Psychol.* 94, 25–32. doi: 10.1037/H0032781

Kirshner, H. S. (2012). Primary progressive aphasia and Alzheimer's disease: brief history, recent evidence. *Curr. Neurol. Neurosci. Rep.* 12, 709–714. doi: 10.1007/S11910-012-0307-2

Levelt, W. J. M. (1993). Speaking: from intention to articulation. Cambridge, MA, USA: MIT Press.

Levelt, W. J. M. (2000). Producing spoken language: a blueprint of the speaker. *The Neurocognition of Language*, Eds. Colin M. Brown and Peter Hagoort, Oxford University Press 82–122.

Madden, D. L., Sale, M. V., and Robinson, G. A. (2019). Age-related differences in idea generation and selection for propositional language. *Neuropsychol. Dev. Cogn. Sec. B Aging Neuropsychol. Cognit.* 26, 486–506. doi: 10.1080/13825585.2018.1476668

Mascali, D., DiNuzzo, M., Serra, L., Mangia, S., Maraviglia, B., Bozzali, M., et al. (2018). Disruption of semantic network in mild Alzheimer's disease revealed by resting-state fMRI. *Neuroscience* 371:38. doi: 10.1016/J.NEUROSCIENCE.2017.11.030

Maseda, A., Lodeiro-Fernández, L., Lorenzo-López, L., Núñez-Naveira, L., Balo, A., and Millán-Calenti, J. C. (2014). Verbal fluency, naming and verbal comprehension: three aspects of language as predictors of cognitive impairment. *Aging Ment. Health* 18, 1037–1045. doi: 10.1080/13607863.2014.908457

Medina, L. D., Rodriguez-Agudelo, Y., Geschwind, D. H., Gilbert, P. E., Liang, L. J., Cummings, J. L., et al. (2011). Propositional density and apolipoprotein E genotype among persons at risk for familial Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.* 32, 188–192. doi: 10.1159/000333023

Mitzner, T. L., and Kemper, S. (2003). Oral and written language in late adulthood: findings from the nun study. *Exp. Aging Res.* 29, 457–474. doi: 10.1080/03610730303698

Mortensen, L., Meyer, A. S., and Humphreys, G. W. (2006). Age-related effects on speech production: a review. *Lang. Cogn. Process* 21, 238–290. doi: 10.1080/01690960444000278

Mueller, K. D., Koscik, R. L., Turkstra, L. S., Riedeman, S. K., Larue, A., Clark, L. R., et al. (2016). Connected language in late middle-aged adults at risk for Alzheimer's disease. *J. Alzheimer's Dis.* 54, 1539–1550. doi: 10.3233/JAD-160252

Norman, S., Kemper, S., Kynette, D., Cheung, H., and Anagnopoulos, C. (1991). Syntactic complexity and adults' running memory span. *J. Gerontol.* 46:P346-51. doi: 10.1093/GERONJ/46.6.P346

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., and Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* 17, 299–320. doi: 10.1037/0882-7974.17.2.299

Pushkar, D., Basevitz, P., Arbuckle, T., Nohara-LeClair, M., Lapidus, S., and Peled, M. (2000). Social behavior and off-target verbosity in elderly people. *Psychol. Aging* 15, 361–374. doi: 10.1037/0882-7974.15.2.361

Radvansky, G. A., and Dijkstra, K. (2007). Aging and situation model processing. *Psychon. Bull. Rev.* 14, 1027–1042. doi: 10.3758/BF03193088

Robinson, G. A. (2013). Primary progressive dynamic aphasia and parkinsonism: generation, selection and sequencing deficits. *Neuropsychologia* 51, 2534–2547. doi: 10.1016/J.NEUROPSYCHOLOGIA.2013.09.038

Robinson, G., Blair, J., and Cipolotti, L. (1998). Dynamic aphasia: an inability to select between competing verbal responses? *Brain* 121, 77–89. doi: 10.1093/BRAIN/121.1.77

Robinson, G., Shallice, T., Bozzali, M., and Cipolotti, L. (2010). Conceptual proposition selection and the LIFG: neuropsychological evidence from a focal frontal group. *Neuropsychologia* 48, 1652–1663. doi: 10.1016/J.NEUROPSYCHOLOGIA.2010.02.010

Robinson, G., Shallice, T., and Cipolotti, L. (2006). Dynamic aphasia in progressive supranuclear palsy: a deficit in generating a fluent sequence of novel thought. *Neuropsychologia* 44, 1344–1360. doi: 10.1016/J.NEUROPSYCHOLOGIA.2006.01.002

Sachs, J. S. (1967). Recopition memory for syntactic and semantic aspects of connected discourse. *Percept. Psychophys.* 2, 437–442. doi: 10.3758/BF03208784

Salis, C. (2011). Understanding of auditory discourse in older adults: the effects of syntax and working memory. *Aphasiology* 25, 529–539. doi: 10.1080/02687038.2010.527998

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Dev. Rev.* 10, 101–124. doi: 10.1016/0273-2297(90)90006-P

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103, 403–428. doi: 10.1037/0033-295X.103.3.403

Sand Aronsson, F., Kuhlmann, M., Jelic, V., and Östberg, P. (2021). Is cognitive impairment associated with reduced syntactic complexity in writing? Evidence from automated text analysis. *Aphasiology* 35, 900–913. doi: 10.1080/02687038.2020.1742282

Smolík, F., Stepankova, H., Vyhnálek, M., Nikolai, T., Horáková, K., and Matějka, S. (2016). Propositional density in spoken and written language of Czech-speaking patients with mild cognitive impairment. *J. Speech Lang. Hear. Res.* 59, 1461–1470. doi: 10.1044/2016_JSLHR-L-15-0301

Snowdon, D. A., Greiner, L. H., and Markesbery, W. R. (2000). Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease: findings from the nun study. *Ann. N. Y. Acad. Sci.* 903, 34–38. doi: 10.1111/J.1749-6632.2000.TB06347.X

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the nun study. *JAMA* 275, 528–532. doi: 10.1001/JAMA.275.7.528

Soares, F. C., de Oliveira, T. C. G., de Macedo, L. D. E. D., Tomás, A. M., Picanço-Diniz, D. L. W., Bento-Torres, J., et al. (2014). CANTAB object recognition and language tests to detect aging cognitive decline: an exploratory comparative study. *Clin. Interv. Aging* 10, 37–48. doi: 10.2147/CIA.S68186

Spencer, E., Craig, H., Ferguson, A., and Colyvas, K. (2012). Language and ageing - exploring propositional density in written language - stability over time. *Clin. Linguist. Phonet.* 26, 743–754. doi: 10.3109/02699206.2012.673046

Stine, E. A. L., and Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging Neuropsychol. Cognit.* 1, 2–16. doi: 10.1080/09289919408251446

Stine-Morrow, E. A. L., Miller, L. M. S., and Hertzog, C. (2006). Aging and self-regulated language processing. *Psychol. Bull.* 132, 582–606. doi: 10.1037/0033-2909.132.4.582

Turner, A., and Greene, E. (1977). The construction and use of a propositional text base. Technical Report No. 63. Boulder: Institute for the Study of Intellectual Behavior, University of Colorado.

Van Dijk, T. A., and Kintsch, W. (1983). Strategies of discourse comprehension. Available at: https://www.academia.edu/download/77815914/Teun_20A_20van_20Dijk_20__20Walter_20Kintsch_20-_20Strategies_20of_20Discourse_20Comprehension.pdf

Véliz, M., Hernández, M., Sáez, K., Riffo, B., and Sáez, Y. (2013). Sentence production by elder and young adults under a controlled situation[Oraciones producidas por adultos mayores y adultos jóvenes en una situación controlada]. *Onomazein* 27, 241–257. doi: 10.7764/onomazein.27.16

Venneri, A., Jahn-Carta, C., De Marco, M., Quaranta, D., and Marra, C. (2018). Diagnostic and prognostic role of semantic processing in preclinical Alzheimer's disease. *Biomark. Med* 12, 637–651. doi: 10.2217/BMM-2017-0324

Venneri, A., Mitolo, M., and De Marco, M. (2016). Paradigm shift: semantic memory decline as a biomarker of preclinical Alzheimer's disease. *Biomark. Med* 10, 5–8. doi: 10.2217/BMM.15.53

Webster, J., Morris, J., Howard, D., and Garraffa, M. (2018). Reading for meaning: what influences paragraph understanding in aphasia? *Am. J. Speech Lang. Pathol.* 27, 423–437. doi: 10.1044/2017_AJSLP-16-0213

Williams, J. M. G., Barnhofer, T., Crane, C., Hermans, D., Raes, F., Watkins, E., et al. (2007). Autobiographical memory specificity and emotional disorder. *Psychol. Bull.* 133, 122–148. doi: 10.1037/0033-2909.133.1.122

Wingfield, A., and Stine-Morrow, E. A. L. (2000). "Language and speech" in The handbook of aging and cognition. eds. F. I. M. Craik and T. A. Salthouse. *2nd* ed (Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers), 359–416.

Zahn, R., Juengling, F., Bubrowski, P., Jost, E., Dykierek, P., Talazko, J., et al. (2004). Hemispheric asymmetries of hypometabolism associated with semantic memory impairment in Alzheimer's disease: a study using positron emission tomography with fluorodeoxyglucose-F18. *Psychiatry Res.* 132, 159–172. doi: 10.1016/J.PSCYCHRESNS.2004.07.006

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 26 September 2024
DOI 10.3389/fpsyg.2024.1436710

OPEN ACCESS

EDITED BY
Antonio Benítez-Burraco,
University of Seville, Spain

REVIEWED BY
Lynne Sanford Koester,
University of Montana, United States
Zhichao Xia,
University of Connecticut, United States

*CORRESPONDENCE
Laurent Lesecq
✉ lesecq.laurent@chu-amiens.fr

RECEIVED 22 May 2024
ACCEPTED 05 September 2024
PUBLISHED 26 September 2024

CITATION
Lesecq L, Querne L, Gornes J, Buffo L, Corbel L, Le Moing AG, Berquin P and Bourdin B (2024) Do gifted children without specific learning disabilities read more efficiently than typically developing children? *Front. Psychol.* 15:1436710. doi: 10.3389/fpsyg.2024.1436710

Do gifted children without specific learning disabilities read more efficiently than typically developing children?

Laurent Lesecq[1,2]*, Laurent Querne[1,3], Julie Gornes[4], Laura Buffo[4], Louise Corbel[4], Anne Gaelle Le Moing[1,3], Patrick Berquin[1,3] and Béatrice Bourdin[2]

[1]Department of Pediatric Neurology and Referral Center for Language and Learning Disorders, Amiens-Picardie University Medical Center, Amiens, France, [2]Psychology Research Center - Cognition, Psychism, Organizations, PRC-CPO (EA7273), Jules Verne University of Picardie, Amiens, France, [3]Groupe de Recherches sur l'Analyse Multimodale de la Fonction Cérébrale (GRAMFC) INSERM U1105, Amiens, France, [4]University Department of Education and Training in Speech Therapy, Paul Sabatier University, Toulouse, France

Introduction: There are no published data on the written language skills of gifted children (GC). The objective of the present study was to evaluate reading abilities of GC vs. normative data from typically developing French children (TDC). Like English, French is considered to be an opaque language.

Method: GC completed the Wechsler Intelligence Scales and a battery of language tests. Only children with a score two standard deviations (SD) above the norm were included. GC with current or past academic difficulties or specific learning disorders were excluded. The GC's scores were compared with TDC's normative scores for language tests in a chi-square-test and corrected for multiple comparisons.

Results: Forty-five GC were included. The highest GC's mean scores were for the WISC's Verbal Comprehension Index (VCI) and the lowest for the Processing Speed Index (from more than two SDs to one SD higher above the TDC's normative scores). GC were between 1.3 and 4.7 times more likely than TDC to achieve a high score. After correction, the distributions of the GC's and TDC's scores differed significantly with regard to spoonerism, phoneme deletion, and rapid automatic naming ($p<0.001$), word and sentence repetition ($p\leq0.007$), and the reading of meaningful text ($p=0.03$). GC and TDC did not differ significantly for reading meaningless texts and spelling accuracy.

Discussion: As described in the literature, the GC in the present study had heterogeneous scores on the Wechsler Intelligence Scales. The GC performed better than TDC in assessments of the underlying skills of reading and when reading of meaningful texts. This advantage was lost in the absence of context, as shown by the lack of significant GC vs. TDC differences for reading meaningless texts and for spelling accuracy. Hence, GC presented a heterogeneous profile with regard to the underlying skills of reading and reading abilities. The present data should help to improve our understanding of GC's reading skills. In particular, it is now essential to determine which written language tests and which score thresholds are appropriate for identifying specific learning disorders in GC.

KEYWORDS

gifted children, Reading abilities, specific learning disabilities, pathological threshold, Wechsler scales

Introduction

Over the years, numerous models and definitions of giftedness have been proposed (Worrell et al., 2019), but to date no consensus has been reached on the precise meaning of giftedness. However, intellectual giftedness is frequently identified by standardized measures of intelligence. According to the American Psychological Association's Dictionary of Psychology (American Psychological Association, 2018), giftedness is defined as a Full-Scale Intelligence Quotient (FSIQ) of 130 or more (corresponding to two standard deviations (SDs) above the population average). Although the FSIQ are frequently used to identify intellectual giftedness, there is no consensus among experts on the best methods and criteria for identifying and assessing superior cognitive abilities (Hodges et al., 2018). For example, the use of single index vs. FSIQ is subject to debate in the literature (Pereira-Fradin et al., 2010; Liratni and Pry, 2007, 2012; Farmer et al., 2021; Watkins and Canivez, 2022). Furthermore, inter-indexes differences increase with intellectual performance and might mask the identification of some gifted children (GC; Labouret and Gregoire, 2018). The levels of performance for reading and writing are at least partly related to the level of intellectual ability. Specifically, research suggests that vocabulary skills (Stanovich, 2000; Gough and Tunmer, 1986; Gavard et al., 2023) and the knowledge of text structure (Duke and Cartwright, 2021) enhance the speed of reading.

Over the past two decades, scientific research has considerably enriched our understanding of how GC operate on the cognitive level. However, few studies have examined the language skills of GC. As noted in the recent review by Bucaille et al. (2022), GC have a higher lexical capacity than their typically developing peers. Similarly, there are few literature data on the reading skills of GC. Indeed, to the best of our knowledge, GC's reading skills have not previously been comprehensively studied and a few studies have focused solely on the reading skills of dyslexic GCs (Kranz et al., 2024; Van Viersen et al., 2015, 2016). Most of these studies were conducted in English, even though it is known that the characteristics of a language have an important impact on reading skills. Indeed, the identification of written words in the reading process depends on the written language's degree of opacity (degree of correspondence between the spelling and the phonology of the language). In their written forms, English and French are both considered to be opaque languages (Caravolas et al., 2019; Landerl et al., 2022; Paulesu et al., 2001). Learning to read in English appears to be more difficult than in other European languages (Seymour et al., 2003), and grapheme-phoneme decoding skills are less effective in English dyslexics than in German dyslexics, for example (Ziegler et al., 2003). Arffa (2007) found that only 28% of the variance in the reading scores of typically developing children (TDC) was explained by intelligence and emphasized the need to further investigate this complex relationship. Although the reading level is known to be related to intelligence, a lack of research on this topic means that there is a significant gap in our understanding of GC's ability to read in English and French.

The Simple View of Reading (SVR) model (Gough and Tunmer, 1986) considers that reading ability (i.e., the ability to understand written language) has two fundamental components: written word recognition (i.e., decoding) and language comprehension (i.e., oral language skills). Sprenger-Charolles and Ziegler's (2019) adaptation of the SVR model (see Figure 1) distinguishes between various cognitive skills involved in decoding and in listening comprehension.

As pointed by Kranz et al. (2024) in their recent review of reading and reading disorders in GC, there are no detailed literature data on reading skills. Kranz et al. pointed out that this type of data is essential for accurately diagnosing Specific Learning Disability (SLD) and precisely understanding the children's reading profiles. The objective of the present study (conducted in France) was to investigate the reading skills and cognitive profiles of GC without learning disabilities. Lastly, we discuss our results with regard to the diagnosis of reading disorders in GC.

Method

Participants

Study participants were recruited through institutions offering educational programs for GC, associations for GC, and healthcare professionals. Each child and his/her legal guardian(s) received a study information sheet and gave their written, informed consent. The study was approved by the local investigational review board [*CPP Nord-Ouest II* (Amiens, France)]; reference: PI2021_843_0098. The study database was registered with the French National Data Protection Commission [*Commission nationale de l'informatique et des libertés* (Paris, France); reference: 2208336 v 0].

The children included in the study were recruited from specific classes for GC children in private schools. All the participating GC completed the Wechsler Intelligence Scale for Children, 4th or 5th edition (WISC-IV or V) or the Wechsler Preschool and Primary Scale of Intelligence, 4th edition (WPPSI-IV). In line with the recommendations of Goldschmidt and Brasseur (2021) and Grégoire (2021), only children who scored 130 or more for at least one of the following reasoning indexes were included in the study: the Verbal Comprehension Index (VCI), the Visual Spatial Index (VSI), and the Fluid Reasoning Index (FRI) for the WISC-V and WPPSI-IV, and the VCI and the Perceptual Reasoning Index (PRI) for the WISC-IV. Children with ongoing or past academic difficulties, SLD, or psychiatric or neurological disorders were not included in the study.

Material

The GC's passive vocabulary and listening comprehension skills were evaluated using the French adaptation of the revised

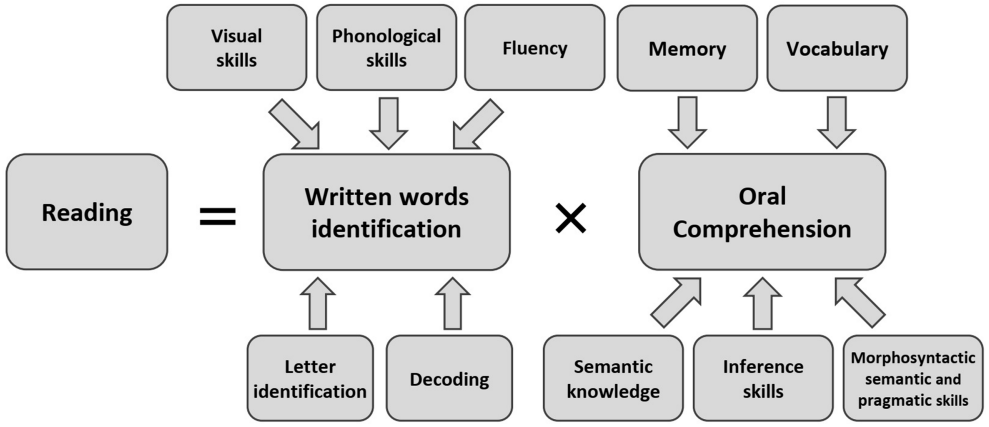


FIGURE 1
The Simple View of Reading (SVR) model by Gough and Tunmer (1986), revised by Sprenger and Ziegler (2019).

Peabody Picture Vocabulary Test [*Evaluation du Vocabulaire en Images Peabody* (EVIP); Dunn and Theriault-Whalen, 1993; Figure 1]. The children were asked to name one of four items that best corresponded to the word spoken by the examiner. Performance was reported as a standardized score (mean = 100, SD = 15).

The *EVAluation du Langage Ecrit et du langage Oral 6-15 ans* (EVALEO) is currently the most comprehensive, standardized, computerized battery for the assessment of written and spoken French language ability in children aged 6 to 15 (Launay et al., 2018a). Each year group corresponds to an average of 145 schoolchildren, ranging from *cours préparatoire* (Year 2/first grade) to *troisième* (Year 10/ninth grade). The distribution of EVALEO scores do not conform to a Gaussian distribution but range from S1 (abnormal) to S7 (very above-average) and correspond to the following percentile intervals: S.1 < 7%, S.2 [7–20%], S3 [21–38%], S.4 [39–62%], S.5 [63–80%], S.6 [81–93%], and S.7 > 93%.

The EVALEO tests were chosen to assess all the cognitive domains in the SVR model, in accordance with the guidelines for good practice in the assessment, prevention and remediation of written language disorders published by the French College of Speech Therapy (Leloup et al., 2022). The tests assessed various aspects of underlying reading skills such as spoonerism, phoneme deletion, and speed and accuracy of rapid color denomination (i.e., rapid automatic naming (RAN)). The tests also assessed the ability to read pseudowords, logatoms (i.e., meaningless words), and meaningless and meaningful texts, as well as spelling accuracy in a dictation. Although these markers cannot be used as individual diagnostic criteria for SLD, they are nonetheless integral elements of learning disability assessment protocols (Saksida et al., 2016; Colé and Sprenger-Charolles, 2021). Hence, they were included to provide standards for these skills within our sample. Furthermore, children with stronger cognitive reasoning abilities (particularly in the verbal domain) achieve higher scores in metaphonology tests than children with weaker cognitive reasoning abilities (McBride-Chang and Manis, 1996). Within the EVALEO population (around 1,500 children), males represent 47% of the sample, females 53%. Three types of pathology were taken into account in the EVALEO sample: dyslexia, dysorthographia and specific language impairment, and their possible combinations. Around 8% of children tested had one or more pathologies, including 6% of girls and 9% of boys. This figure of 8% corresponds to a percent commonly found in France. The distribution of socio-professional categories of parents of children from EVALEO is quite similar to the national statistics provided by the Institut National de la Statistique et Etudes Economiques (INSEE; Launay et al., 2018b) with the exception of an over-representation of the categories "manager, higher intellectual profession" and an under-representation of the "worker" category.

Statistical analysis

Inter-index differences in the VCI, VSI, PRI, Working Memory Index (WMI) and Processing Speed Index (PSI) were tested with Student's t-test for repeated measures or (if the latter could not be applied) Wilcoxon's test.

The EVALEO test results were grouped together, as follows: S.1 with S.2 (S.12 ≤ 20%), S.3 with S.4 and S.5 (S.345 [21–80%]), and S.6 with S.7 (S.67 > 80%). Given that an average of 145 schoolchildren per school year completed the EVALEO battery, we considered that the S.12 group contained 29 TDC, S.345 contained 87 TDC, and S.67 contained 29 TDC children.

Firstly, we calculated the proportion ratios (*p*-ratios) [95% confidence interval (CI)] of GC vs. TDC for the EVALEO S.12 and S.67 scores in order to quantify the expected superiority of GCs over the TDCs of the EVALEO (EVALEO-TDC). Hence, the *p*-ratio for S.12 was defined as [$n_{(GC)}$ with S.12/$all_{(GC)}$]/[$n_{(EVALEO\text{-}TDC)}$ with S.12/$all_{(EVALEO\text{-}TDC)}$], and the *p*-ratio for S.67 was defined as [$n_{(GC)}$ with S.67/$all_{(GC)}$]/[$n_{(EVALEO\text{-}TDC)}$ with S.67/$all_{(EVALEO\text{-}TDC)}$].

Secondly, comparisons of proportions between GC vs. EVALEO-TDC for S.12, S.345 and S.67 were conducted with the chi-squared test.

The Benjamin-Holchberg test was used to check the alpha risk inflation for multiple comparisons in the WISC/WPPSI index (i.e., three GC intragroup comparisons) and the S.12, S.345 and S.67 *p*-ratios (i.e., 12 GC/EVALEO-TDC intergroup comparisons). All the

p-values below are reported after Benjamini-Holchberg correction (statistical-tests: k = 15; significance: $p < 0.05$).

Results

Forty-seven children (age range: 8 to 15) were initially included in the study. Two of the 47 were excluded due to low reading scores; hence, 45 children [34 boys (75.6%) and 11 girls (24.4%)] were included in the analysis.

Intellectual and passive vocabulary efficiency

Regarding intellectual efficiency, the mean scores of GC were more than two SDs higher than the TDC norms for the VCI (135.5, 95%-CI [132.4–138.7]). Compared with the VCI, the scores were progressively lower for the VSI/PRI (128.8, 95%-CI [125.3–132.3], difference VCI vs. VSI/PRI: $p = 0.006$), the WMI (126.1, 95%-CI [123.9–130.7], difference VSI/PRI vs. WMI: nonsignificant) and the PSI (115.8, 95%-CI [111.9–119.7], difference WMI vs. PSI: $p < 0.001$; Figure 2), in that order. The GC's mean PSI was nevertheless one SD above the normative value.

For the passive lexicon, the GC's mean [95%CI] EVIP score (128.7 [125.2–132.2]) was almost two SDs greater than the normative TDC's score.

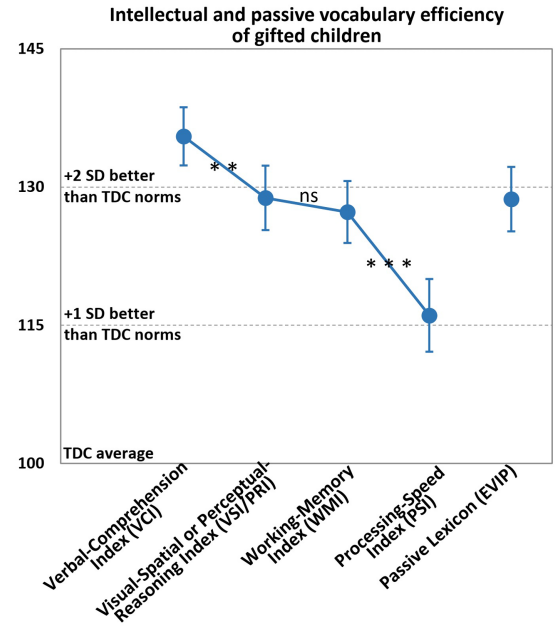


FIGURE 2
Mean scores obtained by GC to the WISC or WPPSI and EVIP. Vertical bars represented the 95%-CI. The *p*-values were given after Benjamini-Hochberg corrections for multiple comparisons with k = 15 (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$). SD: standard deviation; TDC, typically developmental children; ns, non significant.

The GC/EVALEO-TDC *p*-ratio for high and low EVALEO scores

The GC/EVALEO-TDC *p*-ratio showed that the GC were between 4.7 and 3.5 times more likely than EVALEO-TDC to obtain a score of S.67 for RAN accuracy, word repetition, and the metaphonologic test of spoonerism; the 95%CI ranged from 2.4 to 6.5 (Figure 3). In contrast, the GC were 4.5 times less likely to obtain a score of S.12 than the EVALEO-TDCs; the *p*-ratio ranged from 0 to 0.22, and the highest value of the upper boundary of the 95%CI was 0.90.

For the metaphonologic test of phoneme deletion, sentence repetition, and meaningful text reading, spelling accuracy of writing sentences, and RAN speed, the *p*-ratio for S.67 ranged from 1.8 to 2.4 and the lower boundary of the 95%CI ranged from 1.1 and 3.8. For S.12, the GC/EVALEO-TDC *p*-ratio ranged from 0.44 to 0.78 but the upper boundary of the 95%CI was always greater than 1.

For pseudoword repetition and the ability to read words, pseudowords and meaningless texts, the GC/EVALEO-TDC *p*-ratio for S.67 ranged from 1.33 to 1.44 and the lower boundary of the 95%CI was always below 1. The GC/EVALEO-TDC *p*-ratio for S.12 ranged from 0.33 to 0.56, and the upper boundary of the 95%CI was always above 1.

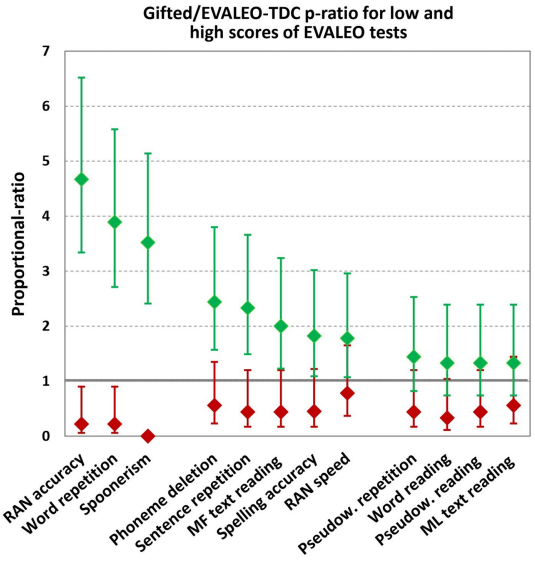


FIGURE 3
Proportional-ratio (*p*-ratio) of GC vs. EVALEO-TDC norms for scores 1 with 2 (S.12) and scores 6 with 7 (S.67) for the different speech and reading tasks of the EVALEO battery. From left to right, tasks were ordered from the largest to the smallest difference in the GC/EVALEO-TDC *p*-ratio. Vertical bars represent the 95%-CI for low scores (S.12) in red and for high scores (S.67) in green. A 95%-CI *p*-ratio for S.12 of less than one indicates that GC were significantly less likely to score 1 or 2 than EVALEO-TDC, while a 95%-CI *p*-ratio for S.67 greater than 1 indicates that GC were significantly more likely to score 6 or 7 than EVALEO-TDC (95%-CI *p*-ratio were not corrected for multiple comparisons). In example for the RAN accuracy, the GC were 4.7 more likely to achieve S.67 scores than EVALEO-TDC (*p*-ratio = 4.7, 95%-CI between 3.34 and 6.52), while they were 4.5 less likely to achieve S.12 scores than EVALEO-TDC (*p*-ratio = 0.22, 95%-CI between 0.06 and 0.90). MF: meaningful; ML: meaningless; Pseudow., pseudowords.

Pairwise comparisons of GC and EVALEO-TDC for the EVALEO test results

For verbal tests that assessed underlying skills of reading, statistical analysis indicated that GC scored significantly better than EVALEO-TDC for spoonerism and phoneme deletion ($p < 0.001$) and for word and sentence repetition ($p < 0.001$ and $p = 0.007$, respectively). However, the GC's and EVALEO-TDC's scores for pseudoword repetition did not differ significantly (Figures 4A,B).

For the RAN, our analysis indicated that GC were significantly better than EVALEO-TDC in terms of accuracy ($p < 0.001$). The GC were better than EVALEO-TDC in terms of reading speed but the difference was not statistically significant ($p = 0.09$; Figure 4C).

For reading words, pseudowords and meaningless texts, the GC and EVALEO-TDC had similar scores ($p =$ NS). For reading meaningful texts, the GC's scores were significantly better than the EVALEO-TDC's scores ($p = 0.03$; Figures 4D,E).

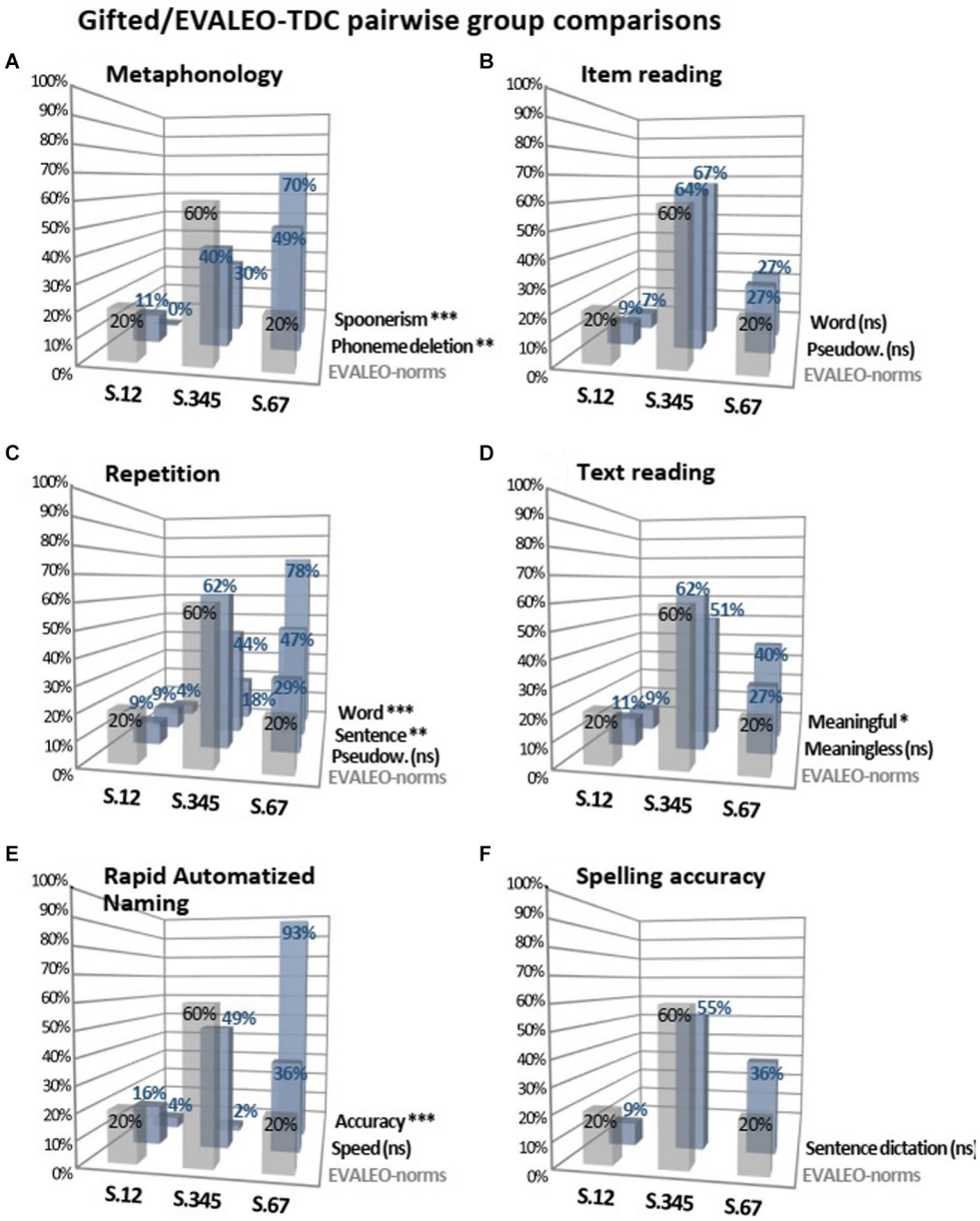


FIGURE 4
Percentage of GC (blue vertical bars) and EVALEO-TDC norms (grey vertical bars) that obtained S.12, S.345, and S.67 for speech and reading tasks of the EVALEO battery. Groups comparisons were tested with the ki-square and *p*-values were given after Benjamini-Hochberg corrections for multiple comparisons with k = 15 (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$). ns, non significant.

Lastly, with regard to spelling accuracy, the GC obtained better sentence dictation scores than the EVALEO-TDC but the difference was not statistically significance ($p = 0.07$; Figure 4F).

Discussion

The results indicated that the GC performed significantly better than the norms in tasks involving automated rapid naming, phoneme deletion, spoonerism, and repetition of words and sentences. Our findings were consistent with those of Van Viersen's study of the Dutch langage (Van Viersen et al., 2015) and indicated that the GC performed better than TDC in almost all the phonological and metaphonological tests of the components underlying reading skills. According to the SVR model (Sprenger-Charolles and Ziegler, 2019), phonological abilities are involved in the first level of reading processes in general and in the decoding of written words in particular. Previously published studies had demonstrated a strong relationship between metaphonology and working memory (Brady, 1991; Gindri et al., 2007). Thus, the GC's strong performance in metaphonologic tests might result from their high working memory capacity (see also Aubry et al., 2021). Indeed, the GC's mean score for the WMI was almost two SDs above the TDC's mean score.

In the present study, the GC were more efficient than TDC'normative scores in reading meaningful texts. However, in contrast to Van Viersen's findings, GC were not significantly better than TDC (i) at reading meaningless texts, isolated pseudowords, and complex words, and (ii) with regard to spelling accuracy. The SVR model suggests that reading involves more than just identifying words: it ultimately requires comprehension of the meaning formed by the words identified during the decoding. This last comprehension step involves vocabulary, semantics, and morphosyntactic knowledge, in addition to working memory. Previously published studies have shown strong positive correlations between the passive lexicon and the VCI and even the FSIQ (Hodapp and Gerken, 1999; Caroff et al., 2006). Furthermore, it is known that a rich vocabulary is related to verbal comprehension (Schelstraete, 2012; Haft et al., 2016).

The GC in the present study had a larger passive vocabulary than the TDC's normative scores, with a EVIP mean score nearly 2 SD higher than the age average. The GC also presented very high semantic knowledge and verbal inference scores; for the latter, the GC's VCI in the WISC or WPPSI was more than two SDs greater than that of the TDC's normative scores. Compared with TDC, the GC's greater semantic and verbal inference abilities allowed them to make much better use of the semantic context of a text, as demonstrated by their excellent ability to read meaningful texts. Our results support the hypothesis whereby decoding speed is linked to comprehension skills (as in the SVR model; Gough and Tunmer, 1986) and also knowledge of the text's structure (as in the Active View of Reading model; Duke and Cartwright, 2021). In contrast, the GC performing no better than TDC with regard to semantic abilities that cannot be used, i.e., reading meaningless texts. Similarly, there is no semantic context to be leveraged when reading isolated words or pseudowords. Indeed, GC did not outperform TDC's normative scores when reading meaningless texts, isolated words, and pseudowords. With regard to spelling accuracy, French language rules are complex, often arbitrary, and thus difficult to predict (Fayol and Jaffré, 2008). Once again, it appears that good semantic and verbal inference skills are not linked to spelling accuracy, as shown by GC's near-normal performance in a sentence dictation.

Limitations

Firstly, the study population was relatively small. Nevertheless, given that our GC's cognitive profile for the WISC was similar to those found in other studies, the sample was probably representative of GC in general. Indeed, the GC's VCI was two SDs higher than the TDC's VCI. Although the GC's other indexes were lower in absolute terms, they were significantly better than the TDC's normative scores. The GC's lowest score (the PSI) was still one SD greater than the equivalent for the TDC (Liratni and Pry, 2007; Terriot, 2018). The details of the socio-economic status of the GC studied were not known. However, the children in the study were recruited from private schools, in which we can expect a socio-economic status bias quite similar to that of EVALEO population: probable over-representation of executives and higher intellectual professions and under-representation of workers.

We did not assess reading comprehension because this complex cognitive activity is extremely difficult to measure in a standardized test (Bianco, 2019). Nevertheless, many processes involved in reading comprehension are also involved in verbal comprehension (Leloup et al., 2022). In their review of the literature, Snowling and Melby-Lervåg (2016) concluded that a written language comprehension disorder is associated with a verbal language comprehension disorder. However, this was not the case for our GC, who scored highly for the WISC VCI and the EVIP vocabulary test.

The French guidelines on good practice in the assessment, prevention and remediation of written language disorders in children and adults were based on the SVR model (Leloup et al., 2022). Recently, more dynamic reading models (such as the Active View of Reading model) have been built on the SVR model but take account of an overlap between listening comprehension and decoding (Duke and Cartwright, 2021). These models describe more complex, intricate processes.

Conclusion

The present study covered most reading skills; the results indicated that GC performed better than TDC in tests that assess the underlying skills of reading (i.e., phonology and metaphonology). This difference was probably due to the GC's highly efficient working memory. The GC were better able to read meaningful texts because they leveraged the semantic context more effectively than TDC did. This was probably due to the GC's excellent verbal inference abilities and richer vocabulary, compared with TDC. In tests that did not involve underlying reading skills or semantic aspects, GC did not perform any better than TDC; this was observed for reading meaningless texts, isolated words and isolated pseudowords and for spelling accuracy (the rules for which are extremely complex and arbitrary in French). Thus, GC did not have uniformly excellent language and reading skills; some skills were well above-average, and others were within the norm. GC also show heterogeneity in various intellectual domains, which significant differences between scores for verbal reasoning, non-verbal reasoning, working memory, and processing speed.

Some researchers have argued that the threshold for abnormal reading ability should be modified for GC with SLD. For example,

Habib (2018) argued that the reading norm for TDC is abnormal threshold for GC with SLD. However, the results of the study argued against this suggestion because our GC (without SLD) and TDC did not differ significantly with regard to certain reading skills of GC. Nevertheless, our results call into question the relevance of the usual reading disorder thresholds for GC. It appears that strict application of the standard abnormal thresholds for GC would result in a large number of false negatives; these children would not receive appropriate care and would not have their disability recognized for school and examination purposes. Our results emphasize the importance of using specific thresholds to assess the intellectual abilities of GC and the need for reference data for reading disorders in this population. Our findings strongly supported Kranz et al.'s conclusion (2024) whereby specific thresholds should be established in this population.

Data availability statement

In accordance with French legislation, the parents were given written information about the study and were free to refuse the inclusion of their child in the study. The datasets cannot be shared for ethical reasons; sharing has not been authorized by the French National Data Protection Commission (Commission nationale de l'informatique et des libertés). Requests to access the datasets should be directed to lesecq.laurent@chu-amiens.fr.

Ethics statement

The studies involving humans were approved by the local investigational review board [CPP Nord-Ouest II (Amiens, France)]; reference: PI2021_843_0098. The study database was registered with the French National Data Protection Commission [Commission nationale de l'informatique et des libertés (Paris, France); reference: 2208336 v 0]. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

LL: Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Investigation, Supervision, Methodology. LQ: Writing – original draft, Writing – review & editing, Conceptualization, Software, Methodology, Investigation, Formal analysis. JG: Writing – review & editing, Investigation, Supervision, Funding acquisition. LB: Writing – review & editing, Funding acquisition, Investigation. LC: Writing – review & editing, Funding acquisition, Investigation. AL: Writing – review & editing, Supervision. PB: Writing – review & editing, Supervision. BB: Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Investigation, Supervision, Methodology.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

American Psychological Association (2018). APA dictionary of psychology. Coping strategy: Washington: Retrieved from American Psychological Association.

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Arch. Clin. Neuropsychol.* 22, 969–978. doi: 10.1016/j.acn.2007.08.001

Aubry, A., Gonthier, C., and Bourdin, B. (2021). Explaining the high working memory capacity of gifted children: contributions of processing skills and executive control. *Acta Psychol.* 218:103358. doi: 10.1016/j.actpsy.2021.103358

Bianco, M. (2019). "Vers un enseignement de la compréhension des textes" in L'apprentissage de la lecture. eds. A. Bentolila and B. Germain (Paris: Nathan), 129–143.

Brady, S. A. (1991). "The role of working memory in reading disability" in Phonological processes in literacy: A tribute to Isabelle Y. Liberman. eds. S. A. Brady and D. P. Shankweiler (Mahwah: Lawrence Erlbaum Associates, Inc), 129–151.

Bucaille, A., Jarry, C., Allard, J., Brochard, S., Peudenier, S., and Roy, A. (2022). Neuropsychological profile of intellectually gifted children: a systematic review. *J. Int. Neuropsychol. Soc.* 28, 424–440. doi: 10.1017/S1355617721000515

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., and Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and Reading comprehension ability. *Sci. Stud. Read.* 23, 386–402. doi: 10.1080/10888438.2019.1580284

Caroff, X., Jouffray, C., Jilinskaya, M., and Fernandez, G. (2006). Identification multidimensionnelle du haut potentiel: Mise au point d'une version française des échelles d'évaluation des caractéristiques de comportement des élèves surdoués. *Bullet. Psychol. Numéro* 485:5. doi: 10.3917/bupsy.485.0469

Colé, P., and Sprenger-Charolles, L. (2021). Lecture et dyslexie: approche cognitive. Malakoff: Dunod.

Duke, N. K., and Cartwright, K. B. (2021). The science of Reading progresses: communicating advances beyond the simple view of Reading. *Read. Res. Q.* 56:25–44. doi: 10.1002/rrq.411

Dunn, L. M., and Theriault-Whalen, C. M. (1993). Échelle de vocabulaire en images Peabody (EVIP), Adaptation française du Peabody Picture Vocabulary Test-Revised. Manuel pour les formes A et B, Toronto: Psycan, 147.

Farmer, R. L., McGill, R. J., Dombrowski, S. C., and Canivez, G. L. (2021). Why questionable assessment practices remain popular in school psychology: instructional materials as pedagogic vehicles. *Can. J. Sch. Psychol.* 36, 98–114. doi: 10.1177/0829573520978111

Fayol, M., and Jaffré, J.-P. (2008). Orthographier. Paris: Presses universitaires de France.

Gavard, E., Chanoine, V. C., Geringswald, F., Tan, Y., Cavalli, E. C., Anton, J. L., et al. (2023). "Semantic and syntactic prediction in reading and statistical learning: an fMRI study in dyslexia. Society for the Neurobiology of Language." in *15th Annual Meeting, Marseille (13)*, France. hal-04292818.

Gindri, G., Keske-Soares, M., and Mota, H. B. (2007). Memória de trabalho, consciência fonológica e hipótese de escrita. *Pró-Fono Rev. Atualização Científica* 19, 313–322. doi: 10.1590/S0104-56872007000300010

Goldschmidt, I., and Brasseur, S. (2021). "Au-delà du QI: donner du sens au bilan intellectuel" in Psychologie du haut potentiel. eds. N. Clobert and N. Gauvrit (Louvain la neuve: De boeck Supérieur), 221–232.

Gough, P. B., and Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial Spec. Educ.* 7, 6–10. doi: 10.1177/074193258600700104

Grégoire, J. (2021). "Chapitre 1. Le WISC-V, outil d'examen clinique de l'intelligence" in WISC-V: Examen clinique de l'intelligence de l'enfant. Bruxelles: Mardaga. 13–58. Available at: https://www.cairn.info/wisc-v-examen-clinique-de-l-intelligence-de-l-enfa--9782804709594-p-13.htm.

Habib, M. (2018). La constellation des dys: Bases neurologiques de l'apprentissage et de ses troubles. *2e éd* Edn: De Boeck supérieur.

Haft, S. L., Myers, C. A., and Hoeft, F. (2016). Socio-emotional and cognitive resilience in children with reading disabilities. *Curr. Opin. Behav. Sci.* 10, 133–141. doi: 10.1016/j.cobeha.2016.06.005

Hodapp, A. F., and Gerken, K. C. (1999). Correlations between scores for Peabody picture vocabulary test—III and the Wechsler intelligence scale for children—III. *Psychol. Rep.* 84, 1139–1142. doi: 10.2466/pr0.1999.84.3c.1139

Hodges, J., Tay, J., Maeda, Y., and Gentry, M. (2018). A Meta-analysis of gifted and talented identification practices. *Gifted Child Quart.* 62, 147–174. doi: 10.1177/0016986217752107

Kranz, A. E., Serry, T. A., and Snow, P. C. (2024). Twice-exceptionality unmasked: a systematic narrative review of the literature on identifying dyslexia in the gifted child. *Dyslexia* 30:e1763. doi: 10.1002/dys.1763

Labouret, G., and Gregoire, J. (2018). La dispersion intra-individuelle et le profil des scores dans les QI élévés. *ANAE* 154, 271–279.

Landerl, K., Castles, A., and Parrila, R. (2022). Cognitive precursors of Reading: a cross-linguistic perspective. *Sci. Stud. Read.* 26, 111–124. doi: 10.1080/10888438.2021.1983820

Launay, L., Maeder, C., Roustit, J., and Touzin, M. (2018a). EVALEO 6-15 – Batterie d'évaluation du 387 langage oral et du langage écrit chez les sujets de 6 à 15 ans. Ortho-edition.

Launay, L., Maeder, C., Roustit, J., and Touzin, M. (2018b). EVALEO 6-15 – Batterie d'évaluation du langage oral et du langage écrit chez les sujets de 6 à 15 ans. *Rééducation Orthop.* 273, 93–135.

Leloup, G., Launay, L., and Witko, A. (2022). Recommandations de Bonne Pratique d'Évaluation, de Prévention et de Remédiation des troubles du langage écrit chez l'enfant et l'adulte. Recommandations par consensus formalisé. Collège Français d'Orthophonie. (CFO, UNADREO, FNO). Available at: https://www.college-francais-orthophonie.fr/recommandations-de-bonne-pratique-en-langage-ecrit/ (Accessed April 12, 2024).

Liratni, M., and Pry, R. (2007). Psychométrie et WISC IV: Quel avenir pour l'identification des enfants à haut potentiel intellectuel? *Neuropsychiatr. Enfance Adolesc.* 55:55, 214–219. doi: 10.1016/j.neurenf.2007.06.003

Liratni, M., and Pry, R. (2012). Profils psychométriques de 60 enfants à haut potentiel au WISC IV. *Prat. Psychol.* 18, 63–74. doi: 10.1016/j.prps.2011.01.006

McBride-Chang, C., and Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: a test of the double deficit hypothesis. *Read. Writ.* 8, 323–339. doi: 10.1007/BF00395112

Paulesu, E., Démonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., et al. (2001). Dyslexia: cultural diversity and biological Unity. *Science* 291, 2165–2167. doi: 10.1126/science.1057179

Pereira-Fradin, M., Caroff, X., and Jacquet, A.-Y. (2010). Le WISC-IV permet-il d'améliorer l'identification des enfants à haut potentiel? *Enfance* 2010:2010:11. doi: 10.4074/S0013754510001035

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J.-F., Bricout, L., et al. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Dev. Psychol.* 52, 1503–1516. doi: 10.1037/dev0000184

Schelstraete, M. A. (2012). "Relations entre langage oral et langage écrit dans les troubles spécifiques du développement du langage oral" in Les dysphasies: de l'évaluation à la rééducation. eds. C. Maillart and M. A. Schelstraete (Paris: Masson).

Seymour, P. H. K., Aro, M., and Erkskine, J. M. (2003). Foundation Literacy Acquisition in European Orthographies *British Journal of Psychology*, 94:143:174. doi: 10.1348/000712603321661859

Snowling, M. J., and Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: a meta-analysis and review. *Psychol. Bull.* 142, 498–545. doi: 10.1037/bul0000037

Sprenger-Charolles, L., and Ziegler, J. (2019). "Apprendre à lire: contrôle, automatismes et auto-apprentissage" in L'apprentissage de la lecture. eds. A. Bentollila and B. Germain (Nathan).

Stanovich, K. E. (2000). Progress in understanding Reading: Scientific foundations and new Frontiers. New York: Guilford Press.

Terriot, K. (2018). De la définition théorique du haut potentiel intellectuel aux conséquences pratiques. *ANAE* 154, 265–270.

Van Viersen, S., De Bree, E. H., Kroesbergen, E. H., Slot, E. M., and De Jong, P. F. (2015). Risk and protective factors in gifted children with dyslexia. *Ann. Dyslexia* 65, 178–198. doi: 10.1007/s11881-015-0106-y

Van Viersen, S., Kroesbergen, E. H., Slot, E. M., and De Bree, E. H. (2016). High Reading skills mask dyslexia in gifted children. *J. Learn. Disabil.* 49, 189–199. doi: 10.1177/0022219414538517

Watkins, M. W., and Canivez, G. L. (2022). Assessing the psychometric utility of IQ scores: a tutorial using the Wechsler intelligence scale for children–fifth edition. *Sch. Psychol. Rev.* 51:Article 5. doi: 10.1080/2372966X.2020.1816804

Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., and Dixson, D. D. (2019). Gifted students. *Annual Review of Psychology*, 70, 551–576. doi: 10.1146/annurev-psych-010418-102846

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., and Schulte-Körne, G. (2003). Developmental dyslexia in different languages: language-specific or universal?. *Journal of experimental child psychology*, 86, 169–193. doi: 10.1016/s0022-0965(03)00139-5

frontiers | Frontiers in Psychology

TYPE Mini Review
PUBLISHED 15 November 2024
DOI 10.3389/fpsyg.2024.1489997

OPEN ACCESS

EDITED BY
Thomas L. Spalding,
University of Alberta, Canada

REVIEWED BY
Kuan-Lin Chen,
National Cheng Kung University, Taiwan

*CORRESPONDENCE
Lonneke Janssen
lo.janssen@kentalis.nl

RECEIVED 02 September 2024
ACCEPTED 15 October 2024
PUBLISHED 15 November 2024

CITATION
Janssen L, Scheper A and Vissers C (2024) Controlling the narrative: the relationship between narrative ability and executive functioning in children with developmental language disorder.
Front. Psychol. 15:1489997.
doi: 10.3389/fpsyg.2024.1489997

Controlling the narrative: the relationship between narrative ability and executive functioning in children with developmental language disorder

Lonneke Janssen[1,2]*, Annette Scheper[1,2] and Constance Vissers[1,2]

[1]Royal Dutch Kentalis, Sint-Michielsgestel, Netherlands, [2]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

Children with developmental language disorder (DLD) experience problems in language comprehension and/or production. In particular, storytelling or narrative ability is often impaired, as this type of discourse involves all domains of language. These problems may lead to a lower quality of social interaction and mental health. Moreover, problems in oral narrative ability during early development have a negative effect on later literacy. However, telling a story involves more than language alone. Executive functioning is thought to play an important part in stimulating narrative ability, as linguistic utterances need to be planned in a temporal and causal order, and switching is needed between multiple characters and events in the story. Research has shown that children with DLD experience problems with executive functioning, independent of their language ability. Thus, the difficulties in storytelling may be caused by both impaired language and executive functioning, as both domains follow hierarchical developmental paths during the early childhood years. In this review, we discuss three components of narrative ability (comprehension, production of macrostructure and production of microstructure) and how they may be interconnected to the three core components of executive functioning (working memory, switching and inhibition) and attention. This review shows that updating and monitoring information in working memory plays an important part in all three components of narrative ability, across multiple studies. This result may give direction in the development of narrative assessment and intervention, and urge further research to disentangle the interplay between language and executive control in DLD.

KEYWORDS

narrative ability, executive functioning, developmental language disorder, children, neuropsychology

1 Introduction

1.1 Narrative ability and executive functions

Narration is a competence that has communicative, social and academic importance. To tell a coherent story children have to learn to produce series of utterances which are temporally, logically and causally related (Berman and Slobin, 1994). Young children are quickly expected to demonstrate their narrative ability by talking about their experiences, feelings and life events in social settings. In language development, narrative ability can be seen as the most complex

linguistic task, as children need to integrate rich vocabulary into grammatical utterances that vary in complexity. Moreover, utterances must be cohesive and coherent. Singular utterances in successive order must form a logical structure to communicate appropriately. Executive functions (EF) are thought to play an important role in narration (Büttner, 2016). EF are defined as 'general-purpose control mechanisms that modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of human cognition' (Miyake et al., 2000, p. 50). EFs are regarded as top-down processes that control and regulate goal-oriented behavior needed in day-to-day activities. Miyake et al. (2000, p. 50) identified three core components of EF: the ability to switch flexibly between different tasks (switching), update and monitor working memory (WM), and ignore distracting information or suppress automatic responses (inhibition). According to Büttner (2016), the prerequisites for comprehension and production of narratives are related to specific EFs. These relationships include the causal-temporal ordering of events (updating and monitoring), taking into account the mental states of different characters and switch perspectives between them (switching), and suppressing redundant details or other characters' perspectives (inhibition).

EF and language (e.g., narrative skills) develop concurrently and it is theorized that the development of language and EF occurs hierarchically. The development of the core EFs are preceded by the development of sustained attention, which starts at infancy and is defined as the ability to focus on a specific target for prolonged time periods (Garon et al., 2008; Marini et al., 2020). The core components of EF start developing before 3 years old but reach a critical period between 3 and 5 years old (Garon et al., 2008). The ability to keep information active in WM and inhibitory control predict shifting, which suggests that shifting is a more complex EF ability that builds on other EF abilities such as WM and inhibition (Garon et al., 2014). The development of narrative skills follows a similar path from simple to more complex structures. Narratives at 2 years old consist of descriptions of character actions without a central theme (Scionti et al., 2023). From 3 to 4 years, children include causal connections between these actions. But it is not until 6 or 7 years old that children use "true coherent narratives," in which utterances are temporally, logically and causally related. However, this continues to develop until at least 10 years old (Scionti et al., 2023).

Research shows that if the development of different complex abilities such as language and EF happens concurrently, interaction takes place between functions (facilitation of inhibition) (Bernstein and Waber, 2018; Nilsson and de López, 2016; Carlson et al., 2002; Vissers et al., 2015). In the case of EF and narrative skills, the ability to keep information active in WM might facilitate connecting different events in the story temporally and causally, later supplemented with linking them to a central theme.

1.2 Developmental language disorder

The development of narrative skills does not occur without effort for all children, however. Children with developmental language disorder (DLD) have difficulty acquiring and using language (Bishop et al., 2017). DLD is defined as a neurocognitive developmental disorder that presents itself primarily in the language domain, such as in comprehension, morpho-syntax, semantics, phonology or pragmatics (e.g., narrative ability). The difficulties in acquisition and production of language may not be explained through any intellectual, biomedical or severe hearing impairment (Bishop et al., 2014). The narratives of children with DLD are shorter, show problems in complexity, grammaticality, coherence and fluency and contain a lower information or plot value compared to the narratives of typically developing (TD) peers (Bergmann et al., 2017; Bliss and Pierre, 1997; Christensen, 2019; Cleave et al., 2010; Colozzo et al., 2011; Duinmeijer et al., 2012; Guo et al., 2008; Pearce et al., 2010; Rodríguez et al., 2017).

Moreover, research has shown that the difficulties observed in children with DLD are not limited to the language domain, but also include difficulties in the executive domain, for instance in phonological and visuospatial WM, cognitive flexibility, and inhibition, even while controlling for language ability (Pauls and Archibald, 2016; Vissers et al., 2015; Vugs et al., 2013, 2017).

1.3 The current literature review

Tomas and Vissers (2019) proposed a model for the developmental interconnections between language and other higher cognitive functions. According to them, the interplay between perception, attention, EF and language might underlie the observed problems in communication in children with DLD (Tomas and Vissers, 2019). Through the observed problems in both domains, and the fact that development of these domains happens concurrently and may be facilitated or inhibited by each other, it could be hypothesized that this interconnection is present for narrative ability and EF.

In the current review, we zoom in on the literature regarding the relationship between narrative ability and EF through peer reviewed experimental research on children with DLD aged 4–7 years old. In a recent meta-analysis conducted for TD children and adolescents, it was found that narrative ability and EF are weakly associated (Scionti et al., 2023). However, they also found that the associations between narrative ability and EF are stronger in children with atypical development between 3 and 7 years old than for their TD peers. Atypical development included children with different neurocognitive developmental disorders, including DLD. This highlights the need for differentiation in the way we investigate these problems in children with and without developmental difficulties. The weaker narrative ability of children with DLD in particular may have a negative impact on their social and academic development. Their participation in daily communication may be less frequent and of a lesser quality. Indeed, children with DLD show lower social–emotional well-being (Goh et al., 2021; Maggio et al., 2014). Moreover, narrative competence contributes to academic success, because adequate oral narrative ability predicts later reading and writing ability (Reese et al., 2010; Schaughency et al., 2017). Academic success is often lower for children with DLD (Conti-Ramsden, 2008). Identification of specific interconnections between narrative ability and EF may inform clinical practice to make more evidence-based assessment, intervention decisions and recommendations.

In this review, narrative ability is divided into three components: narrative comprehension, production of macrostructure (story-level) and production of microstructure (sentence-level). For EF, we will discuss the three core components identified by Miyake et al. (2000). We also include attention, as controlled attention is taken as a common requirement in executive tasks and constitutes the first step in the hierarchical development of EF (Garon et al., 2008, 2014; Miyake et al., 2000). In Table 1 the available literature regarding the

TABLE 1 Overview of discussed papers with children's age, method of narrative and cognitive assessments and their scoring.

Paper	Age range M (SD in months)	Assessment of executive functioning or attention	Domains of executive functioning or attention	Assessment of narrative ability	Narrative comprehension scoring	Narrative macrostructure scoring	Narrative microstructure scoring
Dawes et al. (2018)	5;7 (range 5;2-6;2)	1. Comprehensive test of Phonological Processing (Wagner et al., 1999) 2. Repeating Sentences task of the Test of Language Development – Primary, Third edition (Hamill and Newcomer, 1997) 3. Bear/dragon task and grass/snow task (Carlson, 2005)	1. Working memory – phonological loop 2. Working memory – episodic buffer 3. Inhibition	Squirrel Story Narrative Comprehension Assessment (Dawes et al., 2006)	Comprehension questions with pictures as reference	Retelling: story structure and story content rating	Retelling: level of language, syntax and vocabulary ratings
Dicataldo et al. (2023)	6;6 (3.7)	1. Digit span task of the WISC-IV (Wechsler, 2003) 2. The Day & Night Test (Usai et al., 2017) 3. Modified Wisconsin Card Sorting Test (Cianchetti et al., 2007)	1. Verbal working memory 2. Inhibition 3. Cognitive flexibility	Test for Listening Comprehension (Levorato and Roch, 2007)	Comprehension questions regarding explicit and implicit information	Not applicable (NA)	NA
Dodwell and Bavin (2008)	6;7 (3)	1. The Number Recall of the Kaufman Assessment Battery for Children (Kaufman and Kaufman, 1983) 2. Word list recall (Adams et al., 1999) 3. Central executive task (Gaulin and Campbell, 1994) 4. Recalling sentences of the CELF-3 (Semel et al., 1995) 5. Auditory Continuous Performance Test (Hanson and Montgomery, 2002)	1. Working memory – phonological memory span 2. Phonological memory 3. Working memory – central executive 4. Working memory – episodic buffer 5. Sustained attention	1. The Birthday Story (BS; Culatta et al., 1983) 2. ERRNI (Bishop, 2004)	1. BS: Comprehension questions invoking literal information and inferencing 2. ERRNI: Comprehension questions invoking literal information and inferencing	ERRNI (generation and recall): total amount of listed story ideas (max. 24 ideas)	NA
Duinmeijer et al. (2012)	7.35 (1.05)	1. Sustained attention subtest (TEA-Ch, Manly et al., 1998) 2. Digit span task of the WISC, Dutch version (Kort et al., 2005) 3. Word list recall task, a Dutch version of the California Verbal Learning Test, Children's version (Kalverboer and Deelman, 1964)	1. Sustained attention 2. Verbal working memory 3. Verbal working memory	1. Retelling: Bus Story Test (Renfrew, 1997; Jansonius et al., 2014) 2. Generation: Frog Story Test (Mayer, 1969)	NA	1. Retelling: plot score (max. 25 elements) 2. Generation: plot score (max. 19 elements)	Retelling and generation: mean length of utterance, mean length of the five longest utterances, grammatical accuracy, syntactic complexity, non-fluency
Kalliontzi et al. (2022)	4;5 (2.2)	1. Verbal and non-verbal n-back type task (Yang and Gray, 2016) 2. Verbal and non-verbal Flanker type task (Yang and Gray, 2016) 3. Verbal and non-verbal sorting card task (Yang, 2015)	1. Updating working memory 2. Inhibition 3. Cognitive flexibility	Logometro (Mouzaki et al., 2017)	Comprehension questions	Generation and retelling: score for the event, the problem and the potential solution	Generation and retelling: references in the agents

(Continued)

10.3389/fpsyg.2024.1489997

10.3389/fpsyg.2024.1489997

TABLE 1 (Continued)

Paper	Age range M (SD in months)	Assessment of executive functioning or attention	Domains of executive functioning or attention	Assessment of narrative ability	Narrative comprehension scoring	Narrative macrostructure scoring	Narrative microstructure scoring
Marini et al. (2020)	5.19 (0.03)	1. Forward and backward digit recall test of the Wechsler Scales (Wechsler, 1993) 2. Developmental Neuropsychological Assessment (Urgesi et al., 2011)	1. Updating verbal working memory 2. Inhibition	Battery for the assessment of language in children aged 4 to 12 (BVL 4-12; Marini et al., 2015)	NA	NA	Lexical informativeness and global coherence
Smolak et al. (2020)	7;3 (range 6;0 – 8;0)	1. Track-it Task (Erickson et al., 2015; Fisher et al., 2013) 2. Odd One Out task (Henry, 2001)	1. Sustained attention 2. Visuo-spatial working memory	Test for Narrative Language, edition 1 or 2 (Gillam and Pearson, 2004, 2017)	Comprehension questions	Generation: producing key thematic elements	NA

relationship between narrative ability and EF is presented. For each study, the investigated domains and methods are described in a detailed overview.

2 Narrative comprehension and EF

The comprehension of narratives is usually measured as the ability to correctly answer comprehension questions about a model story. These questions can refer to explicit and implicit information. Children either recall explicit information from the story or are required to integrate information in the story with general, real-world knowledge that is implicit (Dicataldo et al., 2023).

2.1 Narrative comprehension and attention

Evidence that confirms a relationship between comprehension and attention was not found. To the contrary, Smolak et al. (2020) found that sustained attention did not correlate significantly with narrative comprehension. Sustained attention was measured by a tracking task with two conditions: a homogeneous and heterogeneous condition, where in the latter the distractor shapes are all the same but different from the target shape, and in the former all distractor shapes are different. Children with DLD performed worse than TD in both conditions, the homogeneous condition did not facilitate tracking of the target shape. However, the ability to track and identify target shapes was not correlated with narrative comprehension for children with DLD.

2.2 Narrative comprehension and working memory, switching and inhibition

The relationship between narrative comprehension and updating and monitoring in WM has been researched intensely. More often than not, associations are found between narrative comprehension and WM. Comprehension and WM are related in digit or word span tasks and non-word repetition tasks, which are seen as verbal measures (Dawes et al., 2018; Dicataldo et al., 2023; Dodwell and Bavin, 2008). The relationship between verbal measures of EF and comprehension in children that have difficulty in language development seems unsurprising. However, the relationship is also present where the WM task is non-verbal in nature (Kalliontzi et al., 2022). In this particular study a relationship was found between visuo-spatial WM and narrative comprehension in younger children with DLD, highlighting that language problems are related to more domain-general cognitive functions as well. Importantly, non-verbal visuo-spatial WM predicted narrative comprehension (along with switching) with an explained variance of 21%.

Moreover, a relationship exists between narrative comprehension and switching in children with DLD (Dicataldo et al., 2023; Kalliontzi et al., 2022). Importantly, both of these studies not only found correlations between both domains, but also that switching was a predictor for narrative comprehension. In Kalliontzi et al. (2022) this predictor was significant alongside visuo-spatial WM.

Narrative comprehension also seems to relate to inhibition (Kalliontzi et al., 2022). In this study, children needed to push a button

according to the direction the middle fish in a row of five fish was looking in. There was a congruent condition (all fish were looking in the same direction) and an incongruent condition (the middle fish was looking in a different direction). It was found that increased reaction times on this inhibition task were associated with more incorrect comprehension questions for children with DLD ($r = -0.29$).

3 Narrative macrostructure and EF

Narrative macrostructure can be defined as the global organization or structure of a story, wherein it is important to establish coherence between sentences and in the narrative as a whole. The way narrative macrostructure is operationalized varies between studies. Some studies use story grammar (Stein and Glenn, 1979). Here, narrative macrostructure includes producing the key story grammar elements such as introduction of the main character, setting, problem, internal emotional response, actions and the resolution. However, some studies use other measures, such as errors in coherence or referents, and informativeness (based on lexical diversity).

3.1 Narrative macrostructure and attention

Even though a relationship was not found for comprehension, sustained attention does play a role in producing narrative macrostructure (Duinmeijer et al., 2012; Smolak et al., 2020). Duinmeijer et al. (2012) measured narrative macrostructure in two narrative genres (mean age 7.4 years). Story retelling was measured through the *Bus Story Test* (Bolk and Scheper, 2024; Jansonius et al., 2014; Renfrew, 1997) and story generation through the *Frog Story Test* (Mayer, 1969; Scheper and Blankenstijn, 2013). To measure sustained attention, children had to count the number of sounds they heard silently until the end of the item. It was found that sustained attention correlated significantly with narrative macrostructure (i.e., the number of story grammar elements produced) in the story generation task, but not in the story retelling task.

Smolak et al. (2020) used a composite score of both narrative macro- and microstructure to investigate the relationship with attention. Narrative macrostructure was scored as producing key thematic elements. The ability to track the target shape in both conditions (homogeneous and heterogeneous) was associated with narrative language production as a whole.

3.2 Narrative macrostructure and working memory, switching and inhibition

There is evidence for a relationship between macrostructure and WM in children with DLD (Dodwell and Bavin, 2008; Duinmeijer et al., 2012; Marini et al., 2020), similar to the relationship found for narrative comprehension. In all three studies, the WM tasks were verbal, namely a retelling sentences task, word list recall and digit recall. It seems that keeping information active and being able to manipulate this information in WM is important in producing narrative macrostructure.

Moreover, a relationship exists between macrostructure and switching (Kalliontzi et al., 2022). It was found that macrostructure in a self-generated narrative correlated significantly with verbal switching, and in the retelling task with non-verbal switching. The difference between the two switching tasks was small, however. To assess verbal switching, children performed a sorting task in which they saw images of scissors on a red background and glasses on a blue background. The child had to sort cards by object or color by pushing a button on a keyboard (Kalliontzi et al., 2022, p. 5). In the non-verbal condition the only difference was that the images of the scissors and glasses were replaced with images of polygons on two different bags and children were asked to perform the same sorting task. Both tasks did not need the child to give any verbal response to an item.

Evidence has been found for a relationship between narrative macrostructure and inhibition as well (Marini et al., 2020). Lexical informativeness correlated significantly with the inhibition measure, in which children first named the shapes of circles and squares or the up and down direction of arrows (naming condition). In the inhibition condition, children were asked to provide the opposite naming response on the same stimuli. However, it can be argued whether lexical informativeness is an adequate measure for narrative macrostructure, as it relates to informativeness on the word level, instead of the global organization or structure of a story as a whole.

4 Narrative microstructure and EF

Narrative microstructure relates to the within-sentence level. It can be operationalized in a variety of ways. Studies may measure the mean length of utterances (MLU), the level of grammatical accuracy, embedding of subordinate clauses (a measure for syntactic complexity), fluency of utterances, use of conjunctions and so on. Whether microstructure is measured with just a single variable or multiple, varies between studies.

4.1 Narrative microstructure and attention

A relationship between narrative microstructure and attention exists (Duinmeijer et al., 2012; Smolak et al., 2020). In Duinmeijer et al. (2012), the mean length of utterances (MLU) and the mean length of the five longest utterances (MLU5) correlated significantly with sustained attention in the story retelling task. A discrepancy exists between the two narrative genres that were investigated, because in story generation, MLU and MLU5 did not correlate with sustained attention, but rather with WM. Evidently, both genres place a different demand on cognitive functions. In story retelling, children were asked to listen to the model story first, which puts a higher demand on sustained attention and is therefore needed in producing longer utterances.

As discussed previously, Smolak et al. (2020) investigated the relationship between attention and narrative language production as a whole. Narrative microstructure was measured by counting causal conjunctions (e.g., *because, as, since*), grammatical accuracy and character dialogue. Narrative language production, which included these microstructural variables, correlated significantly with sustained attention.

4.2 Narrative microstructure and working memory, switching and inhibition

There is evidence for a relationship between narrative microstructure and WM (Duinmeijer et al., 2012). Emphasizing once more the difference between narrative genres, MLU and MLU5 in the story generation task correlated significantly with digit recall (forward and backward), and MLU5 correlated significantly with word list recall as well. The ability to recall digits and manipulate the order in which to recall them plays a role in producing longer utterances. Moreover, MLU5 is an indirect measure of complexity, as complex utterances are often longer than simple ones. It seems that the ability to remember words is more important for producing even longer and possibly more complex sentences.

Kalliontzi et al. (2022) found a relationship between 'narrative language' (macro- and microstructure taken together) and switching in both story retelling and story generation.

5 Discussion

In this review, we discussed the evidence for interconnections between narrative comprehension and production, three core components of EF (updating and monitoring in WM, switching and inhibition) and attention in children with DLD. The evidence points to an important role of WM in narrative comprehension, macrostructure and microstructure. The ability to hold, update and manipulate information in WM is involved in understanding narratives, in creating a global coherent organization and in producing longer, and indirectly, more complex utterances in narratives (Dawes et al., 2018; Dicataldo et al., 2023; Dodwell and Bavin, 2008; Duinmeijer et al., 2012; Kalliontzi et al., 2022; Marini et al., 2020). A relationship between narrative ability and switching or inhibition was less clear-cut. In any case, Dicataldo et al. (2023) and Kalliontzi et al. (2022) did find that switching was a significant predictor for narrative comprehension. This provides some evidence of the role of switching besides WM in narrative comprehension.

We did not find evidence for relationships between all aspects of EF and narrative ability. The literature on narrative microstructure is sparse. It seems that narrative macrostructure is more often the topic of research interest. This is understandable as the unique characteristic of a story is the global organization of coherent language into a unified construct, with a beginning, middle and end. However, the ability to produce grammatical and complex utterances is a prerequisite to produce adequate macrostructure in a narrative, for instance through the use of obligatory syntactic arguments, conjunctions and embedded clauses.

As we have found evidence for a relationship between narrative ability and updating and monitoring information in WM in children with DLD aged 4–7 years, it is also important to understand how these relationships change over time. As Scionti et al. (2023) found in their meta-analysis, the relation between narrative ability and EF in atypical development (which included DLD) was stronger in the ages 3–7 years than in TD children of the same age. Importantly, after 7 years old, effect sizes between both groups did not differ. Across all children with (a)typical development, it was found that relationships between narrative ability and EF decrease from age seven and up. This may be related to the finding that EF components reach a critical development period between 3 and 5 years old, while the development of narratives is still underway, and continues to until 10 years of age (Garon et al., 2008; Scionti et al., 2023). Perhaps other cognitive and environmental factors influence language development at this developmental stage. In a recent review by Smit et al. (2022), it was stated that children aged 6–12 years old learn the ability "to predict not only what someone else thinks or feels, but also what another person thinks of how someone else feels or thinks" (i.e., second-order theory of mind (ToM); Smit et al., 2022, p. 2). This ability may be important for perspective-taking in telling narratives. Indeed, in two of the articles discussed in this review, ToM seemed to play a role, at least in narrative comprehension. In Dawes et al. (2018), ToM correlated stronger with comprehension than WM. Moreover, in Dicataldo et al. (2023), verbal ToM predicted narrative comprehension. The influence of cognitive factors other than EF must be taken into account in researching theoretical underpinnings of (a)typical language development. With respect to Vissers and Tomas' neuropsychological approach to DLD, to reach better informed decisions in narrative intervention, we need to take a broader perspective than the linguistic domain alone. Tapping into the interplay between EF development (WM in particular) and language development (narrative ability in particular) while assessing and treating children with DLD might lead to more academic success and more meaningful social relationships.

Author contributions

LJ: Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. AS: Investigation, Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. CV: Conceptualization, Investigation, Supervision, Writing – review & editing, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Royal Dutch Kentalis funded this review with the aim to investigate contributing factors in the development of language ability in children with DLD. Royal Dutch Kentalis provides treatment and education for children and adolescents with communication difficulties.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adams, A., Burke, L., and Willis, C. (1999). Working memory and spoken language comprehension in young children. *Int. J. Psychol.* 34, 364–373. doi: 10.1080/002075999399701

Bergmann, L., van Hout, R., and Klatter-Folder, J. (2017). "SLI diagnostics in narratives: exploring the CLARIN-NL VALID data archive" in CLARIN in the low Countries. eds. J. Odijk and A. Hessen (London: Ubiquity Press), 167–180.

Berman, R. A., and Slobin, D. I. (1994). Relating events in narrative: a crosslinguistic developmental study. London: Psychology Press.

Bernstein, J. H., and Waber, D. P. (2018). "Executive capacities from a developmental perspective" in Executive function in education: From theory to practice. ed. L. Meltzer. *2nd* ed (New York, NY: The Guilford Press), 57–81.

Bishop, D. V. M. (2004). Expression, Reception and Recall of Narrative Instrument ERRNI. London: Psychological Corporation.

Bishop, D. V. M., Nation, K., and Patterson, K. (2014). When words fail us: insights into language processing from developmental and acquired disorders. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20120403. doi: 10.1098/rstb.2012.0403

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., and Greenhalgh, T.The CATALISE-2 Consortium (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *J. Child Psychol. Psychiatry* 58, 1068–1080. doi: 10.1111/jcpp.12721

Bliss, L. S., and Pierre, K. S. (1997). Personal narrative structures of children with specific-language impairment. *J. Child. Commun. Dev.* 18, 11–18. doi: 10.1177/152574019701800202

Bolk, E., and Scheper, A. (2024). Handleiding bus verhaal test. Amsterdam: UMC & Koninklijke Kentalis.

Büttner, J. (2016). "Neurolinguistic view into narrative processing" in Perspectives on Narrativity and narrative Perspectivization. eds. N. Igl and S. Zeman (Amsterdam: John Benjamins Publishing).

Carlson, S. (2005). Developmentally sensitive measures of executive function in preschool children. *Dev. Neuropsychol.* 28, 595–616. doi: 10.1207/s15326942dn2802_3

Carlson, S. M., Moses, L. J., and Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant Child Dev.* 11, 73–92. doi: 10.1002/icd.298

Christensen, R. V. (2019). Grammatical characteristics of narratives told by Danish children with developmental language disorder. *Int. J. Lang. Commun.* 49, 130–148.

Cianchetti, C., Corona, S., Foscoliano, M., Contu, D., and Sannio-Fancello, G. (2007). Modified Wisconsin card sorting test (MCST, MWCST): normative data in children 4–13 years old, according to classical and new types of scoring. *Clin. Neuropsychol.* 21, 456–478. doi: 10.1080/13854040600629766

Cleave, P. L., Girolametto, L. E., Chen, X., and Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *J. Commun. Disord.* 43, 511–522. doi: 10.1016/j.jcomdis.2010.05.005

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., and Johnston, J. R. (2011). Content and form in the narratives of children with specific language impairment. *J. Speech Lang. Hear. Res.* 54, 1609–1627. doi: 10.1044/1092-4388(2011/10-0247)

Conti-Ramsden, G. (2008). "Heterogeneity of specific language impairment in adolescent outcomes" in Understanding developmental language disorders: From theory to practice. eds. C. F. Norbury, J. B. Tomblin and B. DVM (New York, NY: Psychology Press), 115–129.

Culatta, B., Page, J. L., and Ellis, J. (1983). Story retelling as a communicative performance screening tool. *J. Speech Hear. Serv.* 14, 66–74. doi: 10.1044/0161-1461.1402.66

Dawes, E., Leitão, S., and Claessen, M. (2006). The squirrel story: narrative comprehension assessment (NCA). Available at: https://www.blacksheeppress.co.uk/product/squirrel-story-narrative-comprehension-assessment-nca/ (Accessed October 02, 2024).

Dawes, E., Leitão, S., Claessen, M., and Kane, R. (2018). A profile of the language and cognitive skills contributing to oral inferential comprehension in young children with developmental language disorder*. *Int. J. Lang. Commun. Disord.* 53, 1139–1149. doi: 10.1111/1460-6984.12427

Dicataldo, R., Moscardino, U., Mammarella, I. C., and Roch, M. (2023). Comprehension of explicit and implicit information in prereaders: the role of maternal education, receptive vocabulary, executive functions, and theory of mind. *Discourse Process.* 60, 163–180. doi: 10.1080/0163853X.2023.2185406

Dodwell, K., and Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *Int. J. Lang. Commun. Disord.* 43, 201–218. doi: 10.1080/13682820701366147

Duinmeijer, I., De Jong, J., and Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *Int. J. Lang. Commun. Disord.* 47, 542–555. doi: 10.1111/j.1460-6984.2012.00164.x

Erickson, L. C., Thiessen, E. D., Godwin, K. E., Dickerson, J. P., and Fisher, A. V. (2015). Endogenously and exogenously driven selective sustained attention: contributions to learning in kindergarten children. *J. Exp. Child Psychol.* 138, 126–134. doi: 10.1016/j.jecp.2015.04.011

Fisher, A., Thiessen, E. D., Godwin, K. E., Kloos, H., and Dickerson, J. P. (2013). Assessing selective sustained attention in 3-to 5-yearold children: evidence from a new paradigm. *J. Exp. Child Psychol.* 114, 275–294. doi: 10.1016/j.jecp.2012.07.006

Garon, N., Bryson, S. E., and Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychol. Bull.* 134, 31–60. doi: 10.1037/0033-2909.134.1.31

Garon, N., Smith, I. M., and Bryson, S. E. (2014). A novel executive function battery for preschoolers: sensitivity to age differences. *Child Neuropsychol.* 20, 713–736. doi: 10.1080/09297049.2013.857650

Gaulin, C., and Campbell, T. (1994). Procedure for assessing verbal working memory in normal schoolage children: some preliminary data. *Percept. Mot. Skills* 79, 55–64. doi: 10.2466/pms.1994.79.1.55

Gillam, R. B., and Pearson, N. A. (2004). Test of narrative language. Austin, TX: Pro-ed.

Gillam, R. B., and Pearson, N. A. (2017). (TNL-2) test of narrative language–second edition. Available at: https://www.wpspublish.com/tnl-2-test-of-narrative-languagesecond-edition (Accessed October 02, 2024).

Goh, S. K. Y., Griffiths, S., and Norbury, C. F.the SCALES Team (2021). Sources of variability in the prospective relation of language to social, emotional, and behavior problem symptoms: implications for developmental language disorder. *J. Abnorm. Psychol.* 130, 676–689. doi: 10.1037/abn0000691

Guo, L., Tomblin, J. B., and Samelson, V. (2008). Speech disruptions in the narratives of English-speaking children with specific language impairment. *J. Speech Lang. Hear. Res.* 51, 722–738. doi: 10.1044/1092-4388(2008/051)

Hamill, D. D., and Newcomer, P. L. (1997). Test of language development—Third edition (TOLD-3). Austin, TX: Pro-ed.

Hanson, R. A., and Montgomery, J. W. (2002). Effects of general processing capacity and sustained selective attention on temporal processing performance of children with specific language impairment. *Appl. Psycholinguist.* 23, 75–93. doi: 10.1017/S0142716402000048

Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory* 9, 233–247. doi: 10.1080/09658210042000085

Jansonius, K., Ketelaars, M., Borgers, M., Den, E., Van, H., Roeyers, H., et al. (2014). Renfrew Taalschalen—Nederlandse aanpassing (RTNA). London: Garant Uitgevers.

Kalliontzi, E., Ralli, A. M., Palikara, O., and Roussos, P. (2022). Examining the relationship between oral language skills and executive functions: evidence from Greek-speaking 4–5-year-old children with and without developmental language disorder. *Res. Dev. Disabil.* 124:104215. doi: 10.1016/j.ridd.2022.104215

Kalverboer, A. F., and Deelman, B. G. (1964). Voorlopige selectie van enkele tests voor het onderzoeken van geheugenstoornissen van verschillende aard. Groningen: Groningen University.

Kaufman, A., and Kaufman, N. (1983). Kaufman assessment battery for children. Circle Pines, MN: American Guidance Service.

Kort, W., Schittekatte, M., Dekker, P. H., Verhaeghe, P., Compaan, E. L., Bosmans, M., et al. (2005). Wechsler intelligence scale for children-III-NL. Amsterdam: Harcourt Test Publishers.

Levorato, M. C., and Roch, M. (2007). Test di valutazione della comprensione del testo orale. Florence: O. S. Organizzazioni Speciali.

Maggio, V., Grañana, N. E., Richaudeau, A., Torres, S., Giannotti, A., and Suburo, A. M. (2014). Behavior problems in children with specific language impairment. *J. Child Neurol.* 29, 194–202. doi: 10.1177/0883073813509886

Manly, T., Robertson, I. H., and Nimmo-Smith, I. (1998). Test of everyday attention for children (TEA-Ch). San Antonio, TX: Pearson Clinical

Marini, A., Marotta, L., Bulgheroni, S., and Fabbro, F. (2015). Batteria per la Valutazione del Linguaggio in Bambini dai 4 ai 12 anni. Firenze: Giunti OS.

Marini, A., Piccolo, B., Taverna, L., Berginc, M., and Ozbič, M. (2020). The complex relation between executive functions and language in preschoolers with developmental language disorders. *Int. J. Environ. Res. Public Health* 17:1772. doi: 10.3390/ijerph17051772

Mayer, M. (1969). Frog, where are you? New York, NY: Dial Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The Unity and Diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Mouzaki, A., Ralli, A. M., Antoniou, F., Diamanti, V., and Papaioannou, S. (2017). Logometro: A language assessment tool for preschool and early school years children. Athens: Intelearn.

Nilsson, K. K., and de López, K. J. (2016). Theory of mind in children with specific language impairment: a systematic review and Meta-analysis. *Child Dev.* 87, 143–153. doi: 10.1111/cdev.12462

Pauls, L. J., and Archibald, L. M. D. (2016). Executive functions in children with specific language impairment: a Meta-analysis. *J. Speech Lang. Hear. Res.* 59, 1074–1086. doi: 10.1044/2016_JSLHR-L-15-0174

Pearce, W. M., James, D. G. H., and McCormack, P. F. (2010). A comparison of oral narratives in children with specific language and non-specific language impairment. *Clin. Linguist. Phon.* 24, 622–645. doi: 10.3109/02699201003736403

Reese, E., Suggate, S., Long, J., and Schaughency, E. (2010). Children's oral narrative and reading skills in the first 3 years of reading instruction. *Read. Writ.* 23, 627–644. doi: 10.1007/s11145-009-9175-9

Renfrew, C. (1997). The bus story: A test of narrative language. Bicester: Winslow.

Rodríguez, V. A., Santana, G. M. R., and Expósito, S. H. (2017). Executive functions and language in children with different subtypes of specific language impairment. *Neurologia* 32, 355–362. doi: 10.1016/j.nrl.2015.12.018

Schaughency, E., Suggate, S., and Reese, E. (2017). Links between early oral narrative and decoding skills and later reading in a New Zealand sample. *Aust. J. Learn. Diffic.* 22, 109–132. doi: 10.1080/19404158.2017.1399914

Scheper, A., and Blankenstijn, C. (2013). Handleiding Frog Story Test. Eindhoven & Oegstgeest: Koninklijke Kentalis en Curium-LUMC.

Scionti, N., Zampini, L., and Marzocchi, G. M. (2023). The relationship between narrative skills and executive functions across childhood: a systematic review and Meta-analysis. *Children* 10:1391. doi: 10.3390/children10081391

Semel, E., Wiig, E., and Secord, W. (1995). Clinical evaluation of language fundamentals—3. London: Psychological Corporation.

Smolak, E., McGregor, K. K., Arbisi-Kelm, T., and Eden, N. (2020). Sustained attention in developmental language disorder and its relation to working memory and language. *J. Speech Lang. Hear. Res.* 63, 4096–4108. doi: 10.1044/2020_JSLHR-20-00265

Smit, L., Knoors, H., Rabeling-Keus, I., Verhoeven, L., and Vissers, C. (2022). Measuring theory of mind in adolescents with language and communication problems: An ecological perspective. *Front. Psychol.* 13:761434. doi: 10.3389/fpsyg.2022.761434

Stein, N., and Glenn, C. (1979). "An analysis of story comprehension in elementary school children" in New directions in discourse processing. ed. R. O. Freedle (New York, NY: Ablex), 53–120.

Tomas, E., and Vissers, C. (2019). Behind the scenes of developmental language disorder: Time to call neuropsychology back on stage. *Front. hum. neurosci.* 12:517. doi: 10.3389/fnhum.2018.00517

Urgesi, C., Campanella, F., and Fabbro, F. (2011). La Versione Italiana Della NEPSY-II per la Valutazione Neuropsicologica ad Ampio Raggio del Bambino da 3 a 16 Anni. Firenze: Giunti OS.

Usai, M., Viterbori, P., Gandolfi, E., and Traverso, L. (2017). E-PS 2–6: Batteria per la Valutazione delle Funzioni Esecutive in Età Prescolare. Trento: Erickson.

Vissers, C., Koolen, S., Hermans, D., Scheper, A., and Knoors, H. (2015). Executive functioning in preschoolers with specific language impairment. *Front. Psychol.* 6. doi: 10.3389/fpsyg.2015.01574

Vugs, B., Cuperus, J., Hendriks, M., and Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: a meta-analysis. *Res. Dev. Disabil.* 34, 2586–2597. doi: 10.1016/j.ridd.2013.05.014

Vugs, B., Hendriks, M., Cuperus, J., Knoors, H., and Verhoeven, L. (2017). Developmental associations between working memory and language in children with specific language impairment: a longitudinal study. *J. Speech Lang. Hear. Res.* 60, 3284–3294. doi: 10.1044/2017_JSLHR-L-17-0042

Wagner, R., Torgesen, J., and Rashotte, C. (1999). Comprehensive test of phonological processing (CTOPP). Austin, TX: Pro-ed.

Wechsler, D. (1993). WISC-R: Scala d'intelligenza Wechsler per Bambini Riveduta. Firenze: Giunti OS.

Wechsler, D. (2003). Wechsler intelligence scale for children–fourth edition (WISC-IV). London: Psychological Corporation.

Yang, H. C. (2015). Executive function in preschoolers with primary language impairment. A dissertation presented in partial fulfillment of the requirements for the degree doctor of philosophy. ProQuest number: 3718816. Arizona State University.

Yang, H. C., and Gray, S. (2016). Executive function in preschoolers with primary language impairment. *J. Speech Lang. Hear. Res.* 60, 379–392. doi: 10.1044/2016_JSLHR-L-15-0267

frontiers | Frontiers in Psychology

TYPE Mini Review
PUBLISHED 25 November 2024
DOI 10.3389/fpsyg.2024.1522168

OPEN ACCESS

EDITED BY
Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Omid Khatin-Zadeh,
University of Electronic Science and Technology of China, China

*CORRESPONDENCE
Zilong Zhong
✉ zhongzilong1106@bfsu.edu.cn

RECEIVED 04 November 2024
ACCEPTED 13 November 2024
PUBLISHED 25 November 2024

CITATION
Li Y and Zhong Z (2024) Visual insights into translation: demystifying trends of adopting eye-tracking techniques in translation studies. *Front. Psychol.* 15:1522168. doi: 10.3389/fpsyg.2024.1522168

Visual insights into translation: demystifying trends of adopting eye-tracking techniques in translation studies

Yin Li[1] and Zilong Zhong[2]*

[1]Nanyang Institute of Technology, Nanyang, China, [2]Beijing Foreign Studies University, Beijing, China

Introduction: The increasing use of eye-tracking techniques in translation studies offers valuable insights into cognitive processes and behavioral strategies of translators, reflecting a significant trend within cognitive linguistics and translator training methodologies.

Methods: This review harnesses quantitative bibliometric analysis through Bibliometrix R-package with qualitative content assessment to evaluate the trajectory and thematic evolution of eye-tracking research in translation studies. Through a dataset from the Web of Science, 56 articles were analyzed, revealing distinct thematic dimensions and trend dynamics.

Results: The analysis revealed that eye-tracking is increasingly pivotal in exploring the cognitive and technological dimensions of translation. Central themes include interactions with translation tools, machine translation, and human-computer interaction, highlighting the importance of cognitive research in technology-driven translation. Niche areas such as English-Chinese translation and online consultation suggest specialized topics that warrant further investigation. Additionally, emerging themes like cognitive load and sight translation demonstrate a shift toward exploring real-time translation processing. Declining traditional topics, such as broader translation theories, indicate a growing integration of cognitive research with technological advancements.

Conclusion: These findings elucidate the growth and diversification of eye-tracking applications in translation studies, emphasizing the method's importance in both academic research and practical applications, thereby informing future studies and enhancing translator training programs.

KEYWORDS

eye-tracking techniques, translation studies, machine translation, human-computer interaction, cognitive effort

1 Introduction

Eye tracking techniques have emerged as powerful tools for understanding cognitive processes across various domains, including in translation studies (Kornacki, 2019). These techniques, which include both screen-based and mobile eye trackers, measure eye positions and movements, providing valuable insights into how individuals allocate visual attention (Chang, 2009). In the context of translation, eye tracking allows researchers to explore cognitive load and attention distribution, shedding light on the mental strategies employed by translators during the translation process (Płużyczka, 2013). By elucidating these dynamics, eye tracking can significantly enhance translation training methodologies and improve overall

translation quality (Hansen, 2013). Previous studies have successfully investigated specific aspects of translation processes adopting eye tracking, such as attention distribution and reading strategies (Gambier and van Doorslaer, 2010). Despite the growing body of research employing eye-tracking techniques in translation studies, a notable gap exists in the form of review within this specialized area. This lack of an overview hampers a holistic understanding of the field's evolution, research trends, and emerging patterns.

The current review aims to address this gap by providing an analysis of translation studies that utilize eye-tracking techniques, covering the period from 1996 to 2024. By synthesizing findings from multiple studies, this research intends to map out current hotspots and identify emerging trends within the field. Utilizing Bibliometrix, an R-based bibliometric software package (Aria and Cuccurullo, 2017), this study analyzes 56 relevant articles to reveal insights into thematic areas and research trends. This methodological approach allows for providing a clearer understanding of the contributions and development of eye-tracking methodologies in translation research. By highlighting key trends and identifying gaps, the findings are expected to inform both scholars and practitioners in translation studies and cognitive linguistics.

2 Data collection

A comprehensive collection of published papers related to translation studies that employed eye-tracking techniques was gathered from the Web of Science (WoS) Core Collection. This collection utilized multiple components of the WoS Core Collection, specifically the Science Citation Index Expanded (SCI-Expanded), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), and the Emerging Sources Citation Index (ESCI). The search strategy implemented was: "TI = ('eye track*' OR 'eye movement*') AND ('translat*' OR 'interpret*')." Inclusion criteria were clearly defined, as outlined in Figure 1. Data collection occurred on September 30, 2024, resulting in a total of 56 records obtained from 22 journals across 36 WoS categories.

3 Results and discussion

3.1 Analysis of research themes

The thematic map of keywords (Figure 2) provides a framework for understanding the dynamics of eye-tracking techniques in translation studies. By utilizing the concepts of density and centrality, we can categorize the identified themes into four distinct quadrants, each representing a unique aspect of the research landscape. This categorization not only aids in identifying core themes but also sheds light on emerging trends and foundational concepts that warrant further exploration.

3.1.1 Motor themes

In the first quadrant, known as the motor themes, we find the most influential topics that exhibit high development and strong connections within the field. Key themes including "eye tracking," "machine translation," and "human-computer interaction (HCI)" occupy this space, indicating their pivotal role in advancing the discourse surrounding translation processes. The presence of terms like "acceptability" and "usability" further emphasizes the practical implications of these core themes, highlighting the importance of user-centered approaches in the application of eye-tracking techniques. As a foundational technique in translation studies, "eye tracking" has become a motor theme due to its significant development and integration into various research paradigms (Kruger, 2013). It provides empirical data on cognitive processes during translation, offering insights into how translators interact with text. The robustness of this theme indicates its central role in understanding translation behaviors and strategies. "Machine translation" reflects the intersection of technology and translation practices (Kasperavičienė et al., 2020). With advances in artificial intelligence (AI) and natural language processing, machine translation has garnered attention for its implications on translation quality and efficiency. The strong connections to other themes underscore its relevance in contemporary translation studies, especially as researchers explore user interactions with machine-generated outputs. The prominence of "HCI" in the thematic map highlights the growing recognition of user experience in translation technologies (Doherty and O'Brien, 2014). This theme is crucial for examining how translators engage with digital tools and the cognitive implications of such interactions. As translation increasingly relies on technology, understanding HCI becomes essential for improving usability and acceptability of translation systems. "Acceptability" pertains to the perceived quality and appropriateness of translations, particularly in the context of machine-assisted tools (Cui and Zheng, 2021). Its presence among motor themes indicates a maturing discourse around the criteria by which translations are evaluated. Acceptability connects with cognitive load and usability, emphasizing the need for translators to critically assess the outputs generated by automated systems. Closely related to acceptability, "usability" focuses on the effectiveness, efficiency, and satisfaction with which users can achieve their goals in using translation tools (Doherty and O'Brien, 2014). This theme underscores the importance of user-centered design in translation technologies. A robust understanding of usability can inform the development of eye-tracking studies aimed at improving translator workflows. These motor themes suggest a well-established nexus of research that not only informs current practices but also guides future inquiries into the intricacies of translation.

3.1.2 Niche themes

The second quadrant encapsulates niche themes that, while highly developed, remain somewhat isolated from broader discussions. "English-Chinese translation" and "online consultation" are significant here, indicating specialized areas of focus that demonstrate maturity but may not yet be fully integrated into the larger research framework. "English-Chinese translation" suggests a specialized focus on a particular language pair, reflecting its unique challenges and contexts in translation studies (Cui and Zheng, 2021). Despite its high development, its isolation from broader themes may indicate a need for more integration into the general discourse on translation practices. Researchers could explore cross-linguistic comparisons or the application of eye tracking in understanding specific cultural nuances. "Online consultation" encompasses the dynamics of remote collaboration in translation processes, particularly in the context of professional settings (Cui and Zheng, 2022). Its development suggests that online tools are becoming

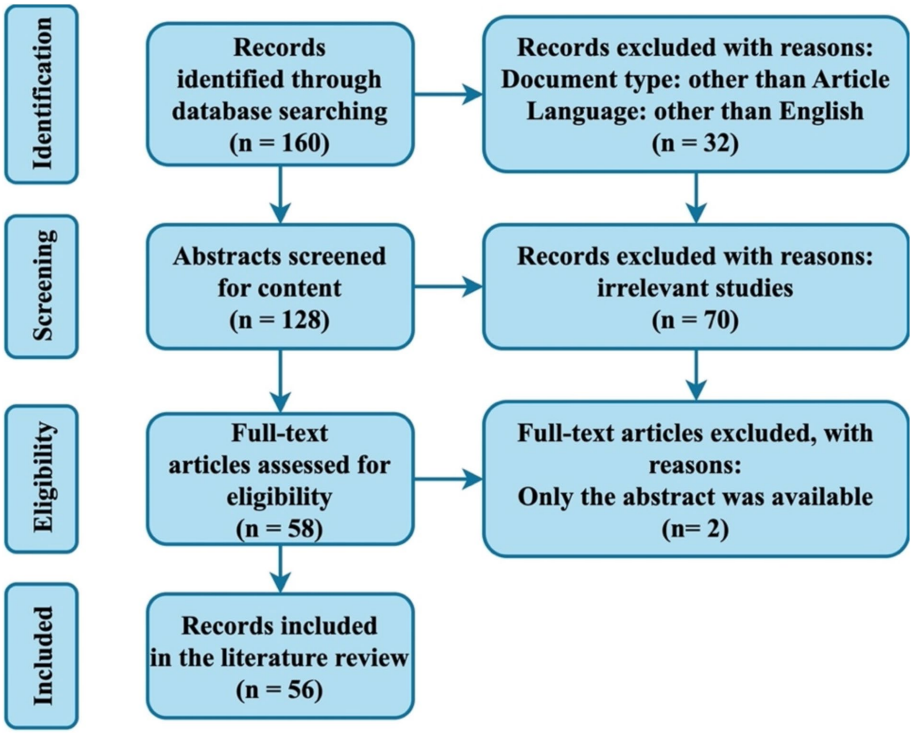


FIGURE 1
PRISMA flow diagram.

increasingly significant in translation practice. However, its niche status implies that further exploration is required to understand how online environments impact cognitive processes and decision-making in translation. The isolation of these themes suggests potential opportunities for interdisciplinary collaboration and the exploration of their intersections with broader topics in eye tracking and translation. Researchers might consider investigating how these niche areas can inform and enhance the understanding of user interactions within machine translation systems.

3.1.3 Emerging or declining themes

The third quadrant reveals emerging or declining themes, including "translation process research" and "cognitive effort." These themes indicate a transitional phase in the research landscape, where certain concepts are gaining attention, while others may be losing relevance. Positioned in the emerging or declining quadrant, "translation process research" reflects an ongoing evolution in understanding how translations are produced (Su and Li, 2019). While it has gained traction, there may be fluctuations in its emphasis within the field. Future research could delve into the cognitive mechanisms involved in real-time translation processes, potentially integrating eye-tracking methodologies to uncover nuanced behaviors. "Cognitive effort" indicates a growing interest in the mental demands placed on translators during the translation process (Schaeffer et al., 2019). As researchers increasingly consider cognitive aspects, this theme's position suggests that it may be in a transitional phase, possibly overshadowed by more technologically oriented discussions. Exploring the relationship between cognitive effort and eye-tracking data could yield valuable insights into translator performance and efficiency. Researchers should remain vigilant to shifts in this quadrant, as they often signal changing priorities within the field and potential areas for innovation.

3.1.4 Basic themes

The fourth quadrant is characterized by basic themes that, despite being lowly developed, hold the potential to emerge as future research hotspots. Themes such as "eye movements," "fixation," "cognitive load," and "sight translation" represent foundational concepts that are crucial for understanding the mechanisms underlying translation processes. While they may currently lack the robust connections seen in motor themes, their basic nature suggests that they are ripe for exploration and could form the basis for new lines of inquiry. "Eye movements" is foundational for understanding how visual processing impacts translation (Shaikh et al., 2005). Although currently low in development, it presents a rich area for future research. Investigating eye movements in various translation contexts could illuminate patterns that inform both theory and practice. Similar to eye movements, "cognitive load" is a critical concept in translation studies (Ma et al., 2021). Its basic status indicates that while it is recognized, it may not yet be fully exploited in empirical research. Exploring cognitive load through eye-tracking studies could enhance our understanding of translator stress and efficiency. "Sight translation" relates to the instantaneous translation of spoken or written texts without prior preparation (Su and Li, 2019). As a basic theme, it suggests an area ripe for further investigation, particularly in how eye-tracking can reveal cognitive processes involved in sight translation tasks. The theme of "reading" is fundamental to all translation activities. Its basic nature reflects its essential role in the translation process, yet it may require deeper integration with more applied studies. Researching reading patterns

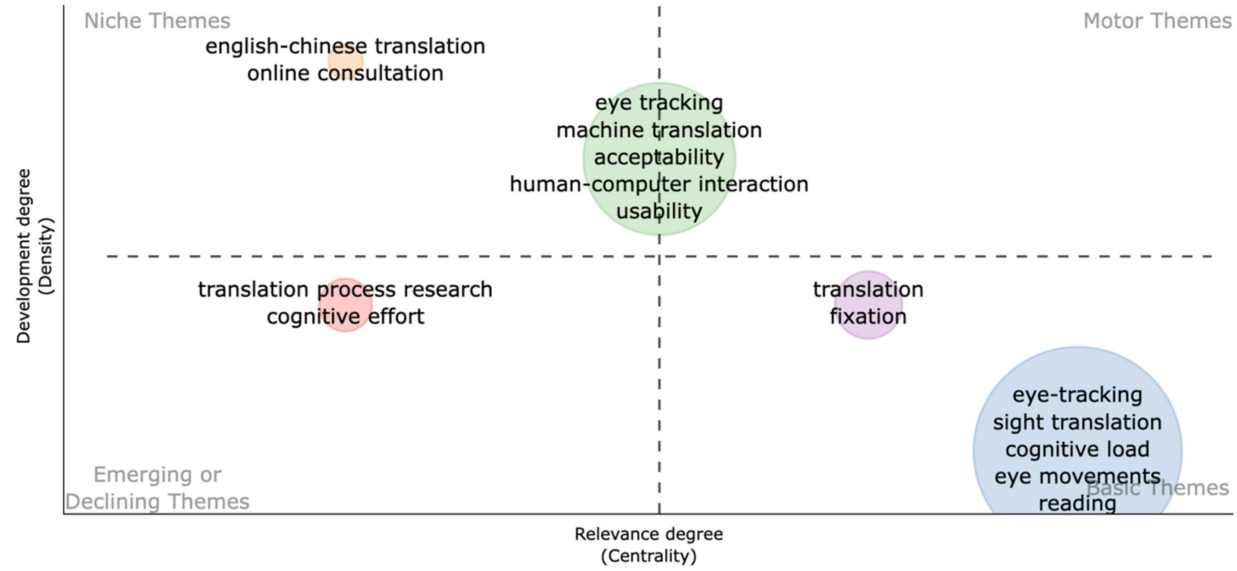


FIGURE 2
Thematic map of keywords in translation studies employing eye-tracking techniques.

through eye-tracking could provide insights into how translators process information in real-time (Lim and Christianson, 2015). As a broad and foundational theme, "translation" represents the core of the field. However, its low development status suggests that there is a need for more specific investigations that can connect traditional translation theories with empirical data gathered through eye-tracking methods. "Fixation" points during reading and translation tasks reveal where attention is directed. As a basic theme, this concept offers a fundamental understanding of visual attention in translation (Kruger, 2013). Investigating fixation patterns can lead to better insights into cognitive strategies employed by translators. Researchers should consider how advancing techniques in eye tracking can illuminate these basic themes and lead to new insights in translation studies.

The findings of the research theme analysis suggest that eye-tracking techniques are poised to play a pivotal role in advancing both the cognitive and technological dimensions of translation studies. The prominence of motor themes, such as eye tracking, machine translation, and HCI, indicates that these areas have become central to the discourse, highlighting the need for continued research on user interactions with translation tools and their cognitive impacts. Meanwhile, niche themes like English-Chinese translation and online consultation point to specialized areas that could benefit from further integration with broader research on translation processes and technology. Emerging and declining themes suggest a shifting focus toward understanding cognitive effort and the translation process, though these areas may require renewed attention to fully explore their potential. Lastly, basic themes related to eye movements, fixation, and cognitive load offer fertile ground for future research, particularly in relation to sight translation and real-time cognitive processing. Overall, these findings underscore the necessity of a more holistic approach that combines cognitive research with advancements in translation technology, paving the way for more effective, user-centered tools and deeper insights into the mental strategies underlying translation tasks.

3.2 Analysis of trend topics

The exploration of trend topics in translation studies incorporating eye-tracking methodologies illustrates an evolution of research interests and thematic shifts over time (see Figure 3). This analysis provides a comprehensive overview of the changing dynamics within the field, identifying key areas of sustained focus, emerging interests, and potential declines in specific research topics.

3.2.1 Sustained and expanding interest

"Eye-tracking" stands out as the dominant research topic in the field, with significant growth and sustained interest, particularly from 2020 onwards. With a notable presence in 22 publications, the use of eye-tracking techniques in translation studies has gained momentum, reaching its peak median mentions in 2021 and continuing to expand through 2023. This indicates that eye-tracking has evolved from a specialized method to a central tool in translation research, enabling a deeper understanding of cognitive processes, attention distribution, and decision-making strategies during translation tasks.

The sustained focus on eye-tracking highlights both the increasing availability and sophistication of eye-tracking technologies and their growing applicability to various facets of translation. Researchers have leveraged eye-tracking to explore complex translation phenomena such as translator training, workflow optimization, and cognitive effort, which were previously harder to quantify. The expanding scope reflects advancements in the technology itself, such as the development of mobile eye-tracking devices and more powerful data analysis techniques, allowing for more nuanced and sophisticated studies that address real-world translation scenarios. As translation practices continue to evolve, particularly with the advent of machine translation and other assistive technologies, eye-tracking might become one of the key tools in examining the cognitive dynamics at play in these contexts.

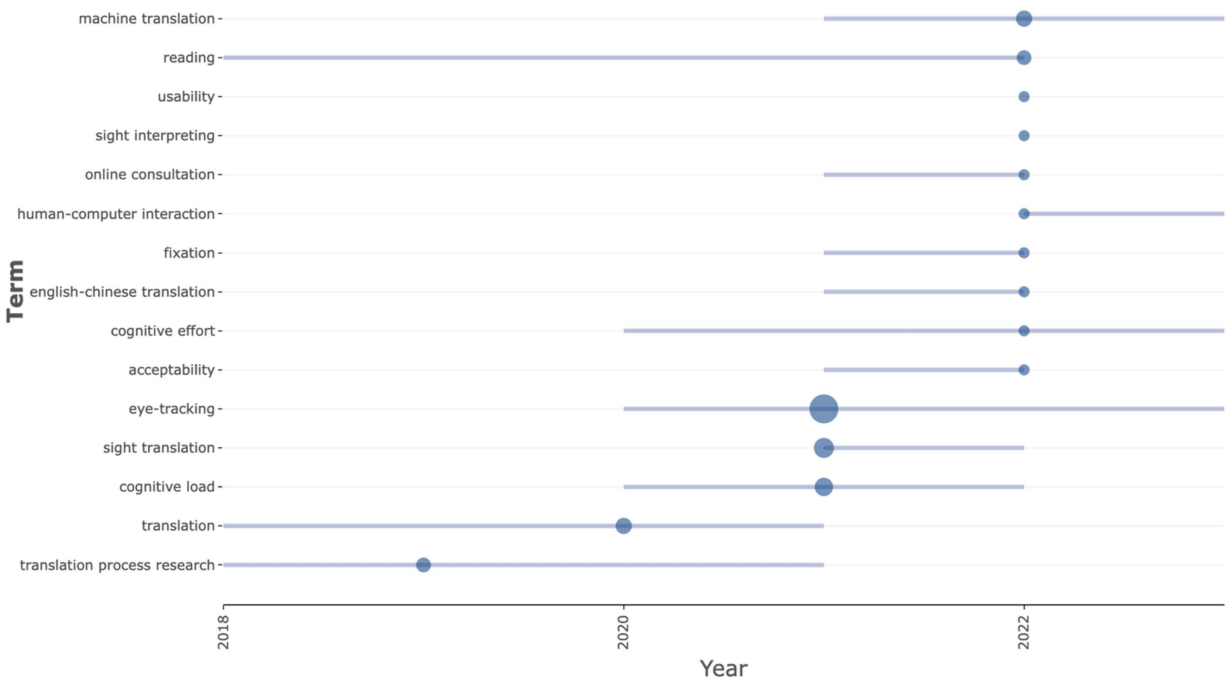


FIGURE 3
Trend topics by keywords of translation studies adopting eye-tracking techniques.

3.2.2 Niche and specialized topics

Two specialized topics, "sight translation" and "cognitive load," have attracted significant attention within the broader field of eye-tracking in translation studies. Both have shown a peak in activity around 2021, suggesting a phase of consolidation in research.

Sight translation, which involves translating orally or writing down a text without prior preparation, has become a focal point for exploring the real-time cognitive processes involved in simultaneous translation tasks. This topic has garnered increasing attention, with eight publications dedicated to the subject, reflecting a growing recognition of its importance in both educational and professional settings. Research in this area has yielded significant insights into how translators manage rapid decision-making, prioritize information, and allocate cognitive resources during sight translation tasks. The continued interest in sight translation is indicative of a broader trend toward examining high-pressure translation tasks that require quick thinking and cognitive flexibility. Cognitive load, another critical area within the field, peaked in 2021 as well. Cognitive load research explores the mental effort involved in translation and how it impacts performance and accuracy. The slight decline in its activity after 2021 suggests that foundational theories related to cognitive load are now well-established, and the field is shifting toward more applied or integrated studies. Specifically, researchers are increasingly interested in examining how cognitive load affects real-world translation tasks, particularly in complex or high-stress scenarios. Although the research focus on cognitive load has slightly diminished, its theoretical implications remain crucial in understanding translator efficiency and mental fatigue.

3.2.3 Emerging research frontiers

"Machine translation" and "HCI" are two emerging topics that are gaining traction in the field of translation studies. Both themes highlight the growing intersection between technology and translation, marking a shift toward more digitally integrated research.

Machine translation has emerged as a key area of focus, with a notable increase in publications exploring how automated translation tools influence the translation process. Appearing in four publications, machine translation studies examine how translators interact with and assess machine-generated outputs, as well as the cognitive and psychological effects of relying on these tools. As machine translation systems improve through AI and machine learning, the need to understand how translators engage with these systems becomes increasingly important. Eye-tracking allows researchers to explore real-time user interaction with machine translations, providing insights into how machine-generated translations affect decision-making, cognitive load, and overall translation quality. HCI in translation has also seen a rise, with two publications in recent years. The focus on HCI reflects an increasing interest in understanding how translators interact with translation tools and digital interfaces. As translation technologies become more sophisticated, understanding how users engage with these systems is crucial for optimizing workflows, improving usability, and minimizing cognitive load. Eye-tracking can provide valuable data on how translators navigate translation interfaces, track their gaze patterns, and assess the usability of machine-assisted translation tools. This growing interest in HCI underscores the broader trend toward user-centered design in translation technologies and the need for tools that facilitate the cognitive processes involved in translation.

3.2.4 Consolidation and decline

Some traditional topics, such as "translation process research" and "translation" itself, have shown a peak in research activity in earlier years, followed by a noticeable decline in recent mentions. This shift

indicates a broader pivot within the field from foundational process studies to more technologically driven research, which integrates cognitive theories with the tools and methods of modern translation practices.

Translation process research has been a longstanding area of interest, exploring how translators work through texts, manage decision-making, and engage in cognitive processing. However, as more specific and technology-driven research areas (such as eye-tracking in machine translation and HCI) gain momentum, there has been a reduction in the focus on traditional translation process research. The decline suggests that many of the foundational questions about translation processes have been addressed, and future research may focus more on integrating traditional theories with emerging technologies. Similarly, the theme of translation as a broad concept, while central to the field, has seen a decrease in research activity in recent years. This decline likely reflects a saturation of fundamental studies and a shift toward more specialized investigations that incorporate new technologies or interdisciplinary approaches. The decline in the focus on "translation" as a topic suggests that scholars may now be focusing on more specific aspects of the translation process, such as cognitive load, machine translation, and HCI, rather than revisiting broader theoretical discussions.

3.2.5 Recent interest peaks

Several topics, including "reading," "acceptability," "English-Chinese translation," "fixation," and "online consultation," have shown a peak in activity around 2022. These peaks suggest that certain research areas are gaining traction due to new methodologies or specific shifts in industry or educational needs.

Reading has long been a foundational component of translation studies, but recent research has started to delve deeper into how reading patterns and eye movements can shed light on the cognitive dynamics of translation. This growing interest reflects a broader trend toward incorporating empirical data into reading processes, especially in real-time translation tasks. Acceptability studies, which focus on the quality and appropriateness of translations, have become increasingly important. This rise may be linked to the need for higher translation quality, especially in professional and public-facing contexts, such as media, localization, and legal translation. Eye-tracking can provide insights into how translators assess acceptability during the translation process and how cognitive load affects their judgment. English-Chinese translation remains a prominent topic, reflecting the particular challenges posed by this language pair and its importance in the global translation market. Research in this area highlights the unique cognitive processes involved in translating between languages with vastly different syntactic, lexical, and cultural structures. Fixation and online consultation have also seen increased interest, with fixation studies focusing on visual attention patterns during translation tasks, and online consultation reflecting the growing demand for remote translation services. The increase in interest in these topics may be linked to the rise of digital platforms and online collaboration tools that are reshaping the translation industry, necessitating further research into how these factors influence cognitive processing and decision-making.

The findings of trend topic analysis suggest a significant shift toward a more integrated, technology-driven approach to translation studies. The sustained focus on eye-tracking, coupled with emerging themes like machine translation and HCI, highlights the growing importance of cognitive research in understanding how translators engage with digital tools. As eye-tracking technologies continue to evolve, they offer valuable insights into real-time decision-making, cognitive load, and the impact of machine-generated translations, which are crucial for improving translation efficiency and quality. The decline in traditional topics such as translation process research and broader concepts of translation suggests that foundational theories are being increasingly integrated with cutting-edge technologies, marking a pivot toward applied, interdisciplinary studies. Moreover, the rise of specialized topics, such as sight translation and cognitive load, points to a more nuanced understanding of the mental and cognitive demands of translation tasks. These trends indicate that future research will likely continue to focus on the intersection of cognition and technology, with an emphasis on user-centered design and the cognitive dynamics involved in real-world translation practices.

4 Future directions

Given the dynamic evolution and the distinct trends identified in the analysis of research themes and trend topics, as well as their temporal distribution in translation studies employing eye-tracking techniques, several promising directions for future research can be proposed (see Table 1). These suggestions aim to build on current strengths, address emerging gaps, and anticipate evolving needs within the field. (1) Advanced integration of eye-tracking with machine translation systems. Future studies could explore deeper integrations of eye-tracking technologies with machine translation tools to reveal how translators interact with AI-generated translations. Research could focus on the cognitive load experienced by translators and how it affects translation quality and efficiency, aiming to optimize human-AI collaboration in translation workflows. (2) Real-time eye-tracking in HCI. With the growing interest in HCI, further research could investigate how real-time eye-tracking data can enhance the usability and design of translation software interfaces. Studies could examine how translators navigate different interface designs and how these impact their cognitive processes and translation strategies. (3) Eye-tracking in sight translation. While sight translation shows a recent peak in research activity, further exploration into how eye movements correlate with cognitive strategies during sight translation could provide deeper insights. This could include studies across different language pairs, particularly those involving complex syntactic structures or scripts. (4) Sector-specific online consultation studies. With online consultation emerging as a key area, future research could focus on specialized contexts such as legal, medical, or technical translations. Eye-tracking could be employed to study how consultants handle on-demand translation, and the cognitive efforts involved, potentially improving remote translation services. (5) Longitudinal studies on cognitive effort and fatigue. Given the importance of cognitive effort and emerging concerns about translator fatigue, longitudinal studies using eye-tracking could assess how cognitive load varies over extended periods and under different working conditions. Such research might lead to better management strategies for cognitive resources in professional translation settings. (6) Cross-linguistic studies on fixation and reading patterns. Research could expand into comparative studies of eye movements across different linguistic systems, examining how script type (e.g., alphabetic vs. logographic) affects reading strategies and fixation patterns during translation tasks.

TABLE 1 Research gaps and future directions.

	Research gaps	Future directions
1	Limited integration of eye-tracking with machine translation systems.	Advanced integration of eye-tracking with machine translation tools to study cognitive load and its effects on translation quality and efficiency.
2	Underexplored use of real-time eye-tracking in HCI for translation software.	Investigate how real-time eye-tracking data can enhance usability and design of translation software interfaces, examining translators' navigation and cognitive processes.
3	Insufficient research on eye movements during sight translation, especially for complex language pairs.	Explore the correlation between eye movements and cognitive strategies in sight translation, focusing on different language pairs with complex syntactic structures or scripts.
4	Lack of studies on eye-tracking in sector-specific translation contexts (e.g., legal, medical, technical).	Study eye-tracking in specialized translation contexts such as legal, medical, or technical translations, focusing on cognitive efforts in on-demand translation tasks.
5	Limited longitudinal research on cognitive effort and translator fatigue.	Conduct longitudinal studies using eye-tracking to assess cognitive load over extended periods and under various conditions, aiming to develop better fatigue management strategies.
6	Scarcity of cross-linguistic studies on eye movements and reading patterns.	Conduct cross-linguistic studies examining how script types (e.g., alphabetic vs. logographic) affect eye movements, reading strategies, and fixation patterns during translation tasks.
7	Lack of research on the use of eye-tracking in translation education.	Assess the effectiveness of eye-tracking in teaching and assessing translation skills, exploring its correlation with student performance and learning outcomes.
8	Need for more studies on acceptability and quality control in translation.	Use eye-tracking to investigate how different translation choices affect reader perceptions and satisfaction, contributing to the development of quality control frameworks.

This could inform targeted training programs to enhance reading efficiency and translation accuracy. (7) Eye-tracking in translation education. Future studies could assess the effectiveness of eye-tracking as a tool in educational settings, examining how it can be used to teach and assess translation skills. Research could explore how eye-tracking data correlates with student performance and learning outcomes. (8) Acceptability and quality control. As acceptability becomes increasingly critical, especially in high-stakes translations, further research using eye-tracking could explore how different translation choices impact reader perceptions and satisfaction. Such studies could help develop more rigorous quality assessment frameworks for translations.

5 Conclusion

This review has charted the evolution and integration of eye-tracking techniques in translation studies, highlighting significant strides and identifying fertile areas for future research. The analysis confirms the application of eye-tracking across a spectrum of themes, from the well-established, such as "eye tracking" and "machine translation," to more isolated yet mature topics like "English-Chinese translation." It is evident that eye-tracking technology has not only enriched our understanding of the cognitive and linguistic processes in translation but also influenced translation training and practice. The thematic maps and trend analyses point to a dynamic field where traditional processes are increasingly interfaced with cutting-edge technology, underlining the potential for eye-tracking to enhance the efficacy and quality of translation outputs. Furthermore, the emergence of niche and specialized themes suggests a gradual but noticeable shift toward addressing specific linguistic pairings and settings, indicating a maturation within the field that could lead to more targeted and context-specific research. Meanwhile, the rise of themes related to technology use in translation, such as "HCI" and "machine translation," signals a trend toward digital transformation in the field. These insights not only guide academic research but also impact practical applications, ensuring that translation practices evolve to meet contemporary demands.

In conclusion, the review demonstrates the expansive utility of eye-tracking in translation studies and sets the stage for its continued evolution as a pivotal research tool. By shedding light on both dominant and emergent themes, this review not only reflects the current landscape but also serves as a call to action for researchers to explore underrepresented areas. As the field progresses, it is crucial to leverage these insights to refine translation methodologies and to harness the full potential of technology in translation studies, thereby enhancing both theoretical knowledge and practical applications in this interdisciplinary domain.

Author contributions

YL: Writing – original draft, Writing – review & editing. ZZ: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Fundamental Research Funds for the Central Universities (2024JX063).

Acknowledgments

We would like to express our sincere gratitude to the reviewer for their valuable feedback and insightful suggestions, which significantly improved the quality of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aria, M., and Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *J. Informet.* 11, 959–975. doi: 10.1016/j.joi.2017.08.007

Chang, C. Y. (2009). Testing applicability of eye-tracking and fMRI to translation and interpreting studies: An investigation into directionality (Doctoral dissertation, Imperial College, London).

Cui, Y., and Zheng, B. (2021). Consultation behaviour with online resources in English-Chinese translation: an eye-tracking, screen-recording and retrospective study. *Perspectives* 29, 740–760. doi: 10.1080/0907676X.2020.1760899

Cui, Y., and Zheng, B. (2022). Extralinguistic consultation in English–Chinese translation: a study drawing on eye-tracking and screen-recording data. *Front. Psychol.* 13:891997. doi: 10.3389/fpsyg.2022.891997

Doherty, S., and O'Brien, S. (2014). Assessing the usability of raw machine translated output: a user-centered study using eye tracking. *Int. J. Hum. Comp. Interact.* 30, 40–51. doi: 10.1080/10447318.2013.802199

Gambier, Y., and Van Doorslaer, L. (2010). Handbook of translation studies, vol. *1*. Amsterdam: John Benjamins Publishing Company.

Hansen, G. (2013). "The translation process as object of research" in *The routledge handbook of translation studies*. eds. C. Millán and F. Bartrina (Abingdon, England: Routledge), 88–101.

Kasperavičienė, R., Motiejūnienė, J., and Patašienė, I. (2020). Quality assessment of machine translation output: cognitive evaluation approach in an eye tracking experiment. *Texto livre* 13, 271–285. doi: 10.35699/1983-3652.2020.24399

Kornacki, M. (2019). The application of eye-tracking in translator training. Available at: https://www.intralinea.org/specials/article/2421 (Accessed December 20, 2020).

Kruger, H. (2013). Child and adult readers' processing of foreignised elements in translated south African picturebooks: an eye-tracking study Target. *Int. J. Transl. Stud.* 25, 180–227. doi: 10.1075/target.25.2.03kru

Lim, J. H., and Christianson, K. (2015). Second language sensitivity to agreement errors: evidence from eye movements during comprehension and translation. *Appl. Psycholinguist.* 36, 1283–1315. doi: 10.1017/S0142716414000290

Ma, X., Li, D., and Hsu, Y. Y. (2021). Exploring the impact of word order asymmetry on cognitive load during Chinese–English sight translation: evidence from eye-movement data. *Targets* 33, 103–131. doi: 10.1075/target.19052.ma

Płużyczka, M. (2013). "Eye-tracking supported research into sight translation. Lapsological conclusions," in *Translation studies and eye-tracking analysis*. eds. S. Grucza, M. Pluzycka, and J. Alnajjar (Berlin, Germany: Peter Lang Verlag), 105–138.

Schaeffer, M., Nitzke, J., Tardel, A., Oster, K., Gutermuth, S., and Hansen-Schirra, S. (2019). Eye-tracking revision processes of translation students and professional translators. *Perspectives* 27, 589–603. doi: 10.1080/0907676X.2019.1597138

Shaikh, A. G., Ghasia, F. F., Dickman, J. D., and Angelaki, D. E. (2005). Properties of cerebellar fastigial neurons during translation, rotation, and eye movements. *J. Neurophysiol.* 93, 853–863. doi: 10.1152/jn.00879.2004

Su, W., and Li, D. (2019). Identifying translation problems in English-Chinese sight translation: an eye-tracking experiment. *Transl. Interpret. Stud.* 14, 110–134. doi: 10.1075/tis.00033.su

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 08 January 2025
DOI 10.3389/fpsyg.2024.1471108

OPEN ACCESS

EDITED BY
Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Ozge Razi,
Cyprus International University, Cyprus
Amal Zakaria Hal,
Port Said University, Egypt

*CORRESPONDENCE
Yuqing Chen
✉ neuwy2021@163.com

RECEIVED 26 July 2024
ACCEPTED 11 December 2024
PUBLISHED 08 January 2025

CITATION
Hui X and Chen Y (2025) The mediating role of English learning motivation between socioeconomic status and pragmatic awareness.
Front. Psychol. 15:1471108.
doi: 10.3389/fpsyg.2024.1471108

The mediating role of English learning motivation between socioeconomic status and pragmatic awareness

Xiaomeng Hui[1] and Yuqing Chen[2]*

[1]College of Foreign Studies, Northeastern University, Shenyang, China, [2]School of Foreign Languages, Southeast University, Nanjing, China

Socioeconomic status (SES) has received great attention in learning a second or foreign language (SL/FL). However, little research has investigated the association between SES and SL/FL pragmatic learning, let alone the influencing pathways of SES on SL/FL pragmatic awareness (PA). Therefore, this research aimed to address the link between learners' SES and PA with the mediating effects of learning motivation based on the L2 motivation self-system (L2MSS) theory by surveying 292 Chinese EFL university students. Structural equation modeling analyses indicated that: (1) SES had no significant effect on ought-to L2 self and intended learning efforts; (2) ought-to L2 self and intended learning efforts had significant predictive effects on PA; (3) SES positively and directly predicted EFL learners' PA; and (4) ideal L2 self, attitudes toward L2 community, and attitudes toward learning English mediated the relationship between SES and Chinese EFL learners' PA significantly.

KEYWORDS

L2 motivational self-system (L2MSS), pragmatic awareness (PA), socioeconomic status (SES), motivation, ideal L2 self

1 Introduction

The importance of pragmatic awareness (PA) has been increasingly acknowledged in recent years as it is an important part of pragmatic competence (Ren, 2015; Yang and Ren, 2020), which in turn determines the success or failure of human interaction (García-Gómez, 2022; Yang, 2022). The development of foreign or second language pragmatic awareness requires the support of motivation to assist students in noticing target pragmatic norms and recognizing pragmatic inappropriateness or errors of an utterance (Takahashi, 2005; Chiravate, 2012; Yamato et al., 2013; Yang and Ren, 2020). Yang and Ren (2020) elaborated that students' L2 motivation influences their pragmatic awareness, while the potential effect of motivation on pragmatic awareness in SLA has been largely underrepresented (Taguchi and Roever, 2017; Yang and Ren, 2020). Furthermore, as the antecedent of motivation, socioeconomic status (SES) has been demonstrated to play a prominent role in second/foreign language (SL/FL) learning. That is, socioeconomically privileged students tend to have more access to learning environments and resources (Butler, 2014; Ghorbani and Golparvar, 2020), gain more effective language input (Huang et al., 2018), are more willing to make efforts (Shin and So, 2018), display a higher level of motivation (Lee and Lee, 2023) and self-efficacy (Kormos and Kiddle, 2013), are more adapted to autonomous language learning with technology (Ghorbani and Golparvar, 2020), and use more cognitive, meta-cognitive, compensatory, and social strategies (Shin and So, 2018) in SL/FL learning.

Compared with research on influencing factors of pragmatic awareness in viewing learners as a homogeneous group, research focusing on the impact of individual differences on learners' pragmatic awareness has gained great attention recently (Yang and Ren, 2020; Wang and Ren, 2023). SL/FL learning happens under the influence of a wider range of internal and external factors, such as learners' motivation, learning environment, and learning resources (Oxford, 2016; Mercer, 2018; Paradowski and Jelińska, 2024), which are more accessible to learners from higher SES families (Butler, 2014; Ghorbani and Golparvar, 2020; Lee and Lee, 2023) and contribute to their SL/FL pragmatic awareness (Xu et al., 2009; Yang and Ren, 2020). Although it is widely accepted that SES is correlated with SL/FL pragmatic competence, much less is known about the pathways by which SES exerts its influence on pragmatic awareness in SL/FL learning.

Besides, research in the FL learning context has shown that the effect of SES on learning outcomes is indirect and mediated by other variables (e.g., Huang et al., 2018; Lee and Lee, 2023), calling to attach importance to the mediating role of motivation in SES and FL learning outcomes (Huang et al., 2018). Indeed, SES was discovered to have significant effects on English learners' motivation (Lamb, 2012; Kormos and Kiddle, 2013; Butler, 2015; Butler and Le, 2018), which has been found to influence foreign language learning to notice of pragmatic strategy (Takahashi, 2005) and awareness (Chiravate, 2012; Yang and Ren, 2020). However, there is currently no research directly exploring the relationship between SES, motivation, and pragmatic awareness in foreign language learning contexts, disclosing a necessary field of investigation yet to be explored.

The current study aims to address these gaps by exploring the structural relations between SES and pragmatic awareness among Chinese EFL learners and investigating the possible mediating role of their motivation between SES and pragmatic awareness. The research questions are: (1) Does EFL students' SES directly affect their motivation? (2) Is there a direct effect of EFL students' motivation on their pragmatic awareness? (3) Is there a direct effect of EFL students' SES on their pragmatic awareness? (4) Does EFL students' learning motivation significantly mediate the relationship between their SES and pragmatic awareness? The present study is expected to expand the effect of SES on pragmatic awareness from the first language to the second or foreign language field and clarify the relationship between SES, motivation, and pragmatic awareness, as well as the mediating role of motivation variables. In addition, the present study can reveal how EFL students with different types of motivation (e.g., ideal L2 self vs. ought-to L2 self) handle their English pragmatic learning, which will provide information for English pragmatic pedagogy.

2 Literature review

2.1 SES and motivation

A student's SES refers to an individual or a group's ranking in social hierarchy according to some valued commodities they accessed, like wealth, power, and social status (Mueller and Parcel, 1981; Sirin, 2005), which is often operationalized as a combination of their parents' educational background, parents' occupation, and family income in educational research (e.g., Fan, 2011; Ensminger and Fothergill, 2014; Whitney and Bergin, 2018; OECD, 2019; Xu et al., 2021; Zheng and Mei, 2021; Ma et al., 2022). It has been widely recognized that SES plays an important role in students' academic achievement (Pace et al., 2017; Nikolov and Csapó, 2018; Sanjurjo et al., 2018; Chen et al., 2021; Luo et al., 2021; Liu et al., 2023; Sanfo and Malgoubri, 2023), whereas research on the association between SES and students' learning process and outcome in the FL context is quite recent (Butler and Le, 2018; Huang et al., 2018; Ghorbani and Golparvar, 2020). Most studies have showcased the positive role of students' SES background in foreign language learning (e.g., Csapó and Nikolov, 2009; Butler and Le, 2018; Shin and So, 2018; Huang, 2022; Ma et al., 2022). Specifically, high SES contributes to improving students' achievement in foreign language learning, which is achieved by influencing students' language input (Huang et al., 2018), parental educational behavior (Butler, 2014), language learning with technology (Ghorbani and Golparvar, 2020), cognitive skills (Liu et al., 2020), and motivation (Lee and Lee, 2023). In summary, SES background positively contributes to the process and outcomes of foreign language learning, and the positive impact on outcomes is generally achieved indirectly through other factors such as motivation (Huang et al., 2018).

Language learning motivation greatly impacts SL/FL learning (Yang and Ren, 2020; Vonkova et al., 2021; Jia and Cheng, 2022). It has been widely acknowledged that learners with high levels of motivation perform better in second/foreign language acquisition than learners with low levels of motivation (Dörnyei, 2005; Papi, 2018; Gong et al., 2020; Sudina, 2021; Wang and Liu, 2022; Xu et al., 2022; Li and Han, 2024). Extensive research has explored the influencing factors of foreign language learning motivation, such as buoyancy (Jia and Cheng, 2022), SES (Kormos and Kiddle, 2013; Butler, 2015; Lee and Lee, 2023), self-efficacy (You et al., 2016), social support (Papi and Hiver, 2020; Trigueros et al., 2020; Jia and Cheng, 2022), and teaching approaches (Liu and Lan, 2016; Önal et al., 2019). Moreover, research has documented the predictive role of motivation in learners' attitudes (Huong et al., 2017), emotions (Saito et al., 2018), willingness to communicate (Lin, 2019), engagement (Oga-Baldwin et al., 2017; Li and Han, 2024), and pragmatic awareness (Yang and Ren, 2020) in foreign language learning. However, the specific effect of motivation in the link between learners' SES and pragmatic awareness in foreign language learning is still unclear.

The predictive role of learners' socioeconomic background in foreign language learning motivation has been supported by previous studies (Muñoz, 2008; Lamb, 2012; Kormos and Kiddle, 2013; Butler, 2015, 2017; Shin and So, 2018; Lee and Lee, 2023). Among multiple motivational variables, the ideal L2 self has been investigated as a crucial motivational factor closely related to social class. Lamb (2012) concluded the ideal L2 self is an important factor only for the urban population. Specifically, due to the lack of role models or social contact with respected others, rural learners are less likely to develop strong possible self-images, resulting in holding less favorable views of their ideal selves than urban learners (Lamb, 2012; Kormos and Kiddle, 2013). Besides, empirical studies have reported a positive association between SES and English learning motivation among primary (e.g., Butler, 2015), secondary (e.g., Kormos and Kiddle, 2013), and college school students (e.g., Lee and Lee, 2023) and inferred that the impact of family resources on students' English learning motivation varies depending on grades (Butler, 2015, 2017). Specifically, a higher family socioeconomic status is more conducive to the development of

children's motivation in English learning as the grade level increases. Hence, SES may have a strong influence on university students' motivation to learn foreign languages.

> *H1:* SES positively predicts EFL learners' motivational variables.

2.2 L2 motivational self-system

Dörnyei (2005) L2 motivation self-system (L2MSS) is one of the leading theories widely used in the study of English learning motivation (Yousefi and Mahmoodi, 2022). This framework considers learners' future self-images as the driving force behind their motivation. When learners perceive discrepancies between their present and future selves (i.e., ideal and ought-to selves), they are motivated to make efforts to bridge the gap. The L2MSS comprises three components: the ideal L2 self, the ought-to L2 self, and the L2 learning experience.

According to Dörnyei (2019), the ideal L2 self represents the learner as a proficient and skillful user of the target language, while the ought-to L2 self reflects external expectations, social or familial obligations, and the desire to avoid negative outcomes in language learning. Among these two kinds of selves, the ideal L2 is widely recognized as having higher predictive validity for L2 learning. Research has demonstrated that the ideal L2 self significantly and positively influences motivational intensity, persistence, intended efforts, and achievement in L2 learning (Al-Hoorie, 2018; Feng and Papi, 2020; Yousefi and Mahmoodi, 2022).

In contrast, while the ought-to L2 self has been found to be significantly and positively correlated with L2 motivational intensity (Feng and Papi, 2020) and intended effort (Al-Hoorie, 2018), it shows no significant correlation with L2 learning achievement (Al-Hoorie, 2018) and is negatively correlated with persistence in L2 learning (Feng and Papi, 2020).

The conflicting results regarding the ought-to L2 self may be explained by its nature. While meeting others' expectations, avoiding negative outcomes, and fulfilling obligations can initially stimulate learners' motivation and encourage their willingness to study (Papi et al., 2019; Feng and Papi, 2020), this effort may not necessarily translate into sustained or effective learning behavior.

On the other hand, learners driven by this kind of motivation aim to achieve the minimum goal of avoiding negative outcomes, resulting in a non-significant relationship with L2 achievement. Moreover, driven by this kind of motivation, learners are less likely to sustain long-term engagement and enthusiasm in L2 learning (Al-Hoorie, 2018). Thus, further exploration is needed to explore the impact of the L2 selves on specific L2 learning variables.

The present study identified L2MSS as the theoretical framework to explore the mediating role of motivation between Chinese EFL learners' SES and pragmatic awareness firstly because it holds that L2 learners' ideal and ought-to L2 self are constantly evolving and changing due to individual factors (e.g., learner's socio-economic background and language proficiency) and environmental factors (e.g., instruction pattern, social background), which in turn stimulate or inhibit the formation and maintenance of motivation (Dörnyei, 2005), secondly because L2MSS has shown significant explanatory power in exploring the relationship between L2 motivation and multiple aspects of language learning achievement (Csizér and Gyula, 2017; Sasaki et al., 2017; Yang and Ren, 2020).

2.3 Motivation and PA

Pragmatic awareness has recently attracted considerable attention as it is an important part of pragmatic competence (Ren, 2015; Yang and Ren, 2020). It has been defined as "conscious, reflective, and explicit knowledge about pragmatics" (Alcón and Safont Jordà, 2008, p. 193). Bardovi-Harlig and Dörnyei (1998) pioneered the study of pragmatic awareness. The pragmatic awareness test they developed has been widely used by researchers to assess English learners' noticing of pragmatic infelicities (Yang and Ren, 2020; Lv et al., 2021). Previous researchers have found that learning environment (Niezgoda and Röver, 2001), language proficiency (Bardovi-Harlig and Dörnyei, 1998), motivation (Chiravate, 2012), classroom instruction, length of residence, and L2 community attitudes (Yang and Ren, 2020) are the contributing factors to the different levels of pragmatic awareness. However, the relationship between L2 motivation and pragmatic awareness has been largely underexplored (Yang, 2022). It is necessary to explore the impact of individual differences (such as SES and L2 motivation) on the pragmatic awareness of second language learners (Yang and Ren, 2020; Lv et al., 2021).

In exploring the relationship between SL/FL learning motivation and pragmatic awareness, researchers have found that motivation may play a significant role in students' pragmatic awareness development in SL/FL learning. Specifically, highly motivated learners tend to exhibit higher levels of pragmatic awareness, that is, succeeding more in recognizing pragmatic inappropriateness or errors than less motivated learners (Chiravate, 2012; Yamato et al., 2013; Yang and Ren, 2020). Among the few studies investigating the impact of L2 learning motivation on pragmatic awareness, Yang and Ren (2020) research has made a significant contribution to this study, as they discovered the significant predictive role of attitudes toward the L2 community, attitudes toward learning English and intended learning efforts on Chinese EFL learners' pragmatic awareness level. Besides, intrinsic motivation (like ideal L2 self) and communication-oriented motivation (attitudes toward the L2 community) were found to be more closely associated with pragmalinguisitic awareness (Yamato et al., 2013; Takahashi, 2015). However, the existing research design could be further improved in terms of the multifacetedness of motivational variables and group diversity (Botes et al., 2020; Yang and Ren, 2020). Thus, the present study took five motivational variables, that is, ideal L2 self, ought-to L2 self, attitudes toward L2 community, attitudes toward learning English, and intended learning efforts into consideration and invited EFL students from different universities and majors to investigate the predictive effect of motivation on L2 pragmatic awareness.

> *H2:* Motivational variables positively predict EFL learners' PA.

2.4 SES, motivation, and PA

There is a relative lack of empirical research on how SES affects pragmatic awareness of foreign languages. On the one hand, research in first language learning has provided preliminary evidence for this study, demonstrating that due to the disadvantage of in-home educational resources (Pace et al., 2017), learners with low SES consistently lag behind their more affluent peers in first language pragmatic development (Pace et al., 2017; Fannin et al., 2018; Qasem

et al., 2022). Compared with first language learning, foreign language learning may be more closely associated with learning resources due to the lack of access to the daily linguistic environment for foreign language learners (Bardovi-Harlig and Hartford, 1996; Kasper, 1997; Yang and Ren, 2020). Therefore, this study could reasonably infer that SES significantly affects FL learners' pragmatic awareness. On the other hand, in FL learning, learners' SES strengths will help them to be more motivated and achieve higher language proficiency (Butler, 2014; Butler and Le, 2018; Ghorbani and Golparvar, 2020; Lee and Lee, 2023) and finally contribute to their FL pragmatic awareness level (Schauer, 2009; Xu et al., 2009; Yang and Ren, 2020). As mentioned previously, the effect of SES on FL/L2 learning outcomes is considered to be indirect and mediated by other variables. It is logical to further assume that the effect of SES on FL pragmatic awareness is at least partially due to motivation (Huang et al., 2018). However, there is no research investigating the association between SES and FL pragmatic awareness in a single model, let alone further exploring the mediating role of motivation between them. It is quite illuminating to examine the impact of SES on FL learners' pragmatic awareness and the exact influence pathways.

> *H3:* SES positively and directly predicts EFL learners' PA.
>
> *H4:* Motivational variables significantly mediate the relationship between SES and EFL learners' PA.

2.5 The present study

Based on the literature review, the present study aimed to investigate the relationship among SES, motivation, and pragmatic awareness in EFL learning, particularly the mediating effect of motivation. The result of this study can help educators and teachers design intervention programs to improve students' pragmatic performance, narrow the learning gap caused by SES, and further promote education equity, which is of great practical significance. The current study sought to address the following questions:

> *RQ1:* Is there a direct effect of EFL students' SES on their motivation?
>
> *RQ2:* Is there a direct effect of EFL students' motivation on their pragmatic awareness?
>
> *RQ3:* Is there a direct effect of EFL students' SES on their pragmatic awareness?
>
> *RQ4:* Does EFL students' learning motivation significantly mediate the relationship between their SES and pragmatic awareness?

3 Research design

3.1 Participants and procedures

The present study aimed to investigate the relationship between SES and PA, as well as the mediating role of English learning motivation, by using quantitative research methods. Instruments were set to measure students' SES, motivation, and pragmatic awareness levels. Adopting the convenience sampling method, a questionnaire survey on Chinese EFL learners was conducted. Having obtained permission to conduct the study and use the data only for research purposes, the author of this study contacted university teachers who were asked to forward the questionnaire link to their students. The questionnaires of this study were published through an online survey platform named Wenjuanxing. The teachers introduced the purpose and procedures of the study to all potential participants, and they could voluntarily choose whether to respond to the questionnaires. Excluding questionnaires from participants who submitted incomplete, duplicate, or blank questionnaires and responses of less than 3 min (the minimum time to complete whole questionnaires), the present study collected 292 valid questionnaires voluntarily answered by Chinese EFL students. Participants come from three universities in Beijing, three universities in Hebei, one university in Yunnan, and one university in Guizhou, with a total including 82 male students (28.1%) and 210 female students (71.9%), 87 freshmen (29.8%), 117 sophomores (40.1%), 39 junior students (13.4%), 9 senior students (3.1%), and 40 master students (13.7%), among which 99 students major in foreign language (33.9%), 60 students major in economics and management (20.5%), 51 students major in arts (17.5%), and 82 major in science and engineering (28.1%) (Table 1). Considering the stratification and diversity of the sample, the influence of sampling bias on the generalizability of findings can be greatly reduced.

3.2 Instruments

3.2.1 SES scale

The measurement of SES was adapted from Zheng and Mei (2021) study. Specifically, students' SES was computed based on their parental educational level (maternal and paternal educational levels), parental occupation (maternal and paternal occupation), and family annual income. A minor adjustment was made to make the scale more

TABLE 1 Demographic information.

Measure	Item	Frequency	Percentage (%)
Gender	Male	82	28.1
	Female	210	71.9
Grade	Freshman	87	29.8
	Sophomore	117	40.1
	Junior	39	13.4
	Senior	9	3.1
	Postgraduates	40	13.7
Major	Foreign Language	99	33.9
	Economics and Management	60	20.5
	Arts	51	17.5
	Science and Engineering	82	28.1
Total		292	100

appropriate in Chinese. Following Zheng and Mei (2021) study, the student's final SES score was calculated through the formula "SES score = Education level*0.384 + Occupation*0.371 + Family annual income*0.348". The scale has a good internal consistency (Cronbach's $\alpha = 0.750$).

3.2.2 Motivation scale

This motivational questionnaire attempted to measure the following constructs, i.e., ideal L2 self, ought-to self, attitudes toward the L2 community, attitudes toward learning English, and intended learning efforts. All items were answered on a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree).

Ideal L2 self and ought-to L2 self scales were derived and adapted from Taguchi et al. (2009) and Papi (2010). Attitudes toward the L2 community, attitudes toward learning English, and intended learning efforts scales were adopted from Yang and Ren (2020). Two independent Chinese-English bilingual teachers obtained the Chinese version of the questionnaire using translation and back-translation methods. The ideal L2 self scale consists of five items: "I can imagine myself speaking English as if I were a native speaker of English." The ought-to L2 self consists of six items: "Studying English is important to me because other people will respect me more if I have knowledge of English." Attitudes toward the L2 community scale consist of three items: "I like to travel to English-speaking countries." The adapted attitudes toward learning English scale consists of two items: "I find learning English interesting." The intended learning efforts scale includes three items: "I would like to spend lots of time studying English." Each scale has an ideal reliability (Cronbach's $\alpha = 0.904, 0.911, 0.910, 0.963, 0.900$).

3.2.3 Pragmatic awareness test

An appropriateness judgment task (AJT) was used to assess students' pragmatic awareness. Adopted from Yang and Ren (2020) and Bardovi-Harlig and Dörnyei (1998) studies, the adapted task consists of seven short conversations (e.g., Peter needs directions to the library). He asks another student. A: Hi. P: Hi. P: Tell me how to get to the library, of which four were pragmatically inappropriate, and three were appropriate (controls). Pragmatically inappropriate and appropriate items were developed from responses of nontarget-like learners and native speakers, respectively, both of which were widely accepted and applied in assessing learners' pragmatic awareness (Yang and Ren, 2020), and a detailed development process of AJT could be seen from Bardovi-Harlig and Dörnyei (1998) study. Learners were asked to assess the appropriateness of the last sentence of each item on a 6-point Likert scale ranging from 1 (strongly inappropriate) to 6 (strongly appropriate). The four pragmatically inappropriate items' scores were reversed, so the full score of AJT is 42, and the higher the learners' scores, the higher their pragmatic awareness. The Cronbach's α was 0.611, indicating an acceptable internal consistency of AJT.

3.3 Data analysis

In this study, SPSS 23.0 was used to conduct data standardization, confirmatory factor analysis, common method deviation test, and correlation analysis. Harman's single-factor test confirmed that there was no significant common method bias in the current study (Podsakoff et al., 2003). Then, Smart PLS 3.0 was utilized to perform the data analysis based on the partial least squares (PLS) method, which could estimate complex models with many latent and manifest variables and is suitable for exploratory studies where the relationship between measures has not been explored. The internal consistency, indicator reliability, convergent validity, and discriminant validity of the measurement model were evaluated first. In this study, the Cronbach's alpha (CA) of variables ranged from 0.750 to 0.963, greater than 0.7. The composite reliability (CR) ranged from 0.858 to 0.982 (Table 2), greater than 0.7, indicating a good internal consistency (Hair et al., 2019). The Average Variance Extracted (AVE) value is between 0.650 and 0.964 (Table 2), greater than 0.5, explaining a good convergent validity (Fornell and Larcker, 1981). The discriminant validity of the scale was tested through the Fornell-Larcker criterion and the Heterotrait-Monotrait Ratio (HTMT). According to Table 3, the square roots of AVE values are higher than the correlations between constructs (Fornell and Larcker, 1981), and HTMT values are all under 0.85, further proving the discriminant validity of scales (Clark and Watson, 1995). Then, multiple regression analysis was conducted to verify the causal relationship between SES, motivation, and pragmatic awareness, as well as the mediating role of motivational variables, based on the coefficient of determination (R^2), path coefficients, effect size (f^2), and predictive relevance (Q^2).

4 Results

4.1 Correlation analysis

The Pearson correlation analysis tested the relationship among learners' SES, motivation, and pragmatic awareness. The results showed that learners' SES was significantly and positively correlated with their ideal L2 self, attitudes toward the L2 community, attitude toward English, intended learning efforts, and pragmatic awareness, and is not related to ought-to L2 self. Besides, among five motivational variables, ideal L2 self, attitudes toward the L2 community, attitude toward English, and intended learning efforts were significantly and positively correlated with learners' pragmatic awareness. There was no significant correlation to be found between learners' ought-to self and pragmatic awareness (Table 4).

4.2 The direct effect of EFL students' SES on their motivation

SES was found to have a significantly positive effect on students' ideal L2 self, attitudes toward the L2 community, and attitudes toward

TABLE 2 Results for the measurement model.

	Cronbach's Alpha	CR	AVE
SES	0.750	0.858	0.670
ILS	0.904	0.926	0.718
OLS	0.911	0.914	0.650
ATLC	0.910	0.943	0.847
ATLE	0.963	0.982	0.964
ILE	0.900	0.926	0.760

SES, socioeconomic status; ILS, ideal L2 self; OLS, ought-to L2 self; ATLC, attitudes toward the L2 community; ATLE, attitudes toward learning English; ILE, intended learning efforts.

TABLE 3 Fornell-Larcker criterion and HTMT.

	Model 1			Model 2			Model 3			Model 4			Model 5		
Fornell-Larcker criterion															
	SES	ILS	PA	SES	OLS	PA	SES	ATLC	PA	SES	ATLE	PA	SES	ILE	PA
SES	0.817			0.811			0.815			0.813			0.812		
CMV	0.183	0.847		−0.035	0.805		0.260	0.921		0.210	0.982		0.143	0.871	
PA	0.164	0.253	1	0.169	−0.175	1	0.167	0.233	1	0.168	0.196	1	0.169	0.162	1
HTMT															
SES															
CMV	0.204			0.077			0.306			0.238			0.143		
PA	0.186	0.236		0.186	0.148		0.186	0.243		0.186	0.199		0.186	0.151	

CMV, corresponding mediating variable; SES, socioeconomic status; ILS, ideal L2 self; OLS, ought-to L2 self; ATLC, attitudes toward the L2 community; ATLE, attitudes toward learning English; ILE, intended learning efforts.

TABLE 4 Results of Pearson correlation coefficient.

	SES	ILS	OLS	ATLC	ATLE	ILE	PA
SES	1						
ILS	0.168**	1					
OLS	0.09	0.287**	1				
ATLC	0.253**	0.635**	0.216**	1			
ATLE	0.202**	0.649**	0.231**	0.588**	1		
ILE	0.118*	0.513**	0.270**	0.367**	0.691**	1	
PA	0.161**	0.225**	−0.113	0.232**	0.195**	0.144*	1

* $p < 0.05$,** $p < 0.01$. SES, socioeconomic status; ILS, ideal L2 self; OLS, ought-to L2 self; ATLC, attitudes toward the L2 community; ATLE, attitudes toward learning English; ILE, intended learning efforts; PA, pragmatic awareness.

learning English, and no effect on students' ought-to L2 self and intended learning efforts.

4.3 The direct effect of EFL students' motivation on their PA

All five motivational variables significantly predict EFL students' pragmatic awareness, among which ideal L2 self, attitudes toward L2 community, attitudes toward learning English, and intended learning efforts positively predict learners' pragmatic awareness while ought-to L2 self negatively predicts it.

4.4 The direct effect of EFL students' SES on their PA

SEM was conducted to test the five proposed models, respectively. As Table 5 shows, the regressive analysis indicated that SES directly affected EFL learners' pragmatic awareness. SES has shown a significant and positive effect on English learners' pragmatic awareness level in all five proposed models. However, the effect size is weak (f^2 is around 0.02, the low threshold value indicating the predictor's effect). The proportion of this direct effect to the total effect varies among the five models, ranging from 67.7 to 100%. It could be concluded that SES accounts for a certain portion of the discrepancies among English learners' pragmatic awareness levels.

4.5 The mediating role of EFL students' motivation between their SES and PA

SEM tested the mediating effect of five motivational variables respectively, finding that the ideal L2 self, attitudes toward the L2 community, and attitudes toward learning English play significant mediating roles in the relationship between their SES and pragmatic awareness. Though ought-to L2 self and intended learning effort were found to significantly and negatively predict pragmatic awareness, their mediating role is insignificant. In model 1, the mediating effect of ideal L2 self accounts for 26.1% of the total effect, and this model could predict 7.8% of learners' pragmatic awareness variance. In model 3, the mediating effect of attitudes toward the L2 community accounts for 32.3% of the total effect, and this model could predict 6.6% of learners' pragmatic awareness variance. In model 4, the mediating effect of attitudes toward learning English accounts for 21.1% of the total effect, and this model could predict 5.4% of learners' pragmatic awareness variance. All statistically significant causal relationships found above have substantial but slightly weak explanatory power ($0.19 > R^2 > 0$) and effect size (f^2 near or over 0.02) (Cohen, 1988; Chin, 1998). Besides, all of the five models have good predictive relevance ($Q^2 > 0$) except for the predictive effect of SES on ought-to self ($Q^2 = -0.001 < 0$) (Geisser, 1975) (Figures 1–5).

5 Discussion

5.1 The influence of SES on ought-to L2 self and intended learning efforts

There was no significant relationship found between SES and ought-to L2 self and intended learning efforts in the present study, revealing that SES does not have a predictive effect on the state that English learners achieve to satisfy significant others and the efforts they intended to apply to English learning, which was against H1. This result may be attributed to the compulsoriness and popularity

TABLE 5 The mediating role of motivational variables.

Model	Mediating path	DE	IE	TE	VAF	Pathway	R^2	f^2	Q^2
1	SES → ILS → PA	0.119*	0.042*	0.161**	0.261	SES → ILS	0.033	0.034	0.020
						SES → PA	0.078	0.015	0.061
						ILS → PA		0.056	
2	SES → OLS → PA	0.157*	0.004	0.161**	0	SES → OLS	0.001	0.001	−0.001
						SES → PA	0.055	0.026	0.038
						OLS → PA		0.031	
3	SES → ATLC→PA	0.109*	0.052*	0.161**	0.323	SES → ATLC	0.064	0.069	0.052
						SES → PA	0.066	0.012	0.052
						ATLC→PA		0.042	
4	SES → ATLE→PA	0.127*	0.034*	0.161**	0.211	SES → ATLE	0.041	0.043	0.031
						SES → PA	0.054	0.016	0.032
						ATLE→PA		0.029	
5	SES → ILE → PA	0.142**	0.019	0.161**	0	SES → ILE	0.019	0.019	0.008
						SES → PA	0.046	0.021	0.020
						ILE → PA		0.021	

DE, direct effect; IE, indirect effect; TE, total effect; SES, socioeconomic status; ILS, ideal L2 self; OLS, ought-to L2 self; ATLC, attitudes toward the L2 community; ATLE, attitudes toward learning English; ILE, intended learning efforts; PA, pragmatic awareness. *$p < 0.05$, **$p < 0.01$ (2-tailed).

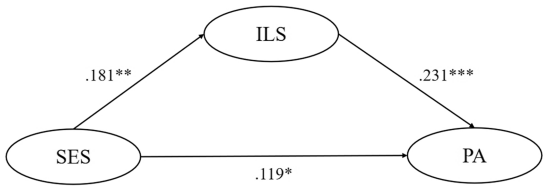


FIGURE 1
The mediating role of ideal L2 self. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SES, socioeconomic status; ILS, ideal L2 self; PA, pragmatic awareness.

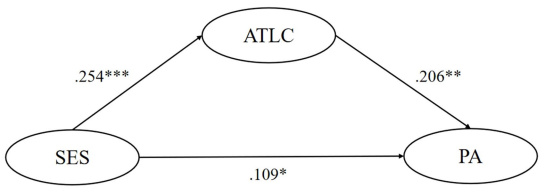


FIGURE 3
The mediating role of attitudes toward the L2 community. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SES, socioeconomic status; ATLC, attitudes toward the L2 community; PA, pragmatic awareness.

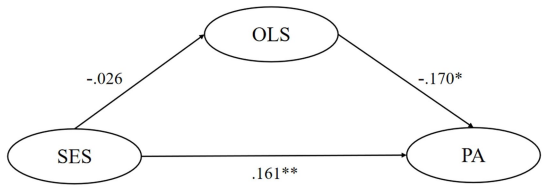


FIGURE 2
The mediating role of ought-to L2 self. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SES, socioeconomic status; OLS, ought-to L2 self; PA, pragmatic awareness.

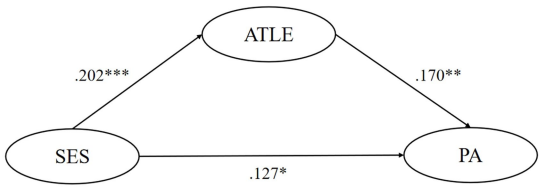


FIGURE 4
The mediating role of attitudes toward learning English. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SES, socioeconomic status; ATLE, attitudes toward learning English; PA, pragmatic awareness.

of the Chinese university English curriculum (Lamb, 2012). In Chinese universities, English courses are generally offered and taken as compulsory subjects. Students develop instrumental motivation toward English learning to meet the expectations of parents and teachers or to avoid negative outcomes like failing exams or letting others down. That is, once students take English courses and exams, they have strong, oblationary feelings to put effort into making it regardless of their family background.

5.2 The influence of ought-to L2 self and intended learning efforts on PA

The predictive role of ought-to L2 self and intended learning efforts on English learners' pragmatic awareness was revealed in the present study. Specifically, a negative correlation was found between ought-to L2 self and pragmatic awareness, which was against H2. That is to say, the stronger a student's ought-to L2 self is, the less conducive it is to their development of English pragmatic awareness, which is

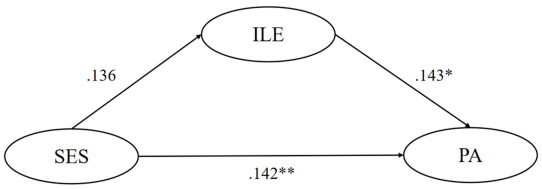


FIGURE 5
The mediating role of intended learning efforts. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SES, socioeconomic status; ILE, intended learning efforts; PA, pragmatic awareness.

similar to previous results documenting the negative effect of ought-to L2 self on L2 learning (Papi and Teimouri, 2014; Peng, 2015; Feng and Papi, 2020). A possible reason for the negative correlation between ought-to L2 self and pragmatic learning outcomes is the mismatch between Chinese students' immediate English learning needs and English pragmatic learning (Yang and Ren, 2020). Unlike the ideal L2 self, the ought-to L2 self is more like an instrumental motivation (Lamb, 2012), which primarily focuses on current learning outcomes and how to satisfy others' expectations (Al-Hoorie, 2018). In the Chinese exam-oriented education system, driven by this extrinsic motivation, students think passing exams and getting teachers' approval are the most important and urgent issues (Yang and Ren, 2020). College English Test (Band 4 or Band 6) is a well-recognized national English level test among university students in China. It mainly examines writing and reading literacy (Taguchi and Roever, 2017). Under this circumstance, students are more motivated to improve receptive skills such as listening and reading comprehension with focuses on grammatical and vocabulary enrichment, while pragmatic learning, which will not be tested, such as appropriate expressions of daily conversation, has been largely neglected, which finally has a negative effect on EFL learners' motivation to develop pragmatic awareness. In addition, a positive and significant relationship between intended learning efforts and pragmatic awareness was found, which supported H2 and Yang and Ren (2020) findings that when students are willing to put effort into English learning, they tend to perform well in pragmatic awareness test, and further confirming the significant role of motivation in L2 pragmatic development.

5.3 The direct predictive effect of EFL students' SES on PA

This study investigated the relationship between SES and English pragmatic awareness among Chinese university students. The results showed that students' SES could positively and directly predict their pragmatic awareness level, which answered our research question 3 and verified H3. The results indicated that EFL students with higher family SES are more likely to notice pragmatic issues and perform better in pragmatic awareness tests. Previous studies have shown the positive effect of SES on children's pragmatic performance in first language learning (Fannin et al., 2018; Qasem et al., 2022). The current study found this same relationship in the FL learning context. The influence existed, although the effect size was between the low and middle ranges. In FL learning, the target language does not exist in the learner's direct environment. In the case of insufficient pragmatic contact in class, extracurricular language access appears particularly important for Chinese EFL students to cultivate their pragmatic competence (Yang and Ren, 2020). Therefore, SES has been attached great importance in FL learning since high SES families are more likely to provide students with learning environments and pragmatic resources helpful for their development of communicative abilities, which contributes to their pragmatic awareness (Niezgoda and Röver, 2001; Schauer, 2009; Butler and Le, 2018; Ghorbani and Golparvar, 2020). However, not all parents can provide such a kind of promoting learning environment and resource for their children. Differences in family SES can lead to discrepancies in students' pragmatic performance. Therefore, exploring the mediating pathways to bridge this gap is even more important.

5.4 The mediating effect of motivation between SES and PA

The study further examined the mediating effect of motivational variables between English learners' SES and pragmatic awareness. The results suggested that the ideal L2 self, attitudes toward the L2 community, and attitudes toward learning English significantly mediate the relationship between SES and pragmatic awareness, indicating that English learners with higher SES levels are more likely to see themselves as fluent English speakers, hold more positive attitudes toward English speaking countries and people and are more interested in learning English, which in turn contribute to their performance in pragmatic awareness test. These findings answered RQ4 and supported H1, H2, and H4. It is worth noticing that these associations are statistically significant but have small explanatory and influential power (Cohen, 1988; Chin, 1998), which may be attributed to some external factors beyond motivation (Schmidt, 1993; Bardovi-Harlig and Dörnyei, 1998; Ren and Han, 2016; Yang and Ren, 2020). According to Dörnyei (2005, 2009), students' immediate language learning experience may greatly affect their motivation to learn the target language. In Yang and Ren (2020) interviews with Chinese university English learners, they reported some common issues that exist in the Chinese English teaching environment, that is, lack of pragmatics instruction, lack of opportunities to practice pragmatic knowledge, and various pragmatic norms that students find difficult to choose one to follow (Yang and Ren, 2020). That is to say, although students develop motivation to learn English, these factors will, to some degree, hinder them from greatly improving their pragmatic achievement (Yang and Ren, 2020), resulting in these significant but weak influences.

5.4.1 The mediating effect of ideal L2 self between SES and PA

Among three motivational mediating variables, the ideal L2 self contributes more to English learners' pragmatic awareness. It is reasonable to assume that when socioeconomically privileged English learners hold an optimistic view toward their future success in learning English, they ought to perform better in pragmatic tasks. This result confirmed previous studies supporting that SES positively influences learners' vision of future success in English learning (Kim and Kim, 2014; Papi and Teimouri, 2014; Khan, 2015). For example, Lamb (2012) and Kormos and Kiddle (2013) found that students from

high-class families were more optimistic about their future language competence than those from low-class families. Oyserman and Fryberg (2006) emphasized the importance of a social environment for learners' future self-guides. Peng (2015) said that learners' ideal L2 self would be triggered or enhanced by successful and pleasant personal experiences. Students with better SES are more likely to encounter role models during the English learning process. Positive social contact with prestigious people is conducive to their development of an ideal L2 self in English learning (Lamb, 2012). In conclusion, learners with high SES are more capable of learning English in such a context where socializing with English speakers, encountering role models, and participating in international competitions are accessible, tending to imagine themselves becoming fluent language users. This practically orientated mental image makes students more persistent and more focused on the actual use of the target language (Feng and Papi, 2020) and finally contributes to their development of pragmatic awareness (Schmidt, 1993; Bardovi-Harlig and Dörnyei, 1998; Kormos and Kiddle, 2013).

Besides, the mediating effect also showed a positive correlation between ideal L2 self and pragmatic awareness, showing that when EFL learners have a positive attitude toward their future English achievements, they will better assess the appropriateness of English speech acts. Specifically, when students aspire to become capable and proficient English speakers, they are keen to reduce discrepancies between their ideal and present selves and focus on the actual use of the target language, particularly successful communication, such as writing emails and engaging in academic exchanges (Taguchi et al., 2009). In this case, the discrepancies between the ideal and present self may facilitate the development of students' pragmatic awareness. This conclusion is consistent with previous research that learners with intrinsic motivation and communication-oriented motivation tend to exhibit higher levels of pragma linguistic awareness (Yamato et al., 2013; Takahashi, 2015) but inconsistent with Yang and Ren (2020) research finding no significant correlations between ideal L2 self and English pragmatic awareness. Students' different interpretations of the ideal self may be a possible reason for inconsistent results. In Yang and Ren (2020) research, the students being interviewed came from a prominent university in China with a prevalent academic atmosphere. Their reported ideal selves were more about their ability to meet immediate academic accomplishments, which led students to focus on listening, reading, and writing literacy rather than practical aspects of English. In contrast, students in the present study came from a wider range of universities. By eliminating the limitation of sample singularity, the obtained results may be closer to objective reality and more referential.

Notably, the ideal L2 self rather than the ought-to L2 self was found to mediate the association between SES and pragmatic awareness. This result is consistent with previous studies based on the L2MSS, demonstrating that the ideal L2 self is a better predictor of learning outcomes than the ought-to L2 self (Al-Hoorie, 2018; Sadoughi et al., 2023). Researchers have found it is ideal L2 self that positively predicted the use of self-regulated learning writing strategies (Xu and Wang, 2022), academic engagement (Sadoughi et al., 2023), and achievement (Dörnyei and Chan, 2013; Al-Hoorie, 2018), rather than ought-to L2 self. These differential results may relate to the promotion focus of the ideal L2 self and the prevention focus of the ought-to L2 self (Dörnyei, 2009). Therefore, we believe that the ideal L2 self is the primary factor affecting EFL learners' pragmatic awareness and demonstrate the more internalized and active role of the ideal self in motivating students and empowering their L2 learning strength.

5.4.2 The mediating effect of attitudes toward L2 community and attitudes toward learning English between SES and PA

Attitudes toward the L2 community and attitudes toward learning English significantly mediate the relationship between English learners' SES and their pragmatic awareness level, indicating that students with higher family SES are typically more likely to have a positive and friendly attitude toward English-speaking countries and people there, and tend to be more proactive and engaged in English learning, which would assist them in succeeding in judging English pragmatic appropriateness. Prior research stated that learners from low social class families typically lacked awareness of the importance of English, thus having their future visions weakly correlated with English (Lamb, 2012). There are still some students who have realized the significance of pragmatic learning but put it on hold, suffering from no resources to learn and no urgent pragmatic needs (neither taking pragmatic exams nor necessity to communicate with foreigners) (Yang and Ren, 2020). Under these circumstances, learners rarely envision their future development linked to English, and they are less motivated to learn English, finding themselves lacking interest and effort in English. In contrast, learners from upper-class families, as Kormos and Kiddle (2013) argued, often believe that they will study abroad or be involved in international competition, where English is an indispensable part of their lives. These future communicative needs will drive them to develop pragmatic motivations. It is logical to suppose that when students come from families with high SES, they place themselves on the international stage, are eager to socialize with English speakers and exert themselves to learn English. Previous research has fully demonstrated the positive correlation between students' SES and English learning motivation (Kormos and Kiddle, 2013; Butler, 2015; Lee and Lee, 2023), and this correlation has become stronger as grades increase. The current study found three motivation variables to be significantly correlated with SES, which to some extent further confirmed the above conclusion among university students. Moreover, this result is confirmed by Yang and Ren (2020) research finding that Chinese students' attitudes toward the L2 community are significantly related to their performance in pragmatic awareness test, as well as to some degree conforming to Schumann (1986) acculturation model, which holds that the psychological distance between learners and the target community has an impact on their L2 learning.

6 Conclusions, implications, limitations, and suggestions

The present study examined the relationship between university EFL students' socioeconomic status and pragmatic awareness with the mediating role of English learning motivation, which demonstrates some motivation variables as pathways in which socio-economic inequality can contribute to students' PA gap. The theoretical contribution of the study lies in the mediating effect of an ideal self, attitudes toward the L2 community, and attitudes toward learning English between EFL

students' family socio-economic background and their appropriate judgment capacities concerning pragmatics. The relationship between socioeconomic status and pragmatic acquisition was also expanded from the first language field to the foreign language learning area and from younger learners such as primary and secondary school students to university students. In terms of practical instruction, with specific clarification of the association between the different variables, it is important to design and develop proper interventions to enhance students' pragmatic judgment abilities. Hopefully, the findings will encourage English educators to consider students' individual differences in pragmatic instruction and provide the foundation for effective intervention in future instruction.

Based on research findings, the study proposed some substantial implications for educators. Teachers should recognize learners' diverse SES backgrounds and needs in class, as well as the different impacts of motivational variables on English pragmatic awareness. Classroom instruction should include stimulating learners' pragmatic interests and needs and providing learners with pragmatic help and practice opportunities. Differentiated teaching strategies (Xu and Feng, 2024) are expected to be implemented if human and material resources permit. On the one hand, the generally current instruction situation that undermines students' pragmatic motivation should be addressed. Incorporating pragmatic instruction into EFL teaching curricula is of the essence (Qin et al., 2024; Yang and Ren, 2020). For example, the design of textual and oral practices that reflect real-life situations in both academic and daily life can allow learners to develop and apply their pragmatic awareness in a specific context (Qin et al., 2024). In addition, increasing learners' knowledge of the effect of their L1 on L2 pragmatic uses and making contrastive analyses are conducive to their notice and judgment of L2 pragmatic inappropriateness (Qin et al., 2024). Besides, sample conversations of native English speakers in specific contexts, such as academic conferences and emails, can serve as effective learning resources for learners to cultivate language intuition and understand the norms of the target language (Qin et al., 2024).

On the other hand, cultivating learners' pragmatically related motivation, that is ideal L2 self, attitudes toward the L2 community, and attitudes toward learning English, will help. By providing positive feedback and constructive suggestions, teachers can assist students in establishing their vision of future success in EFL learning, facilitating their development of pragmatic awareness (Yang and Ren, 2020). An immersive language learning environment, that is, making learners more exposed to the English environment, can increase learners' interest in English learning and the L2 community. The set of pragmatic instruction and practice mentioned above, as well as the inclusion of English movies, music, blogs, and other elements in curricula, are effective measures to increase learners' contact with the target language. Notably, the addition of pragmatic instruction and exercise to the English curriculum might be conducive to the pragmatic learning of students driven by ought-to L2 self. They might attach importance to pragmatic learning because it has become part of their curriculum and tests (Yang and Ren, 2020).

This study also has some limitations. First, adopting the convenience sampling method in preliminary research with a large sample is appropriate to quickly conclude and lay the groundwork for the follow-up study. The appropriateness of the sampling method and stratified and diversified sample guarantee that the results of this study have a certain generalisability among Chinese EFL students. However, what should be acknowledged is that this study still has some sample limitations in terms of region and gender. Future studies ought to supplement qualitative material and collect more comprehensive and diverse data to further verify and enhance the strength and applicability of these conclusions in other populations and contexts. Secondly, this study adopted self-reported measurement to assess students' motivation levels. Discrepancies may exist between their actual conditions and reported conditions. Therefore, future studies should consider more measurements, such as interviews or others' evaluations, to obtain students' most authentic motivation level. In addition, this study examined EFL learners' motivation based on L2MSS theory, which has been well-established in SLA but has also been criticized for overlooking students' important motivation forces, resulting in the study's shortcomings in revealing students' motivation beyond language learning goals. Hence, future studies should consider more novel and comprehensive motivational frameworks to compensate for this deficiency. Last but not least, given that students' L2 motivation is dynamic, longitudinal investigations are needed to reveal the dynamic variation of their motivation during a period of English learning.

Data availability statement

The datasets presented in this article are not readily available because part of the data was involved in other unpublished studies. Requests to access the datasets should be directed to Yuqing Chen, neuwy2021@163.com.

Ethics statement

The studies involving humans were approved by Academic Committee, School of Foreign Studies, Northeastern University at Qinhuangdao. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XH: Funding acquisition, Writing – review & editing, Project administration, Resources, Supervision. YC: Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Software.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The work was supported by the National Social Science Fund of China [Grant No.:24BYY135].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alcón, S. E., and Safont Jordà, M. P. (2008). "Pragmatic awareness in second language acquisition" in Encyclopedia of language and education. eds. J. Cenoz and N. H. Hornberger. *2nd* ed (New York: Springer), 193–204.

Al-Hoorie, A. H. (2018). The L2 motivational self-system: a meta-analysis. *Stud. Second Lang. Learn. Teach.* 8, 721–754. doi: 10.14746/SSLLT.2018.8.4.2

Bardovi-Harlig, K., and Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Q.* 32, 233–262. doi: 10.2307/3587583

Bardovi-Harlig, K., and Hartford, B. S. (1996). Input in an institutional setting. *Stud. Second. Lang. Acquis.* 18, 171–188. doi: 10.1017/S027226310001487X

Botes, E., Gottschling, J., Stadler, M., and Greiff, S. (2020). A systematic narrative review of international posture: what is known and what still needs to be uncovered. *System* 90:102232. doi: 10.1016/j.system.2020.102232

Butler, Y. G. (2014). Parental factors and early English education as a foreign language: a case study in mainland China. *Res. Pap. Educ.* 29, 410–437. doi: 10.1080/02671522.2013.776625

Butler, Y. G. (2015). Parental factors in children's motivation for learning English: a case in China. *Res. Pap. Educ.* 30, 164–191. doi: 10.1080/02671522.2014.891643

Butler, Y. G. (2017). "The dynamics of motivation development among young learners of English in China" in Early language learning: complexity and mixed methods. eds. J. Enever and E. Lindgren (Bristol, Blue Ridge Summit: Multilingual Matters), 167–185.

Butler, Y. G., and Le, V. N. (2018). A longitudinal investigation of parental social-economic status (SES) and young students' learning of English as a foreign language. *System* 73, 4–15. doi: 10.1016/j.system.2017.07.005

Chen, J., Jiang, T., and Liu, M. (2021). Family socioeconomic status and learning engagement in Chinese adolescents: The multiple mediating roles of resilience and future orientation. *Front. Psychol*, 12, 714346. doi: 10.3389/fpsyg.2021.714346

Chin, W. W. (1998). "The partial least squares approach for structural equation modeling" in Modern methods for business research. ed. G. A. Marcoulides (Mahwah: Lawrence Erlbaum Associates Publishers), 295–336.

Chiravate, B. (2012). "The effects of motivation and proficiency on pragmatic and grammatical awareness in foreign language learning" in Perspectives on individual characteristics and foreign language education. eds. W. M. Chan, K. N. Chin, S. K. Bhatt and I. Walker (Berlin: Mouton de Gruyter), 93–114.

Clark, L. A., and Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychol. Assess.* 7, 309–319. doi: 10.1037/1040-3590.7.3.309

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum.

Csapó, B., and Nikolov, M. (2009). The cognitive contribution to the development of proficiency in a foreign language. *Learn. Individ. Differ.* 19, 209–218. doi: 10.1016/J.LINDIF.2009.01.002

Csizér, K., and Gyula, T. (2017). English majors' self-regulatory control strategy use in academic writing and its relation to L2 motivation. *Appl. Linguis.* 38, amv033–amv404. doi: 10.1093/applin/amv033

Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in second language acquisition. New Jersey, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z. (2009). "The L2 motivational self system" in Motivation, language identity and the L2 self. eds. Z. Dörnyei and E. Ushioda (Bristol: Multilingual Matters), 9–42.

Dörnyei, Z. (2019). Toward a better understanding of the L2 learning experience, the Cinderella of the L2 motivational self system. *Stud. Second Lang. Learn. Teach.* 9, 19–30. doi: 10.14746/ssllt.2019.9.1.2

Dörnyei, Z., and Chan, L. (2013). Motivation and vision: an analysis of future L2 self-images, sensory styles, and imagery capacity across two target languages. *Lang. Learn.* 63, 437–462. doi: 10.1111/lang.12005

Ensminger, M. E., and Fothergill, K. (2014). "A decade of measuring SES: what it tells us and where to go from here" in Socioeconomic status, parenting, and child development. eds. M. H. Bornstein and R. H. Bradley (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 13–27.

Fan, W. (2011). Social influences, school motivation and gender differences: an application of the expectancy-value theory. *Educ. Psychol.* 31, 157–175. doi: 10.1080/01443410.2010.536525

Fannin, D. K., Barbarin, O. A., and Crais, E. R. (2018). Communicative function use of preschoolers and mothers from differing racial and socioeconomic groups. *Lang. Speech Hear. Serv. Sch.* 49, 306–319. doi: 10.1044/2017_LSHSS-17-0004

Feng, L., and Papi, M. (2020). Persistence in language learning: the role of grit and future self-guides. *Learn. Individ. Differ.* 81:101904. doi: 10.1016/j.lindif.2020.101904

Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104

García-Gómez, A. (2022). Learning through Whats app: students' beliefs, L2 pragmatic development and interpersonal relationships. *Comput. Assist. Lang. Learn.* 35, 1310–1328. doi: 10.1080/09588221.2020.1799822

Geisser, S. (1975). The predictive sample reuse method with applications. *J. Am. Stat. Assoc.* 70, 320–328. doi: 10.1080/01621459.1975.10479865

Ghorbani, M. R., and Golparvar, S. E. (2020). Modeling the relationship between socioeconomic status, self-initiated, technology-enhanced language learning, and language outcome. *Comput. Assist. Lang. Learn.* 33, 607–627. doi: 10.1080/09588221.2019.1585374

Gong, Y., Ma, M., Hsiang, T. P., and Wang, C. (2020). Sustaining international students' learning of Chinese in China: shifting motivations among New Zealand students during study abroad. *Sustain. For.* 12, 6289–6302. doi: 10.3390/su12156289

Hair, J. F., Risher, J. J., Sarstedt, M., and Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *Eur. Bus. Rev.* 31, 2–24. doi: 10.1108/EBR-11-2018-0203

Huang, J. (2022). Relationships between EFL learners' background, motivational, instructional, ICT-related characteristics, and writing performance: evidence from 13 educational systems. *Psychol. Sch.* 59, 495–514. doi: 10.1002/pits.22626

Huang, B., Chang, Y.-H. S., Niu, M., and Zhi, M. (2018). Examining the effects of socio-economic status and language input on adolescent English learners' speech production outcomes. *System* 73, 27–36. doi: 10.1016/j.system.2017.07.004

Huong, V. T., Casadesus, M., and Marimon, F. (2017). Assessing learner satisfaction by simultaneously measuring learner attitude, motivation, loyalty and service quality in English academies. *Innov. Educ. Teach. Int.* 54, 301–312. doi: 10.1080/14703297.2015.1088397

Jia, Y., and Cheng, L. (2022). The role of academic buoyancy and social support on English as a foreign language learners' motivation in higher education. *Front. Psychol.* 13:892603. doi: 10.3389/fpsyg.2022.892603

Kasper, G. (1997). "The role of pragmatics in language teacher education" in Beyond methods: components of language teacher education. eds. K. Bardovi-Harlig and B. S. Hartford (New York: McGraw-Hill), 113–136.

Khan, M. R. (2015). Analyzing the relationship between L2 motivational selves and L2 achievement: a Saudi perspective. *Int. J. Engl. Lang. Teach.* 2, 68–75. doi: 10.5430/ijelt.v2n1p68

Kim, T. Y., and Kim, Y. K. (2014). A structural model for perceptual learning styles, the ideal L2 self, motivated behavior, and English proficiency. *System* 46, 14–27. doi: 10.1016/j.system.2014.07.007

Kormos, J., and Kiddle, T. (2013). The role of socio-economic factors in motivation to learn English as a foreign language: the case of Chile. *System* 41, 399–412. doi: 10.1016/j.system.2013.03.006

Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Lang. Learn.* 62, 997–1023. doi: 10.1111/j.1467-9922.2012.00719.x

Lee, J. H., and Lee, H. (2023). The role of learners' socioeconomic status and perception of technology use in their second language learning motivation and achievement. *Lang. Aware.* 32, 217–234. doi: 10.1080/09658416.2021.2014510

Lin, Y. T. (2019). Taiwanese EFL learners' willingness to communicate in English in the classroom: impacts of personality, affect, motivation, and communication confidence. *Asia Pac. Educ. Res.* 28, 101–113. doi: 10.1007/s40299-018-0417-y

Liu, S. H. J., and Lan, Y. J. (2016). Social constructivist approach to web-based EFL learning: collaboration, motivation, and perception on the use of Google docs. *Educ. Technol. Soc.* 19, 171–186.

Liu, H., Liu, Q., Du, X., Liu, J., and Schumacker, R. E. (2023). Teacher-student relationship as a protective factor for socioeconomic status, students' self-efficacy and achievement: a multi-level moderated mediation analysis. *Curr. Psychol.* 42, 3268–3283. doi: 10.1007/s12144-021-01598-7

Liu, J., Peng, P., and Luo, L. (2020). The relation between family socioeconomic status and academic achievement in China: a meta-analysis. *Educ. Psychol. Rev.* 32, 49–76. doi: 10.1007/s10648-019-09494-0

Luo, R., Pace, A., Levine, D., Iglesias, A., de Villiers, J., Golinkoff, R. M., et al. (2021). Home literacy environment and existing knowledge mediate the link between socioeconomic status and language learning skills in dual language learners. *Early Childhood Research Quarterly*, 55, 1–14. doi: 10.1016/j.ecresq.2020.10.007

Lv, X., Ren, W., and Li, L. (2021). Pragmatic competence and willingness to communicate among L2 learners of Chinese. *Front. Psychol.* 12:797419. doi: 10.3389/fpsyg.2021.797419

Li, X., and Han, J. (2024). What motivates korean language learners to learn? Motivational Characteristics and their relationships with learning engagement. *The Asia-Pacific Education Researcher*, 33, 273–282.

Ma, L., Xiao, L., Liu, Z., and Liu, J. (2022). Teacher-student relationships moderate the link between socioeconomic status and foreign language learning: evidence from China. *Lang. Teach. Res.*, 1–22. doi: 10.1177/13621688221122236

Mercer, S. (2018). Psychology for language learning: spare a thought for the teacher. *Lang. Teach.* 51, 504–525. doi: 10.1017/S0261444817000258

Mueller, C. W., and Parcel, T. L. (1981). Measures of socioeconomic status: alternatives and recommendations. *Child Dev.* 52, 13–30. doi: 10.2307/1129211

Muñoz, C. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. *International review of applied linguistics. Lang. Teach.* 46, 197–220. doi: 10.1515/IRAL.2008.009

Niezgoda, K., and Röver, C. (2001). "Pragmatic and grammatical awareness: a function of learning environment?" in Pragmatics in language teaching. eds. K. R. Rose and G. Kasper (New York: Cambridge University Press), 63–79.

Nikolov, M., and Csapó, B. (2018). The relationships between 8th graders' L1 and L2 reading skills, inductive reasoning and socio-economic status in early English and German as a foreign language programs. *System* 73, 48–57. doi: 10.1016/j.system.2017.11.001

OECD (2019). PISA 2018 technical report. Paris: OECD Publishing.

Oga-Baldwin, W. Q., Nakata, Y., Parker, P., and Ryan, R. M. (2017). Motivating young language learners: a longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemp. Educ. Psychol.* 49, 140–150. doi: 10.1016/j.cedpsych.2017.01.010

Önal, N., Çevik, K. K., and Şenol, V. (2019). The effect of SOS table learning environment on mobile learning tools acceptance, motivation and mobile learning attitude in English language learning. *Interact. Learn. Environ.* 30, 834–847. doi: 10.1080/10494820.2019.1690529

Oxford, R. L. (2016). "Toward a psychology of well-being for language learners: the 'EMPATHICS' vision" in Positive psychology in SLA. eds. P. D. MacIntyre, T. Gregersen and S. Mercer (Bristol: Multilingual Matters), 10–88.

Oyserman, D., and Fryberg, S. (2006). "The possible selves of diverse adolescents: content and function across gender, race and national origin" in Possible selves: Theory, research and applications. eds. C. Dunkel and J. Kerpelman (New York: Nova Science Publishers), 17–40.

Pace, A., Luo, R., Hirsh-Pasek, K., and Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annu. Rev. Linguist.* 3, 285–308. doi: 10.1146/annurev-linguistics-011516-034226

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: a structural equation modeling approach. *System* 38, 467–479. doi: 10.1016/j.system.2010.06.011

Papi, M. (2018). Motivation as quality: regulatory fit effects on incidental vocabulary learning. *Stud. Second. Lang. Acquis.* 40, 707–730. doi: 10.1017/S027226311700033X

Papi, M., Bondarenko, A., Mansouri, S., Feng, L., and Jiang, C. (2019). Rethinking L2 motivation: the 2×2 model of self-guides. *Stud. Second. Lang. Acquis.* 41, 337–361. doi: 10.1017/S0272263118000153

Papi, M., and Hiver, P. (2020). Language learning motivation as a complex dynamic system: a global perspective of truth, control, and value. *Mod. Lang. J.* 104, 209–232. doi: 10.1111/modl.12624

Papi, M., and Teimouri, Y. (2014). Language learner motivational types: a cluster analysis study. *Lang. Learn.* 64, 493–525. doi: 10.1111/lang.12065

Paradowski, M. B., and Jelińska, M. (2024). The predictors of L2 grit and their complex interactions in online foreign language learning: motivation, self-directed learning, autonomy, curiosity, and language mindsets. *Comput. Assist. Lang. Learn.* 37, 2320–2358. doi: 10.1080/09588221.2023.2192762

Peng, J.-E. (2015). L2 motivational self system, attitudes, and affect as predictors of L2 WTC: an imagined community perspective. *Asia Pac. Educ. Res.* 24, 433–443. doi: 10.1007/s40299-014-0195-0

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879

Qasem, F., Alduais, A., Alfadda, H., Alfadda, N., and Amri, L. A. (2022). A study on the relationship between pragmatic language development and socioeconomic status in Arab preschoolers with and without pragmatic language impairment. *Sustain. For.* 14:6369. doi: 10.3390/su14106369

Qin, W., Jia, R., and Ren, W. (2024). Pragmatic competence in an email writing task: influences of situation, L1 background, and L2 proficiency. *Writ. Commun.* 41, 726–755. doi: 10.1177/07410883241263543

Ren, W. (2015). L2 pragmatic development in study abroad contexts. Bern: Peter Lang.

Ren, W., and Han, Z. (2016). The representation of pragmatic knowledge in recent ELT textbooks. *ELT J.* 70, 424–434. doi: 10.1093/elt/ccw010

Sadoughi, M., Hejazi, S. Y., and Lou, N. M. (2023). How do growth mindsets contribute to academic engagement in L2 classes? The mediating and moderating roles of the L2 motivational self system. *Soc. Psychol. Educ.* 26, 241–261. doi: 10.1007/s11218-022-09744-0

Saito, K., Dewaele, J. M., Abe, M., and In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: a cross-sectional and longitudinal study. *Lang. Learn.* 68, 709–743. doi: 10.1111/lang.12297

Sanfo, J. -B. M. B., and Malgoubri, I. (2023). Teaching quality and student learning achievements in Ethiopian primary education: How effective is instructional quality in closing socioeconomic learning achievement inequalities? *International Journal of Educational Development*, 99, 102759. doi: 10.1016/j.ijedudev.2023.102759

Sanjurjo, J. F., Blanco, J. M. A., and Fernández-Costales, A. (2018). Assessing the influence of socio-economic status on students' performance in content and language integrated learning. *System* 73, 16–26. doi: 10.1016/j.system.2017.09.001

Sasaki, M., Yoko, K., and Steven, J. R. (2017). The impact of normative environments on learner motivation and L2 reading ability growth. *Mod. Lang. J.* 101, 163–178. doi: 10.1111/modl.12381

Schauer, G. A. (2009). Interlanguage pragmatic development: The study abroad context. London: Continuum.

Schmidt, R. (1993). "Consciousness, learning and interlanguage pragmatics" in Interlanguage pragmatics. eds. G. Kasper and S. Blum-Kulka (New York: Oxford University Press), 21–42.

Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development*, 7, 379–392. doi: 10.1080/01434632.1986.9994254

Shin, H. W., and So, Y. (2018). The moderating role of socioeconomic status on motivation of adolescents' foreign language learning strategy use. *System* 73, 71–79. doi: 10.1016/j.system.2017.11.007

Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev. Educ. Res.* 75, 417–453. doi: 10.3102/00346543075003417

Sudina, E. (2021). Study and scale quality in second language survey research, 2009-2019: the case of anxiety and motivation. *Lang. Learn.* 71, 1149–1193. doi: 10.1111/lang.12468

Taguchi, T., Magid, M., and Papi, M. (2009). "The L2 motivational self system among Japanese, Chinese and Iranian learners of English: a comparative study" in Motivation, language identity and the L2 self. eds. Z. Dörnyei and E. Ushioda (Bristol: Multilingual Matters), 66–97.

Taguchi, N., and Roever, C. (2017). Second language pragmatics. Oxford: Oxford University Press.

Takahashi, S. (2005). Pragmalinguistic awareness: is it related to motivation and proficiency? *Appl. Linguis.* 26, 90–120. doi: 10.1093/applin/amh040

Takahashi, S. (2015). The effects of learner profiles on pragmalinguistic awareness and learning. *System* 48, 48–61. doi: 10.1016/j.system.2014.09.004

Trigueros, R., Aguilar-Parra, J. M., Lopez-Liria, R., Cangas, A. J., González, J. J., and Álvarez, J. F. (2020). The role of perception of support in the classroom on the students' motivation and emotions: the impact on metacognition strategies and academic performance in math and English classes. *Front. Psychol.* 10:2794. doi: 10.3389/fpsyg.2019.02794

Vonkova, H., Jones, J., Moore, A., Altinkalp, I., and Selcuk, H. (2021). A review of recent research in EFL motivation: research trends, emerging methodologies, and diversity of researched populations. *System* 103:102622. doi: 10.1016/j.system.2021.102622

Wang, Y., and Liu, H. (2022). The mediating roles of buoyancy and boredom in the relationship between autonomous motivation and engagement among Chinese senior high school EFL learners. *Front. Psychol.* 13:992279. doi: 10.3389/fpsyg.2022.992279

Wang, Y., and Ren, W. (2023). The roles of language mindsets and willingness to communicate in receptive pragmatic competence among Chinese EFL learners. *J. Multiling. Multicult. Dev.* 1–18, 1–17. doi: 10.1080/01434632.2023.2284892

Whitney, S. D., and Bergin, D. A. (2018). Students' motivation and engagement predict reading achievement differently by ethnic group. *J. Genet. Psychol.* 179, 357–370. doi: 10.1080/00221325.2018.1527754

Xu, J., and Feng, X. (2024). Mindsets, resilience and student engagement as predictors of L2 achievement among Chinese English learners: insights from fuzzy-set qualitative comparative analysis. *System* 124:103358. doi: 10.1016/j.system.2024.103358

Xu, W., Case, R. E., and Wang, Y. (2009). Pragmatic and grammatical competence, length of residence, and overall L2 proficiency. *System* 37, 205–216. doi: 10.1016/j.system.2008.09.007

Xu, J., and Wang, Y. (2022). The differential mediating roles of ideal and ought-to L2 writing selves between growth mindsets and self-regulated writing strategies. *System* 110:102900. doi: 10.1016/j.system.2022.102900

Xu, X., Xia, M., and Pang, W. (2021). Family socioeconomic status and Chinese high school students' test anxiety: serial mediating role of parental psychological control, learning resources, and student academic self-efficacy. *Scand. J. Psychol.* 62, 689–698. doi: 10.1111/sjop.12750

Xu, W., Zhang, H., Sukjairungwattana, P., and Wang, T. (2022). The roles of motivation, anxiety and learning strategies in online Chinese learning among Thai learners of Chinese as a foreign language. *Front. Psychol.* 13:962492. doi: 10.3389/fpsyg.2022.962492

Yamato, K., Kenji, T., and Takamichi, I. (2013). "Pragmatic awareness of Japanese EFL learners in relation to individual differences: a cluster analytic approach" in Pragmatics and language learning. eds. G. Tim, H. T. Donna and R. Carsten (Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i), 245–266.

Yang, H. (2022). Second language learners' competence of and beliefs about pragmatic comprehension: insights from the Chinese EFL context. *Front. Psychol.* 12:801315. doi: 10.3389/fpsyg.2021.801315

Yang, H., and Ren, W. (2020). Pragmatic awareness and second language learning motivation: a mixed-methods investigation. *Pragmat. Cogn.* 26, 447–473. doi: 10.1075/pc.19022.yan

You, S., Dang, M., and Lim, S. A. (2016). Effects of student perceptions of teachers' motivational behavior on reading, English, and mathematics achievement: the mediating role of domain specific self-efficacy and intrinsic motivation. *Child Youth Care Forum* 45, 221–240. doi: 10.1007/s10566-015-9326-x

Yousefi, M., and Mahmoodi, M. H. (2022). The L2 motivation self-system: a meta-analysis approach. *Int. J. Appl. Linguist.* 32, 274–294. doi: 10.1111/ijal.12416

Zheng, Y., and Mei, Z. (2021). Two worlds in one city: a sociopolitical perspective on Chinese urban families' language planning. *Curr. Issues Lang. Plan.* 22, 383–407. doi: 10.1080/14664208.2020.1751491

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 17 January 2025
DOI 10.3389/fpsyg.2024.1499994

OPEN ACCESS

EDITED BY
Valentina Cuccio,
University of Messina, Italy

REVIEWED BY
Ying Li,
Max Planck Institute for Human Development, Germany
Mihaela Buzec,
Babeș-Bolyai University, Romania

*CORRESPONDENCE
Antonio Benítez-Burraco
abenitez8@us.es

RECEIVED 22 September 2024
ACCEPTED 05 December 2024
PUBLISHED 17 January 2025

CITATION
Benítez-Burraco A (2025) How (and why) languages became more complex as we evolved more prosocial: the human self-domestication view.
Front. Psychol. 15:1499994.
doi: 10.3389/fpsyg.2024.1499994

How (and why) languages became more complex as we evolved more prosocial: the human self-domestication view

Antonio Benítez-Burraco*

Department of Spanish, Linguistics and Theory of Literature (Linguistics), Faculty of Philology, University of Seville, Seville, Spain

This paper aims to re-examine the problem of the emergence of present-day languages from the specific perspective of the self-domestication account of human evolution. According to this view, our species went through an evolutionary process that parallels the changes experienced by domesticated mammals. Relying on evidence of diverse kind (from paleogenetic to clinical), the paper argues that our self-domestication might have potentiated the cognitive and behavioral features of the human phenotype with an impact on language acquisition and use. Specifically, it might have facilitated the creation of the cultural niche that favors the complexification of languages via a cultural mechanism. The paper further proposes a model of language complexification in the past under the effects of human self-domestication, including the complexification of the structural aspects of language (grammar, prosody, and semantics) and the potentiation of its functional properties (pragmatics). The paper concludes with some suggestions for any future research aimed to test and improve this view.

KEYWORDS

language evolution, language structural complexity, language uses, prehistory, aggression, human self-domestication

1 Introduction

How language evolved and how present-day languages appeared have been a concern for human beings during millennia. In every human culture, one can find mythological accounts of why humans speak and why they speak the languages they speak. The idea that the emergence of language (and of modern-like languages) represents a true evolutionary leap forward, accounting for the success of the human species, is now widely acknowledged. In their famous paper about the complexification of life on Earth, Szathmáry and Smith (1995) regarded the evolution of language as the two final steps in this process: first, the emergence of a protolanguage without a true syntax, later the emergence of present-days languages, endowed with recursive grammars. Nowadays, language evolution is indeed a favorite topic for many disciplines, not only for linguistics, but also for archeology, paleoanthropology, or genetics, to name just a few. The same can be said of the dynamics followed by languages in our remote past, including the putative type of languages spoken by prehistoric societies, or the patterns of language diversity and change at that time. Explaining language evolution is a formidable task, as language does not fossilize. But it is a doable task now that researchers have massively adopted a truly multidisciplinary approach to this issue. Explaining the dynamics of languages during prehistory is also a great challenge. But it is likewise a doable task now that we have better tools for studying human evolution, and particularly, the changes in the physical environment and in social dynamics in the past. Contributing to this exciting enterprise is the main objective of this paper.

The paper is structured as follows. First, I will provide some background discussion about the relationships between language, languages, and uses of language. This is indeed a long-lasting debate in linguistics, psychology, cognitive science, and allied disciplines, but it is also crucial for understanding language change in the past. The field is progressively moving to more nuanced views of this issue, according to which languages coevolve with human cognition and behavior in response to environmental changes. Against this background, in the second part of the paper, which is the bulk of this contribution, I discuss in detail an original evolutionary model for human language(s) under the view that we evolved increasingly prosocial (aka the human self-domestication hypothesis). The paper finishes with some conclusions and prospects for future research.

2 A framework for language evolution studies

Let us begin with some basic clarifications. When we talk about *language evolution* our interest is not put on languages, like Russian, Spanish, or Japanese. It is put instead on our species-specific ability to learn and use these (and many others) languages. More technically, we wish to learn about the evolutionary trajectory of the biological foundations of our species' ability to spontaneously develop mental rule systems that are put to use in thought and communication. We can call this ability, which essentially equates to Saussure's *language*, our *faculty of language*. With time, some other denominations have been coined, like *language-readiness* (to stress that this ability mostly depends on our brain), or the one I will use in this paper, namely *human linguisticality*, a recent term proposed by the German linguist Haspelmath (2020). Under this view, our linguisticality is a cognitive ability, so that our attention should be drawn to our biology as our focus of inquiry. By contrast, languages should be understood as the collection of contingent properties of the communication/thought systems that humans eventually acquire as a result of social interactions. If you socialize with Japanese people, you acquire Japanese; if you are born in Spain, you usually acquire Spanish, and the like. Finally, people use their native language (or languages) to fulfil many different functions, like thinking, sharing information with others, socializing, persuading others, playing, and so on. Typically, this entails using your knowledge of the grammar of your language(s) to create utterances that fulfil such functions.

For many years, the mainstream view in the field of (evolutionary) linguistics has been that language can (and should) be construed as a human-specific cognitive faculty which is homogeneous in the species (pathological instances aside) and that resulted from biological changes mostly (e.g., Berwick and Chomsky, 2016, 2017, 2019). Likewise, most linguists agreed that all languages (present-day languages, but also prehistoric languages) are roughly equal in terms of their basic components, fundamental structure, overall complexity, and main functions (e.g., Dixon, 1997: 65–66; Fromkin et al., 2011: 375–374). The reason is that these core features were hypothesized to depend mostly on how our brain is configured and works. So, under this view, once our species emerged, our distinctive brain also emerged, and ultimately, present-day like languages emerged too. Lastly, as noted, people use their native languages to fulfil many functions. But since human behavior and human societies are not so different after all, these functions (at least the basic ones, like socializing, or conveying information) have been assumed to be quite similar worldwide. Accordingly, under this traditional view, causation goes in one direction only: from human linguisticality to languages to uses of language. The approach to language evolution by Noam Chomsky nicely exemplifies this view. According to him (e.g., Bolhuis et al., 2014; Berwick and Chomsky, 2016), language appeared suddenly as a result of one single gene mutation that caused a brain rewiring that brought about recursion, the distinctive feature of all human languages (and of our cognition). Chomsky further argued that this novel brain configuration has not changed since our inception and hence, that it is shared by all human beings (again, pathological instances aside). According to this view, prehistoric languages can be expected to have been quite similar to present-day languages, at least during the last 100.000 years. This refers, of course, to their fundamental properties, since Chomsky acknowledges that all languages do change with time, as when Spanish emerged from Latin. Here, the use of *change* instead of *evolve* stresses that the fundamental properties of languages are expected not to change historically, and accordingly, that the historical change of languages has no impact on the evolution of our linguisticality (see Mendívil-Giró, 2019 for discussion).

Over the years, however, this view has attracted increasing criticism. To begin with, language features can impact on our cognition. For instance, people speaking languages with an object-verb word order, like Japanese, are better at recalling initial items from a list, whereas speakers of final-object languages, like English, are better at recalling the last items from a list (Amici et al., 2019). A reason is that objects are usually a focus of attention and typically, the most informative part of a sentence. Hence, the language we speak conditions, even if subtly, the way in which we perceive the world and process and storage information about the world. More generally, planning to talk also biases our perception and the way in which we process data, because we need to accommodate the structural features of the language we are using. Still, this effect can be more profound, if the habitual encoding and use of such specific language features results in non-linguistic representational and even behavioral effects. Ultimately, aspects of languages that are more costly to process and learn might favor the creation of "cognitive gadgets" through permanent modifications in learning and data-acquisition mechanisms (Heyes, 2018). For example, one could argue that the cognitive device we use for reading is one of such gadgets. Contrary to language, we did not evolve for reading, but "parasite" instead several neuronal devices fulfilling other functions, most notably, the visual word form area, which recognizes visual patterns, as well as the phonological loop, involved in using sounds for conveying meanings (Dehaene and Cohen, 2007; Wandell and Le, 2017). Potentially, these cognitive gadgets could be "fixed" through, e.g., epigenetic inheritance, but this process takes time. Since it also takes time for such cultural innovations to spread and consolidate, these feedback effects posit a challenge to hypotheses arguing that language evolved abruptly.

Likewise, ample research suggests that languages are sensitive to the environment in which they are spoken. Quantitative approaches to phonological diversity have found, for instance, significant correlations between the degree of vocalism and tree coverage (Maddieson and Coupé, 2015). Accordingly, languages spoken in areas that are rich in forests exhibit a higher proportion of vowels, whereas languages spoken in open areas have more consonants. This is seemingly because sound propagates differently in different physical

environments: also animals adapt their calls to the medium in which they live (Ey and Fischer, 2009). Interestingly too, tonal languages like Chinese or Thai are usually found in tropical and subtropical regions. This uneven distribution is seemingly explained by the perturbations of phonation caused by desiccated ambient air, as typically experienced in drier and colder regions, which make tonality less efficient for conveying linguistic information (Everett et al., 2015; Roberts, 2018). To offer a last example, recent research has also found a positive correlation between sonority and local temperature, so that languages spoken in cold regions have on average more plosive and fricative sounds, whereas languages spoken in warm areas show more sounds with high sonority, like trills or nasals (Wang et al., 2023).

Not surprisingly, the effects of the social environment on language features are stronger. We are familiar with the impact of different social factors on linguistic diversity within a language. Hence, structural and functional differences can be found between the varieties of a language spoken by children vs. adults, by men vs. women, and the like. It is also widely acknowledged that the context of a conversation (who is speaking, what they speak about, what they speak for, and the like) also affects the structure and the pattern of usage of a language. Nonetheless, the effect of these factors on language features of interest from a typological perspective is more controversial. Most linguists would agree that the vocabularies of the world languages differ because vocabularies store relevant cultural features, which diverge from one society to another (Evans, 2003; Sharifian, 2014; Majid, 2015). Also, languages tend to grammaticalize, in different ways and to different degrees, aspects of the environment in which they are spoken. Quantitative approaches to this issue suggest that many grammatical features can be significantly affected by social factors. For example, the index of agglutination (that is, how complex a word is) negatively correlates with population size (Lupyan and Dale, 2010). Over time, diverse sociological, political, and cultural factors have been suggested to impact on the structure of languages, including the number of speakers, the degree of bilingualism, the tightness or the looseness of the social networks, the sociopolitical organization, or the number of adult learners of a language (Wray and Grace, 2007; McWhorter, 2011; Trudgill, 2011; Nettle, 2012; Sampson et al., 2009). A recent study using nearly 100 morphological and syntactic parameters from the World Atlas of Language Structure (WALS), as well as a dozen of cultural and sociopolitical features of human societies retrieved from D-Place, Ethnologue and Glottolog, has found evidence of an inverse correlation between morphological complexity and sociopolitical complexity, as well as a direct correlation between syntactic complexity and sociopolitical complexity (Chen et al., 2024), in line with the view that languages adapt to their social environment (Lupyan and Dale, 2016).

When one considers all the social factors with an impact on language structure together with the language features subject to variation, an interesting pattern emerges. On the one side, the languages spoken by isolated human groups living in small, close-knit communities with high proportions of native speakers usually exhibit larger sound inventories and complex phonotactics, opaque morphologies (with more irregularities and morpho-phonological constraints), limited semantic transparency (with abundance of idioms and idiosyncratic speech), reduced compositional structure, and less sophisticated syntactic devices. On the contrary, large and complex social networks, involving greater rates of inter-group contacts and cultural exchanges, seemingly favor languages with expanded vocabularies and increased syntactic complexity (including greater reliance on recursion). These languages also exhibit greater compositionality and enhanced semantic transparency, as well as simpler sound combinations and more regular morphologies. Overall, the difference between these two types of languages seemingly results from their differential context-dependency. In fact, the same pattern can be expected for different varieties of the same language, as the standard vs. the vernacular. And of course, this difference can be safely expected to be a matter of degree. In Chen et al.'s paper, (2024) the poles of this continuum are characterized, respectively, as esoteric (or S) languages and exoteric (or X) languages, but as noted, one can expect that esoteric and exoteric varieties of the same language, or more generally, esoteric and exoteric types of communication do exist. Linguistic esotericity is thus related to people sharing considerable amounts of knowledge, whereas linguistic exotericity involves using language in decontextualized settings.

Research on other domains of linguistics beyond language typology and sociolinguistics have converged onto this idea that language structure is sensitive to the environment (and particularly, to the social environment), to the extent that even core language features (that is, features thought to be universal and imposed by our cognitive hardware) can result from language learning and use. This is a robust conclusion, for instance, of studies using artificial grammars and involving iterated learning. In these experiments, core properties of language, like morphology, arise from a trade-off between pressures for compressibility and expressivity (Kirby et al., 2015). Compressibility is the tendency to capture systematic regularities in the form of abstract rules. For example, using the same prefix for all the words with a negative meaning, as in *impossible, improbable*, and the like. Expressivity is the capacity of providing a unique and unambiguous signal for every meaning, as with a list of proper nouns. Less compressible languages (like dictionaries) are more expressive, but more costly to learn. Conversely, compressible languages (like languages with a grammar) are easier to learn, but can incur ambiguity. Accordingly, speakers prefer compressible languages, whereas receivers favor expressive languages. As the relative strength of these two pressures typically changes from one social context to another, different social contexts can be expected to result in different language types endowed with different design properties.

Likewise, studies dealing with recently emerged sign languages reinforce the view that language structure, and even key design features of human language, can be sensitive to the social environment and ultimately, result from cultural evolution. In a language like Al-Sayyid Bedouin Sign Language (ABSL), core features like phonology, word order, or even recursion, develop with time in response to environmental triggers, like the kind and amount of input, the size of the community, or the degree of interaction among speakers (Sandler et al., 2005). Interestingly, this parallels what has been found in some oral languages. The celebrated Pirahã language could be a good example. According to the description of the language by Everett, Pirahã lacks recursion in the domain of complex sentence (Everett, 2005), with this reduced grammar complexity resulting from cultural constraints. These findings open the door to using our current knowledge of the social dynamics in the past for inferring basic aspects of the grammars of the languages spoken during prehistory, that are far beyond the limits of linguistic reconstructions as achieved by traditional methods in historical linguistics.

To complete the picture, one could also expect that social dynamics impact on human cognition (and ultimately on language) either

directly or indirectly, through their impact on language structure and use, as sketched above. As for the direct effect, Dunbar (1998, 2009) has argued that the human brain increased in size as we evolved more prosocial and human groups grew larger. Dunbar has further claimed (in, e.g., several of the chapters he authors in Dunbar et al., 2014) that a more sophisticated language capacity (and particularly, advanced storytelling abilities) might have favored the creation of these larger and more complex human groups, since narratives help to reduce social stress. Accordingly, while primates rely on grooming for managing social conflicts, humans have circumvented the limitations of grooming, which is more time-consuming (particularly, in the case of the big social groups we form), and use instead language to resolve conflicts and reinforce bonding. Like grooming, storytelling (but also other activities governed by language, such as feasting or religion) triggers the endorphin system and increases affiliative behaviors. To offer another example, changes resulting in increased joint attention or increased cooperation as human groups evolved larger and more complex can be expected to have improved our pragmatic and conversational abilities, in turn making grammar more sophisticated (see Ferretti, 2022 for details). As for the indirect effects of social dynamics on cognition through their impact on language structure and use, one could hypothesize, for instance, that people speaking S-languages exhibit a potentiated declarative memory compared to their procedural memory abilities. The reason is that the former is typically implicated in vocabulary learning and irregular phenomena across language domains, and it is thus most associated with memorized, opaque, formulaic chunks of language (for instance, idioms and proverbs), which are all more abundant features in this type of languages. By contrast, procedural memory is typically (although not exclusively) implicated in compositional, automated, rule-governed dimensions of language, which are all aspects found potentiated in X-languages (see Benítez-Burraco et al., 2022, and Chen et al., 2023 for more details). Nonetheless, other cognitive differences between the speakers of S and X-languages could be hypothesized to exist too, including differences in working memory, executive function, episodic memory, perception, emotion, or sensorimotor aspects of language.

To finish, the effect on cognition (via its impact on language structure) of the social aspects we are considering here (like network complexity, contacts with other groups, or ways of life) could also result from different social environment potentiating different uses or functions of language. For example, marking identity through language or using language for socializing (as in greetings) can be more important for close-knit groups, whereas sharing de-contextualized information and know-hows with strangers can be more familiar for speakers of X-languages. The structural features of the type of language used to fulfil these functions can be remarkably different. Commands and greetings typically involve short utterances, if not single words, whereas explanations about how the Solar System evolved, as found in a book of astronomy, typically demand long sentences with embedding, passives, and the like. Ultimately, favoring some language functions over others because of sociopolitical reasons can result not just in favoring some structural features of language over others, but also in cognitive differences, since different language functions impose different cognitive demands and entail different patterns of language processing by the brain, also because of the structural differences mentioned above. Typically, referential uses of compositionally-complex language recruit brain areas around the classical language network, which is mostly left-lateralized (Friederici et al., 2017), whereas less compositional linguistic items, like idioms, which are more frequent in informal, emotionally-charged uses of language, involve bilateral activation patterns of the language areas, with the recruitment of right areas (Hertrich et al., 2020). This is also true of figurative language, implicit meanings, background knowledge, discourse contexts, and pragmatic interpretations (Ferstl et al., 2008).

In summary, in sharp contrast with previous views of the emergence of modern languages as a direct outcome of brain changes, the complex links and feedback effects reviewed in this section between our cognition, our linguisticality, our behavior, the languages we speak, the uses we give to our languages, and the physical and the social environments in which we live (Figure 1), are suggestive of a different evolutionary scenario for present-day languages. Accordingly, one can expect that selected changes in our cognitive architecture and our behavior certainly improved the structural and functional properties of the languages spoken by archaic humans. But at the same time, our languages were certainly shaped by changes in our environment broadly construed, this including our physical environment, but particularly, the type of societies in which we lived, and the cultural niches we created. And to some important extent, these changes also impacted on our cognition and behavior, in turn affecting language structure and use.

The latter is a gradualist scenario for the emergence of modern languages. Compared to saltationist views, it is more in line with the gradual evolution of the human body and human behavior, as attested by paleoanthropological and archaeological research. For instance, there is evidence of a progressive globularization of the human skull/brain, with first archaic humans exhibiting anatomies (and presumably, functionalities) similar to those of late Neanderthals (Neubauer et al., 2018; Gunz et al., 2019). Since a globular brain has been related to our distinctive linguisticality (Boeckx and Benítez-Burraco, 2014), these progressive changes can be expected to have impacted on our language faculty, and ultimately, on the languages we spoke in our remote past. Likewise, evidence of modern behavior (like the use of pigments or ornaments, complex hunting strategies, advanced lithic technologies, or intentional burials) have appeared quite recently only (e.g., McBrearty and Brooks, 2000), with these behavioral innovations being suggestive of some cognitive changes too (Langley and Suddendorf, 2022), and with both types of modifications, cognitive and behavioral, impacting on our language abilities, as previously discussed. It is thus not speculative to hypothesize that more complex languages might have emerged in late prehistory in response to the biological and particularly, the cultural changes experienced by our species. The evolutionary framework of human self-domestication (HSD), which will be examined in detail in the next section, can successfully accommodate this progressive evolution of the human body, behavior, and culture, and emerges as a promising framework for the evolution of language as well.

3 Evolving more complex languages as we evolved more prosocial: the self-domestication view

In brief, HSD refers to a recent hypothesis about how our species emerged. It claims that the human distinctiveness is, to a large extent, the outcome of an evolutionary process similar to animal domestication (see Hare, 2017, or Hare and Woods, 2020,

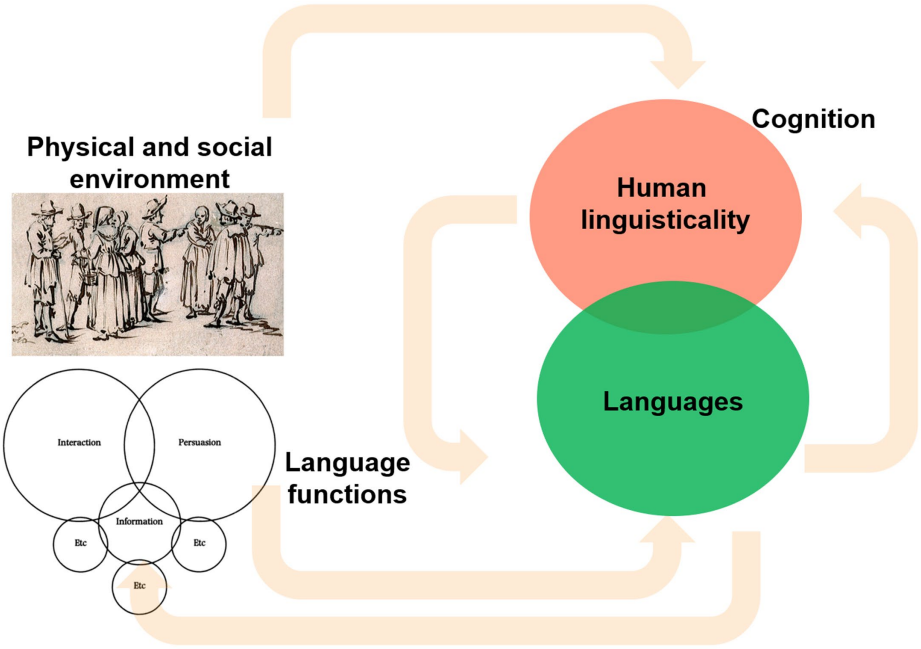


FIGURE 1
The expected links between human behavior and cognition (including our linguisticality), the languages we speak and the uses we give to them, and the physical and the social environments in which we live (own elaboration).

for an overview). In animals, domestication is usually triggered by selection for tameness, and in most cases results in a constellation of distinctive traits that are physical, cognitive, and behavioral by nature: the so-called domestication syndrome. It has been suggested that this is because tameness reduces the input to the neural crest, an embryonic structure that gives rise to many different body parts (Wilkins et al., 2014). Certainly, this view of animal domestication is not uncontroversial, mostly because not all domesticates show the whole suite of features purportedly encompassing the syndrome (see Sánchez-Villagra and van Schaik, 2019 or Lord et al., 2020 for critical views), but also because not all experts on animal domestication would agree that these traits result from the hypofunction of the neural crest (see Sánchez-Villagra et al., 2016 and Lord et al., 2020 for some criticism). That said, the hypothesis of HSD builds on the finding in humans of many of the traits commonly observed in domesticated varieties of mammals, including reduced skulls/brains, childish facial features, less hair, prolonged childhood, more time devoted to play, and particularly, a less aggressive behavior (Shea, 1989; Leach, 2003; Somel et al., 2009; Zollikofer and Ponce de León, 2010; Plavcan, 2012; Fukase et al., 2015; Stringer, 2016). Different factors might have triggered HSD: the rise of community living, the advent of co-parenting, changes in our foraging ecology, climate deterioration, and/or the colonization of new territories (see Pisor and Surbeck, 2019; Brooks and Yamamoto, 2021; Spikins et al., 2021; Raviv et al., 2023 for recent discussions). These factors would have promoted a selection toward less emotionally reactive partners and toward tolerance for extra-group individuals, resulting in increased cooperative behaviors. In turn, the hypothesis follows, the behavioral, cognitive, and even physical changes brought about by HSD would have promoted the emergence of many human distinctive features, including our enhanced social cognition, increased cooperation and extended social networks, and ultimately, our advanced technology and sophisticated culture.

Nonetheless, it was the finding that in some birds, domestication results in more complex communicative signals (e.g., Takahasi and Okanoya, 2010; Okanoya, 2017), that paved the way toward claims that HSD could be valuable in capturing key aspects of the evolution of language, specifically, those resulting from cultural evolution. If one recalls the complex interactions discussed in Section 2 between human cognition, human linguisticality, the languages we speak, and our physical and social environment, one could argue that HSD might have brought about both some of the physical, cognitive and behavioral changes with an impact on language structure and use, and the richer interactional niche favoring the complexification of language via a cultural mechanism. Regarding the physical changes, some of the modifications occurred in the human skull/brain and face during the last 100.000 years (as described by, e.g., Cieri et al., 2014), all with a potential impact on language, do resemble the changes in brain size or the snouts of domesticated animals, which typically show smaller brains and less prominent jaws. Likewise, the pigmentation changes usually associated with domestication could account for our distinctive white sclerae, which favors joint attention and face-to-face interactions (Wacewicz et al., 2022). The cognitive and behavioral changes are far more important for the complexification of language, as we discuss in depth in the subsections below. There, we sketch an evolutionary model for language under the forces of HSD. In truth, we expect this model to apprehend the effects on language structure and use of our contrasted trend toward a more prosocial behavior, even if the HSD hypothesis

eventually turns to be incorrect as such. The main reason is that in most scenarios, increased contacts between individuals can be safely expected to trigger the sort of behavioral and cognitive changes discussed below.

3.1 HSD and grammar

Figure 2 shows an outline of a model of grammar evolution under the effects of HSD (described in detail in Benítez-Burraco and Kempe, 2018; Progovac and Benítez-Burraco, 2019; Benítez-Burraco, 2020; Langley et al., 2020; and particularly, Benítez-Burraco and Progovac, 2020). As noted, the process encompasses four stages.

Stage 1 corresponds to the beginnings of HSD, roughly 300 kya (thousand years ago). Because reactive aggression was still high, communication through language could not have involved patient and cooperative turn-taking, using long utterances, as with modern languages, but just single-word commands, threats, and exclamations, mostly aimed to convey emotions. Notice that this is a characterization of the human languages purportedly spoken during that period, not a depiction of human linguisticality at that time. Even today, people can rely on simpler linguistic structures/systems in some circumstances, as observed in pidgins or restricted languages, but particularly, when they are angry and involved in verbal disputes.

As features of HSD increased, the cultural process that makes languages more complex seemingly increased too. Hence, reduced reactive aggression would have facilitated the establishment of stronger in-group networks, involving more diverse, frequent, and prolonged contacts between members. All these are factors that make language structure and language use more sophisticated. However, the potentiation of our language abilities, and particularly, the complexification of grammar, might have resulted not only from a cultural process like this, but also from selected brain changes brought about by HSD, as advanced above. One important reason is that the brain regions involved in the control of aggression are functionally connected to, or are in some cases partially overlapping with, the areas that are involved in language processing (see Miller et al., 2008 for discussion). Evidence of this is that the abnormal processing of language cues can result in misperceptions of emotional contents that trigger reactive aggression responses (Miller et al., 2008); and conversely, the activation of aggression responses can inactivate selected language areas, giving rise to abnormal language production and comprehension (Barratt et al., 1997). More specifically, there is evidence supporting the view that this increased control on aggressive responses would have been achieved, specifically, by enhancing the connectivity of the subcortical components of the circuit of aggression to selected cortical areas (see Benítez-Burraco and Progovac, 2021 for details). As discussed by, e.g., Lischinsky and Lin (2020), controlled aggression responses (as in learned aggressive actions) result from an increased control of the hypothalamus (part of the 'core aggression circuit') and the striatum (part of the 'learned aggression circuit') by the prefrontal cortex. However, the striatum is also a core component of the procedural memory system, and more generally, of the cortico-subcortical networks involved in grammar processing (Murphy et al., 2022). Therefore, it can be hypothesized that HSD enhanced this sort of functional connections, and even partial overlaps. In brief, the more cortical control of aggressive responses as HSD increased, the more potentiated language processing abilities and the more sophisticated

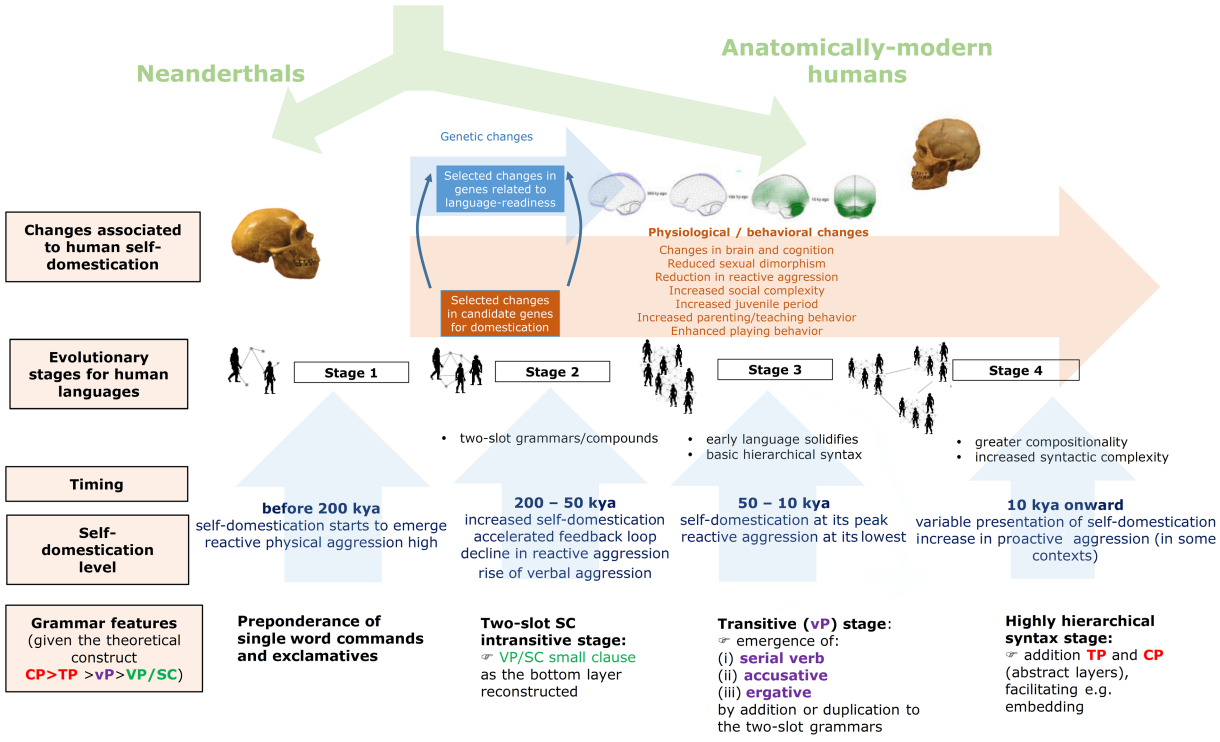


FIGURE 2
An outline of the evolution of grammar under the effect of HSD (reproduced from Benítez-Burraco and Progovac, 2020, Figure 1).

grammar… but also the richer cross-modal thought, as cross-modality also demands more connectivity between distant cortical and subcortical areas. As highlighted by many cognitive scientists, most notably Spelke (2003), human cognition excels at this ability of unifying and combining conceptual units belonging to different core knowledge systems. One intriguing possibility to be explored in the future is that both grammar and cross-modality (underlying, e.g., figurative uses of language, as in metaphors) depend on the same combinatorial ability and that this ability can be equated to the basic combinatorial operation in natural languages, which is called Merge by Chomskyan minimalism (Chomsky, 1995).

Returning to the model, this sort of cognitive and behavioral changes might have favored the transition to Stage 2, which might have spread between 200 kya and 100 kya, when the Last Glaciation began. For this stage, single word utterances might have started to be combined in a pair-wise fashion, leading to rudimentary two-slot grammars that would have employed nouns and verbs to express predications. An important use of these early grammars might have been the creation of derogatory compound expressions, which allowed to replace physical (reactive) aggression by verbal aggression (see Progovac and Benítez-Burraco, 2019 for details). In turn, this might have contributed to accelerate HSD, because of the common neurobiological mechanism supporting these three core dimensions of language: aggression, language processing, and cross-modality, as depicted above.

Around 100 kya, climate deteriorated notably and HSD exacerbated, reaching its peak around 50 kya as evidenced by the paleoanthropological record (Cieri et al., 2014). Increased cooperation would have enabled humans to survive during the Glacial Ages. This is Stage 3 in the model. The extremely low levels of reactive aggression during this period likely facilitated more frequent and more diverse contacts between children and adults, resulting in enhanced opportunities for teaching and learning. Increased HSD would have potentiated as well neotenic features in our species, this resulting in prolonged learning periods and increased play behavior. These are all factors that make language more complex, as noted enough. As a result, more sophisticated forms of grammar might have generalized, specifically, the first hierarchical grammars expressing transitivity. The most confident proxy of the languages spoken during Stage 3 are the languages of present-day hunter-gatherer human groups, and more generally, the Type S-languages discussed in Section 2.

As population size increased in response to cultural innovations and climatic changes, inter-group contacts generalized and extensive social networks emerged, relevant for trading and mating. Consequently, the necessity of exchanging information and know-hows with strangers also increased. This probably favored the advent of the second type of complex languages discussed in Section 2, namely, Type X-languages. This is the Stage 4 in the model, whose starting point it is tentatively situated 10 kya, during the transition period from the Paleolithic to the Neolithic. The advance of X-languages can be linked as well to the emergence of new forms of aggression, specifically, proactive (that is, premeditated) aggression, that became generalized during this period (see Wrangham, 2018, or Sarkar and Wrangham, 2023 for discussion). Since X-languages seem quite fit for conscious planning, they could have contributed to large-scale hostilities and escalated battles, and ultimately, to the emergence of cultural institutions around war and peace in complex societies (see Kissel and Kim, 2019 for a general discussion; see Meijer, 2024 for a HSD view). Neurobiologically, being a form of conscious aggression, proactive aggression demands even more control of the circuits of aggression by the cortex (Zhu et al., 2019, 2022), similarly to sophisticated syntax, a hallmark of Type-X languages. Another factor that might have contributed to the emergence of this type of languages was the increased number of people learning them as a second language. Whereas morphological complexity and morphological irregularities seem to be easy for children (allegedly, because they entail a lot of redundancy), they posit a learning problem for adults, who tend to simplify language morphology (and to compensate the less redundancy with their better pragmatic abilities and their more extensive general knowledge) and to potentiate the syntactic aspects of language (since they have more working memory resources) (see Dahl, 2004; Gil, 2009; Lupyan and Dale, 2010, Atkinson et al., 2018 among many others for further discussion).

In summary, this model of grammar change under the effects of HSD ties the different stages in the evolution of morphology and syntax with changes in the management of aggression, either reactive or proactive, and ultimately with the behavioral and cognitive changes brought about by HSD, with both aspects, namely, language features and HSD, being engaged in a mutually reinforcing feedback loop, since both aspects depend on, and impact on a common neurobiological substrate.

3.2 HSD and phonology

In a recent paper (Benítez-Burraco and Elvira-García, 2023), we have reasoned that this HSD framework could account as well for some other expected changes in the structure of human languages, particularly, in the domain of phonology. Phonology is essentially tied to speech; hence its evolution has been usually explained in terms of changes in the anatomy of speech organs (e.g., Barney et al., 2012; Conde-Valverde et al., 2021) and their control by the brain (e.g., Fitch et al., 2016; Brown et al., 2021). However, phonology is not totally detached from grammar. For instance, languages typically exhibit diverse morphophonological constraints. Also, phonological processes like ablaut or reduplication are commonly used for expressing grammatical features, like plurality or aspect. Our main hypothesis is that prosody, specifically, might have become complexified in parallel (and possibly, in a positive feedback loop) with the complexification of grammar in response to our increased HSD.

Diverse evidence supports this view:

- Although, as noted enough, HSD is more about changes in human interaction patterns than about major modifications in our bodies, HSD can be expected to have affected, even if subtly, the biological infrastructure of speech, with some potential impact on human prosody, since a main target of domestication is the craniofacial area (Geiger et al., 2022).
- As also noted at the beginning of this section, domestication is known to promote the complexification of sound signals in many species, particularly, birds.
- In all cultures, prosody is used for conveying emotional contents and for coordinating with others; as also noted enough, HSD is expected to have impacted both the control of emotions (via its effects on aggression management) and the way in which

we socialize with others (via its effects on interactional patterns between people).

- A neurobiological link between (the control of) prosody and (the control of) aggression can be safely expected. Hence, whereas prosodic anomalies resulting in misperceptions of emotional contents can produce impulsive outbursts, it also happens that a reduced control on aggressive responses impacts negatively on the processing of prosodic cues (see Miller et al., 2008 for details). In truth, the brain areas computing prosodic cues are connected to, and in some cases partially overlap with the regions controlling aggressive responses (see Miller et al., 2008 for details). As with grammar, one could expect a feedback effect between the control of prosody and the control of aggression. More specifically, since increased HSD essentially resulted in more cortical control on selected subcortical areas, the emergence of a true linguistic prosody might have resulted from increased top-down effects of the language-related areas on subcortical areas as we evolved more prosocial (consider that affective intonation is more dependent on subcortical regions, whereas linguistic prosody is more dependent on cortical regions; see Pihan, 2006; Wildgruber et al., 2006 for details). This effect would have been similar to the effect hypothesized for grammar. And in fact, quite an ample overlap exists between phrase-level prosody and syntax, with. e.g., prosodic features marking key sentence constituents and sentence types (Elfner, 2018). Likewise, prosody serves as a scaffolding for the acquisition of grammar (Brooks, 1997).

Overall, one could expect that as grammar gained complexity under the effects of HSD, prosody evolved more complex too, but also that the complexification of prosody contributed to make grammar more complex, mostly via the grammaticalization of selected prosodic features, as discussed below. Figure 3 summarizes our proposal. The model encompasses four stages, which parallel the four evolutionary stages hypothesized for grammar.

As noted in subsection 3.1., for Stage 1 we hypothesized simple languages expressing emotions through commands, threats, and exclamations, all of them consisting in single words. Because of physiological reasons, humans cannot produce strings of sounds that lack prosodic features: even the simplest vocalization shows pitch, duration, intensity and pause. In fact, we seem to be biologically programmed to detect differences in pitch contours, prosodic patterns, and stress patterns (Bhatara et al., 2018). Since all world languages convey emotions through changes in pitch and voice quality (Quinto et al., 2013; Wang and Lee, 2014), the languages spoken during this stage can be expected to have exhibited these prosodic features too. For this stage, one could also expect multisyllabic single-word utterances, since syllables and syllable composites have been found in the calls of other mammals, like bats (Chi et al., 2020). The finding that newborns are able to segment utterances into syllables supports this view too (Teinonen et al., 2009; Fló et al., 2019). For such multisyllabic words some sort of stress patterns could be hypothesized as well. Interestingly, in present-day tonal languages, pitch contours (or pitch ranges) can be employed to convey pragmatic meanings, even in the case of single words (Connell et al., 1983). One intriguing possibility is that some of the pitch and voice quality properties of the one-word utterances predicted for this Stage 1 became grammaticalized at some point, first as markers of the type of emotion conveyed by the utterance, but later as true pragmatic markers. Additionally, typological research has found evidence that complex pitch contours can trigger the creation of new morphemes, as observed in the vocatives of several languages (e.g., Sóskuthy and Roettger, 2020). This opens the door to the possibility that during Stage 1 some pitch contours associated to specific pragmatic functions, like warnings, favored the emergence of especific morphemes, thus contributing to the arrival of the two-slot grammars hypothesized for Stage 2 in the model.

As discussed in Section 3.1. for this Stage 2 we have hypothesized the emergence of the first grammars capable of combining single words into compounds. Since word prosody is found in all present-day languages (Nespor and Vogel, 2007) and because the ability to identify words relying on prosodic cues emerges early during ontogeny (e.g., Shukla et al., 2007; Bosch et al., 2013), the existence of a true word prosody can be confidently hypothesized for this Stage 2. At the same time, we expect that word prosody also contributed to the sophistication of grammar, not only to the emergence of the two-slot grammars hypothesized for this stage, as noted, but also to the transition to the three-slot grammars predicted for Stage 3. Notice that in present-day languages, asymmetric compounds consisting of a head and a dependent typically experience a loss of stress (see Hualde, 2007 and Rao, 2015 for Spanish; or Liberman and Sproat, 1992 for English). Accordingly, during Stage 2, some compounds might have experienced a similar loss of stress, this contributing to the advent of true prosodic phrasal markers and structures, which could have helped the

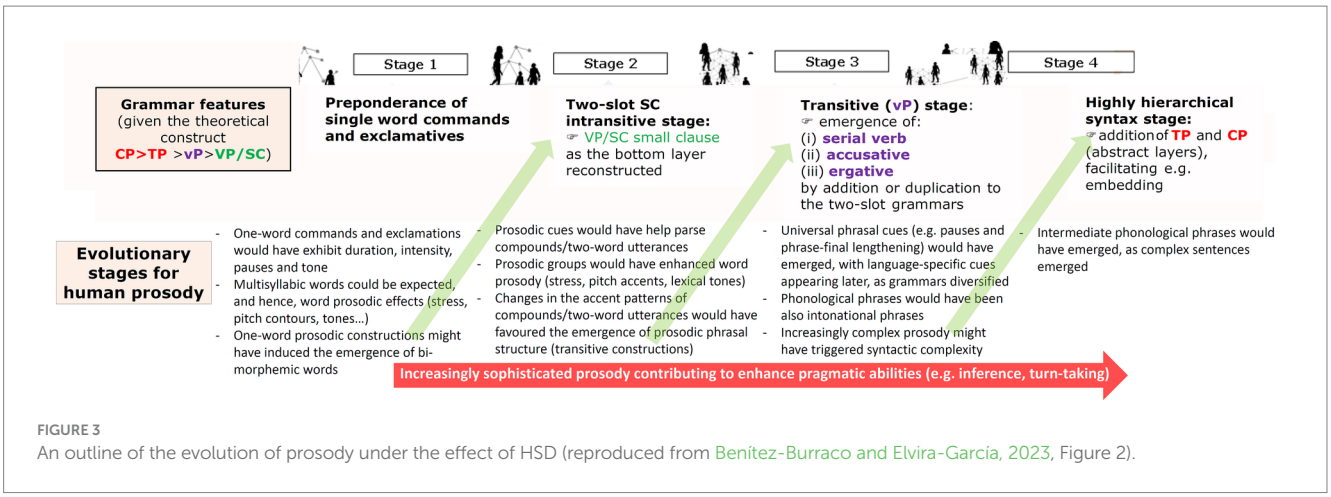


FIGURE 3
An outline of the evolution of prosody under the effect of HSD (reproduced from Benítez-Burraco and Elvira-García, 2023, Figure 2).

development of intransitive constructions, the hallmark of the languages spoken during Stage 2. Nonetheless, by the same reasons, this generalization of prosodic phrasal patterns could have facilitated the emergence of the transitive constructions predicted for the next stage, as we discuss below.

For Stage 3, we predicted the advent of the first hierarchical grammars expressing transitivity. The emergence of these grammars should have been paralleled by the emergence of the type of prosodic cues that present-day languages use for marking transitive sentences, like pauses between main constituents, or a pitch downstepping at the end of the utterance (Gussenhoven, 2002). As noted above, a scaffolding for these more complex prosodic markers might have been the type of phonological phrases emerged during Stage 2 from the prosodic reanalysis of compounds. In turn, a richer prosodic marking might have bootstrapped the more complex syntactic constructions hypothesized for this Stage 3, as observed during language acquisition by the child (Brusini et al., 2018; de Carvalho et al., 2018). That said, prosodic cues might have become more diverse and complex during this Stage 3 because pragmatics complexified during this stage too. As noted in Section 3.1, these Stage 3-languages can be roughly regarded type-S languages. But remember from Section 2, that type-S languages exhibit a notable contextual dependency, with a great amount of meaning being conveyed through idioms, implicatures, and references to the shared knowledge or the common ground. Prosody plays an important role in all this, particularly in distinguishing given or known information from new or pragmatically relevant information, usually through specific pitch patterns (Dennison and Schafer, 2010; Huang and Snedeker, 2018; Roettger et al., 2019). As we will discuss in detail in Section 3.4 below, HSD, which reached its peak during this period, might have contributed in more direct ways to the potentiation of our inferential abilities and to our advanced pragmatic capacities, more generally, through selected impacts on our cognition and behavior. To finish, as also suggested for Stages 1 and 2, this more complex prosody can be expected to have contributed to the sophistication of grammar, specifically to the generalization of embedding, and ultimately to the advent of the sort of languages hypothesized for Stage 4. One reason is that, as observed in some languages like Mohawk, at some initial point prosody can be the only marker of embedding, with full-fledged complement constructions appearing later (see Mithun, 2009 for discussion). Another reason is that prosody usually helps disambiguate between dependent and independent clauses when segmental contents are ambiguous (see Elvira-García et al., 2017 for discussion). Arguably, recursion could have been used during earlier stages, as in rhythmic patterns (morae, syllables or feet), or even in compounding, but our view is that it was only fully exploited at the sentence level during Stage 4.

Notice that in the discussion above we have mostly focused on upper-level categories (like phonological words, phonological phrases, intonational phrases and whole utterances), while we have largely ignored lower-level components, particularly rhythmic categories (like feet, syllables, or morae). The reason is that the latter can be expected to be evolutionarily older, since they have been found in many other species (see, e.g., Mann et al., 2021 on consonant-like and vowel-like sounds in birds).

3.3 HSD and semantics

There is ample evidence that other species can acquire and use symbols for referring to specific aspects of the environment, particularly, if properly trained (Seyfarth et al., 1980; Krause and Beran, 2020). Species closer to us can also use gestures in a context-dependent way for conveying different meanings according to the setting and/or the interlocutor, similarly to how humans use words (Hobaiter et al., 2022). Thinking of extinct hominins, one could thus safely hypothesize that they owned ampler and richer "vocabularies" compared to present-day primates, not only because of their bigger brains, but also because they exhibited more complex behaviors, social lives, and interactions with their environment. It is far beyond the scope of this review paper to discuss the purported features of other hominins' lexicons. Our interest is put instead on the narrower question of the potential impact, if any, of HSD on the semantics of the different languages spoken by anatomically-modern humans (henceforth, AMHs), with a focus on how their lexicons might have changed over time. Being necessarily speculative, the view that HSD did contribute to the diversification of the vocabularies of the languages spoken by our ancestors is not totally baseless. Because of the behavioral and cognitive changes brought about by HSD, of the sort discussed in previous subsections, one could argue that HSD favored, specifically, three processes that make vocabularies more complex: conceptual blending, categorization, and grammaticalization.

As noted enough, HSD seemingly enhanced our cross-modal thinking, which enables us to combine concepts belonging to different core knowledge systems. This can be expected to have enlarged the vocabularies of the first languages spoken by AMHs, particularly, the number of words denoting concepts without a real correlate (e.g., 'lion-man'). Moreover, some preexisting words might have enriched their meanings by gaining connotative senses, particularly, synesthetic values (i.e., idiosyncratic associations with other perceptual domains). The main reason is that synesthesia is a type of cross-modality, or more properly, as Cuskley and Kirby (2013: 871) puts it, of super cross-modal association.

Furthermore, as we reasoned in a recent paper (Benítez-Burraco et al., 2023), increased cross-modality might have contributed as well to the improvement of our categorization abilities, which are also relevant for vocabulary building. In truth, we have advocated for a feedback loop between the sophistication of our categorization abilities and the gradual emergence of syntactic structure, including Merge, of the sort discussed in Section 3.1 above. Our hypothesis is that our enhanced categorization abilities resulting from our increased cross-modal capability did not only result in more diverse categories, but also in more tokens within each category. Both types of diversity are necessary for Merge to take off in a systematic and productive way. In turn, this potentiation of our combinatorial abilities can be expected to have improved our categorization abilities, because of such an increase in the number of both categories and items within each category.

Improved categorization abilities certainly enable us to understand (and describe) the world in more accurate ways. Having said that, one type of categorization concerns word classes, which are a core structural aspect of languages. Over time, new grammatical categories can emerge though grammaticalization. Heine and Kuteva (2007) has proposed an appealing evolutionary model for language according to which the diversity of word classes, as found in present-day languages, might have resulted from labels for things, in essence, from noun(−like) symbols. Our contention here is that our increased HSD might have favored this trend too. Two lines of reasoning (discussed in depth in Benítez-Burraco, 2017) support this view. First, grammaticalization heavily depends on cognitive abilities like inferencing,

metaphorization, or metonymization that were seemingly potentiated by our enhanced cross-modal thinking, in turn resulting from our increased HSD. Second, grammaticalization is also triggered by social factors/needs, like expressing new types of social bounds, or creating new fashionable expressions (see Heine and Kuteva, 2007: 323–329 for details). The richer social environment and the increased sensibility to social cues brought about by our HSD should have favored this trend too.

One example of the emergence (or at least the spread) of one word class under the effects of HSD concerns ideophones. Ideophones are marked words that depict, in vivid and conventionalized ways, sensory images or events (Dingemanse, 2012, 2018, 2019). They are a word class in many languages, and are endowed with distinctive phonological, semantic, morphosyntactic, and pragmatic properties; but overall, they can be regarded as sound-symbolic words, and particularly, synesthetic words. In a recent paper (Di Paola et al., 2024), we have argued that because of this synesthetic-like nature, ideophones could not be very old, while at the same time, might have been more frequent in the past than nowadays. Specifically, and in line with our evolutionary model for language under the effects of HSD, we suggest that ideophones emerged (or at least, generalized) during Stage 2, but became less abundant during Stage 4. The main reason is that synesthesia(−like) effects demand a potentiated cross-modality, which, as discussed enough, might have emerged progressively under the effect of our HSD, which reached its peak during Stage 3. Supporting this view, synesthetes show increased abilities for understanding unfamiliar sound-symbolic words (Bankieris and Simner, 2015). As with other aspects of language, we support the view that the emergence of ideophones might have fostered the complexification of language, mostly because ideophones often come in reduplicative pairs (Dingemanse, 2019), so that they might have scaffolded the two-slot grammars hypothesized for Stage 2.

3.4 HSD and pragmatics

To finish this detailed characterization of the evolution of language under the effects of HSD, we will now provide some reasons why HSD might have contributed as well to the advent of modern uses of language, that is, modern pragmatics. To some extent, our pragmatic abilities can be expected to have improved as more elaborated grammatical, prosodic, and lexical resources emerged under the effects of HSD, as discussed in previous subsections. A more sophisticated grammar and a richer vocabulary allow the expression of one's thoughts in more precise ways, and thus the better defense of one's beliefs against the beliefs of others, this contributing to the optimization of persuasive reciprocity, as it is typically found in human conversations. Likewise, a richer prosody typically results in more abundant and varied pragmatic markers. In turn, as also argued in subsections 3.1 to 3.3, this improvement of our pragmatic capabilities might have helped the sophistication of the structural aspects of language. Nonetheless, in a recent paper (Benítez-Burraco et al., 2021a), we have argued that the potentiation of pragmatics might have also been a direct outcome of the behavioral and cognitive changes brought about by HSD, resulting in more sophisticated turn-takings, as well as more complex inferential capabilities, which are key aspects of our pragmatic abilities. Among the behavioral changes, two of them stand as particularly relevant. First, prolonged face-to-face interactions. Second, more cooperation and increased sensitivity to the needs of one's interlocutor. Both are at the heart of fine-tuned turn-taking. With regards to the cognitive changes, the most important one was likely the full emergence of our social brain. This has been claimed to result from the generalization of pair-bonds to other, non-reproductive relationships (Dunbar, 2009), and/or the potentiation of our evolutionary tendency toward social dependency for survival (Atzil et al., 2018). Both trends can be safely expected to have been fostered by a reduction in reactive aggression. However, the potentiation of cross-modal thinking, as we characterized it earlier, was relevant too, since cross-modality is central to figurative uses of language, such as metaphors and metonyms, and particularly, to pragmatic inferencing. The ultimate consequence of all these behavioral and cognitive changes was, we contend, that face-to-face interactions became more frequent and richer, with richer inferences and with more complex meanings being conveyed by more indirect means.

4 Conclusions and future prospects

In this review paper, we have supported the view that both the structural and the functional aspects of language might have co-evolved gradually in AMHs under the effects of HSD, with changes in aggression types and levels, and in language structure and use being intertwined in a complex feedback loop. Certainly, HSD is not the only factor accounting for the emergence of our distinctive linguisticality and the type of languages we speak nowadays, but it seems to be an important one. Compared to other evolutionary models for language, one strong point of this HSD account is that it acknowledges a stronger continuity between our linguisticality and the cognitive abilities and behaviors exhibited by other species. A second strong point is that it grants cultural niche construction, cultural evolution, and gene-culture co-evolution a more central role than others in the advent of modern languages.

There are several lines of (ongoing) research that could help test and improve this HSD approach to language evolution.

4.1 The timeline of HSD

One first aspect of interest is the timing and the presentation of the HSD phenotype during human history. As noted already several times, there is evidence that some of the physical features of HSD increased over time in our species, reaching its peak during Upper Paleolithic. It is certainly of relevance to know more about the starting point of the HSD process, the environmental factors that triggered and fostered it, and the precise stages it followed. Likewise, it is worth clarifying if other hominin species (particularly, Neanderthals) also went through a HSD process, even if less markedly, since there is evidence that other primates, specifically bonobos, have been self-domesticated (Hare et al., 2012). Neanderthals had quite rich social lives, but also exhibited reduced contacts with non-kin people (Skov et al., 2022; Slimak et al., 2024). Additionally, their physical features are only partially compatible with a self-domesticated phenotype. And they showed violence levels quite similar to those of early AMHs (Zollikofer et al., 2002). One promising way of addressing this question is delving into the molecular mechanism of HSD. In animals,

there is evidence that domestication entails (and is promoted by) selected genetic and epigenetic changes. Past research by Theofanopoulou et al. (2017) found a statistically significant overlap between genes showing evidence of selective sweeps in AMHs compared to Neanderthals and Denisovans, and genes selected in several domesticated species, particularly the dog, the cat, the horse, and the taurine cattle. However, this research involved present-day human genomes, so it was essentially inconclusive about the timing of the selection events. In turn, Benítez-Burraco et al. (2021b), using an improved list of genes selected in domesticated varieties of mammals, as well as European genomes from Late Neolithic/Bronze Age, found that candidates for mammal domestication have been accumulating nonsynonymous mutations during the past 6.000 years. This is a period when important changes in human behavior and culture occurred (including the spread of agricultural practices, sedentism and urbanization), when population density increased notably, and when long-distance trading routes developed. These changes reshaped not only the gene pool of Europe, but also modified its linguistic landscape, resulting in the nearly total replacement of the European hunter-gatherer languages by Indo-European languages. One possibility discussed by these authors is that the observed changes in genes related to HSD are a hallmark (and perhaps favored in some way) the emergence of proactive aggression within these populations and the transition from Type S-languages to Type X-languages in Europe. By the reasons discussed in section 2, acquiring a Type X-language might demand some cognitive adaptation, like the improvement of procedural memory abilities because of their more complex syntax. Interestingly, in these European samples from 6.000 years ago, previous research had found evidence of selection in two pathways related to cognition, particularly, to long-term potentiation, which underlies synaptic plasticity and ultimately, memory and learning abilities (Chekalin et al., 2019). Interestingly too, the domestication candidates showing signals of selection in AMHs compared to Neanderthals and Denisovans, as listed by Theofanopoulou and colleagues, do not overlap with the candidates showing signals of selection in Europeans during the last 6,000 years, as identified by Benítez-Burraco and colleagues. One possible explanation for this is that the former genes account for the milder HSD phenotype exhibited by early AMHs, particularly during Stages 1 and 2 in our model, whereas the latter genes are responsible for more recent HSD traits, as observed during Stage 3 and the transition to Stage 4.

4.2 Animal self-domestication and communication

A second line of research concerns the potential impact of self-domestication on the communication abilities of other species. As noted above, several species others than humans have been claimed to have gone through a self-domestication process. Bonobos are an outstanding example. Compared to chimpanzees, the structure of their vocalizations seems more complex (Taglialatela et al., 2018). Also, they use more indexical cues and acquire better linguistic skills in experimental settings (Savage-Rumbaugh et al., 1996; Gillespie-Lynch et al., 2014; MacLean and Hare, 2015). And there is evidence of a multimodal use of socially-directed calls by bonobos, but not by chimps (Genty et al., 2014). It would be interesting to prove that these distinctive features in the domain of communication correlate with (and can be explained by) the physical, behavioral, and cognitive changes brought about by their self-domestication, similarly to what we have hypothesized in Section 3 for human language. Genetic studies comparing the sequences of candidate genes for mammal domestication in bonobos and chimps would be also worth conducting. Likewise, it would be interesting to check whether the opposite is also true, namely, that species showing abilities that are crucial for developing complex communicative signals also exhibit signals of self-domestication. Actually, this seems to be the case. In an ongoing research, De Reus et al. (2024) have found that vocal learning mammals show most of the behavioral and cognitive hallmarks of self-domestication, including increased prosociality, exploratory behavior, and play behavior; more social tolerance; and sophisticated communication and information sharing abilities. Vocal learning is, of course, a prerequisite for human speech, and a key ability underlying language acquisition and development by children. Additional support to this view results from the examination of elephants. Elephants are well known for their advanced social skills, but they are also vocal leaners. In a recent paper, Raviv et al. (2023) found that, among other features, elephants exhibit reduced reactive aggression (the triggering factor of domestication and self-domestication), cortisol homeostasis sensitive to social factors (with changes in cortisol levels being a reliable biomarker of reactivity to stress), and an extended juvenile period (associated with increased exploratory play and resulting in enhanced social learning). Interestingly, when they compared the set of genes showing evidence of positive selection in elephants with the set of genes involved in mammal domestication, they found nearly 40 overlapping genes. According to the authors, several of them stand out as potential causative factors for some of the self-domestication features observed in elephants, as in humans they are related to communication problems, social and behavioral disturbances, or an altered presentation of HSD features. Interestingly too, they also found that the genes that have been positively selected in the elephant lineage are enriched in pathways likely related to domestication. Specifically, they observed enrichment in pathways involved in socialization and the management of aggression, including serotonin signaling and corticotropin signaling, the latter playing a key role in stress responses.

4.3 Feralization and communication

A third promising line of research concerns feral animals. Feralization is the process by which a once-domesticated animal returns to a wild-like state due to desocialization from humans, or the absence of human socializing pressures. Whereas domestication mostly results from selection for tameness, feralization usually involves the reactivation of mechanisms triggering reactive aggression. As discussed in detail by Niego and Benítez-Burraco (2022), not all domestication features are lost in feral animals, but some of them (particularly, prosocial behavior) are indeed reversed, seemingly because these traits are more sensitive to environmental changes. Significantly, the communication patterns and abilities exhibited by feral animals are halfway between those showed by domesticated animals and the ones observed in their wild counterparts. Typically, their communicative signals are less diverse and versatile, and are used in less varied social situations compared to domestic animals. In their

paper, Niego and Benítez-Burraco also compared the set of genes under positive selection in domesticated mammals with the set of genes positively selected in feral animals. Similarly to what can be observed at the phenotypic level, only a subset of candidate genes for domestication seems to be involved in feralization. Potentially, these genes could be related to the features of domestication (and of HSD) that are more sensitive to environmental factors.

4.4 HSD and language impairment

A fourth aspect of particular interest is the manifestation of features of HSD in cognitive disorders entailing language deficits. A strong link seems to exist between human evolution and human-specific diseases/conditions, with recently evolved phenotypic traits being more sensitive to developmental perturbations because of the reduced resilience exhibited by their biological components (see Gibson, 2009; Gibson and Lacek, 2020; Pattabiraman et al., 2020 for general discussions). Since HSD seems to be a recent human trait, one would expect that it is also particularly sensitive to developmental damage. Actually, this seems to be the case. Consider the case of autism spectrum disorders (ASD). These are notably prevalent conditions in all human populations, present-day and past populations, and entail diverse behavioral and cognitive disturbances (Bailey et al., 1996; Frith and Happé, 2005; Lord et al., 2020). These include problems with structural aspects of language, but also with figurative language, language use in conversational settings, and pragmatics more generally (Rapin and Dunn, 2003; Tager-Flusberg et al., 2005; Tager-Flusberg, 2006; Eigsti et al., 2007). Research has found ample evidence that candidate genes for ASD are enriched in genes positively selected in AMHs after our split from extinct hominins (Polimanti and Gelernter, 2017), supporting the link between human cognitive and behavioral evolution and human cognitive and behavioral impairment. Research by Benítez-Burraco et al. (2016) found, specifically, that features of HSD present attenuated in people within the ASD spectrum. Likewise, many candidates for mammal domestication are among the candidates for ASD and/or exhibit altered expression profiles in the brain of people with these conditions, whereas many candidate genes for ASD show signals of positive selection in domesticated animals.

Different hypotheses about the etiology of ASD have been formulated to date, including an altered Theory of Mind, a reduced sensitivity to social relations and social cues, or the effect of selected environmental factors on brain development (see discussions by Bölte et al., 2019; Steinman, 2020; Yoon et al., 2020; or Sauer et al., 2021, among many others). The same happens, specifically, with the language problems observed in autists, which have been claimed to result from diverse cognitive, behavioral, or even motor underlying deficits (Walenski et al., 2006; Preissler, 2008; Lindgren et al., 2009; Belmonte et al., 2013; Lampri et al., 2024). In a recent paper (Benítez-Burraco and Progovac, 2023), we have hypothesized that the language deficits typically observed in ASD, both structural and functional, might result, at least in part, from the alteration of the biological mechanism underlying HSD. Besides the deficits mentioned above, people within the ASD spectrum also feature an increased aptitude for rule-governed abilities, including language (Baron-Cohen et al., 2009; Ward et al., 2017). This seemingly explains their typically hyper-systemizing behavior, as observed, e.g., in their frequent over-regularization of past-tense forms (Baron-Cohen et al., 2009). Overall, the ASD phenotype could be characterized as "rigid": a rigid behavior, a rigid application of grammar rules, a rigid interpretation of non-literal language. As we discuss in the paper, this "rigidity" is suggestive of an increased striatal function. Nonetheless, typical language processing (and acquisition) results from a delicate balance between the application of rules and patterns (i.e., rigidity), and the ability to suspend such rules when exceptions need to be learnt or when metaphorical extension is needed to understand utterances (i.e., flexibility), with this balance resulting from the intervention of cortical areas on subcortical function (see Benítez-Burraco and Progovac, 2023 for details). Remember from subsection 3.1, that a functional connection, and seemingly a partial overlap exists between the brain mechanisms involved in the modulation of aggression and the mechanisms supporting language. In that subsection, we also highlighted that the feedback loop between HSD and language supported by this partially common neural substrate and resulting in richer language structures and richer pragmatic abilities might have involved an increased cortical control over selected subcortical areas. Accordingly, our proposal about the etiology of language dysfunction in ASD is that the "linguistic rigidity," and ultimately the increased striatal function and the reduced connection with the cortex found in this condition, could be a consequence of the abnormal presentation of HSD features. In our view, this possibility is supported by the finding that in other conditions resulting from striatal damage (like Tourette's syndrome) patients also exhibit rigidity with syntactic rules, as well as problems with figurative language (Eddy et al., 2010; Drury et al., 2018), but also increased reactive aggression (Ganos et al., 2014).

If our hypothesis is correct, it should not only help improve our understanding of the prevalence, etiology, and symptomatology of ASD, but also clarify the evolutionary history of the human linguisticality, because of the link between human disease and human evolution. Very broadly speaking, ASD would result from the attenuation of otherwise beneficial traits for language acquisition and use that were potentiated by our HSD, including enhanced cross-modal thinking and enhanced rule-governed systematicity (important for the evolution of the cognitive hardware of language), as well as enhanced control of aggression (important also for socialization and the cultural evolution of language). ASD could be further construed as a window to the earliest stages in the complexification of languages in our species under the forces of HSD, which featured a high degree of reactive aggression (and thus, a low degree of sociability), together with a diminished cross-modality (and thus, reduced metaphoricity and merging abilities). We have equally reasoned that another prevalent, human-specific cognitive disorder like schizophrenia (SZ) might provide a glimpse into a somewhat later stage in the complexification of human languages, when the higher disinhibited connectivity in the brain networks highlighted before would have resulted in exaggerated, super cross-modality (Stage 3 in our model) (see Benítez-Burraco and Progovac, 2021 for details). The reason is that in people with SZ, features of HSD are presented exaggerated at the anatomical, physiological, and behavioral levels (Benítez-Burraco, 2017). Likewise, nearly 20% of the candidate genes for mammal domestication are candidates for this condition, with around 75% of them being also differentially expressed in brain areas like the frontal cortex, the associative striatum, and the hippocampus, which exhibit structural and functional anomalies in schizophrenics, play a role in language processing, and show differences in domesticated animals

compared to their wild conspecifics (see Benítez-Burraco, 2017 for detailed discussion).

4.5 Animal models of HSD

An additional line of research aimed to provide support to the HSD hypothesis of language evolution concerns the development of animal models of HSD. Animal models allow to study in controlled ways the effects of the myriads of factors involved in development and disease (Phillips and Roth, 2019). Mammals that are known (or hypothesized) to have gone through a self-domestication process (like bonobos or elephants) are not idoneous for this type of research. In a recent paper (Anastasiadi et al., 2022), we have proposed to use the sea bass as an animal model for the genetic, and particularly the epigenetic changes underlying HSD. Sea basses can be bred in captivity very easily and we already have a good understanding of the genetic and epigenetic changes occurred during their domestication, with these changes parallelling the changes observed in domesticated mammals (Anastasiadi and Piferrer, 2019). Certainly, epigenetics can be expected to have played a crucial role in HSD, since epigenetic mechanisms are key to integrate environmental information and ultimately, to generate plastic responses to environmental changes. Most differences between AMHs and extinct hominins indeed pertain to the epigenome, with most of them affecting body parts known to be impacted by HSD and involved in language, particularly, the face (Gokhman et al., 2020). Moreover, epigenetic dysregulation is widely acknowledged as a key etiological factor of cognitive disorders (Gräff and Mansuy, 2009; Peedicayil and Grayson, 2018). However, an important limitation of current attempts to determine the impact of epigenetic changes in HSD is that we have access to fossil bones only. But epigenetic signals are different in the different body tissues. And in the case of HSD, soft tissues like the brain or the organs derived from the neural crest are more relevant than the bones. This is another reason to rely on animal models. In our research, we performed a comparison between known candidates for mammal domestication and genes differentially methylated in the domesticated sea bass. The significant overlap we found reinforces the view that the sea bass is a good animal model for mammal domestication, and arguably, for HSD too. Additional support to this view comes from our other finding, namely that a significant overlap exists between the genes showing epigenetic changes in early domesticates of the European sea bass and the genes exhibiting methylation changes in AMHs compared to Neanderthals and Denisovans. The overlapping genes are involved in processes like limb morphogenesis and in phenotypes like abnormal jaw morphology and hypopigmentation, which are related to the changes observed during domestication (and HSD). Finally, we found a significant overlap between the genes exhibiting epigenetic changes in early domesticates of the European sea bass and genes showing methylation changes in the brain of people with SZ as well as in the brain and the blood of subjects with ASD. Overlapping genes are involved in processes such as neural crest differentiation and ectoderm differentiation. This is suggestive of the sea bass also being a good model of human cognitive conditions entailing, and seemingly resulting from an abnormal HSD process.

4.6 HSD and language diversity

A final line of inquiry of potential interest for testing this HSD account of language evolution pertains the domain of linguistic typology. By the reasons provided in section 2, the languages spoken by present-day isolated human groups can be roughly characterized as strongly Type S-languages. As also noted in that section, such languages can be regarded as rough proxies for the languages spoken during Upper Paleolithic, particularly during Stage 3 in our model, when the effects of HSD seemingly reached their peak. Interestingly, some HSD features have been found to exhibit a variable presentation in present-day human populations (Gleeson and Kushnick, 2018). This is not surprising, since HSD is above all a response to changes in the human social environment. By all these reasons, it would be interesting to check whether, globally, speakers of Type S-languages (and more generally, users of esoteric communication) show more marked features of HSD, for example, a stronger feminization of the skull, as found in Upper Paleolithic populations (Cieri et al., 2014).

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Author contributions

AB-B: Conceptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by grant PID2023-147095NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF/EU.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Amici, F., Sánchez-Amaro, A., Sebastián-Enesco, C., Cacchione, T., Allritz, M., Salazar-Bonet, J., et al. (2019). The word order of languages predicts native speakers' working memory. *Sci. Rep.* 9:1124. doi: 10.1038/s41598-018-37654-9

Anastasiadi, D., and Piferrer, F. (2019). Epimutations in Developmental Genes Underlie the Onset of Domestication in Farmed European Sea Bass. *Mol. Biol. Evol.* 36, 2252–2264. doi: 10.1093/molbev/msz153

Anastasiadi, D., Piferrer, F., Wellenreuther, M., and Benítez Burraco, A. (2022). Fish as Model Systems to Study Epigenetic Drivers in Human Self-Domestication and Neurodevelopmental Cognitive Disorders. *Genes* 13:987. doi: 10.3390/genes13060987

Atkinson, M., Smith, K., and Kirby, S. (2018). Adult learning and language simplification. *Cogn. Sci.* 42, 2818–2854. doi: 10.1111/cogs.12686

Atzil, S., Gao, W., Fradkin, I., and Barrett, L. F. (2018). Growing a social brain. *Nat. Hum. Behav.* 2, 624–636. doi: 10.1038/s41562-018-0384-6

Bailey, A., Phillips, W., and Rutter, M. (1996). Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J. Child Psychol. Psychiatry* 37, 89–126. doi: 10.1111/j.1469-7610.1996.tb01381.x

Bankieris, K., and Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition* 136, 186–195. doi: 10.1016/j.cognition.2014.11.013

Barney, A., Martelli, S., Serrurier, A., and Steele, J. (2012). Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 367, 88–102. doi: 10.1098/rstb.2011.0259

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., and Chakrabarti, B. (2009). Talent in autism: Hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philos. Trans. Royal Soc. B* 364, 1377–1383. doi: 10.1098/rstb.2008.0337

Barratt, E. S., Stanford, M. S., Kent, T. A., and Felthous, A. (1997). Neuropsychological and cognitive psychophysiological substrates of impulsive aggression. *Biol. Psychiatry* 41, 1045–1061. doi: 10.1016/s0006-3223(96)00175-8

Belmonte, M. K., Saxena-Chandhok, T., Cherian, R., Muneer, R., George, L., and Karanth, P. (2013). Oral motor deficits in speech-impaired children with autism. *Front. Integr. Neurosci.* 7:47. doi: 10.3389/fnint.2013.00047

Benítez-Burraco, A. (2017). Grammaticalization and language evolution: focusing the debate. *Lang. Sci.* 63, 60–68. doi: 10.1016/j.langsci.2017.03.003

Benítez-Burraco, A. (2020). The self-domestication hypothesis of language evolution. *Paradigmi* 38, 221–235. doi: 10.30460/97895

Benítez-Burraco, A., Cahuana, C, Chen, S., Gil, D., Progovac, P., Reifegerste, J., et al. (2022) "Cognitive and genetic correlates of a single macro-parameter of crosslinguistic variation", The Evolution of Language, Proceedings of the 14th International Conference (EVOLANG14).

Benítez-Burraco, A., Chekalin, E., Bruskin, S., Tatarinova, T., and Morozova, I. (2021b). Recent selection of candidate genes for mammal domestication in Europeans and language change in Europe: a hypothesis. *Ann. Hum. Biol.* 48, 313–320. doi: 10.1080/03014460.2021.1936634

Benítez-Burraco, A., and Elvira-García, W. (2023). Human self-domestication and the evolution of prosody Language and Linguistics. *Compass* 17:e12485. doi: 10.1111/lnc3.12485

Benítez-Burraco, A., Ferretti, F., and Progovac, L. (2021a). Human Self-Domestication and the Evolution of Pragmatics. *Cogn. Sci.* 45:e12987. doi: 10.1111/cogs.12987

Benítez-Burraco, A., Hoshi, K., and Progovac, L. (2023). The gradual coevolution of syntactic combinatorics and categorization under the effects of human self-domestication: a proposal. *Cogn. Process.* 24, 425–439. doi: 10.1007/s10339-023-01140-6

Benítez-Burraco, A., and Kempe, V. (2018). The Emergence of Modern Languages: Has Human Self-Domestication Optimized Language Transmission? *Front. Psychol.* 9:551. doi: 10.3389/fpsyg.2018.00551

Benítez-Burraco, A., Lattanzi, W., and Murphy, E. (2016). Language Impairments in ASD Resulting from a Failed Domestication of the Human Brain. *Front. Neurosci.* 10:373. doi: 10.3389/fnins.2016.00373

Benítez-Burraco, A., and Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Lang. Commun.* 73, 1–17. doi: 10.1016/j.langcom.2020.03.002

Benítez-Burraco, A., and Progovac, L. (2021). Language evolution: examining the link between cross-modality and aggression through the lens of disorders. *Philos. Trans. R. Soc. B* 376:20200188. doi: 10.1098/rstb.2020.0188

Benítez-Burraco, A., and Progovac, L. (2023). Rigidity in autism spectrum disorder (ASD): A unified (evolutionary) account of linguistic and non-linguistic symptoms. *PsyArXiv*. doi: 10.31234/osf.io/euq5k

Berwick, R. C., and Chomsky, N. (2016). Why Only Us? Language and Evolution. Cambridge, MA: MIT Press.

Berwick, R. C., and Chomsky, N. (2017). Why only us: Recent questions and answers. *J. Neurolinguistics* 43, 166–177. doi: 10.1016/j.jneuroling.2016.12.002

Berwick, R. C., and Chomsky, N. (2019). All or nothing: No half-Merge and the evolution of syntax. *PLoS Biol.* 17:e3000539. doi: 10.1371/journal.pbio.3000539

Bhatara, A., Boll-Avetisyan, N., Höhle, B., and Nazzi, T. (2018). Chapter 3. Early sensitivity and acquisition of prosodic patterns at the lexical level. Amsterdam: John Benjamins, 37–57.

Boeckx, C., and Benítez-Burraco, A. (2014). The shape of the human language-ready brain. *Front. Psychol.* 5:282. doi: 10.3389/fpsyg.2014.00282

Bolhuis, J. J., Tattersall, I., Chomsky, N., and Berwick, R. C. (2014). How could language have evolved? *PLoS Biol.* 12:e1001934. doi: 10.1371/journal.pbio.1001934

Bölte, S., Girdler, S., and Marschik, P. B. (2019). The contribution of environmental exposure to the etiology of autism spectrum disorder. *Cellular Mol. Life Sci.* 76, 1275–1297. doi: 10.1007/s00018-018-2988-4

Bosch, L., Figueras, M., Teixidó, M., and Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: Evidence from infants acquiring syllable-timed languages. *Front. Psychol.* 4:106. doi: 10.3389/fpsyg.2013.00106

Brooks, P. J. (1997). The Prosodic Bootstrapping Hypothesis: How Distributional Analyses of Sound Patterns in Speech Facilitate Syntax Acquisition. *Contemp. Psychol.* 42, 505–506. doi: 10.1037/000292

Brooks, J., and Yamamoto, S. (2021). The founder sociality hypothesis. *Ecol. Evol.* 11, 14392–14404. doi: 10.1002/ece3.8143

Brown, S., Yuan, Y., and Belyk, M. (2021). Evolution of the speech-ready brain: The voice/jaw connection in the human motor cortex. *J. Comp. Neurol.* 529, 1018–1028. doi: 10.1002/cne.24997

Brusini, P., de Carvalho, A., Dautriche, I., Gutman, A., Cauvet, E., Millotte, S., et al. (2018). Chapter 3. Bootstrapping lexical and syntactic acquisition. Amsterdam: John Benjamins Publishing.

Chekalin, E., Rubanovich, A., Tatarinova, T. V., Kasianov, A., Bender, N., Chekalina, M., et al. (2019). Changes in Biological Pathways During 6,000 Years of Civilization in Europe. *Mol. Biol. Evol.* 36, 127–140. doi: 10.1093/molbev/msy201

Chen, S., Gil, D., Gaponov, S., Reifegerste, J., Yuditha, T., Tatarinova, T., et al. (2024). Linguistic correlates of societal variation: a quantitative analysis. *PLoS One* 19:e0300838. doi: 10.1371/journal.pone.0300838

Chen, S., Gil, D., Gaponov, S., Reifegerste, J., Yuditha, T., Tatarinova, T. V., et al. (2023). Linguistic and memory correlates of societal variation: A quantitative analysis. *PLoS ONE*. doi: 10.31234/osf.io/bnz2s

Chi, T., Liu, M., Tan, X., Sun, K., Jin, L., and Feng, J. (2020). Syllable merging during ontogeny in *Hipposideros larvatus*. *Bioacoustics* 29, 387–398. doi: 10.1080/09524622.2019.1610906

Chomsky, N. (1995). The minimalist program. Cambridge, MA: MIT press.

Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., and Hare, B. (2014). Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Curr. Anthropol.* 55, 419–443. doi: 10.1086/677209

Conde-Valverde, M., Martínez, I., Quam, R. M., Rosa, M., Velez, A. D., Lorenzo, C., et al. (2021). Neanderthals and *Homo sapiens* had similar auditory and speech capacities. *Nature Ecol. Evol.* 5, 609–615. doi: 10.1038/s41559-021-01391-6

Connell, B. A., Hogan, J. T., and Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *J. Phon.* 11, 337–351. doi: 10.1016/S0095-4470(19)30834-4

Cuskley, C., and Kirby, S. (2013). "Synesthesia, cross-modality, and language evolution" in Oxford handbook of synesthesia. eds. J. Simner and E. Hubbard (Oxford: Oxford University Press), 869–907.

Dahl, Ö. (2004). The growth and maintenance of linguistic complexity. Amsterdam: John Benjamins.

de Carvalho, A., Dautriche, I., Millotte, S., and Christophe, A. (2018). Chapter 2. Early perception of phrasal prosody and its role in syntactic and lexical acquisition. Amsterdam: John Benjamins Publishing.

de Reus, K., Benítez-Burraco, A., Hersh, T. A., Groot, N., Lambert, M. L., Slocombe, K. E., et al. (2024). Self-domestication traits in vocal learning mammals. In 15th International Conference on the Evolution of Language (EVOLANG) (pp. 105–108).

Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398. doi: 10.1016/j.neuron.2007.10.004

Dennison, H. Y., and Schafer, A. J. (2010). Online construction of implicature through contrastive prosody. Speech Prosody. Available online at: https://par.nsf.gov/biblio/10028989

Di Paola, G., Progovac, L., and Benítez-Burraco, A. (2024). Revisiting the hypothesis of ideophones as windows to language evolution. *Linguistics Vanguard*:127. doi: 10.1515/lingvan-2023-0127

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Lang. Linguistics Compass* 6, 654–672. doi: 10.1002/lnc3.361

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa J. Gen. Linguistics* 3. doi: 10.5334/gjgl.444

Dingemanse, M. (2019). ""Ideophone" as a comparative concept" in Ideophones, Mimetics, Expressives. eds. K. Akita and P. Pardeshi (Amsterdam: John Benjamins), 13–33.

Dixon, R. M. W. (1997). The Rise and Fall of Languages. Cambridge: Cambridge University Press.

Drury, H., Shah, S., Stern, J. S., Crawford, S., and Channon, S. (2018). Comprehension of direct and indirect sarcastic remarks in children and adolescents with Tourette's syndrome. *Child Neuropsychol.* 24, 490–509. doi: 10.1080/09297049.2017.1284777

Dunbar, R. I. M. (1998). The social brain hypothesis†. *Evol. Anthropol.* 6, 178–190. doi: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. (2009). The social brain hypothesis and its implications for social evolution. *Ann. Hum. Biol.* 36, 562–572. doi: 10.1080/03014460902960289

Dunbar, R., Gamble, C., and Gowlett, J. (Eds.) (2014). Lucy to Language. The Benchmark Papers. Oxford: Oxford University Press.

Eddy, C. M., Mitchell, I. J., Beck, S. R., Cavanna, A. E., and Rickards, H. E. (2010). Impaired comprehension of nonliteral language in Tourette syndrome. *Cogn. Behav. Neurol.* 23, 178–184. doi: 10.1097/WNN.0b013e3181e61cb7

Eigsti, I. M., Bennetto, L., and Dadlani, M. B. (2007). Beyond pragmatics: morphosyntactic development in autism. *J. Autism Dev. Disord.* 37, 1007–1023. doi: 10.1007/s10803-006-0239-2

Elfner, E. (2018). The syntax-prosody interface: current theoretical approaches and outstanding questions. *Linguistics Vanguard* 4:20160081. doi: 10.1515/lingvan-2016-0081

Elvira-García, W., Roseano, P., and Fernández-Planas, A. M. (2017). Prosody as a cue for syntactic dependency. Evidence from dependent and independent clauses with subordination marks in Spanish. *J. Pragmat.* 109, 29–46. doi: 10.1016/j.pragma.2016.12.002

Evans, N. (2003). Context, culture, and structuration in the languages of Australia. *Annu. Rev. Anthropol.* 32, 13–40. doi: 10.1146/annurev.anthro.32.061002.093137

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Curr. Anthropol.* 46, 621–646. doi: 10.1086/431525

Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proc. Natl. Acad. Sci.* 112, 1322–1327. doi: 10.1073/pnas.1417413112

Ey, E., and Fischer, J. (2009). The "acoustic adaptation hypothesis"—a review of the evidence from birds, anurans and mammals. *Bioacoustics* 19, 21–48. doi: 10.1080/09524622.2009.9753613

Ferretti, F. (2022). Narrative persuasion. A cognitive perspective on language evolution, vol. *7*. Berlin: Springer Nature.

Ferstl, E. C., Neumann, J., Bogler, C., and von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp.* 29, 581–593. doi: 10.1002/hbm.20422

Fitch, W. T., de Boer, B., Mathur, N., and Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Sci. Adv.* 2:e1600723. doi: 10.1126/sciadv.1600723

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., and Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Dev. Sci.* 22:e12802. doi: 10.1111/desc.12802

Friederici, A., Chomsky, N., Berwick, R. C., Moro, A., and Bolhuis, J. J. (2017). Language, mind and brain. *Nat. Hum. Behav.* 1, 713–722. doi: 10.1038/s41562-017-0184-4

Frith, U., and Happé, F. (2005). Autism spectrum disorder. *Curr. Biol.* 15, R786–R790. doi: 10.1016/j.cub.2005.09.033

Fromkin, V., Rodman, R., and Hyams, N. (2011). An Introduction to Language. *9th* Edn. Boston: Wadsworth, Cengage Learning.

Fukase, H., Kondo, O., and Ishida, H. (2015). Size and placement of developing anterior teeth in immature Neanderthal mandibles from Dederiyeh Cave, Syria: implications for emergence of the modern human chin. *Am. J. Phys. Anthropol.* 156, 482–488. doi: 10.1002/ajpa.22665

Ganos, C., Münchau, A., and Bhatia, K. (2014). The Semiology of Tics, Tourette's, and Their Associations. *Mov. Disord. Clin. Pract.* 1, 145–153. doi: 10.1002/mdc3.12043

Geiger, M., Sánchez-Villagra, M. R., and Sherratt, E. (2022). Cranial shape variation in domestication: A pilot study on the case of rabbits. *J. Exp. Zool. B Mol. Dev. Evol.* 338, 532–541. doi: 10.1002/jez.b.23171

Genty, E., Clay, Z., Hobaiter, C., and Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PLoS One* 9:e84738. doi: 10.1371/journal.pone.0084738

Gibson, G. (2009). Decanalization and the origin of complex disease. *Nat. Rev. Genet.* 10, 134–140. doi: 10.1038/nrg2502

Gibson, G., and Lacek, K. A. (2020). Canalization and Robustness in Human Genetics and Disease. *Annu. Rev. Genet.* 54, 189–211. doi: 10.1146/annurev-genet-022020-022327

Gil, D. (2009). "How much grammar does it take to sail a boat?" in Language complexity as an evolving variable. eds. G. Sampson, D. Gil and P. Trudgill (Oxford: Oxford University Press), 19–33.

Gillespie-Lynch, K., Savage-Rumbaugh, S., and Lyn, H. (2014). "Language learning in non-human primates" in Encyclopedia of Language Development. eds. P. J. Brooks and V. Kempe (Thousand Oaks, CA: SAGE Publications), 334–337.

Gleeson, B. T., and Kushnick, G. (2018). Female status, food security, and stature sexual dimorphism: Testing mate choice as a mechanism in human self-domestication. *Am. J. Phys. Anthropol.* 167, 458–469. doi: 10.1002/ajpa.23642

Gokhman, D., Nissim-Rafinia, M., Agranat-Tamir, L., Housman, G., García-Pérez, R., Lizano, E., et al. (2020). Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* 11:1189. doi: 10.1038/s41467-020-15020-6

Gräff, J., and Mansuy, I. M. (2009). Epigenetic dysregulation in cognitive disorders. *Eur. J. Neurosci.* 30, 1–8. doi: 10.1111/j.1460-9568.2009.06787.x

Gunz, P., Tilot, A. K., Wittfeld, K., Teumer, A., Shapland, C. Y., van Erp, T. G. M., et al. (2019). Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity. *Curr. Biol.* 29, 120–127.e5. doi: 10.1016/j.cub.2018.10.065

Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and Phonology. Available online at: http://20.210-193-52.unknown.qala.com.sg/archive/sp2002/papers/sp02_047.pdf

Hare, B. (2017). Survival of the Friendliest: *Homo sapiens* Evolved via Selection for Prosociality. *Annu. Rev. Psychol.* 68, 155–186. doi: 10.1146/annurev-psych-010416-044201

Hare, B., Wobber, V., and Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Anim. Behav.* 83, 573–585. doi: 10.1016/j.anbehav.2011.12.007

Hare, B., and Woods, V. (2020). Survival of the friendliest: Understanding our origins and rediscovering our common humanity. Random House, NY: Oneworld Publications.

Haspelmath, M. (2020). Human Linguisticality and the Building Blocks of Languages. *Front. Psychol.* 10:3056. doi: 10.3389/fpsyg.2019.03056

Heine, B., and Kuteva, T. (2007). The genesis of grammar: A reconstruction, vol. *9*. Oxford: Oxford University Press.

Hertrich, I., Dietrich, S., and Ackermann, H. (2020). The margins of the language network in the brain. *Front. Commun.* 5:519955. doi: 10.3389/fcomm.2020.519955

Heyes, C. (2018). Cognitive Gadgets: The Cultural Evolution of Thinking. Cambridge: Harvard University Press.

Hobaiter, C., Graham, K. E., and Byrne, R. W. (2022). Are ape gestures like words? Outstanding issues in detecting similarities and differences between human language and ape gesture. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 377:20210301. doi: 10.1098/rstb.2021.0301

Hualde, J. I. (2007). Stress removal and stress addition in Spanish. *J. Portuguese Linguistics* 6:59. doi: 10.5334/JPL.145

Huang, Y. T., and Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cogn. Psychol.* 102, 105–126. doi: 10.1016/j.cogpsych.2018.01.004

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016

Kissel, M., and Kim, N. C. (2019). The emergence of human warfare: Current perspectives. *Am. J. Phys. Anthropol.* 168, 141–163. doi: 10.1002/ajpa.23751

Krause, M. A., and Beran, M. J. (2020). Words matter: Reflections on language projects with chimpanzees and their implications. *Am. J. Primatol.* 82:e23187. doi: 10.1002/ajp.23187

Lampri, S., Peristeri, E., Marinis, T., and Andreou, M. (2024). Figurative language processing in autism spectrum disorders: A review. *Autism Res.* 17, 674–689. doi: 10.1002/aur.3069

Langley, M. C., Benítez-Burraco, A., and Kempe, V. (2020). Playing with language, creating complexity: Has play contributed to the evolution of complex language? *Evol. Anthropol.* 29, 29–40. doi: 10.1002/evan.21810

Langley, M. C., and Suddendorf, T. (2022). Archaeological evidence for thinking about possibilities in hominin evolution. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 377:20210350. doi: 10.1098/rstb.2021.0350

Leach, H. M. (2003). Human domestication reconsidered. *Curr. Anthropol.* 44, 349–368. doi: 10.1086/368119

Liberman, M., and Sproat, R. (1992). "The stress and structure of modified noun phrases in English" in Lexical matters– CSLI Lecture Notes No. 24. eds. I. A. Sag and A. Szabolcsi (Stanford: CSLI), 131–181.

Lindgren, K., Folstein, S., Tomblin, J. B., and Tager-Flusberg, H. (2009). Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives. *Autism Res.* 2, 22–38. doi: 10.1002/aur.63

Lischinsky, J. E., and Lin, D. (2020). Neural mechanisms of aggression across species. *Nat. Neurosci.* 23, 1317–1328. doi: 10.1038/s41593-020-00715-2

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., et al. (2020). Autism spectrum disorder. *Nat. Rev. Dis. Primers* 6:138. doi: 10.1038/s41572-019-0138-4

Lord, K. A., Larson, G., Coppinger, R. P., and Karlsson, E. K. (2020). The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends Ecol. Evol.* 35, 125–136. doi: 10.1016/j.tree.2019.10.011

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5:e8559. doi: 10.1371/journal.pone.0008559

Lupyan, G., and Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends Cogn. Sci.* 20, 649–660. doi: 10.1016/j.tics.2016.07.005

MacLean, E. L., and Hare, B. (2015). Bonobos and chimpanzees exploit helpful but not prohibitive gestures. *Behaviour* 152, 493–520. doi: 10.1163/1568539x00003203

Maddieson, I., and Coupé, C. (2015). Human spoken language diversity and the acoustic adaptation hypothesis. *J. Acoust. Soc. Am.* 138:1838. doi: 10.1121/1.4933848

Majid, A. (2015). Cultural factors shape olfactory language. *Trends Cogn. Sci.* 19, 629–630. doi: 10.1016/j.tics.2015.06.009

Mann, D. C., Fitch, W. T., Tu, H. W., and Hoeschele, M. (2021). Universal principles underlying segmental structures in parrot song and human speech. *Sci. Rep.* 11:776. doi: 10.1038/s41598-020-80340-y

McBrearty, S., and Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J. Hum. Evol.* 39, 453–563. doi: 10.1006/jhev.2000.0435

McWhorter, J. H. (2011). Linguistic simplicity and complexity: Why do languages undress? vol. *1*. Berlin: Walter de Gruyter.

Meijer, H. (2024). Janus faced: The co-evolution of war and peace in the human species. *Evol. Anthropol.* 33:e22027. doi: 10.1002/evan.22027

Mendívil-Giró, J. L. (2019). Did language evolve through language change? On language change, language evolution and grammaticalization theory. *Glossa* 4:895. doi: 10.5334/gjgl.895

Miller, L. A., Collins, R. L., and Kent, T. A. (2008). Language and the modulation of impulsive aggression. *J. Neuropsychiatry Clin. Neurosci.* 20, 261–273. doi: 10.1176/jnp.2008.20.3.261

Mithun, M. (2009). "Re (e) volving complexity: Adding intonation" in Syntactic Complexity: Diachrony, acquisition, neuro-cognition, evolution. eds. T. Givón and M. Shibatani (Amsterdam: John Benjamins Publishing), 53–80.

Murphy, E., Hoshi, K., and Benítez-Burraco, A. (2022). Subcortical syntax: Reconsidering the neural dynamics of language. *J. Neurolinguistics* 62:101062. doi: 10.1016/j.jneuroling.2022.101062

Nespor, M., and Vogel, I. (2007). Prosodic Phonology. Berlin: Walter de Gruyter.

Nettle, D. (2012). Social scale and structural complexity in human languages. *Philos. Trans. R. Soc. B* 367, 1829–1836. doi: 10.1098/rstb.2011.0216

Neubauer, S., Hublin, J. J., and Gunz, P. (2018). The evolution of modern human brain shape. *Sci. Adv.* 4:eaao5961. doi: 10.1126/sciadv.aao5961

Niego, A., and Benítez-Burraco, A. (2022). Are feralization and domestication truly mirror processes? *Ethol. Ecol. Evol.* 34, 557–590. doi: 10.1080/03949370.2021.1975314

Okanoya, K. (2017). Sexual communication and domestication may give rise to the signal complexity necessary for the emergence of language: An indication from songbird studies. *Psychon. Bull. Rev.* 24, 106–110. doi: 10.3758/s13423-016-1165-8

Pattabiraman, K., Muchnik, S. K., and Sestan, N. (2020). The evolution of the human brain and disease susceptibility. *Curr. Opin. Genet. Dev.* 65, 91–97. doi: 10.1016/j.gde.2020.05.004

Peedicayil, J., and Grayson, D. R. (2018). An epigenetic basis for an omnigenic model of psychiatric disorders. *J. Theor. Biol.* 443, 52–55. doi: 10.1016/j.jtbi.2018.01.027

Phillips, N. L. H., and Roth, T. L. (2019). Animal models and their contribution to our understanding of the relationship between environments, epigenetic modifications, and behavior. *Genes* 10:47. doi: 10.3390/genes10010047

Pihan, H. (2006). Affective and linguistic processing of speech prosody: DC potential studies. *Prog. Brain Res.* 156, 269–284. doi: 10.1016/S0079-6123(06)56014-5

Pisor, A. C., and Surbeck, M. (2019). The evolution of intergroup tolerance in nonhuman primates and humans. *Evol. Anthropol.* 28, 210–223. doi: 10.1002/evan.21793

Plavcan, J. M. (2012). Sexual size dimorphism, canine dimorphism, and male-male competition in primates. *Hum. Nat.* 23, 45–67. doi: 10.1007/s12110-012-9130-3

Polimanti, R., and Gelernter, J. (2017). Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genet.* 13:e1006618. doi: 10.1371/journal.pgen.1006618

Preissler, M. A. (2008). Associative learning of pictures and words by low-functioning children with autism. *Autism* 12, 231–248. doi: 10.1177/1362361307088753

Progovac, L., and Benítez-Burraco, A. (2019). From physical aggression to verbal behavior: language evolution and self-domestication feedback loop. *Front. Psychol.* 10:2807. doi: 10.3389/fpsyg.2019.02807

Quinto, L., Thompson, W. F., and Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Front. Psychol.* 4:184. doi: 10.3389/FPSYG.2013.00184/BIBTEX

Rao, R. (2015). On the phonological status of Spanish compound words. *Word Structure* 8, 84–118. doi: 10.3366/WORD.2015.0074

Rapin, I., and Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain Dev.* 25, 166–172. doi: 10.1016/S0387-7604(02)00191-2

Raviv, L., Jacobson, S. L., Plotnik, J. M., Bowman, J., Lynch, V., and Benítez-Burraco, A. (2023). Elephants as an animal model for self-domestication. *Proc. Natl. Acad. Sci. USA* 120:e2208607120. doi: 10.1073/pnas.2208607120

Roberts, S. G. (2018). Robust, causal, and incremental approaches to investigating linguistic adaptation. *Front. Psychol.* 9:166. doi: 10.3389/fpsyg.2018.00166

Roettger, T. B., Mahrt, T., and Cole, J. (2019). Mapping prosody onto meaning–the case of information structure in American English. *Lang. Cogn. Neurosci.* 34, 841–860. doi: 10.1080/23273798.2019.1587482

Sampson, G., Gil, D., and Trudgill, P. (Eds.) (2009). Language complexity as an evolving variable, vol. *13*. Oxford: Oxford University Press.

Sánchez-Villagra, M. R., Geiger, M., and Schneider, R. A. (2016). The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. *R. Soc. Open Sci.* 3:160107. doi: 10.1098/rsos.160107

Sánchez-Villagra, M. R., and van Schaik, C. P. (2019). Evaluating the self-domestication hypothesis of human evolution. *Evol. Anthropol.* 28, 133–143. doi: 10.1002/evan.21777

Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proc. Natl. Acad. Sci.* 102, 2661–2665. doi: 10.1073/pnas.0405448102

Sarkar, A., and Wrangham, R. W. (2023). Evolutionary and neuroendocrine foundations of human aggression. *Trends Cogn. Sci.* 27, 468–493. doi: 10.1016/j.tics.2023.02.003

Sauer, A. K., Stanton, J. E., Hans, S., and Grabrucker, A. M. (2021). "Autism Spectrum Disorders: Etiology and Pathology" in Autism spectrum disorders. ed. A. M. Grabrucker (Brisbane City, QLD: Exon Publications).

Savage-Rumbaugh, E. S., Williams, S. L., Furuichi, T., and Kano, T. (1996). "Language perceived: *pan paniscus* branches out" in Great Ape Societies. eds. C. Mcgrew, L. F. Marchant and T. Nishida (Cambridge: Cambridge University Press), 173–184.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* 28, 1070–1094. doi: 10.1016/S0003-3472(80)80097-2

Sharifian, F. (2014). "Language and culture: Overview" in The Routledge handbook of language and culture (New York: Routledge), 3–17.

Shea, B. (1989). Heterochrony in human evolution: the case for neoteny reconsidered. *Am. J. Phys. Anthropol.* 32, 69–101. doi: 10.1002/ajpa.1330320505

Shukla, M., Nespor, M., and Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cogn. Psychol.* 54, 1–32. doi: 10.1016/j.cogpsych.2006.04.002

Skov, L., Peyrégne, S., Popli, D., Iasi, L. N. M., Devièse, T., Slon, V., et al. (2022). Genetic insights into the social organization of Neanderthals. *Nature* 610, 519–525. doi: 10.1038/s41586-022-05283-y

Slimak, L., Vimala, T., Seguin-Orlando, A., Metz, L., Zanolli, C., Joannes-Boyau, R., et al. (2024). Long genetic and social isolation in Neanderthals before their extinction. *Cell Genomics* 4:100593. doi: 10.1016/j.xgen.2024.100593

Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., et al. (2009). Transcriptional neoteny in the human brain. *Proc. Natl. Acad. Sci. USA* 106, 5743–5748. doi: 10.1073/pnas.0900544106

Sóskuthy, M., and Roettger, T. B. (2020). When the tune shapes morphology: The origins of vocatives. *J. Lang. Evol.* 5, 140–155. doi: 10.1093/JOLE/LZAA007

Spelke, E. S. (2003). "What makes us smart? Core knowledge and natural language" in Language in Mind: Advances in the Study of Language and Thought (Cambridge, MA: MIT Press), 277–311.

Spikins, P., French, J. C., John-Wood, S., and Dytham, C. (2021). Theoretical and methodological approaches to ecological changes, social behaviour and human intergroup tolerance 300,000 to 30,000 BP. *J. Archaeol. Method Theory* 28, 53–75. doi: 10.1007/s10816-020-09503-5

Steinman, G. (2020). The putative etiology and prevention of autism. *Prog. Mol. Biol. Transl. Sci.* 173, 1–34. doi: 10.1016/bs.pmbts.2020.04.013

Stringer, C. (2016). The origin and evolution of *Homo sapiens*. *Philos. Trans. Royal Society B Biol. Sci.* 371:20150237. doi: 10.1098/rstb.2015.0237

Szathmáry, E., and Smith, J. M. (1995). The major evolutionary transitions. *Nature* 374, 227–232. doi: 10.1038/374227a0

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clin. Neurosci. Res.* 6, 219–224. doi: 10.1016/j.cnr.2006.06.007

Tager-Flusberg, H., Paul, P., and Lord, C. (2005). "Language and communication in autism" in Handbook of autism and pervasive developmental disorders. eds. F. R. Volkmar, A. Klin, R. Paul and D. J. Cohen. *3rd* ed (Hoboken, NJ: Wiley), 335–364.

Taglialatela, J. P., Milne, S. C., and Evans, R. E. (2018). "A Comparison of the Socio-communicative Behavior in Chimpanzees and Bonobos" in Evolution of Primate Social Cognition. Interdisciplinary Evolution Research. eds. L. D. Di Paolo, F. Di Vincenzo and F. De Petrillo, vol. *5* (Cham: Springer).

Takahasi, M., and Okanoya, K. (2010). Song learning in wild and domesticated strains of white-rumped munia, *Lonchura striata*, compared by cross-fostering procedures: domestication increases song variability by decreasing strain-specific bias. *Ethology* 116, 396–405. doi: 10.1111/j.1439-0310.2010.01761.x

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10:21. doi: 10.1186/1471-2202-10-21

Theofanopoulou, C., Gastaldon, S., O'Rourke, T., Samuels, B. D., Martins, P. T., Delogu, F., et al. (2017). Self-domestication in *Homo sapiens*: Insights from comparative genomics. *PLoS One* 12:e0185306. doi: 10.1371/journal.pone.0185306

Trudgill, P. (2011). Sociolinguistic typology: social determinants of linguistic complexity. Oxford: Oxford University Press.

Wacewicz, S., Perea-García, J. O., Lewandowski, Z., and Danel, D. P. (2022). The adaptive significance of human scleral brightness: an experimental study. *Sci. Rep.* 12:20261. doi: 10.1038/s41598-022-24403-2

Walenski, M., Tager-Flusberg, H., and Ullman, M. (2006). "Language in autism" in Understanding Autism: From Basic Neuroscience to Treatment. eds. S. O. Moldin and J. L. R. Rubenstein (Boca Raton: Taylor and Francis Books), 175–203.

Wandell, B. A., and Le, R. K. (2017). Diagnosing the Neural Circuitry of Reading. *Neuron* 96, 298–311. doi: 10.1016/j.neuron.2017.08.007

Wang, T., and Lee, Y. (2014). Does restriction of pitch variation affect the perception of vocal emotions in Mandarin Chinese? *J. Acoust. Soc. Am.* 137, EL117–EL123. doi: 10.1121/1.4904916

Wang, T., Wichmann, S., Xia, Q., and Ran, Q. (2023). Temperature shapes language sonority: Revalidation from a large dataset. *PNAS Nexus* 2:pgad384. doi: 10.1093/pnasnexus/pgad384

Ward, J., Hoadley, C., Hughes, J. E. A., Smith, P., Allison, C., Baron-Cohen, S., et al. (2017). Atypical sensory sensitivity as a shared feature between synaesthesia and autism. *Sci. Rep.* 7:41155. doi: 10.1038/srep41155

Wildgruber, D., Ackermann, H., Kreifelts, B., and Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* 156, 249–268. doi: 10.1016/S0079-6123(06)56013-3

Wilkins, A. S., Wrangham, R. W., and Fitch, W. T. (2014). The "domestication syndrome" in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics* 197, 795–808. doi: 10.1534/genetics.114.165423

Wrangham, R. W. (2018). Two types of aggression in human evolution. *Proc. Natl. Acad. Sci. USA* 115, 245–253. doi: 10.1073/pnas.1713611115

Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005

Yoon, S. H., Choi, J., Lee, W. J., and Do, J. T. (2020). Genetic and Epigenetic Etiology Underlying Autism Spectrum Disorder. *J. Clin. Med.* 9:966. doi: 10.3390/jcm9040966

Zhu, W., He, L., and Xia, L. X. (2022). The brain correlates of state proactive aggression. *Neuropsychology* 36, 231–242. doi: 10.1037/neu0000791

Zhu, W., Zhou, X., and Xia, L. X. (2019). Brain structures and functional connectivity associated with individual differences in trait proactive aggression. *Sci. Rep.* 9:7731. doi: 10.1038/s41598-019-44115-4

Zollikofer, C. P. E., and Ponce de León, M. S. (2010). The evolution of hominin ontogenies. *Semin. Cell Dev. Biol.* 21, 441–452. doi: 10.1016/j.semcdb.2009.10.012

Zollikofer, C. P., Ponce De Leon, M. S., Vandermeersch, B., and Leveque, F. (2002). Evidence for interpersonal violence in the St. Cesaire Neanderthal. *Proc. Natl. Acad. Sci. USA* 99, 6444–6448. doi: 10.1073/pnas.082111899

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 27 January 2025
DOI 10.3389/fpsyg.2025.1499563

OPEN ACCESS

EDITED BY
Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Maity Siqueira,
Federal University of Rio Grande do Sul, Brazil
Niamh A. O'Dowd,
University of Oslo, Norway

*CORRESPONDENCE
Mocheng Lin
✉ 1207573637@qq.com

RECEIVED 21 September 2024
ACCEPTED 07 January 2025
PUBLISHED 27 January 2025

CITATION
Sun Y and Lin M (2025) A bibliometric analysis of metonymy in SSCI-indexed research (2000–2023): retrospect and prospect. *Front. Psychol.* 16:1499563. doi: 10.3389/fpsyg.2025.1499563

A bibliometric analysis of metonymy in SSCI-indexed research (2000–2023): retrospect and prospect

Yi Sun and Mocheng Lin*

Fuzhou University of International Studies and Trade, Fuzhou, Fujian, China

Introduction: Metonymy has gained increasing attention for its role in shaping language, thought, and communication. Despite its prominence, the thematic evolution and future directions of metonymy research remain underexplored. This study seeks to address this gap by analyzing metonymy research published between 2000 and 2023, providing a comprehensive overview of its key trends and emerging themes.

Methods: A bibliometric analysis was conducted using data sourced from the Social Science Citation Index (SSCI) within the Web of Science Core Collection. Co-citation and co-word analysis were employed alongside k-means clustering techniques to identify research themes. Predictive modeling, including ARIMA and LSTM approaches, was used to forecast future research topics based on keyword trends.

Results: The analysis identified 11 key research clusters, highlighting the central role of cognitive and conceptual linguistics in metonymy research, along with its applications in semantics, pragmatics, and multimodal contexts. Predictive modeling suggested the emergence of seven new research themes for 2024–2028, including the interaction between metonymy and discourse, its role in multimodal communication, and its application in social and cultural narratives.

Discussion: This study underscores the interdisciplinary nature of metonymy research, bridging linguistic, cognitive, and social dimensions. The findings highlight promising areas for future exploration, namely, its integration into digital communication and its impact on cultural identity construction. The methodological approach offers a robust framework for analyzing and predicting research trends, paving the way for innovative contributions to the field.

KEYWORDS

metonymy, bibliometric analysis, Python, metaphor, figurative language

1 Introduction

Conceptual Metaphor Theory (CMT), pioneered by Lakoff and Johnson (1980), heralded a new chapter in the exploration of the essence of metaphors. This theory posits that metaphors are not merely linguistic embellishments but are deeply ingrained in everyday language use. Originating from inquiries into the cognitive foundations of language practice, CMT research has progressively permeated various interdisciplinary fields, encompassing metaphor and culture (Kövecses, 2005, 2006), metaphor and communicative discourse (Zinken and Mussolf, 2009), and the neuroscientific foundations of metaphor (Feldman and Narayanan, 2004; Grady and Ascoli, 2017).

While metaphors have garnered significant scholarly attention, an equally important cognitive-linguistic phenomenon, metonymy, has not been explored as extensively. Metonymy, often studied alongside metaphor, is part of the everyday way of thinking, grounded in experience and governed by systematic principles that structure our thought and actions (Gibbs, 1994, pp. 324–333). The precise definition of metonymy has been

a topic of scholarly debate, with two main perspectives emerging. The first considers metonymy as an intra-domain conceptual mapping (Lakoff and Johnson, 1980), involving representational relationships within a single conceptual domain. For example, in the sentence "The guitar has been drinking heavily," the guitar stands for the guitarist. The second perspective views metonymy as a "reference point" phenomenon (Langacker, 1993; Kövecses and Radden, 1998), as in "He has a Picasso," where the artist serves as a reference point for his artwork. Panther and Radden (1999) pointed out that "metonymy is a cognitive phenomenon that may be even more fundamental than metaphor" (p. 1). Subsequent examinations by scholars like Barcelona (2003/2000), Dirven and Pörings (2002), Panther and Tornburg (2003), and Panther et al. (2009) have confirmed Panther and Radden's (1999) speculation. Based on this, research has extended to various facets of metonymy, including its cognitive operations (El Yamlahi and Cortés de los Ríos, 2022), pragmatic functions (Pannain, 2017), and its intersection with other linguistic phenomena (Yurchenko et al., 2020). A landmark contribution is Littlemore's (2015) comprehensive overview metonymy research, emphasizing its significance in cognitive and discourse studies and highlighting its pervasive influence in language and communication.

Despite the valuable insights provided by these studies, there remains a conspicuous gap in the literature regarding a holistic analysis of the thematic evolution and future trends of metonymy research. To address this gap, this study aims to systematically review and analyze metonymy research from 2000 to 2023, using data sourced from the Social Science Citation Index (SSCI) within the Web of Science Core Collection. SSCI offers a more targeted selection of journals with a robust focus on social science disciplines, ensuring that the literature surveyed in this study represents the most relevant and influential contributions to the cognitive and linguistic dimensions of metonymy research. By employing advanced bibliometric and time series analysis techniques, this study seeks to provide a comprehensive overview of metonymy research, tracing its development over the past two decades and identifying emerging trends and future directions. Specifically, this study aims to answer the following research questions:

1. What have been the research focuses on metonymy over the past two decades?
2. What are the prospective research topics for the future development of metonymy research?
3. What are the evolutionary trends in metonymy research?

2 Methodology

Bibliometric analysis is a well-established quantitative method in academic research that facilitates the systematic assessment of scholarly literature. This method employs a variety of techniques, including citation analysis, co-citation analysis, and keyword co-occurrence analysis, to uncover patterns, trends, and relationships within a given corpus of literature. These methodologies enable researchers to delineate the intellectual structure of a field, monitor its development over time, and forecast potential future research trajectories. By examining publication patterns, citation networks, and co-authorship relationships, bibliometric methods offer quantitative insights into research trends, key contributors, and thematic evolutions (Börner et al., 2003). The versatility of bibliometric approaches allows them to encompass a wide array of scholarly disciplines, providing both micro and macro-level perspectives (Van Raan, 2005; Xiao and Li, 2021). Additionally, bibliometrics provides a statistical means to evaluate and quantify research output and growth trends in specific academic fields (Chen et al., 2021).

The present study aims to forecast the evolution of metonymy research topics by analyzing the relationship between topics and keywords. Established research topics are typically characterized by specific combinations of multiple keywords, and shifts in these combinations often signal the emergence and development of new research areas. As topics evolve, new keyword combinations emerge, driving recurring cycles of conceptual deconstruction and reconstruction. This dynamic relationship between research topics and keywords is observable across different academic fields (Liang et al., 2023). The literature on a specific research topic, once organized and processed, forms a topic-keyword representation that conveys its core essence. By selecting and clustering keywords, researchers can abstract a set of keywords to form keyword groups that, in conjunction with word frequency analysis, represent the research topic. Throughout this process, research topics exhibit continuity, evolving from original topics to new topics through the reorganization and reinterpretation of keyword groups. Consequently, changes in keyword combinations can reflect both the evolution of existing topics and the emergence of new ones.

Furthermore, the development of research topics typically follows a predictable life cycle, comprising stages such as emergence, growth, maturity, stabilization, and decline. Topics do not appear or disappear abruptly; instead, their developmental trajectories are often traceable. Therefore, short-term predictive analysis of topic trends using time series data is both feasible and valuable. Based on this understanding, the study hypothesizes that the temporal evolution of metonymy research topics is continuous and influenced by preceding stages. This continuity suggests that it is possible to construct time series models based on historical data to predict future trends in metonymy research.

As illustrated in Figure 1, the methodology for predicting research trends and topic evolution in metonymy research involves three primary steps: data collection, identifying overarching topics, and forecasting future topic trend. This approach begins by gathering relevant research articles, then identifies common themes and patterns based on keywords and citations, and finally uses historical data to anticipate future developments in these themes.

Each step is detailed as follows:

2.1 Data collection

For the systematic review of metonymy research, data were sourced from the Social Science Citation Index (SSCI) within the Web of Science (WOS) Core Collection, provided by Clarivate Analytics. The Web of Science repository offers access to high-impact publications and their citation data across the natural and social sciences. Utilizing the Core Collection ensured the inclusion

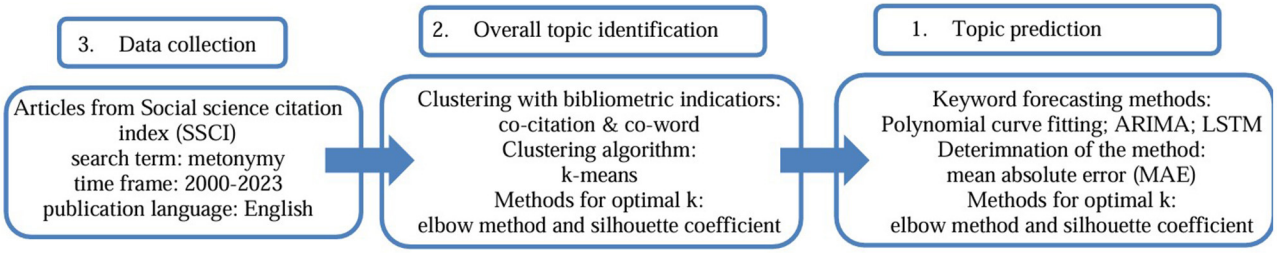


FIGURE 1
Research procedure.

of high-quality materials and detailed citation data. The systematic literature retrieval strategy was as follows:

1. "Metonymy" was selected as the core search term, with a temporal boundary set from 1 January 2000 to 31 December 2023.
2. The search was restricted to articles, and only publications in English were considered to ensure consistency and comparability in the analysis.
3. After the initial online retrieval, a manual screening process was conducted to exclude publications that were unrelated to metonymy or lacked keywords.

The initial search resulted in 589 publications. After the filtering process, a final total of 499 publications were included in the dataset for subsequent analysis. The distribution of publications over time is depicted in Figure 2, which shows a steady increase in the number of publications, with an acceleration in cumulative counts over time. This trend suggests that the field of metonymy has transitioned from its early stages into a mature phase characterized by rapid growth and increased scholarly attention.

2.2 Overall topic identification

Co-citation analysis and co-word analysis are two fundamental approaches in bibliometric research that provide complementary perspectives for understanding the structure and development of scientific fields. Co-citation analysis, developed from Kessler's (1963) concept of bibliographic coupling and formalized by Small (1973), identifies relationships between documents based on their shared references in a third document. Despite some criticism, Small (1974) argued that co-citation patterns, particularly multiple citation connections, are significant indicators of research specialties and disciplines (Small and Greenlee, 1980; Small and Crane, 1979). At a higher level of abstraction, Price (1965) utilized ISI data to theorize about the structure of science itself, exploring networks of scientific papers to identify research fronts. Cozzens (1985) further observed that co-citation studies appear to confirm Price's (1970) hypotheses regarding significant intellectual focus areas, as evidenced by referencing patterns within active specialty groups. In the study of research specialties, various works have explored reference networks (Baldi and Hargens, 1997; Price, 1965), the codification and accumulation of knowledge in various fields (Cozzens, 1985; Lewis, 1980), and the use of journal-to-journal citation data to identify the emergence and transformation of specialties (Van den Besselaar and Leydesdorff, 1996). Co-word analysis, introduced by Callon et al. (1983), maps the co-occurrence of specific terms across documents, revealing how concepts cluster together within and across fields. Co-word analysis has been applied to study various fields, including biotechnology, artificial intelligence, cancer research, polymer chemistry, and acidification research (Rip and Courtial, 1984; Courtial and Law, 1989; Oehler et al., 1989; Callon et al., 1991; Law and Whitaker, 1992; Courtial, 1994; Ding et al., 2001; Coulter et al., 1998). While co-word analysis has faced criticism due to the evolving nature of language (Leydesdorff, 1997), it remains a powerful tool for tracking scientific change and development (Courtial, 1998).

To maximize the insights from both approaches, clustering methods were developed to integrate co-citation and co-word data. Clustering approaches, widely used in bibliometric research, are often designed to address specific needs, making the adaptation of generic clustering techniques to different tasks complex (Jain et al., 1999). Consequently, clustering techniques, such as k-means, hierarchical clustering, and topic models, leverage indicators like word co-occurrence, co-citation patterns, and bibliographic coupling to group related research topics (Zhang et al., 2017; Funk and Owen-Smith, 2016; Li et al., 2014; Zhao and Strotmann, 2014). Various combinations of clustering algorithms and bibliometric indicators have been evaluated across multiple datasets and tasks. For example, Boyack et al. (2011) assessed the accuracy of five clustering approaches on biomedical articles from Medline; Ding and Chen (2014) compared the effectiveness of topic models, co-word analysis, and co-citation analysis for topic detection and tracking; Zhang et al. (2016) explored the usefulness of k-means, hierarchical clustering, and topic models in analyzing academic proposals granted by the National Science Foundation; Klavans and Boyack (2017) tested the ability of directional citations, bibliographic couplings, and co-citations to accurately represent scientific and technical knowledge taxonomies. K-means, a widely used clustering method, remains popular due to its simplicity and low computational complexity, despite being one of the oldest clustering methods (Jain, 2010).

In the context of this study, we utilized Python to conduct a comprehensive topic identification process. The analysis began with constructing two essential matrices: a co-citation matrix and a TF-IDF keyword matrix. These matrices formed the foundation for the subsequent clustering process using the k-means algorithm.

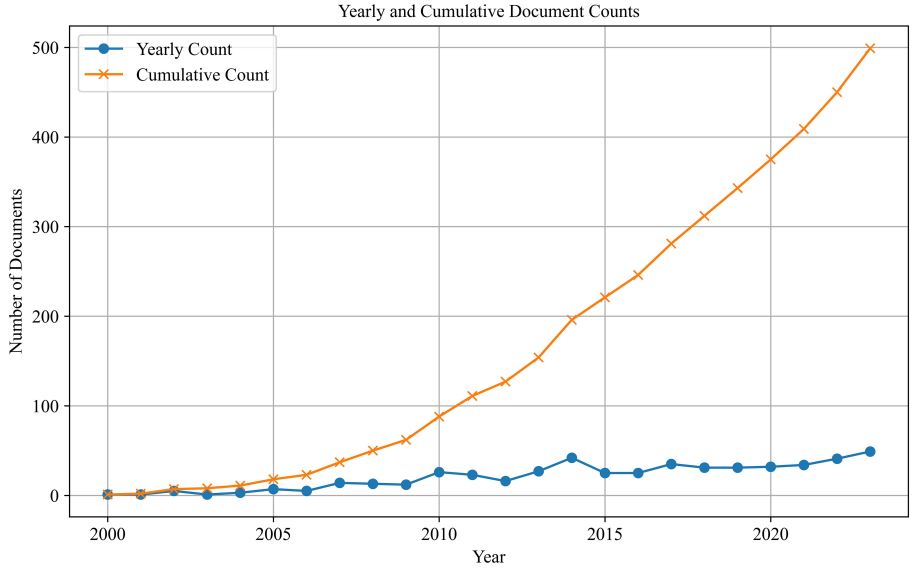


FIGURE 2
Yearly and cumulative article counts.

The first step involves preprocessing the keywords extracted from research articles. We used NLTK (nltk) for preprocessing, including converting keywords to lowercase and applying lemmatization via the WordNetLemmatizer to standardize different word forms. Keywords appearing fewer than three times were filtered out using Pandas (pandas) to retain only the most relevant terms. Simultaneously, we constructed a co-citation matrix by creating a network graph using NetworkX (networkx), where nodes represent references, and edges represent the co-citation relationships between articles, indicating the strength of the relationship. For keyword analysis, we generated a TF-IDF matrix using Scikit-learn's (sklearn) TfidfVectorizer. This matrix captured the importance of each keyword by adjusting for its frequency within the entire dataset.

The second step involves the integration of co-citation and keyword information. Both matrices were standardized to ensure their equal contribution to the clustering process. This is accomplished using Scikit-learn's StandardScaler. The standardized co-citation matrix and TF-IDF matrix were then combined into a single feature matrix that integrates both citation relationships and semantic content. This combined matrix offered a comprehensive representation of each document, capturing content and citation-based similarities.

Third, to prepare the data for clustering, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction, using TSNE from Scikit-learn. This technique reduced the complexity of the combined feature matrix, making it easier to visualize the clusters in a two-dimensional space. Next, we determined the optimal number of clusters using two approaches:

1. The Elbow method, which is implemented to calculate the Sum of Squared Errors (SSE) for different numbers of clusters.
2. Silhouette analysis, using silhouette_score from Scikit-learn, to measure the quality of clustering.

Once the optimal number of clusters is identified, we use the k-means algorithm to partition the dataset.

Finally, we conducted an analysis of each cluster to identify the top keywords that characterize it. To achieve this, we created a topic-keyword probability matrix, which was constructed using the clustered keyword data. This matrix quantified the likelihood of each keyword occurring within a cluster and is normalized across the dataset to provide accurate comparisons. The top 20 keywords for each cluster were extracted based on their occurrence probabilities, offering a refined view of the most significant terms associated with each topic. The Matplotlib (matplotlib.pyplot) and Seaborn (seaborn) libraries were used to generate bar plots and visualizations, which depict:

1. The distribution of documents across clusters.
2. The frequency of the top keywords within each cluster.

2.3 Topic prediction

The evolution of research topics typically follows a pattern of continuity, with keyword trends reflecting underlying inertia. Keyword frequency serves as one of the most direct and effective external indicators of a topic's state. Compared to manually set indicators for predicting topic states, raw word frequency indicators are inherently more objective and accurate, boasting strong scientific validity and broad applicability. In predictive research, these indicators help minimize errors, thereby enhancing the reliability of the predictions. Accordingly, this study employed time series analysis to forecast future keyword frequencies, which were subsequently used to calculate vector adjustment coefficients.

To comprehensively and accurately capture the trends in word frequency evolution, a time window of 1 year was employed for multi-step forecasting.

Three forecasting methods were utilized to predict keyword frequencies: polynomial curve fitting, ARIMA modeling, and LSTM modeling. The study adopted a recursive prediction approach, wherein the frequency of keywords for year n+1 is predicted based on the data from period (2000 to year n). The predicted value for year n+1 was then integrated back into the original dataset, which was subsequently used to predict the data for year n+2, and so forth.

To minimize prediction errors, the mean absolute error (MAE) was employed as the error evaluation metric, determining the final prediction model and forecasting method. Following this, the k-means algorithm, implemented in Python, was applied once again to cluster high-frequency keywords, thereby identifying the predicted topics. This approach, compared to qualitative methods, is more scientific and better reflects the temporal trends and inertia of topics, with Python's capabilities ensuring robust and reproducible results.

We began by constructing a Keyword-Year Frequency Matrix, using Pandas (pandas) to extract keyword frequencies for each year in our dataset (2000–2023). Missing values were handled using linear interpolation from Pandas, ensuring continuity across the years. After filtering out less significant keywords based on a predefined minimum occurrence threshold, the dataset was standardized using StandardScaler from Scikit-learn (sklearn). This step transformed the data to have a mean of zero and a standard deviation of one, facilitating accurate modeling by eliminating bias from differing scales. Following preprocessing, the keyword frequency data was standardized using Python's StandardScaler. This standardization transformed the data to have a mean of zero and a standard deviation of one, facilitating effective modeling and ensuring that all variables contribute equally to the prediction process.

The next phase involved the use of three distinct recursive prediction models to forecast future keyword frequencies:

1. Polynomial Recursive Model: This involved fitting polynomial regression models of varying degrees to historical keyword data using Python's scikit-learn library. The model with the highest R^2 value was selected, and future keyword frequencies are recursively predicted.
2. ARIMA Recursive Model: Python's pmdarima library was used to automatically select the best-fit ARIMA model parameters via the auto_arima function. The chosen ARIMA model was then used to recursively forecast keyword frequencies, ensuring non-negative predictions.
3. LSTM Recursive Model: Data was reshaped to fit the input requirements of LSTM models using Python's numpy and keras libraries. Hyperparameters for the LSTM model, such as the number of units and learning rate, were optimized using a Random Search within the Keras Tuner framework. The best-tuned LSTM model was then used to recursively predict future keyword frequencies.

These predictive models were evaluated based on their Mean Absolute Error (MAE) over the predicted periods. For each model, Python's scikit-learn library is used to compute the MAE for each future year to assess prediction accuracy. The predicted frequencies were then converted back to their original scale using inverse standardization, ensuring comparability with the original data.

Finally, the k-means algorithm were applied to cluster the predicted keywords and identify distinct research themes. This approach ensured a robust and scientifically valid method of forecasting keyword trends, providing valuable insights into the future trajectory of research topics in academic fields.

3 Results

3.1 Clustering metonymy research topics using co-citation and co-word analysis

By combining co-citation and co-word analysis with the k-means algorithm, clusters were formed, and keywords for each cluster were extracted. The optimal number of clusters (k) was determined using the elbow method and silhouette scores, as shown in Figure 3. This method plots the Sum of Squared Errors (SSE) against the number of clusters. The plot reveals a noticeable "elbow" point at $k = 11$, where the rate of decrease in SSE significantly slows down. This inflection point suggests that 11 clusters strike a balance between underfitting and overfitting, providing a meaningful partitioning of the data. Silhouette analysis, depicted in Figure 4, further validated the choice of the optimal number of clusters. The silhouette score evaluates the quality of clustering by measuring how similar each point is to others within its cluster compared to points in other clusters. Higher silhouette scores indicate better-defined clusters. The analysis showed that $k = 11$ yielded a relatively high silhouette score, supporting the choice of 11 clusters for subsequent analysis.

The clustering result with $k = 11$ was visualized, as shown in Figure 5. The plot presents the clusters in a two-dimensional space, with each color representing a different cluster and centroids marked with red circles. This visualization confirms the separation and cohesion of the clusters, indicating distinct research topics within the metonymy studies.

To ensure topic completeness and interpretability, the top 20 words with the highest probability distribution in each scientific topic were extracted and listed in descending order of frequency within each cluster, as shown in Table 1, to represent the topic content. This table presents the detailed keyword analysis and article distribution of each cluster in metonymy research. Thus, the research topics and topic-keyword sets in the field of "metonymy" from 2000 to 2023 were obtained. The top keywords reflect the primary research interests and thematic focus within each cluster, demonstrating the diversity and specific areas of study in metonymy research. The identified topics encapsulate diverse areas of focus in metonymy research, characterized by shared keywords and co-citation patterns. These clusters provide a structured overview of the research landscape, highlighting the evolution and differentiation of topics within metonymy studies over the specified period.

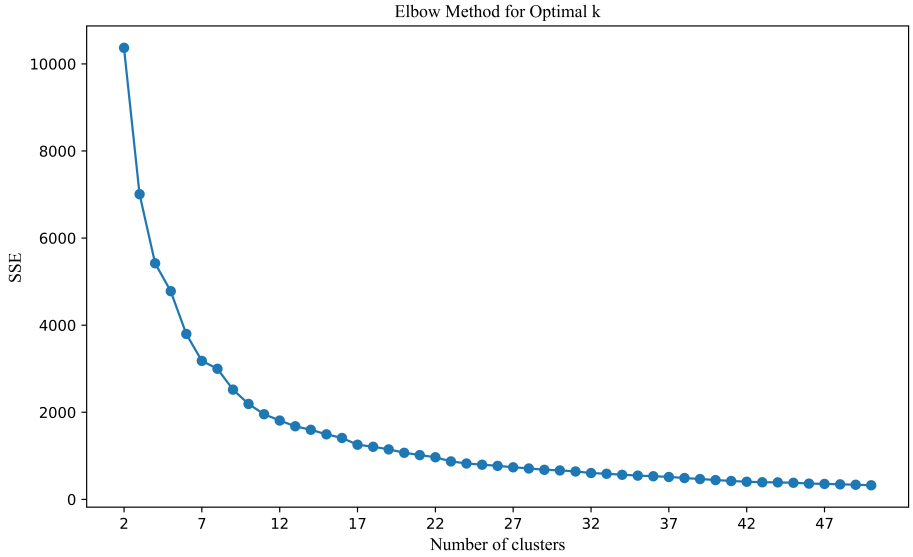


FIGURE 3
Elbow method for optimal K.

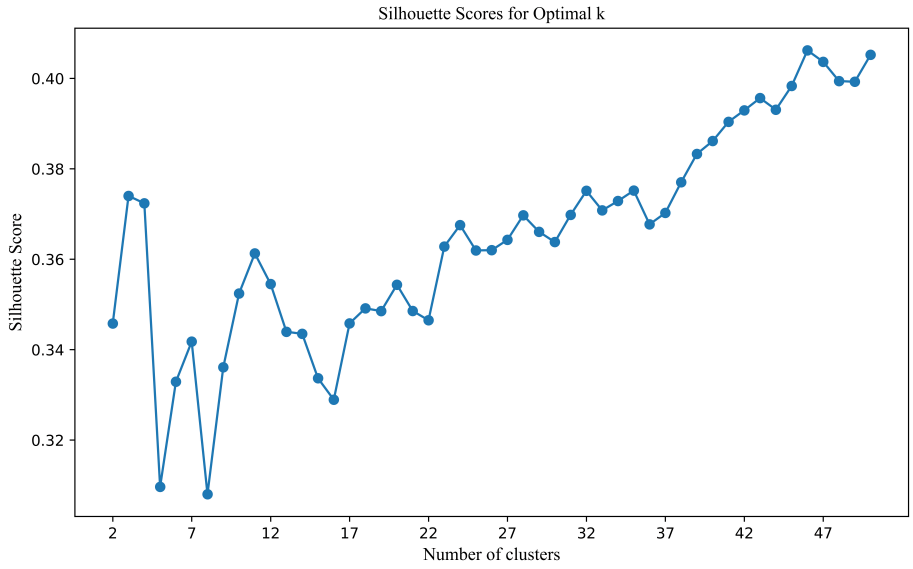


FIGURE 4
Silhouette scores for optimal K.

3.2 Metonymy research topic forecast

Building on the previously generated topic-keyword set, 80 keywords were identified. Among these, 36 keywords (45%) have a length greater than two words, and 6 keywords (7.5%) have a length greater than three words. To determine the most accurate forecasting method, the mean absolute error (MAE) of three methods—polynomial curve fitting, ARIMA, and LSTM—was compared, as illustrated in Figure 6. The results indicate that the polynomial curve fitting method performed worse than both ARIMA and LSTM. Although LSTM and ARIMA exhibited similar predictive performance, the ARIMA model generally yielded a lower MAE, suggesting superior forecasting accuracy. Consequently, the ARIMA model was selected for predicting word frequency trends.

Using the ARIMA model, keyword trends for the next 5 years (2024–2028) were forecasted. To further analyze the predicted topics, 51 keywords with an average frequency above the median of the original dataset were selected for clustering. These keywords are presented in descending order of frequency in Table 2. The same clustering methodology as previously described was applied. The optimal number of clusters was determined using a combination of the Elbow method (Figure 7) and the Silhouette coefficient

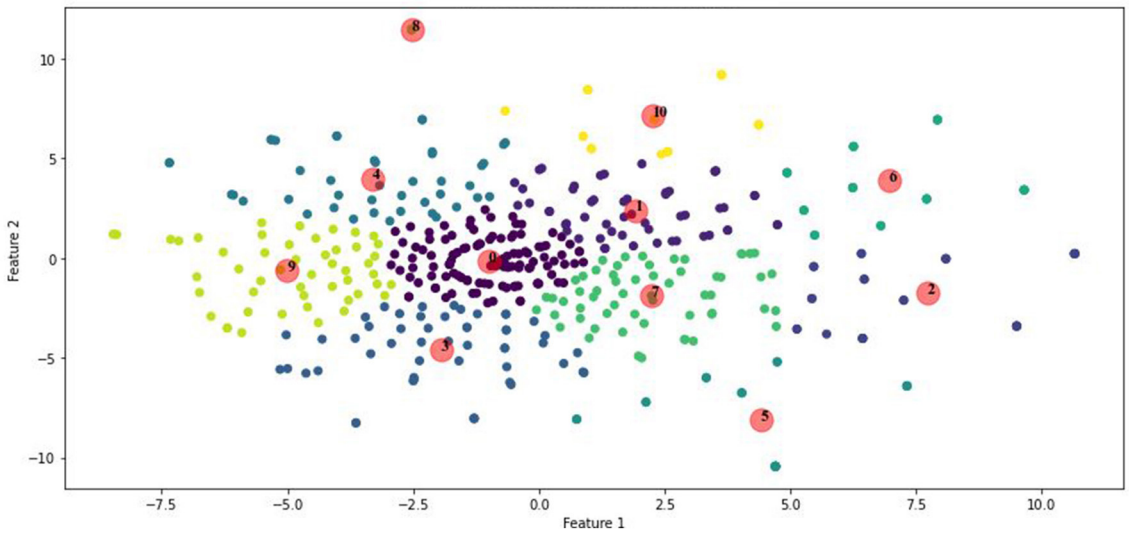


FIGURE 5
K-means clustering with optimal K.

TABLE 1 Summary of keywords of clusters.

Cluster-ID	Size	Top keywords
0	128	metonymy; metaphor; cognitive linguistics; conceptual metonymy; emotion; semantic change; syntax; visual metaphor; embodiment; figurative language; construction; rhetoric; irony; polysemy; gender; adjective; conceptual integration; visual metonymy; experimental pragmatic; n400
1	54	metonymy; metaphor; conceptual metaphor; cognitive linguistics; conceptual metonymy; pragmatic; figurative language; language; polysemy; visual metonymy; visual metaphor; conceptualization; cultural model; cognitive operation; autism; emotion; English; euphemism; conceptual metaphor theory; development
2	31	metonymy; polysemy; metaphor; homonymy; lexical ambiguity; construction; conceptual metaphor; pragmatic; idiom; metaphor and metonymy; construal; n400; blending; cognitive operation; speech act; multimodality; conceptual metonymy; euphemism; semiotics; high-level metonymy
3	52	metonymy; metaphor; semantics; corpus linguistics; discourse; pragmatic; iconicity; cognitive linguistics; lexical semantics; conceptual metaphor theory; conceptual metonymy; cognitive grammar; idiom; figurative language; coercion; conceptual metaphor; motivation; construal; construction; construction grammar
4	43	metonymy; metaphor; emotion; figurative language; indexicality; iconicity; semantics; conceptual metaphor; coercion; frame; motivation; conceptual metonymy; visual metonymy; cognitive grammar; contiguity; adjective; polysemy; cognitive semantics; novel metaphor; blending
5	23	metonymy; metaphor; figurative language; euphemism; persuasion; italian; conceptual blending; cognitive semantics; autism spectrum disorder; iconicity; polysemy; image schema; indexicality; cognitive linguistics; English; high-level metonymy; conceptual metaphor; embodiment; political cartoon; emotion
6	26	metonymy; metaphor; autism spectrum disorder; williams syndrome; figurative language; pragmatic; semantic change; autism; novel metaphor; inference; conceptual metonymy; frame; corpus; language; idiom; image schema; metaphor and metonymy; creativity; word-formation; cognitive linguistics
7	58	metonymy; metaphor; polysemy; figurative language; pragmatic; cognitive linguistics; image schema; creativity; conceptual metonymy; multimodal metaphor; homonymy; English; autism spectrum disorder; emotion; novel metaphor; cognitive model; cognitive semantics; development; semantics; narrative
8	11	metonymy; metaphor; multimodality; picture book; cognitive operation; cognitive linguistics; multimodal metaphor; corpus; creativity; visual metonymy; cognitive model; gender; contiguity; embodiment
9	51	metonymy; metaphor; construction; cognitive linguistics; figurative language; analogy; rhetoric; emotion; trope; Chinese; metaphor and metonymy; conceptual metonymy; conceptual metaphor; creativity; persuasion; discourse; cognitive semantics; compound; frame; logical metonymy
10	22	metonymy; metaphor; visual metaphor; conceptual metaphor; political cartoon; speech act; cognitive linguistics; cognitive semantics; analogy; irony; cultural model; visual metonymy; image schema; multimodality; discourse; motivation; high-level metonymy; corpus linguistics; conceptual metonymy; cognitive model

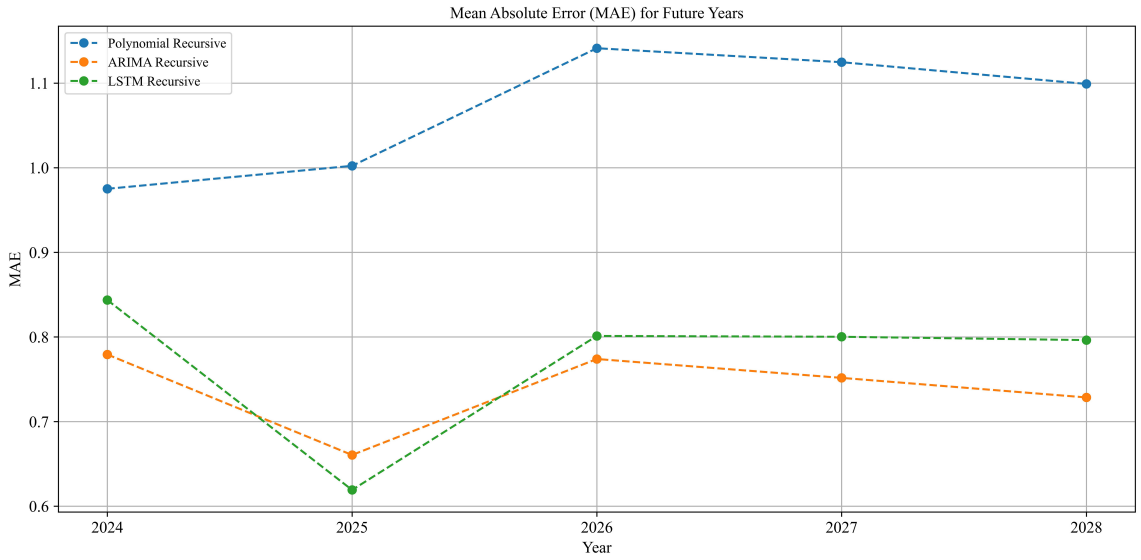


FIGURE 6
Mean absolute error (MAE) comparison.

(Figure 8). The analysis revealed that clustering with $k = 7$ provided the highest silhouette coefficient, corroborated by the Elbow method, thus $k = 7$ was chosen for the clustering of predicted topics.

The clustering results for the predicted topics (PT) are summarized in Table 3, with each cluster characterized by a set of keywords representing anticipated focus areas in metonymy research from 2024 to 2028.

4 Discussion

The above clustering analysis during identified 11 distinct research topics, each representing key areas of interest within metonymy research from 2000 to 2023. In addition, a forecast for the next 5 years (2024–2028) predicts seven emerging topics that are likely to shape the future trajectory of the field. This section discusses the primary focuses of the 11 clusters from 2000 to 2023, the temporal distribution of research activities, and the trends and prospects for future research based on the predicted clusters.

4.1 Research focuses on metonymy from 2000 to 2023

The clustering analysis identified 11 distinct research topics within the field of metonymy. Each cluster represents a specific area of focus, reflecting the interdisciplinary approaches and multifaceted nature of metonymy research. These include theoretical constructs, specific genres, investigative domains, and socio-cultural constructs. The shared keywords of these clusters reveal two important trends in the field of metonymy research. First, almost all clusters include keywords related to cognitive linguistics and figurative language, indicating a strong foundation in understanding metonymy through cognitive frameworks. The consistent presence of terms like "metonymy," "metaphor," and "cognitive linguistics" underscores the importance of these concepts in metonymy research, highlighting the interplay between metonymy and metaphor. This further confirms our previous claim that these two figurative devices are often studied together. Terms like "conceptual metonymy," "conceptual metaphor," and "blending" appear frequently, reflecting ongoing interest in how these mechanisms interact within language and thought. Second, pragmatic and semantic aspects are prominently featured, with keywords such as "pragmatic," "semantics," and "lexical semantics" appearing in multiple clusters. This indicates a focus on how metonymy operates at the level of meaning and use in different contexts. The sizes of the clusters vary significantly, with Cluster 0 being the largest (128 articles) and Cluster 8 the smallest (11 articles). This disparity suggests that certain areas, such as cognitive linguistics and conceptual metonymy, have received more extensive research attention compared to more specialized topics like multimodal metonymy. However, specialized topics can also provide several insights into current research focuses of metonymy research.

1) Cognitive and Conceptual Linguistics

A significant portion of metonymy research is anchored in cognitive and conceptual linguistics. This includes studies on how metonymy operates within cognitive frameworks, examining mental processes and conceptual integration. For example, Cluster 0 (Cognitive Linguistics and Conceptual Metonymy) underscores the centrality of cognitive linguistics in metonymy research. This cluster also reflects a broader engagement with metonymy as a cognitive process that intersects with other theoretical constructs and applied domains. Terms like "syntax," "visual metaphor," and "embodiment" indicate a strong interest in cognitive and syntactic frameworks, as well as visual and embodied manifestations. The prominence of "experimental pragmatic" and

TABLE 2 Summary of forecast keywords.

1	Metonymy	21	Semantics	41	Autism spectrum disorder
2	Metaphor	22	Homonymy	42	Inference
3	Figurative language	23	Corpus linguistics	43	Motivation
4	Conceptual metonymy	24	Cognitive semantics	44	High-level metonymy
5	Cognitive linguistics	25	Advertising	45	Race
6	Persuasion	26	Coercion	46	Conceptual blending
7	Multimodality	27	Compound	47	Metaphor and metonymy
8	Conceptual metaphor	28	Lexical ambiguity	48	Memory
9	Polysemy	29	Semantic change	49	Irony
10	Creativity	30	Rhetoric	50	Logical metonymy
11	Language	31	Gender	51	Multimodal metaphor
12	Visual metonymy	32	Construction		
13	Visual metaphor	33	Conceptualization		
14	Discourse	34	Construal		
15	Development	35	Conceptual integration		
16	Emotion	36	Contiguity		
17	Embodiment	37	Subjectivity		
18	Iconicity	38	Blending		
19	Word-formation	39	Euphemism		
20	Corpus	40	Indexicality		

"n400" suggests engagement with experimental and neurocognitive approaches, reflecting a trend toward empirical validation of theoretical constructs.

2) Semantic and Pragmatic Dimensions

Metonymy research extensively covers semantic and pragmatic dimensions, focusing on how metonymic expressions function at the level of meaning and use. Cluster 2 (Polysemy and Lexical Ambiguity) highlights the relationship between metonymy, polysemy, and lexical ambiguity. The focus on terms like "homonymy," "idiom," and "blending" suggests an exploration of how metonymy contributes to multiple meanings and complex word forms. Additionally, Cluster 3 (Semantics and Corpus Linguistics) emphasizes detailed examinations of metonymy within language use and structure, integrating metonymic and metaphoric frameworks.

3) Multimodal and Visual Metonymy

The expansion of metonymy research into multimodal and visual genres is a notable trend. Cluster 8 (Multimodal Metonymy in Picture Books) focuses on the use of metonymy in multimodal contexts, particularly in picture books. This cluster suggests an interest in how metonymy functions in multimodal storytelling and educational materials. Similarly, Cluster 10 (Visual Metaphor and Political Cartoons) highlights the intersection of visual metaphor and metonymy in political cartoons and other visual media, focusing on the complex interplay between verbal and visual elements in conveying political and cultural messages.

4) Social and Cultural Applications

Research has increasingly applied metonymy theory to social and cultural contexts, examining its role in shaping societal discourses and identity constructions. Cluster 5 (Persuasion and Political Communication) focuses on the persuasive aspects of figurative language, including metonymy. The inclusion of terms like "euphemism," "persuasion," and "political cartoon" points to an interest in the rhetorical uses of metonymy. Meanwhile, Cluster 6 (Metonymy and Neurodiversity) addresses metonymy in relation to autism spectrum disorder and novel metaphor, exploring cognitive differences in neurodiverse populations.

5) Corpus and Data-Driven Approaches

Methodological innovations have played a significant role in advancing metonymy research. Cluster 3 (Semantics and Corpus Linguistics) underscores the use of corpus methodologies to study semantic change and the intersection of metonymy with social issues such as gender and race. This empirical foundation supports more precise and comprehensive analyses of how metonymy functions in natural language use.

6) Emotional and Iconic Dimensions

The emotional and iconic dimensions of metonymy are also prominent research areas. Cluster 4 (Emotional and Iconic Metonymy) centers on the emotional and iconic dimensions of metonymy, highlighting how metonymy conveys emotional states and its relationship with iconic signs. This suggests a nuanced investigation into how metonymy interacts with other semiotic resources to create meaning.

4.2 Trends and prospects in metonymy research

The k-means clustering analysis provides valuable insights into the future directions and emerging themes within metonymy research. These predicted topic clusters highlight key areas of focus and potential developments, reflecting the evolving landscape of metonymy studies.

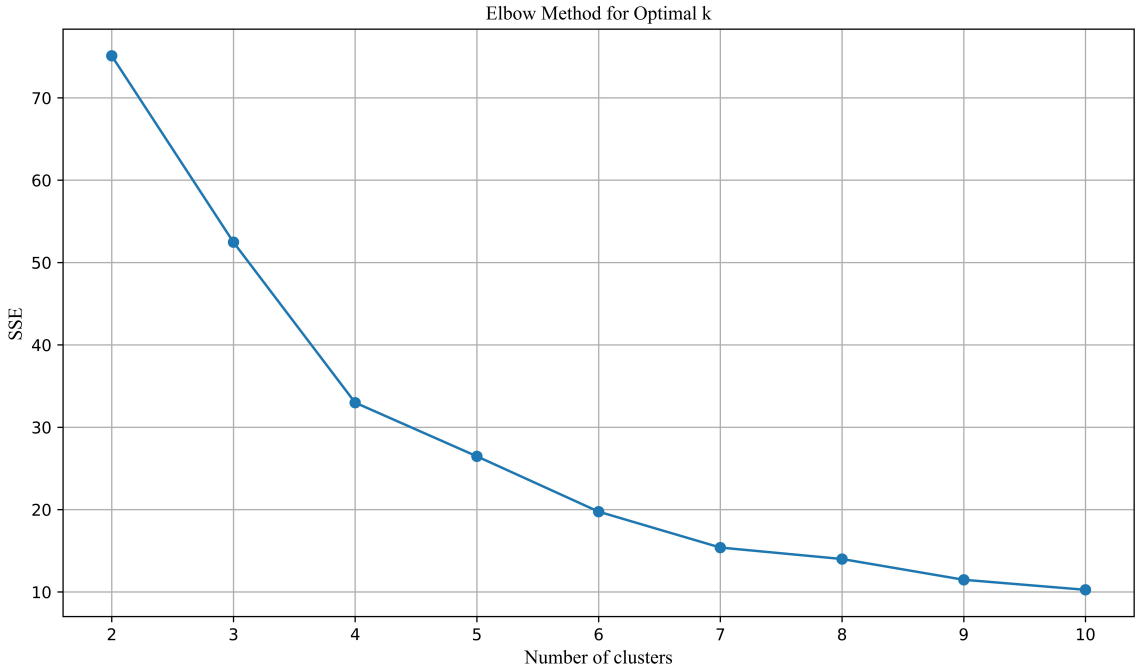


FIGURE 7
Elbow method for optimal K in PT clustering.

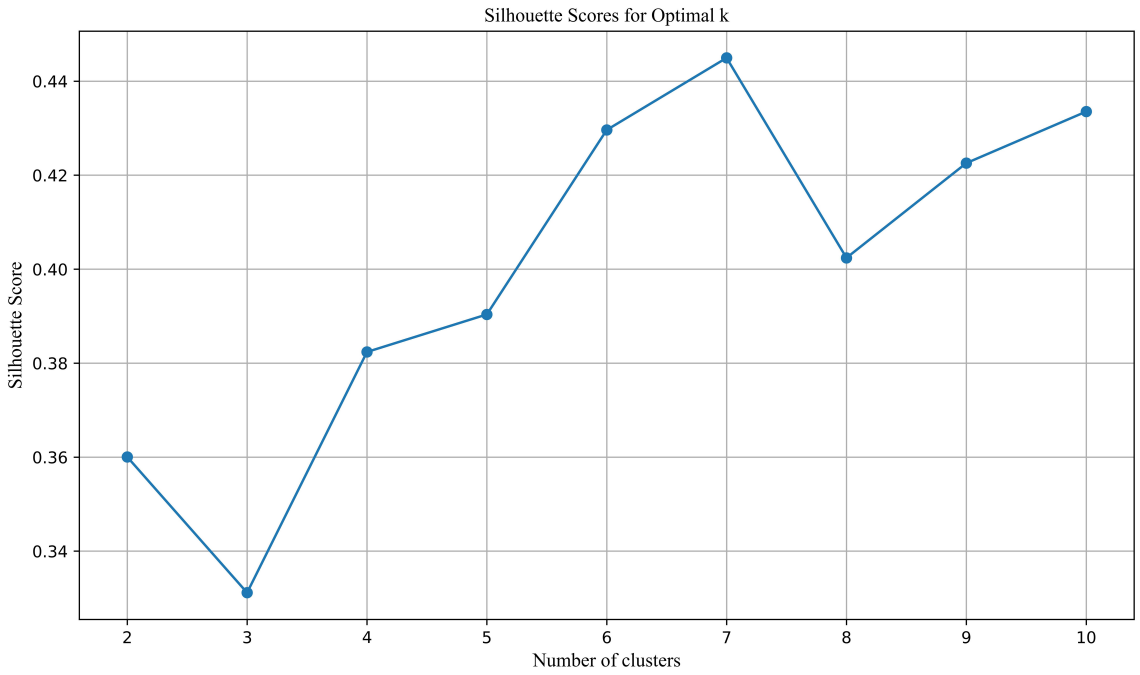


FIGURE 8
Silhouette scores for optimal K in PT clustering.

1) Semantic Complexity and Cognitive Processes

The predicted clusters reveal a continued interest in semantic complexity and cognitive processes. PT Cluster 0 focuses on "polysemy," "homonymy," and "lexical ambiguity," indicating ongoing exploration into how metonymy contributes to semantic richness and ambiguity in language. The inclusion of "construction" and "construal" suggests a focus on the cognitive processes involved in constructing and interpreting metonymic meanings,

TABLE 3 Summary of keywords of predicted topic.

PT Cluster ID	Keywords
0	polysemy, homonymy, lexical ambiguity, construction, construal
1	figurative language, language, development, word-formation, advertising, conceptualization, autism spectrum disorder, inference
2	persuasion, blending, euphemism, high-level metonymy, conceptual blending
3	corpus, corpus linguistics, compound, semantic change, rhetoric, gender, conceptual integration, contiguity, subjectivity, race, memory, irony
4	discourse, iconicity, semantics, cognitive semantics, coercion, indexicality, motivation
5	metonymy, conceptual metonymy, cognitive linguistics, visual metonymy, emotion, embodiment, logical metonymy
6	metaphor, multimodality, conceptual metaphor, creativity, visual metaphor, metaphor and metonymy, multimodal metaphor

reflecting the foundational role of cognitive linguistics in metonymy research.

2) Developmental and Conceptual Applications

PT Cluster 1 highlights the application of metonymy in developmental and conceptual contexts. The presence of keywords such as "figurative language," "language development," and "word-formation" points to a focus on how metonymy is acquired and utilized across different stages of language development. The mention of "autism spectrum disorder" suggests an interest in how metonymic processes may differ in neurodiverse populations, while "advertising" and "conceptualization" indicate an exploration of metonymy's role in shaping concepts and influencing communication strategies.

3) Rhetorical and Persuasive Dimensions

The rhetorical and persuasive dimensions of metonymy are emphasized in PT Cluster 2, which includes keywords like "persuasion," "blending," and "euphemism." This cluster suggests a focus on the strategic use of metonymy in achieving persuasive communication goals, highlighting its rhetorical power in various discourses. The inclusion of "high-level metonymy" and "conceptual blending" underscores the complex cognitive operations involved in crafting persuasive messages.

4) Social and Cultural Contexts

PT Cluster 3 underscores the relevance of metonymy in social and cultural contexts. Keywords such as "corpus," "semantic change," "rhetoric," "gender," and "race" indicate a focus on how metonymy reflects and influences social dynamics and cultural narratives. This cluster suggests that metonymy research will continue to engage with issues of identity, power, and social change, leveraging corpus methodologies to analyze large datasets and uncover patterns in metonymic usage.

5) Multimodal and Iconic Aspects

The predicted topics also highlight the expansion of metonymy research into multimodal and iconic domains. PT Cluster 6 includes terms like "metaphor," "multimodality," "creativity," and "visual metaphor," pointing to an interest in how metonymy and metaphor interact across different communicative modes. This reflects a trend toward integrating visual and multimodal analysis into traditional linguistic studies, examining how metonymic expressions function in visual art, film, and digital media. PT Cluster 4 emphasizes "discourse," "iconicity," and "cognitive semantics," indicating a focus on the interaction between metonymy and discourse, particularly in its iconic and indexical dimensions. This cluster suggests that researchers will explore how metonymic relationships are visually and iconically represented, enhancing our understanding of metonymy's role in creating meaning across different semiotic landscapes.

6) Core Cognitive and Emotional Themes

PT Cluster 5 focuses on core cognitive and emotional themes, with keywords such as "metonymy," "conceptual metonymy," "cognitive linguistics," and "emotion." This cluster highlights the central role of cognitive and emotional processes in metonymy research, emphasizing the foundational aspects of metonymy that continue to drive inquiry into its cognitive underpinnings and emotional impacts.

4.3 Analysis of the evolution of predicted topics

This section compares the relationship between the predicted topics (PT) and the original topics (OT) to understand how metonymy research is expected to evolve. The predicted topics result from the recombination and splitting of keyword groups from the original topics. By analyzing these relationships, we can identify significantly changed predicted topics, which may represent newly emerging scientific topics within the field. The degree of change in predicted topics is measured by the overlap of keyword groups between the predicted and original topics. Overlap is defined as complete lexical matches or semantic equivalence, with the degree of overlap calculated as the ratio of overlapping words to the total number of words in the theme. A threshold of 0.75 is used to differentiate between insignificant and significant changes. Predicted topics with an overlap >0.75 are considered to have insignificant changes, while those with an overlap <0.75 are seen as significantly changed and are potentially aggregates of multiple original topics.

Figure 9 illustrates the relationships between the original topics and the predicted topics. Each circle represents a Predicted Topic (PT), while each square represents an Original Topic (OT). The lines connecting the circles (PTs) to the squares (OTs) indicate the relationship between specific publication topics and the broader

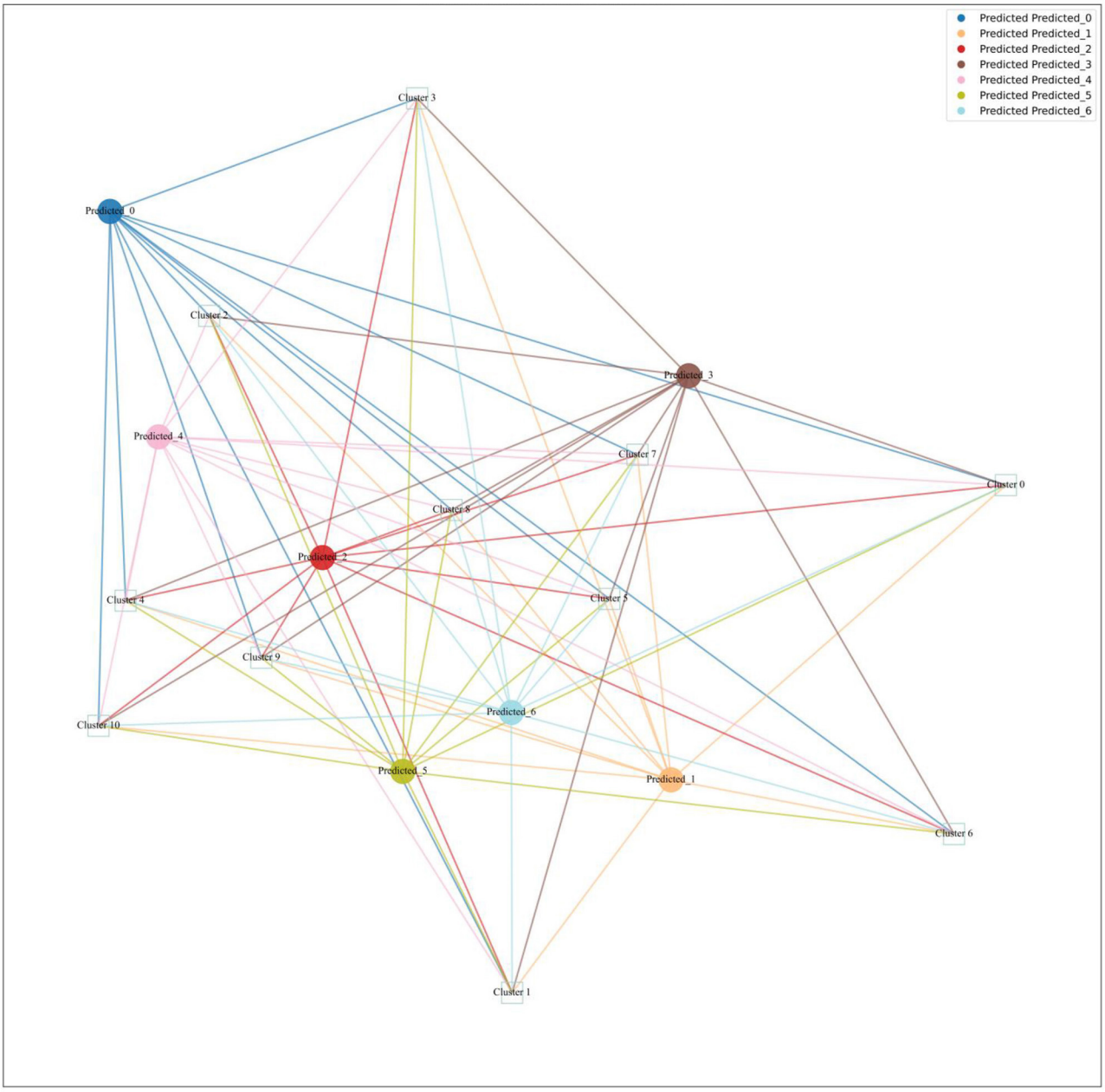


FIGURE 9
Mapping of predicted topic to original topic.

research themes. The proximity of the lines reflects the strength or closeness of the relationship between the PT and OT. Shorter lines suggest a stronger or more direct connection, while longer lines indicate a weaker or more distant relationship.

1) Predicted Topic with Insignificant Changes

PT Cluster 0 aligns with OT Cluster 2 from the original topics, emphasizing semantic complexity through a focus on "homonymy," "polysemy," and "lexical ambiguity." This suggests a sustained interest in exploring how metonymy contributes to semantic richness and complex language construction, reflecting a stable and enduring area of research. Similarly, PT Cluster 1 overlaps with OT Cluster 6, maintaining a focus on language development, conceptualization, and the role of metonymy in neurodiverse contexts such as autism spectrum disorder. This indicates ongoing research into how metonymy aids language acquisition and conceptual growth, highlighting its importance in cognitive and educational perspectives. Meanwhile, PT Cluster 2 aligns with OT Cluster 5, continuing to explore the persuasive uses of metonymy, including "high-level metonymy," "conceptual blending," and "persuasion." This reflects a consistent interest in the rhetorical and strategic applications of metonymy in communication, emphasizing its role in shaping discourse and

cognitive blending processes. Lastly, PT Cluster 3 shows strong alignment with OT Cluster 0, focusing on empirical approaches using corpus methodologies to analyze metonymy's role in social and cultural contexts. The continued emphasis on "rhetoric," "conceptual integration," and "corpus" indicates a stable research interest in understanding how metonymy influences societal dynamics and cultural narratives, leveraging corpus-based analyses to uncover patterns and variations in metonymic expressions across large datasets. Overall, these themes demonstrate the ongoing relevance and impact of established research areas within metonymy, providing a solid foundation for continued exploration and deeper insights into its role in human communication.

2) Predicted Topic with Significant Changes

The significantly changed predicted topics reveal a dynamic shift in metonymy research, reflecting the merging and recombination of elements from multiple original topics. These predicted topics represent potential new directions and highlight the evolving landscape of metonymy studies. PT Cluster 4 primarily originates from OT Cluster 4, which initially focused on the emotional and iconic dimensions of metonymy. This cluster integrates insights from OT Cluster 3 (Semantics and Corpus Linguistics) and OT Cluster 10 (Visual Metaphor and Political Cartoons), reflecting a new focus on "discourse," "iconicity," and "cognitive semantics." This evolution signifies an interest in exploring how metonymic relationships are represented in discourse, particularly their iconic and indexical dimensions, suggesting a shift toward understanding metonymy within broader communicative and semiotic contexts. PT Cluster 5 emerges from OT Cluster 8 (Multimodal Metonymy in Picture Books) and Cluster 10 (Visual Metaphor and Political Cartoons), with contributions from OT Cluster 5 (Persuasion and Political Communication). This cluster emphasizes "conceptual metonymy," "visual metonymy," and emotional processes, building on the original topics' focus on cognitive and visual aspects. The predicted theme represents an expansion toward examining the interplay between cognitive processes and emotional impacts, highlighting a more nuanced understanding of metonymy's role in human cognition and communication. PT Cluster 6 is significantly influenced by OT Cluster 10 and OT Cluster 8, with input from Cluster 9 (Construction Grammar and Analogy). This cluster focuses on "creativity," "metaphor," and "multimodal metaphor," indicating a shift toward integrating multimodal and metaphoric analyses. This evolution reflects a broader interest in exploring how metonymy and metaphor interact across different media, emphasizing creativity and visual communication, and highlighting the trend toward studying metonymy in diverse communicative modes. PT Cluster 3 primarily draws from Cluster 0 (Cognitive Linguistics and Conceptual Metonymy), with additional influences from OT Cluster 3 and OT Cluster 9. It highlights the use of corpus methodologies to analyze metonymy in social contexts, focusing on "rhetoric," "gender," "race," and other social constructs. This shift reflects an emerging interest in how metonymy influences societal dynamics and cultural narratives, leveraging corpus methodologies to provide insights into metonymic expressions across large datasets. Together, these significantly changed predicted topics illustrate the expanding and evolving nature of metonymy research, suggesting a vibrant and interdisciplinary future for the field.

5 Conclusion

This study offers a comprehensive analysis of metonymy research, specifically focusing on articles published from 2000 to 2023, identifying key trends and evolving themes within the field. Using bibliometric analysis and clustering techniques, we explored established and emerging areas of metonymy research, highlighting the dynamic interplay between cognitive, semantic, and multimodal dimensions.

5.1 Key findings

Our analysis reveals that cognitive and conceptual linguistics remain foundational to metonymy research, with a strong emphasis on understanding the cognitive mechanisms driving the use and interpretation of figurative language. This is evidenced by the consistent presence of keywords such as "metonymy," "metaphor," and "cognitive linguistics" across multiple clusters. These findings align with the work of Barcelona (2000) and Kövecses (2002), who argue that metonymy and metaphor often interact in particular linguistic behaviors. These findings reflect the sustained interest in the cognitive processes underlying metonymic expressions and highlight how metonymy research often appears alongside metaphor studies or as a complementary area of inquiry. The close relationship between these two figurative devices indicates their intertwined nature in understanding language and thought, suggesting that they are frequently studied together to provide a more comprehensive picture of figurative language, reinforcing the claims made by Radden and Kövecses (2007) that their overlapping cognitive bases provide an integrated framework for analyzing figurative language. This also supports the possibility of a unified theoretical framework encompassing metonymy, metaphor, and other figures of speech, as proposed by Ruiz de Mendoza (2020).

The study also points to the importance of semantic and pragmatic dimensions, focusing on how metonymy functions at the level of meaning and use. This trend builds distinguished pragmatic types of metonymies (see Panther and Thornburg 1998). The exploration of semantic complexity and lexical ambiguity, particularly in PT Cluster 0, indicates a still ongoing interest in how metonymy contributes to semantic richness and ambiguity in language. A notable trend in metonymy research is the expansion into multimodal and visual contexts, as highlighted in PT Cluster 6. This reflects a growing interest in how metonymy operates across different communicative modes, including visual art (Uno et al., 2019), film (Feng, 2017), and digital media (Bolognesi et al., 2019). The integration of linguistic analysis with modern technological tools suggests a recognition of the importance of studying metonymy in visual and digital contexts. Additionally, the study highlights the application of metonymy theory in social and cultural contexts, exploring its role in shaping societal discourses and identity constructions. PT Cluster 3 emphasizes the relevance

of metonymy in social and cultural domains, a theme that resonates the work of Kövecses (2005, 2006), focusing on how metonymy reflects and influences societal dynamics and cultural narratives.

Finally, the predicted topics indicate potential future directions for metonymy research, emphasizing the evolving landscape of the field. Significantly changed themes, such as those in PT Clusters 4 and 5, suggest new areas of inquiry, including the interaction between metonymy and discourse and the exploration of cognitive and emotional processes.

5.2 Strengths and limitations

Our methodological approach, which integrates bibliometric analysis with clustering techniques, proves to be highly effective in mapping the intellectual structure of metonymy research. The use of co-citation and co-word analysis, combined with advanced clustering algorithms such as k-means, allows for the identification of distinct research themes and the tracking of their evolution over time. This approach provides a robust framework for exploring large datasets, enabling a detailed understanding of the field's development and the interconnections between various research areas. Moreover, the predictive modeling using ARIMA and other forecasting methods offers valuable insights into potential future trends, guiding researchers in identifying emerging areas of interest.

However, it is important to acknowledge the limitations of this methodological approach. The analysis focused exclusively on articles indexed in the SSCI, excluding books published during the same period (2000–2023), which may also contain valuable insights into metonymy research. While bibliometric and clustering techniques are powerful tools for analyzing research trends, they rely heavily on the quality and scope of the underlying data. The selection of keywords, the accuracy of citation databases, and the inherent biases in publication practices can all influence the results. Additionally, the focus on quantitative analysis may overlook nuanced qualitative aspects of metonymy research, such as the depth of theoretical discussions or the subtleties of interdisciplinary integration. Future research could benefit from combining these quantitative methods with qualitative analyses to provide a more comprehensive understanding of the field.

5.3 Implications for future research

Overall, this analysis provides a roadmap for future studies, encouraging continued exploration and discovery within the diverse and multifaceted realm of metonymy. By building on the insights gained from this study, researchers can further expand the boundaries of metonymy research, ensuring its continued relevance and contribution to linguistic, cognitive, and social sciences. To provide more precise guidance for future research, it would be beneficial to explore specific applications of metonymy in digital communication platforms, assess the impact of cultural differences on metonymic usage, and develop innovative computational models to analyze large-scale data sets. Additionally, future studies could employ mixed-method approaches, combining ethnographic methods with quantitative data analysis, to gain deeper insights into the usage of metonymy across different languages and cultures.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YS: Conceptualization, Project administration, Resources, Supervision, Writing – review & editing. ML: Data curation, Software, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baldi, S., and Hargens, L. L. (1997). Re-examining Price's conjectures on the structure of reference networks: Results from the special relativity, spatial diffusing modeling and role analysis literature. *Soc. Stud. Sci.* 27, 669–687. doi: 10.1177/030631297027004004

Barcelona, A. (2000). *Metaphor and Metonymy at the Crossroads*. New York: Mouton de Gruyter.

Barcelona, A. (ed.). (2003/2000). *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. New York: Mouton de Gruyter. doi: 10.1515/9783110894677

Bolognesi, M., Despot, K. Š., and Brdar, M. (eds). (2019). *Metaphor and Metonymy in the Digital Age*. Amsterdam/Philadelphia: John Benjamins. doi: 10.1075/milcc.8

Börner, K., Chen, C., and Boyack, K. W. (2003). Visualizing knowledge domains. *Ann. Rev. Inform. Sci. Technol.* 37, 179–255. doi: 10.1002/aris.1440370106

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., et al. (2011). Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches. *PLoS ONE* 6:e18029. doi: 10.1371/journal.pone.0018029

Callon, M., Courtial, J. P., and Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics* 22, 155–205. doi: 10.1007/BF02019280

Callon, M., Courtial, J. P., Turner, W. A., and Bauin, S. (1983). From translations to problematic networks: an introduction to co-word analysis. *Soc. Sci. Inform.* 22, 191–235. doi: 10.1177/053901883022002003

Chen, X., Zou, D., Xie, H. R., and Su, F. (2021). Twenty-five years of computer-assisted language learning: a topic modeling analysis. *Lang. Learn. Technol.* 25, 151–185. doi: 10.29140/9781914291012-8

Coulter, N., Monarch, I., and Konda, S. (1998). Software engineering as seen through its research literature: a study in co-word analysis. *J. Am. Soc. Inform. Sci. Technol.* 49, 1206–1223. doi: 10.1002/(SICI)1097-4571(1998)49:13<1206::AID-ASI7>3.0.CO;2-F

Courtial, J. P. (1994). A co-word analysis of scientometrics. *Scientometrics* 31, 251–260. doi: 10.1007/BF02016875

Courtial, J. P. (1998). Comments on Leydesdorff's article. *J. Am. Soc. Inform. Sci.* 49:98. doi: 10.1002/(SICI)1097-4571(199801)49:1<98::AID-ASI14>3.0.CO;2-1

Courtial, J. P., and Law, J. (1989). A co-word study of artificial intelligence. *Soc. Stud. Sci.* 19, 301–311. doi: 10.1177/030631289019002005

Cozzens, S. E. (1985). Using the archive: Derek Price's theory of differences among the sciences. *Scientometrics* 7, 431–441. doi: 10.1007/BF02017159

Ding, W., and Chen, C. (2014). Dynamic topic detection and tracking: a comparison of HDP, C-word, and cocitation methods. *J. Assoc. Inform. Sci. Technol.* 65, 2084–2097. doi: 10.1002/asi.23134

Ding, Y., Chowdhury, G. G., and Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Inf. Process. Manag.* 37, 817–842. doi: 10.1016/S0306-4573(00)00051-0

Dirven, R., and Pörings, R. (eds.). (2002). *Metaphor and Metonymy in Comparison and Contrast*. Berlin; New York, NY: Mouton de Gruyter. doi: 10.1515/9783110219197

El Yamlahi, F., and Cortés de los Ríos, M. E. (2022). Examining the metonymic relation between a brand name and a product: a case study of Moroccan cosmetic brand names. *Ibérica* 44, 391–416. doi: 10.17398/2340-2784.44.391

Feldman, J., and Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain Lang.* 89: 385–392. doi: 10.1016/S0093-934X(03)00355-9

Feng, W. D. (2017). Metonymy and visual representation: towards a social semiotic framework of visual metonymy. *Visual Commun.* 16, 441–466. doi: 10.1177/1470357217717142

Funk, R. J., and Owen-Smith, J. (2016). A dynamic network measure of technological change. *Manage. Sci.* 63, 791–817. doi: 10.1287/mnsc.2015.2366

Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.

Grady, J., and Ascoli, G. (2017). "Sources and targets in primary metaphor theory: Looking back and thinking ahead," in *Metaphor. Embodied Cognition and Discourse*, ed. B. Hampe (Cambridge: Cambridge University Press), 27–45. doi: 10.1017/9781108182324.003

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* 31, 651–666. doi: 10.1016/j.patrec.2009.09.011

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Comp. Surv.* 31, 264–323. doi: 10.1145/331499.331504

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *Am. Document.* 14, 10–25. doi: 10.1002/asi.5090140103

Klavans, R., and Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *J. Assoc. Inform. Sci. Technol.* 68, 984–998. doi: 10.1002/asi.23734

Kövecses, Z. (2002). *Metaphor: A Practical Introduction*. Oxford: Oxford university press. doi: 10.1093/oso/9780195145113.001.0001

Kövecses, Z. (2005). *Metaphor in Culture: Universality and Variation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511614408

Kövecses, Z. (2006). *Language, Mind, and Culture: A Practical Introduction*. Oxford: Oxford University Press.

Kövecses, Z., and Radden, G. (1998). Metonymy: developing a cognitive linguistic view. *Cogn. Linguist.* 9, 37–77. doi: 10.1515/cogl.1998.9.1.37

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Langacker, R. W. (1993). Reference-point constructions. *Cogn. Linguist.* 4, 1–38. doi: 10.1515/cogl.1993.4.1.1

Law, J., and Whitaker, J. (1992). Mapping acidification research: a test of the co-word method. *Scientometrics* 23, 417–461. doi: 10.1007/BF02029807

Lewis, G. L. (1980). The relationship of conceptual development to consensus: an exploratory analysis of three subfields. *Soc. Stud. Sci.* 10, 285–308. doi: 10.1177/030631278001000302

Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *J. Am. Soc. Inform. Sci.* 48, 418–427. doi: 10.1002/(SICI)1097-4571(199705)48:5<418::AID-ASI4>3.0.CO;2-Y

Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., et al. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). *Res. Policy* 43, 941–955. doi: 10.1016/j.respol.2014.01.012

Liang, J., Jianlin, Y., and Wei, W. (2023). Research on scientific topic prediction from the perspective of knowledge unit reorganization. *J. Inform. Sci.* 5, 511–524. doi: 10.3772/j.issn.1000-0135.2023.05.001

Littlemore, J. (2015). *Metonymy*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107338814

Oehler, K., Snizek, W. E., and Mullins, N. C. (1989). Words and sentences over time: how facts are built and sustained in a specialty area. *Sci. Technol. Human Values* 14, 258–274. doi: 10.1177/016224398901400304

Pannain, R. (2017). Metonymy in numerals. *Rev. Cognit. Linguistics* 15, 102–120. doi: 10.1075/rcl.15.1.05pan

Panther, K. U., and Radden, G. (eds.). (1999). *Metonymy in Language and Thought*. Amsterdam; Philadelphia, PA: John Benjamins. doi: 10.1075/hcp.4

Panther, K. U., Thornburg, L. L., and Barcelona, A. (eds.). (2009). *Metonymy and Metaphor in Grammar*. Amsterdam; Philadelphia, PA: John Benjamins. doi: 10.1075/hcp.25

Panther, K. U., and Tornburg, L. L. (eds.). (2003). *Metonymy and Pragmatic Inferencing*. Amsterdam; Philadelphia, PA: John Benjamins. doi: 10.1075/pbns.113

Price, D. J. D. (1965). Networks of scientific papers. *Science* 149, 510–515. doi: 10.1126/science.149.3683.510

Price, D. J. D. (1970). "Citation measures of hard science, soft science, technology and nonscience," in *Communication among Scientists and Engineers*, eds. C. E. Nelson and D. K. Pollock (Massachusetts: Heath-Lexington Books), 3–15.

Radden, G., and Kövecses, Z. (2007). "Towards a theory of metonymy," in *Metonymy in Language and Thought: A Cognitive Perspective*, eds. K. U. Panther and G. Radden (Amsterdam & Philadelphia: John Benjamins), 21—59.

Rip, A., and Courtial, J. P. (1984). Co-word maps of biotechnology: an example of cognitive scientometrics. *Scientometrics* 6, 381–400. doi: 10.1007/BF02025827

Ruiz de Mendoza, F. (2020). Understanding figures of speech: dependency relations and organizational patterns. *Lang. Commun.* 71, 16–38. doi: 10.1016/j.langcom.2019.12.002

Small, H. G. (1973). Cocitation in scientific literature: new measure of relationship between 2 documents. *J. Am. Soc. Inform. Sci.* 24, 265–269. doi: 10.1002/asi.4630240406

Small, H. G. (1974). Multiple citation patterns in scientific literature: the circle and hill models. *Inform. Storage Retrieval* 10, 393–402. doi: 10.1016/0020-0271(74)90046-1

Small, H. G., and Crane, D. (1979). Specialties and disciplines in science and social science: an examination of their structure using citation indexes. *Scientometrics* 1, 445–461. doi: 10.1007/BF02016661

Small, H. G., and Greenlee, E. (1980). Citation context analysis of a co-citation cluster: recombinant DNA. *Scientometrics* 2, 277–301. doi: 10.1007/BF02016349

Uno, R., Matsuda, E., and Indurkhya, B. (2019). Analyzing visual metaphor and metonymy to understand creativity in fashion. *Front. Psychol.* 9:2527. doi: 10.3389/fpsyg.2018.02527

Van den Besselaar, P., and Leydesdorff, L. (1996). Mapping change in scientific specialties: a scientometric reconstruction of the development of artificial intelligence. *J. Am. Soc. Inform. Sci.* 47, 415–436. doi: 10.1002/(SICI)1097-4571(199606)47:6<415::AID-ASI3>3.0.CO;2-Y

Van Raan, A. F. J. (2005). For your citations only? Hot topics in bibliometric analysis. *Measure. Interdiscipl. Res. Perspect.* 3, 50–62. doi: 10.1207/s15366359mea0301_7

Xiao, H., and Li, L. (2021). A bibliometric analysis of critical discourse analysis and its implications. *Discourse Soc.* 32(, 482–502. doi: 10.1177/0957926521992150

Yurchenko, A., Lopukhina, A., and Dragoy, O. (2020). Metaphor is between metonymy and homonymy: evidence from event-related potentials. *Front. Psychol.* 11:2113. doi: 10.3389/fpsyg.2020.02113

Zhang, Y., Zhang, G., Chen, H., Porter, A. L., Zhu, D., and Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: methodology and a case study focusing on big data research. *Technol. Forecast. Soc. Change* 105, 179–191. doi: 10.1016/j.techfore.2016.01.015

Zhang, Y., Zhang, G., Zhu, D., and Lu, J. (2017). Science evolutionary pathways: identifying and visualizing relationships for scientific topics. *J. Assoc. Inform. Sci. Technol.* 68, 1925–1939. doi: 10.1002/asi.23814

Zhao, D., and Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: an author cocitation and bibliographic coupling analysis. *J. Assoc. Inform. Sci. Technol.* 65, 995–1006. doi: 10.1002/asi.23027

Zinken, J., and Mussolf, A. (2009). *Metaphor and Discourse*. Basingstoke: Palgrave Macmillan.

frontiers | Frontiers in Psychology

TYPE Review
PUBLISHED 30 January 2025
DOI 10.3389/fpsyg.2024.1507858

OPEN ACCESS

EDITED BY
Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Claudia Mazzuca,
Sapienza University of Rome, Italy
Claude H. Miller,
University of Oklahoma, United States

*CORRESPONDENCE
Emma A. Renström
✉ emma.renstrom@hkr.se

RECEIVED 08 October 2024
ACCEPTED 03 December 2024
PUBLISHED 30 January 2025

CITATION
Renström EA (2025) The implementation of neo- and nonbinary pronouns: a review of current research and future challenges. *Front. Psychol.* 15:1507858. doi: 10.3389/fpsyg.2024.1507858

The implementation of neo- and nonbinary pronouns: a review of current research and future challenges

Emma A. Renström*

Kristianstad University, Kristianstad, Sweden

This review explores the current state of research on attitudes toward and the use of neo- and nonbinary pronouns, as well as their effects on gender conceptualization. Due to the limited scope of existing studies, this review focuses on Swedish and English. Additionally, I will examine resistance to gender-inclusive language and linguistic gender reforms, with a particular emphasis on nonbinary pronouns and the politicization of such reforms, which represents a significant barrier to the adoption of gender-inclusive language. More research is needed to explore attitudes toward, usage of, and the consequences of neo- and nonbinary pronouns across a wide range of languages. Moreover, it is crucial to investigate the politicized polarization surrounding these reforms to better understand when and why people (do not) use nonbinary pronouns and the broader implications of these pronouns for gender conceptualization in the future. This review is structured as follows: I will first describe the general function of pronouns and discuss the interplay between language, gender, and cognition in relation to pronouns. Then, I will examine the implementation of gender-inclusive pronouns in Swedish and English, the dual nature of many gender-inclusive pronouns, and how this duality influences attitudes and usage. I conclude by discussing future research venues that I see, mainly connected to better understanding the politicization of gender-inclusive language and how this politicization and polarization influences attitudes to and use of nonbinary pronouns and effects of nonbinary pronouns in gender conceptualization.

KEYWORDS

pronouns, attitudes, gender conceptualization, neopronouns, nonbinary pronouns, gender-inclusive language, politicization

Introduction

Language is a tool for transferring cultural knowledge, for example about gender (Beukeboom and Burgers, 2019). Research shows that language influences cognitive processes (Lucy, 1992; Samuel et al., 2019; Fiedler, 2008). Words can influence how events in the world and other people are perceived, leading to initiatives to create a more inclusive language. In line with this, words activate mental representations and evaluations both implicitly and explicitly, and hence, word choices can trigger exclusion, stereotypes, discrimination, and harassment related to gender (Sczesny et al., 2016). Consequently, the research field of gender-fair language is dedicated to better understanding how language can be used to create linguistic inclusion and visibility instead of exclusion.

In line with the idea to make language more inclusive, nonbinary pronouns constitute a relatively recent linguistic development that aims to make language more inclusive by adding a third personal pronoun singular to languages that already have pronouns representing

women and men (e.g., English, French, and Dutch). Nonbinary pronouns are sometimes also neopronouns—newly invented words. However, these pronouns are sometimes revamped versions of older pronouns, such as singular *they* in English. Hence, even though referred to as nonbinary, these pronouns often have a dual meaning: a *nonbinary* meaning, referring to individuals with nonbinary gender identities, and a *generic* meaning, referring to anyone regardless of gender (Renström et al., 2022a).

Swedish was the first language to add a nonbinary pronoun, *hen*, to the official dictionary (SAOL, 2015). Since then, other languages have followed suit. In English, singular *they* is the most common nonbinary pronoun known to the majority of English speakers. Another alternative is *ze*, which is relatively unknown (Lindqvist et al., 2019; Renström et al., 2023a). In French, there is a recent initiative to add the nonbinary pronoun *iel* (Wagener, 2021). In Danish, *hen* and singular *de* (they) are suggested, and in Dutch, both *hen* and *die* are proposed (DeCock et al., 2024; Hjorth-Nebel Miltersen et al., 2022). While these are examples of a global trend, this review will focus on Swedish and English due to a lack of published research (in English) on attitudes, use, and consequences of nonbinary pronouns in other languages.

Common to many of these initiatives is a strong resistance against them as they challenge societal structures and deeply rooted identities related to the view of gender as a binary construct (Renström and Klysing, 2024; Hekanaho, 2020; Morgenroth et al., 2020). The politicization of nonbinary pronouns likely influences how the public receives them in terms of attitudes and use, but also how they influence the construction of gender—how the concept of gender is perceived as (non)binary.

The predominant view of gender constructs it as a binary social category derived from two biological sexes, which are generally seen as two mutually exclusive categories (Hyde et al., 2019; Morgenroth and Ryan, 2018). Both feminists and sexual minority individuals have challenged this view in different ways, including through gender-inclusive language (Gustafsson Sendén et al., 2015, 2021).

Third-person pronouns have become important markers of gender-ideological positions by their relation to gender identity, including the potential for signaling to have a gender identity that defies a binary gender system (Hekanaho, 2020, 2022). Besides expanding gender categories, some nonbinary pronouns can also function to de-emphasize gender in language by removing gender information altogether. A better term for such pronouns would be gender-inclusive pronouns. Gender-inclusive pronouns often serve two functions: to emphasize the existence of multiple gender identities or to decrease the overall salience of gender in language.

The binary conceptualization of gender functions prescriptively to specify what is desired of women and men and by proscriptively specifying what is not accepted (Morgenroth et al., 2020; Prentice and Carranza, 2002). Individuals who violate these expectations are socially punished (Rudman et al., 2012), which contributes to the marginalization of sexual and gender minority groups (Thoma et al., 2021). Two strategies have been suggested to challenge the binary conceptualization of gender and battle such problematic effects (Morgenroth and Ryan, 2020). *De-gendering strategies* aim to remove or minimize the salience of gender altogether, while *multi-gendering strategies* aim to draw attention to the fact that gender is not binary. These two strategies correspond to the two ways in which many gender-inclusive pronouns, such as singular *they*, are used. The active strategy type is defined by the context in which the gender-inclusive pronoun occurs (Renström et al., 2022a). When a gender-inclusive pronoun is used to anonymize or to replace the paired pronoun form *he/she* in generic texts, it functions to decrease the influence of gender, but when a gender-inclusive pronoun is used to refer to individuals with nonbinary gender identities, it serves to increase the salience of other gender identities than the binary woman/man.

Gender-inclusive language reforms, including the use of gender-inclusive pronouns, face resistance (Blaubergs, 1980; Parks and Roberton, 1998; Bradley, 2020; Bradley et al., 2019; Hekanaho, 2020; Vergoossen et al., 2020a). This resistance is largely grounded in gender-ideological convictions about gender being an essential and binary category and a desire to keep gendered power structures intact (Douglas and Sutton, 2014; Parks and Roberton, 2005; Hekanaho, 2022). However, as de-gendering and multi-gendering strategies to some extent challenge core convictions of different ideologies, the origins of resistance against gender-inclusive pronouns may also differ depending on which strategy is salient (Morgenroth et al., 2020; Renström et al., 2022a), which poses a challenge to the study of neo- and nonbinary pronouns.

Neo- and nonbinary pronouns

In this article, I use the terms neopronouns and nonbinary pronouns, but what terms to use is not straightforward. In fact, when referring to pronouns that have a dual function, such as singular *they*, nonbinary is not correct since a nonbinary pronoun should specifically refer to a pronoun used for individuals with nonbinary gender identity in the same way that *she* should be used about individuals identifying as women.

The definition of a neopronoun is that it is new. It is a word that did not exist in the language before and has been created with a specific meaning. The definition of a nonbinary pronoun is that it applies to individuals with nonbinary gender identities. Nonbinary refers to individuals who do not define themselves as either of the traditional binary genders, woman or man. Hence, a nonbinary pronoun may or may not be a neopronoun. The most illustrative example of this is the use of singular *they* as a nonbinary pronoun. However, singular *they* has a long history of being used to avoid gender cues or when the gender of a referent is unknown, and thus, singular *they* cannot be considered a neopronoun. Because singular *they* is used differently, it is not entirely correct to call singular *they* a nonbinary pronoun either. Examples of nonbinary neopronouns in English are, for instance, *ze* or *xe*, which should be considered both neopronouns and nonbinary pronouns.

Another example is the Swedish pronoun *hen*, which is new and nonbinary. However, *hen* specifically has two meanings—the nonbinary meaning and the generic meaning. Again, due to the dual meaning, it is not entirely correct to label *hen* a nonbinary pronoun. Hence, another term that would fit better for both *hen* and singular *they* is *gender-inclusive,* meaning the inclusion of all genders regardless of what they may be. The same is true for the Danish initiative singular *de* (they) (Hjort Miltersen, 2020).

Other terms that are used to refer to these kinds of pronouns are gender-neutral or gender-fair pronouns. The idea behind the term "neutral" is that the removal of gender cues should make perceivers less biased because there is no linguistically gendered cue to prime

cognitive processing. However, neutralizing runs the risk of assumptions that all genders are equal in every sense. While it is true that, for instance, cognitive abilities do not greatly differ between women and men (Hyde et al., 2019), the experiences of being a woman, a man, or a transperson differ greatly, which impacts behavior. This relates to how gender stereotypes foster gendered expectations and behavior. Hence, gender-neutral language reforms should be used and implemented with caution.

In the social psychological literature on language and gender, the term *gender-fair* language is often used. While gender-fair language theoretically aims to create broader gender categories that could encompass individuals of any gender (Sczesny et al., 2016), in practice, this research field has mainly been devoted to promoting fairness for women relative to men. However, the field is evolving, with more scholars attempting to adopt a broader perspective on gender in language. Within this literature, two principal strategies for achieving gender-fair language are frequently discussed: neutralization and balancing (also referred to as feminization).

However, as mentioned, neutralization might not be beneficial in the long run, and balancing implies the linguistic representation of different genders, which mainly has been women—neither of these strategies is, in fact, fair to all genders. The unfairness of balancing becomes very clear with the example of how pronoun use has evolved from a generic *he* to the balanced ("gender-fair") paired form *he/she*. It is safe to say that *he/she* does not represent all genders and reinforces gender binarity (Bigler and Leaper, 2015; Lindqvist et al., 2019; Renström et al., 2023a). Hence, balancing, as it has been used in the binary sense, is not gender-fair.

Again, the term I suggest to better encompass what gender-fair language aims to achieve is *gender-inclusive*. Hence, I will use *gender-inclusive pronouns* when referring to both neopronouns and nonbinary pronouns when they aim to include all genders. It should be noted that this is not always the case. This is the case for the English pronoun singular *they*, which can be used to include all genders. However, regarding specifically nonbinary pronouns such as English *xe*, this is not the case—*xe* is supposed to be used for individuals with nonbinary gender identities in the same sense that *she* is assumed to be used for women and *he* for men.

The implementation of neo- and nonbinary pronouns

Swedish was the first language to officially implement a third pronoun (Bäck et al., 2018) in the Swedish Academy Glossary, constituting a non-official guide to the Swedish language (SAOL, 2015). However, *hen* was mentioned in 1966 in a newspaper and suggested by the linguist Rolf Dunås as a replacement for the paired form *he/she*. Dunås was inspired by the Finnish language that does not have pronouns representing women and men (i.e., *she* and *he*), but only a gender-inclusive pronoun, *hän* (Ledin and Lyngfelt, 2013). However, at this time, *hen* did not have any noticeable impact. Instead, in the early 2010s, *hen* started being used as a nonbinary pronoun in some LGBTQIA communities. The public breakthrough for *hen* came in 2012 when a children's book was published where the main character was referred to as *hen* and a debate article was written by the book's publishers and a feminist linguist advocating for *hen* as a gender-inclusive pronoun. This action was coordinated and cleverly focused on children, guaranteeing an impactful public debate. The debate article discussed that gendered pronouns hinder children, and by using *hen* and gender-neutral language, all children could identify with the main character (Milles, 2013). This coordinated action sparked a heated, polarized, and politicized debate about the nature of gender and the rights of parents and children. While the debate was quite negative, it made all Swedish speakers aware of the pronoun *hen* (Gustafsson Sendén et al., 2015).

Not all neo- and nonbinary pronouns are implemented in this way. When it comes to singular *they*, there was mainly a change in how the pronoun was used, which subsequently led to a change in the Merriam-Webster dictionary regarding the definition of singular *they*. Singular *they* is traced back to the late 1300s when it appeared in sentences replacing *he* or *she* (Oxford English Dictionary).[1] However, its use as specifically referring to individuals with nonbinary gender identities is relatively new. For instance, Merriam-Webster appointed *they* as Word of the Year in 2015, which gained quite a lot of attention and helped bring awareness to the word. In 2022, almost all American English-speaking participants in a study on singular *they* knew that the word could be used nonbinary (Renström et al., 2023a).

Singular *they* is the most well-known gender-inclusive pronoun in English. Singular *they* has been a long-standing candidate in English for expressing gender neutrality (Balhorn, 2004). Before, some guides (e.g., Strunk and White, 1972) suggested that *he* should be used generically and argued against singular *they* due to it being a plural form, which by default cannot refer to a single entity. Generic *he* was opposed based on arguments that it is associated with masculinity and men, even though it is supposed to be neutral (Moulton et al., 1978).

There are many more initiatives in English, such as *ze* and *xe,* to mention a few. However, these pronouns have not received the same public attention and are relatively unknown. In 2022, about 25% of American English speakers knew that *ze* was a nonbinary pronoun (Renström et al., 2023a). Similarly, a Dutch study recently found that only half of the participants had knowledge about gender-inclusive pronouns in Dutch (DeCock et al., 2024). This is problematic because if people are unaware of the use of a pronoun, they are unlikely to use it, and it may also influence how a text with the pronoun is processed. Hekanaho (2020), for instance, found that when encountering *ze*, some people simply thought it was a typo, and it should be *he*.

Renström et al. (2024) have explored general attitudes to *hen* as a "gender-neutral" pronoun in a series of studies. As expected, participants were quite negative in the early studies of 2012 and 2015 (Gustafsson Sendén et al., 2015). In one study analyzing changes in attitudes between 2015 and 2018, Gustafsson Sendén et al. (2021) found that time was the strongest predictor of attitudes, even when controlling for several other important factors. They observed the main change among younger people. In a follow-up study in 2021, Renström et al. (2024) found that the change from 2018 to 2021 is smaller, but this time, the change mainly lies with middle-aged and older people. This data suggests that time has the potential to influence attitudes in a longer perspective, and people become more positive or at least less negative (Renström et al., 2024). However, one caveat with this data is that there is no separation between the different meanings of *hen*—that is, when participants answered the questions about their

1 https://www.oed.com/discover/a-brief-history-of-singular-they?tl=true

attitudes to *hen*, we did not know if they were thinking about *hen* as nonbinary or as a generic pronoun replacing *he/she*. This distinction is consequential because there may be different origins of resistance against gender-inclusive pronouns depending on how their meaning is perceived (Renström et al., 2022a).

Language and cognition

Languages are dynamic constructs, constantly changing to accommodate variations in the world and how people perceive objects and events. However, although language reflects how the world is constructed, language also *influences* the construction of the world (Fiedler, 2008; Lucy, 1992). This means that language can modify how individuals perceive objects and other people—a phenomenon known as the Whorfian hypothesis or the linguistic relativity hypothesis (Whorf, 1956). This idea is also captured in the term conceptual engineering—the striving for change in linguistic practices by, for instance, the use of novel words (Koch and Lupyan, 2024).

The linguistic relativity hypothesis states that language influences how the world is conceptualized (Samuel et al., 2019; Whorf, 1956). In its original form, this hypothesis claimed that language could completely change concepts, but a softer version - with more empirical support - claimed that language should be seen as a way to influence perceptions (Lucy, 1992). For example, a definite article defining the feminine/masculine gender of a certain noun appears to influence how that noun is perceived. In one study, bilingual participants took part in an experiment conducted in English and were asked to describe a bridge. Because the experiment was performed in English, no grammatical gender cues were present. Yet German-speaking participants used mainly stereotypically feminine terms to describe a bridge (e.g., slender, elegant) while Spanish-speaking participants used mainly stereotypically masculine terms (e.g., big, sturdy). The conclusion was that because the noun bridge is feminine in German (*Die brücke*) but masculine in Spanish (*el puente*), grammatical gender bias perceptions of non-gendered objects (Boroditsky et al., 2003). Consequently, it can be assumed that, in general, German and Spanish-speaking individuals have different prototypes of the noun bridge.

A prototype can be defined as an exemplar of a category that best represents the category and is contingent upon individual experiences and culture. Different exemplars can have varying degrees of belonging to a conceptual category, making some exemplars more representative than others (Rosch, 1973). In this sense, we can assume that in most languages, people have clear prototypes of the concept of gender that include normative cis-women and -men with a certain appearance and having certain qualities. Even though such prototypes, or the concept of gender, may vary as a function of context (for a discussion, see, for instance, Mazzuca et al., 2020, 2024), the predominant view of gender in Western countries is binary. When the prototype is activated, exemplars that fit better with this prototype will likely be more positively evaluated than exemplars with a worse fit. Research shows that when women and men display gender-stereotype congruent behavior, they are more positively evaluated than when the stereotype and behavior is incongruent (Eagly and Karau, 2002).

More recently, a study using natural language processing showed a similar connection between the grammatical gender of an inanimate noun and the verbs and adjectives associated with that noun (Williams et al., 2021). These studies show that language influences human conceptualization (Flaherty, 2001) and that gender cues in language have the power to influence how gender is constructed. Following the linguistic relativity hypothesis, a substantial change in how a concept (e.g., gender) is grammatically defined (e.g., through personal pronouns) should influence that conceptualization (e.g., the construction of gender) (see also Borghi and Mazzuca, 2023 for a discussion on properties of a concept that facilitate linguistic relativity).

It should be noted that some studies have not found the expected link between grammatical gender in language and conceptualization (e.g., Elpers et al., 2022), and some scholars question whether grammatical gender is a useful tool to investigate linguistic relativity (Samuel et al., 2019). Regardless, research on the negative consequences of gender cues in language is vast. Gender is more or less grammatically salient in different languages. Grammatical language structures can be categorized by how nouns and pronouns are gendered regarding feminine and masculine grammatical gender (Gygax et al., 2019; Prewitt-Freilino et al., 2012). Three language groups have been categorized: languages with natural gender where pronouns but not nouns are gendered (e.g., English, Norwegian); gendered languages where both nouns and pronouns are gendered (e.g., French, Russian); and genderless languages where neither nouns nor pronouns are gendered (e.g., Finnish, Turkish) (Prewitt-Freilino et al., 2012; Siewierska, 2013; Stahlberg et al., 2007).

Nouns and pronouns with feminine/masculine grammatical genders activate gender categorization, stereotypes, and prejudice (Bigler and Leaper, 2015; Lindqvist et al., 2019). The grammatical structure of languages relates to prejudice against women, sexism, and national-level gender equality (DeFranza et al., 2020; Prewitt-Freilino et al., 2012). For example, the more the grammar system in a language distinguishes between feminine and masculine gender, the fewer women are represented in business power positions (Santacreu-Vasut et al., 2014) and the larger the wage gaps between women and men are (Shoham and Lee, 2018).

In many grammatically gendered languages, the generic and plural form of social roles is often masculine (although exceptions exist). Some roles and nouns have suffixes indicative of men and masculinity, such as chair*man*, but also hu*man*. These suffixes provide linguistic gender cues, which are associated with men (McConnell and Fazio, 1996). Note that there are exceptions such as female suffixes for role nouns as well, for example, mid*wife* in English, but also *sjuksyster* (Swedish) or *Krankenswester* (German), both of the latter meaning "sister for the sick," which is the word for a nurse. Regardless of whether the suffix is masculine or feminine, such suffixes indicate that one gender is better suited than another for the role, which biases perceivers and perpetuates gender stereotypes.

Furthermore, it is more common with masculine suffixes for roles and nouns that are not stereotypically associated with men or women, such as chair*man* or ombuds*man*. Using such masculine generics influences cognitive processing and bias retrieval of masculine exemplars to a greater extent than feminine exemplars, resulting in a male bias (Hellinger and Bussman, 2002; Stahlberg et al., 2007). In such cases, the male bias is evoked due to linguistic cues associated with men.

A male bias is also observed when non-gendered words are associated with masculinity (Lindqvist et al., 2019; Liu et al., 2018). For example, an undefined person is often perceived as a man (Bailey

and LaFrance, 2017; Bem, 1993; Hegarty and Buechel, 2006). This bias has been found for linguistically neutral words (e.g., the word "the applicant") that do not carry any grammatical or semantical gender cues (Lindqvist et al., 2019). That such neutral terms are associated with masculinity exemplifies an androcentric worldview where men constitute the norm (Eagly and Kite, 1987). There is also a whiteness bias, meaning that the perception of an undefined person is that of a White person (Bailey and LaFrance, 2017). Moreover, there is a heteronormative bias where the perception of an undefined woman or man is that of a heterosexual woman or man (Klysing, 2023). Together, these biases in supposedly "neutral" words imply that the mental associations of non-defined categories are filled by the norm.

Even in languages without gendered pronouns, such as Finnish and Turkish, their respective pronouns have a male bias (Renström et al., 2023b). In a recent article, general descriptions of an individual in Finnish or Turkish referred to as *hän* (i.e., the Finnish third-person singular pronoun) or *o* (i.e., the Turkish third-person singular pronoun) resulted in more associations with a man than a woman. It is interesting that this was the case in both countries, which vary greatly in national-level gender equality (as assessed by the World Economic Forum, 2021). One conclusion is that androcentrism is deeply rooted in the human mind and possibly associated with a patriarchal culture (Renström et al., 2023b).

Given that most Western societies are patriarchal, there also exists a "male as norm" effect, which shows in varying ways in language and perpetuates the precedence of men and masculinity over other gender identities. Men and masculine terms are often more prevalent than women, and feminine terms and men are presented first. For instance, the paired pronoun form *he/she* presents the masculine form before the feminine form. Such practices lead to perceptions of men as the norm and women as the deviation from the norm. This matters in hiring processes, for instance. If the norm is masculine and perceivers have a man in mind for a position, a woman will be more negatively evaluated, while a man appears more suitable because he constitutes a match. For instance, when masculine generic terms are used in job advertisements, women become less motivated to apply because they do not feel targeted (Bem and Bem, 1973; Gaucher et al., 2011). Thus, androcentrism, not only in language, influences the demand side of hiring processes by influencing who is suitable for a position but also influences who will apply for the position, that is, the supply side. Hence, language is consequential on both micro and macro levels.

In sum, gender cues in language, whether grammatical, linguistic, or based on the order of presentation, tend to favor men over women or other genders and often function to increase binary gender stereotypes about women and men. Hence, gender in language is problematic, which has long been noted by women's activists and gender-interested scholars from varying disciplines.

Decreasing the negative consequences of gender in language

In psychology, gender-fair language describes a research field dedicated to resolving how language can be used to increase equality and inclusion—that is, to activate broader, rather than narrower, category boundaries. Previously, gender-fair language initiatives have mainly focused on decreasing linguistic androcentrism (i.e., the conflation of humanity with masculinity) by either increasing the linguistic salience of femininity or neutralizing the salience of gender in language altogether (Sczesny et al., 2016). Which strategy that has been predominant in a language is mainly contingent upon the grammatical structure. In languages with gendered nouns and pronouns (i.e., gendered languages), the reform balancing or feminization has dominated. The background is found in the widespread use of masculine generics, and research shows that such forms are male-biased. Hence, increasing women's visibility has been an important step in these languages. In languages with gendered pronouns but not nouns (i.e., natural gendered languages), a neutralization strategy is more often employed, implying the removal of gendered information in language.

To increase linguistic salience of femininity, feminine word forms can be included in addition to masculine generics, such as the use of *he/she* instead of generic *he* or including feminine forms in occupational titles (e.g., *Lehrer/Lehrerinnen*, meaning male/female teacher in German). Such reforms make women linguistically visible (Sczesny et al., 2016). However, pairing feminine/masculine words can reinforce the notion of gender/sex as a binary construct (Butler, 1988; Lindqvist et al., 2020; Morgenroth et al., 2020) when the paired forms are presented as a unit highlighting only two genders. Hence, individuals with intersex variations and/or nonbinary gender identities become invisible (Hyde et al., 2019; Lindqvist et al., 2020).

While balancing strategies aim to visualize feminine gender, neutralization aims to decrease the influence of gender on cognition by reducing the frequency of gendered words in a language. For example, using *chair* instead of *chairman* activates broader and more gender-inclusive characteristics associated with that position. Other examples are to use *firefighters* instead of *firemen* or *police officers* instead of *policemen*. As expected, avoiding masculine generics in professional titles and job adverts makes women more interested in applying for that position (Sczesny et al., 2016).

These strategies, however, pose some problems. First, balancing (or feminization) is obviously not balancing in an all-encompassing sense since it refers to the balancing of feminine and masculine forms. This strategy thus diminishes the visibility of other gender identities. Moreover, balancing feminine and masculine forms by presenting them as two parts of a whole—two mutually exclusive categories—reinforces a binary view of gender (Lindqvist et al., 2020) and binary gender stereotypes about women and men (Bigler and Leaper, 2015). Another problem with this strategy is order effects. When a word pair is presented, the word that is presented first is often processed as the dominating, more important, or hierarchically higher one (Hegarty et al., 2016). When word pairs are presented, they are most often presented with the masculine form first (e.g., *he/she*, or in the case of role nouns, *studenten/studentinnen* in German), although there are exceptions such as *ladies and gentlemen*. Gabriel and Gygax (2008) found that the word presented first was given more attention than the word presented second. When the word businesswoman was presented before the word businessman, the woman-form was seen as more central than the man-form (Kesebir, 2017). However, this presentation format with the woman-form first is less common.

Regarding neutralization, role forms are often not gendered but neutral. However, this is a modification of the truth because in many of these languages, such as Norwegian, the historical masculine form is the form most often used (Swan, 1992). In such languages, the feminine form has more or less been completely dropped from the language. While not considered a masculine generic in that the

masculine form is used generically simultaneously with a feminine form existing, it is nevertheless a masculine form. When role names are not clearly stereotypically gendered, such as nurse or pilot, the historically and grammatically masculine forms are male biased (Gabriel and Gygax, 2008). This means that when neutralized forms that most often are historically masculine are used, people tend to associate these with men and masculinity unless the neutralized form is heavily stereotypical.

Both strategies proposed to make the language more inclusive and ensure the visibility of different genders have notable flaws. In a review article, Gabriel et al. (2018) discussed these two strategies. Modifying gendered role nouns, such as replacing gendered suffixes with neutral ones, is relatively straightforward—at least from the perspective of language production (Gabriel et al., 2018). For instance, changing *fireman* to *firefighter* poses no significant linguistic or grammatical challenges. However, resistance may arise due to ideological views. In contrast, for languages with more grammaticalized gender, implementing such changes is more complex, as it requires altering entire sentence structures.

Toward a new terminology: de-gendering and multi-gendering strategies

Instead of the terms balancing/feminization and neutralization, which are closely connected to linguistic reforms, I use the broader terms proposed by Morgenroth et al. (2020) and Morgenroth and Ryan (2020)—*de-gendering* and *multi-gendering*. These strategies aim to counter the general binary conceptualization of gender and are not limited to language. De-gendering strategies aim to decrease the influence of gender by removing gender cues. While not limited to linguistic reforms, when applied to language, this practice equates to neutralization. Multi-gendering strategies, such as balancing in language, aim to emphasize the diversity of gender identities but explicitly include identities beyond the binary genders. While balancing has traditionally been used to increase the visibility of women in language, multi-gendering strategies explicitly incorporate nonbinary identities into linguistic practices.

To illustrate how de-gendering and multi-gendering strategies can be implemented, Klysing et al. (2021) showed organization descriptions to participants in an experiment, where the descriptions contained different equal employment opportunity statements, which either emphasized binary gender, gender as diverse (multi-gender), or gender as irrelevant (de-gender). In a control condition, participants were not shown any statement. The results showed that gender minority participants felt more secure in organizations that used either a de-gendering or multi-gendering statement, which increased organizational attractiveness. The type of statement had no effect on gender majority participants. These results clearly illustrate the consequentiality of both de-gendering and multi-gendering strategies.

As previously mentioned, gender-inclusive pronouns, such as neo- and nonbinary pronouns, may carry a dual meaning. Depending on the context in which they are used, these pronouns can align with either de-gendering or multi-gendering strategies (Renström et al., 2022a). Before exploring the dual meaning of such pronouns and consequences, I will first discuss the relationship between pronouns and the construction of gender as a binary or nonbinary concept.

Pronouns and the construction of gender

Gender (sometimes referred to as sex/gender) is most often constructed as a binary concept. The binary conceptualization of gender prescribes appropriate and proscribes inappropriate traits and behaviors for women and men (Morgenroth et al., 2020; Prentice and Carranza, 2002; Renström et al., 2023a). When these expectations are violated, individuals are often socially sanctioned, which explains the marginalization of gender minority groups (Thoma et al., 2021) but also contributes to the stereotyping of cis-gender women and men (Rudman et al., 2012). Thus, decreasing the binary conceptualization is imperative.

Pronouns denote gender identity. Although *she* and *he* are not synonymous with feminine or masculine gender identities, pronouns have increasingly become markers of gender identity (Hekanaho, 2020). In everyday language, pronouns facilitate communication by allowing stereotypical inferences about individuals when pronouns representing women and men are used (i.e., *she* and *he*). If a person is referred to as *he*, social perceivers assume that this target is a man and should perform stereotypically masculine tasks and roles and have corresponding attitudes (Eagly, 1987). When meeting or hearing about a new person, an impression is formed that provides a mental framework for understanding the target's behavior (Riggio and Friedman, 1986). By using the available information, such as pronouns or gender, social perceivers can effectively categorize targets into meaningful categories. While such inferences are based on stereotypes and thus should not be assumed to be valid for every category member, they facilitate everyday interactions and communications by decreasing cognitive load (Macrae et al., 1994). This implies that the use of gendered pronouns influences social perception, which may have far-reaching consequences. However, what happens when gender-inclusive pronouns, such as neopronouns or nonbinary pronouns, are used? How does such use influence social perceivers and the construction of gender?

In one experiment, Lindqvist et al. (2019) tested to what extent different labels, including the paired pronoun form *he/she* and the gender-inclusive pronoun *hen* in Swedish, were associated with a male bias. Participants were shown a description of a candidate for a job as a real estate agent (which is gender balanced when it comes to women and men, according to Swedish official statistics), ostensibly written by a professional recruiter. The candidate was referred to using the labels *the applicant*, *NN* (short for Latin Nomen Nescio, translating to "do not know the name," which is sometimes used to anonymize), *he/she,* or *hen*. Participants were then asked to indicate who they thought they had read about by selecting one out of several photos showing cis-gender women and men. The results showed that both *the applicant* and *NN* were male-biased, with participants selecting photos of men to a significantly larger degree than what would have been expected had they selected randomly. However, both *he/she* and *hen* led to a more even distribution of selected women and men.

Moreover, the experiment was replicated in English using *the applicant*, the paired form *he/she*, the nonbinary neopronoun *ze,* and the gender-inclusive pronoun singular *they*. The results showed that *the applicant* was again male-biased, but singular *they* was also male-biased. While singular *they* has been quite successfully launched as a gender-inclusive, nonbinary pronoun, social

perceivers categorize singular *they* as a man. *he/she* and *ze* resulted in an even distribution of women and men targets, thus not being male biased. One difficulty with *ze* is that many participants had not heard about *ze* as a nonbinary pronoun. Hekanaho (2020) found that some people may understand *ze* as a misspelled *he*. That appears unlikely in this case since that should have led participants to select a photo of a man.

Nevertheless, the fact that many participants (approximately 60% in this study) had no knowledge about *ze* as a pronoun cannot be disregarded, and we do not know how this influences social perceivers' gender conceptualization. These results indicate that neutral terms are associated with men and masculinity, even without grammatical or linguistic cues. This could be based on a patriarchal and androcentric worldview. Hence, when social perceivers hear about a person referred to using a non-gendered term in a neutral context, they infer masculinity (Gabriel and Gygax, 2008).

While *he/she* was not associated with a male bias, there are problems using paired pronouns as discussed above, such as order effects, the presentation of a coherent unit consisting of two mutually exclusive categories, and stereotypes associated with women and men. To explore the consequences of using the paired form *he/she* on gender conceptualization, Renström et al. (2023a) explored a *normative gender bias*. A normative gender bias is when a word (or word pair) is associated with individuals with normative cis-gender appearances. This was explored in three studies performed in both Swedish and English. In experiment 1, participants were shown a sentence composed of a target (e.g., *The person*) and a pronoun (*he/she* or *hen*) referring to the target. Participants were then asked to select a photo of the person they thought the sentence was about. Photos depicted not only cis-normative looking women and men but also more queer and non-normative looking individuals. The results showed that participants that had read the paired pronoun form *he/she* tended to select photos of normative looking individuals to a larger extent that non-normative looking individuals. In the *hen*-condition, there was no difference. That is, participants who read about a target person referred to as *hen*, tended to select photos of normative and non-normative looking individuals randomly. In a second experiment, the set-up was similar to the previously discussed experiment testing a male bias with the exception that the job now described a candidate for a position as a train attendant and that participants now could select photos of non-normative looking individuals. These results confirmed that *he/she* was associated with normative looking women and men, while *hen* was not. These results clearly indicate that paired forms, or feminization/balancing strategies make nonbinarity invisible. In support of this, Mirabella et al. (2024) show that in a grammatically gendered language (Italian), individuals with nonbinary gender identities report difficulties in expressing their identity, which could be because there is no readily available nonbinary pronoun (Koch and Lupyan, 2024). Similar results have been found for French (Knisely, 2020).

In a replication in English, the pronouns singular *they*, *ze,* and *he/she* were compared. The results showed that *he/she* was normatively biased as expected and that singular *they* was associated with normative-looking individuals. Hence, singular *they* is both male-biased and normatively biased. The neopronoun ze was not associated with any specific gender expression. Yet again, people may be unfamiliar with using *ze* as a nonbinary pronoun.

To date, the results paint a relatively complicated picture. The neopronoun *hen* in Swedish appears to broaden gender categories beyond the binary. However, in English, the story is more complicated. The most popular nonbinary pronoun, singular *they*, appears to be connected not only to masculinity but also to normativity. One reason may be that singular *they* has historical roots where its use has been generic, when gender is unknown or unimportant (Balhorn, 2004). As previously discussed, such neutrality is often associated with masculinity (Lindqvist et al., 2019; Bailey and LaFrance, 2017). While the neo- and nonbinary pronoun *ze* appears to perform better in that *ze* appears to include both women and non-normativity, there is still uncertainty about how readers perceive this word if they have no previous knowledge.

Dual meaning of gender-inclusive pronouns

As previously mentioned, some gender-inclusive pronouns function in two ways. First, they can increase the visibility of gender identities outside the traditional binary categories when used as referents for a specific individual with a nonbinary gender identity—that is they are used to multi-gender. Second, gender-inclusive pronouns can be used generically, which instead decreases the influence and visibility of gender in language altogether—that is, they are used to de-gender. In this section, I will discuss attitudes and acceptability of the different meanings and how this dual meaning influences perceptions of the pronouns and conceptualization of gender.

Attitudes and acceptance of different meanings of gender-inclusive pronouns

As mentioned, general attitudes toward the gender-inclusive pronoun *hen* in Swedish have become increasingly positive, or at least less negative, over 10 years (Gustafsson Sendén et al., 2015, 2021; Renström et al., 2024). General attitudes were the focus in these surveys, not attitudes toward different meanings.

In a study by Renström et al. (2022a), the authors assessed indirect attitudes toward *hen* in de-gendering and multi-gendering contexts by asking participants to rate sentences in both contexts. Participants were shown sentences with the paired pronoun *he/she* or *hen* in de-gendering contexts such as: "When a train attendant is sick, hen [he/she] should stay home," or in a multi-gendering context referring to a specific individual: "Lex took a nap; she [he, hen] was very tired." The participants rated the sentences on grammatical correctness, reading difficulty, and negative valence. The results showed that *hen* used in generic contexts (de-gendering) was more accepted, as shown in higher ratings on grammaticality and lower ratings of reading difficulty and negative valence, compared to *hen* in specific contexts (multi-gendering).

Similar findings are reported for singular *they* in English (Bradley et al., 2019; Bradley, 2020). Sentences with generic singular *they* were evaluated more positively (more grammatically correct and less offensive) compared to sentences with specific singular *they*. Similarly, Renström and Klysing (2024) found that singular *they* was more favorably evaluated in generic contexts than in specific contexts using a comparable sentence rating paradigm. Hekanaho (2020) found in a

survey that attitudes toward nonbinary singular *they* were more positive than those toward other neopronouns, aligning with the findings of Renström et al. (2024).

However, many participants also disagreed with the nonbinary use of singular *they* as well, even when they accepted its use as a generic pronoun. One conclusion from this research was that many participants viewed gender as a binary construct, which led to their opposition to nonbinarity (Hekanaho, 2020).

Taken together, these studies indicate that people tend to have relatively positive attitudes about the use of gender-inclusive pronouns in generic, de-gendering contexts but not in nonbinary contexts.

To better gauge the acceptance of gender-inclusive pronouns, in another experiment (Renström et al., 2022a), participants were shown similar sentences as in the sentence rating paradigm described above, but with the pronoun missing and asked to fill in the missing word. The results showed that *hen* was overwhelmingly popular in generic contexts, even more so than the paired form *he/she*. In specific contexts, however, the participants preferred to assign a binary gender, even though the name used was neutral. The same result was found for singular *they* (Renström and Klysing, 2024): In generic contexts, singular *they* was preferred over binary pronouns, but in specific contexts, participants assigned a binary gender to the target by using *he* or *she*. These usage results align with what Hekanaho (2020) found, that singular *they* was more popular to use in generic contexts than binary gendered pronouns.

Regarding other pronouns, such as *ze*, there is less research to report. One reason is that *ze* is relatively unknown to English speakers in general (Hekanaho, 2020; Lindqvist et al., 2019; Renström et al., 2023a), which explains why some people may believe *ze* is misspelled *he* (Hekanaho, 2020). This implies that results from research using *ze* without specifying its use to the participants may suffer substantial measurement errors. Renström et al. (2024), using the same sentence rating paradigm and *ze* in generic and specific contexts, found that the nonbinary meaning of *ze* was seen as more grammatically correct than the generic meaning. There were no differences in reading difficulty and negative valence between using *ze* as a generic or a nonbinary pronoun. This could indicate that people interpret *ze* as a nonbinary pronoun designated for nonbinary people; hence, it should not be used generically. Hekanaho (2020) found that this is how participants reflected on *ze* in her study—*ze* should be seen as a gendered pronoun in a similar way that *he* and *she* is and hence is equally unsuitable to use generically, as, for instance, generic *he* is.

In sum, it seems that the generic meaning of gender-inclusive pronouns such as singular *they* and *hen,* are accepted, while the nonbinary meaning is still resisted. An empirical question is how these pronouns when specified in the different contexts influence gender conceptualizations.

To better understand how gender-inclusive pronouns in different meanings influence gender conceptualization, Renström et al. (2023a) performed the experiment described earlier where participants read about a candidate for a job position and were asked to select among a set of photos whom they thought was described in the text. This time, they added information about why the pronoun, in this case, *ze* and singular *they*, was used, thus eliminating ambiguous interpretations that may have been present in the earlier studies. That is, participants were in different conditions informed that the pronoun was used because the person had a nonbinary gender identity (multi-gendering), to anonymize (de-gendering), or they were not provided any information in a control condition. The results showed that when participants were explicitly informed that the pronoun was used because the referent had a nonbinary gender identity, the earlier normative bias was reversed. Participants tended to select a non-normative looking individual to a larger extent than a normative looking individual. This result was particularly strong for *ze* but, to a lesser degree, was also present for singular *they*. Interestingly, when participants did not receive any information in the control condition and were informed that the pronoun was used generically, the results were the same, showing a normative bias for both singular *they* and *ze*. This implies that when not informed about the use and when the context is ambiguous – that is, the pronoun could be used either to de-gender or to multi-gender, readers tend to infer a generic use, and this also leads to a normative bias.

Finally, because *ze* has been shown to be fairly unknown to the English-speaking population (Lindqvist et al., 2019), participants were asked if they had knowledge about the use of different pronouns as being nonbinary (Renström et al., 2023a). If people do not have knowledge about the use of a certain word, they will reasonably have problems implementing and interpreting the word. Because singular *they* was known to almost all participants (98%), analyses of differences in knowledge were not possible to run. However, for *ze*, knowledge about *ze* as a nonbinary pronoun dramatically decreased the normative bias, although there was still a tendency to select a normative-looking individual. In total, 204 participants had knowledge about *ze* (24% of the sample), meaning that this is quite a small sample to draw any definite conclusions from, but it appears that knowledge, at least when it comes to neopronouns, is conducive to its influence on gender conceptualization. However, an open question is why this knowledge appeared to be inconsequential for gender conceptualization for singular *they,* given that singular *they* was heavily associated with normative gender expressions even though the majority of the participants did know that *they* was a nonbinary pronoun (as well as a generic).

The greater acceptability of gender-inclusive pronouns' use in generic compared to nonbinary contexts is likely rooted in the way that nonbinarity poses a challenge to cis-normativity (Morgenroth et al., 2020; Renström and Klysing, 2024). As mentioned, this duality may be related to different motives behind resistance against gender-inclusive pronouns (Renström et al., 2022a; Renström and Klysing, 2024), which I will discuss further in the next section.

Politicization of gender and pronouns

Gender, along with other contentious issues such as the environment and race, has become increasingly politicized and, consequently, polarized. Consequently, gender-inclusive language has also been politicized. Politicization can be understood as a shift from discussing how things are to a debate driven by parties or partisans seeking to advance a political agenda. Mazzuca and Santarelli (2023) suggest this process can occur through various features. For instance, the framing of an issue—particularly the idea that things could be different—is a hallmark of politicization, though issues can also be reconstructed in the process.

A clear consequence of the politicization of the Swedish pronoun *hen* is that when *hen* was first introduced, one of the largest newspapers in Sweden forbid its journalists from using *hen* as it would send signals about political standings.

Moreover, the Language Council of Sweden, providing unofficial recommendations on language use, recommended that *hen* be avoided because it was so highly politically charged.[2] Even though this recommendation was later revoked, the need for caution in using *hen* was still emphasized. In line with this, extensive research shows that gender-inclusive language reforms, including the use of gender-inclusive pronouns, face resistance (Blaubergs, 1980; Parks and Roberton, 1998; Bradley, 2020; Bradley et al., 2019; Hekanaho, 2020; Vergoossen et al., 2020a).

This resistance appears largely grounded in gender-ideological convictions about gender being an essential and binary category and a desire to keep gendered power structures intact (Douglas and Sutton, 2014; Parks and Roberton, 2005). Yet, some people oppose such reforms based on more linguistic reasons, such as preferring the linguistic status quo (Vergoossen et al., 2020a) or so-called linguistic prescriptivism (Bradley, 2020). However, according to Hekanaho (2020), such arguments about linguistic awkwardness may reflect a preference for a more socially accepted opposition but be rooted in discomfort with nonbinarity.

In a qualitative analysis of arguments against the use of Swedish *hen*, Vergoossen et al. (2020a) found four dimensions of resistance. Using earlier taxonomies of critical arguments against gender-fair language reforms in the past (i.e., Blaubergs, 1980; Parks and Roberton, 1998), they found that roughly 80% of the arguments against the use of *hen* could be coded into the previously existing categories. Hence, while *hen* is a new word with new implications, the arguments against its use were the same as the arguments against the use of the paired form *he/she* when it was suggested to be used instead of a generic *he,* and these arguments have been consistent over almost 50 years. Two categories of arguments were new and specifically related to *hen* as a gender-inclusive pronoun: gender-inclusive pronouns distract communication, and gender information about a target person is important. Based on the coded categories, four overarching dimensions captured assumptions and beliefs underlying criticism against gender-inclusive or gender-fair language reforms. The dimension with the most arguments (ca. 40%) was a defense of the linguistic status quo, encompassing arguments mainly relating to that change is too difficult or unnecessary. The second dimension was related to the fact that sexism and cisgenderism are acceptable (encompassing ca. 30% of the arguments). Here, arguments about binary biology and hostility against people with nonbinary gender identities were dominant. The third dimension, diminishing of the issue and its proponents (27%), contained disparaging reactions to both gender-inclusive language and people who advocate for it. Hostile, ridiculing, and denigrating comments were common. This dimension is also closely connected to the nonbinary use of *hen*, which is what the arguments were about. Finally, some people claimed they did not want to use *hen* because they perceived it to be distracting in communication (6%). Although relatively small, this latter dimension indicates the political nature of gender-inclusive pronoun use (Vergoossen et al., 2020a).

Taken together, most of the categories of arguments were related to the nonbinary use of *hen*. But, what predicts these negative attitudes? In their general attitude surveys, Gustafsson Sendén et al. (2015, 2021) and Lindqvist et al. (2016) found that some predictors appear relatively stable. For instance, individuals with a more left-leaning ideology, those identifying as women or nonbinary, younger people, and those with a general interest in gender issues tend to hold more positive attitudes toward *hen*.

However, as de-gendering and multi-gendering strategies to some extent challenge core convictions of different ideologies, the origins of resistance against gender-inclusive pronouns may also differ depending on which strategy is salient (Morgenroth et al., 2020; Renström et al., 2022a), necessitating further scrutiny of different predictors for different meanings of gender-inclusive pronouns.

Ideological origins of resistance

De-gendering strategies aim to remove or minimize the salience of gender altogether, while multi-gendering strategies aim to draw attention to the fact that gender is not binary (Morgenroth and Ryan, 2020). The duality of gender-inclusive pronouns may be related to different origins of resistance (Renström et al., 2022a). Because de-gendering strategies remove gender cues, it may lead individuals to not think about gender and, therefore, not question its binary nature (Morgenroth et al., 2020). In relation, singular *they*, used in a de-gendering way, is associated with normative gender expressions even though most participants reported that they were knowledgeable about singular *they* as a nonbinary pronoun (Renström et al., 2023a).

The use and perception of gender-inclusive language reforms are not just a matter of personal preference or identity. They also relate to the motivation to defend a hierarchical and traditional binary gender system as expressed in language. Social dominance orientation (SDO) and right-wing authoritarianism (RWA) are two status-legitimizing ideologies, meaning that they entail the endorsement of worldviews where social inequalities between groups are seen as legitimate (Major and Kaiser, 2017). Both SDO and RWA predict gendered prejudice, including sexism (Duckitt and Sibley, 2010; Van Assche et al., 2019) and homophobia (Crawford et al., 2016). It therefore appears plausible that both SDO and RWA might predict negativity to gender-inclusive pronouns through their role in challenging conservative beliefs about group hierarchies. However, there are important differences between SDO and RWA that may influence resistance against gender-inclusive pronouns differently depending on the active strategy.

Social Dominance Orientation expresses the motivational goal for group-based dominance and superiority (Duckitt, 2001; Pratto et al., 1994; Sidanius and Pratto, 1999; Sidanius et al., 2004). People high in SDO see the world as a competitive jungle, which entails a struggle for resources (Duckitt et al., 2002)—a zero-sum game. In terms of gender relations, this means a belief that if sexual minority groups gain status and privileges, the higher-status gender groups lose out, making people high in SDO motivated to counteract minority rights progress (Poteat and Mereish, 2012). Individuals high in SDO should be primarily concerned with keeping a linguistic structure highlighting the order and construction of gender hierarchies in society, such as

2 The language council of Sweden is the official organ in Sweden that deals with language care and language politics. It is part of the Institute for language and folk lore, https://www.isof.se/other-languages/english/about-the-institute.

using the paired pronoun form *he/she*. Moreover, this pronoun use also positions men as the dominant group, which could explain why men often are more negative toward gender-inclusive language that diminishes group saliency than women (Douglas and Sutton, 2014; Lindqvist et al., 2016). As de-gendering strategies function to remove gender cues, individuals high in SDO may be more skeptical of such procedures and thus particularly dislike the use of singular *they* in de-gendering contexts. Individuals high in SDO might not be overly concerned about singular *they* in multi-gendering contexts because this highlights a "third" gender group in society that the ingroup can dominate, be it men or women. In a study using the sentence rating paradigm and the fill-in-the-blanks task described earlier, Renström and Klysing (2024) found that individuals high in SDO were less positive when evaluating sentences with singular *they* in a de-gendered context, but there was no effect of SDO on evaluations of sentences in the multi-gendered contexts. This could reflect that individuals who prefer a clear gender structure and hierarchy do not mind an extra gender group since that group will be a minority group and, hence, a group that could be dominated. Recently, the triple form *he/she/they* has started to emerge. It is possible that individuals high in SDO might prefer this form even more than *he/she* since it positions a "third" group. Additionally, given the importance of ordering—such as presenting *he* before *she*—the triple form might be more appealing to women with high levels of SDO.

Right-wing authoritarianism (RWA) is an ideologically based personality feature where individuals who are high in RWA desire tradition and conformity to conventional ways and are characterized by an emphasis on submission to authority and upholding norms of social order (Altemeyer, 1981). Therefore, a high level of RWA leads to dislike of individuals who violate these norms (Peresman et al., 2021). Individuals high in RWA perceive norm-violating social groups, such as gender non-conforming people, as threats to the ingroup and, therefore, become motivated to protect the ingroup against these threats (Renström et al., 2022b; Sidanius et al., 2004). One expression of a traditional and conservative belief system is the view of gender as an essential and binary category (Hyde et al., 2019; Tee and Hegarty, 2006).

The multi-gendering strategy challenges these traditional gender roles and norms by implying that there are more than two genders, thereby questioning the validity of binary gender as a system of societal organization (Morgenroth and Ryan, 2020). For instance, gender essentialism has been linked to prejudice against individuals who violate gender norms, such as women in leadership positions (Skewes et al., 2018). Similarly, binary views of gender have been associated with increased prejudice toward nonbinary individuals (Morgenroth et al., 2020).

In the context of singular *they*, transcendent views of gender (i.e., lower endorsement of strict gender roles) were positively associated with attitudes toward singular *they* when used in a nonbinary, multi-gendering context but not when used in a de-gendering context (Bradley et al., 2019). This indicates that traditional views and beliefs about gender as binary are more strongly tied to the understanding of gender-inclusive pronouns in nonbinary contexts. As a general preference for traditions and norms, RWA correlates with the endorsement of traditional gender roles (Peterson and Zurbriggen, 2010). Dislike of gender norm violators (Perez-Arche and Miller, 2021) should, therefore, lead to stronger resistance to the multi-gendered strategy due to it questioning the validity of a traditional, binary gender system.

Renström and Klysing (2024) found that RWA predicted negativity toward singular *they* in the multi-gendered contexts. In Study 1, people high in RWA were less likely to use singular *they* in multi-gendered contexts. In Study 2, RWA predicted negative evaluations of singular *they* in the multi-gendered contexts but less so in de-gendered contexts. RWA encompasses views of minorities as threatening to the majority's conventional way of life (Renström et al., 2022b; Sidanius et al., 2004). Since highlighting the existence of other gender identities than the binary woman/man might constitute such a threat, individuals high in RWA should be more negative toward such practices.

Taken together, these results point to the importance of considering the different contexts or meanings that some gender-inclusive pronouns have to better understand resistance against them and why this may vary.

Challenges and future directions

Pronouns are considered a closed word class that rarely changes, but recently, several languages have implemented additional third-person pronouns. Pronouns have become important identity markers and have consequences for how individuals relate to themselves and others, meaning we face a unique situation. Researchers should take the opportunity to follow the implementation of gender-inclusive pronouns in other languages, as this is a rare opportunity. In this review, I have tried to cover some aspects of gender-inclusive pronouns that are important to consider in this line of research.

First, when researching nonbinary pronouns specifically, one must consider the general knowledge of the pronoun. For instance, English *ze* is relatively unknown, which influences how results using this pronoun should be interpreted and is likely to lead to large measurement errors and, potentially null effects. In addition, a recent study on Dutch found that half of the sample was not familiar with gender-inclusive pronouns in Dutch (DeCock et al., 2024), indicating problems in measuring attitudes.

Second, the dual meaning of some gender-inclusive pronouns, which are well-known, also runs the risk of compromising interpretation and introducing measurement error if the research is not designed with this dual meaning in mind. Given that different ideological convictions may underlie resistance against using gender-inclusive pronouns in different contexts (i.e., corresponding to the de-gendering and multi-gendering strategies), failure to include this as a factor will lead to difficulties in drawing conclusions.

Relatedly, but also separate from the dual meaning and the underlying ideological resistance, is the politicization of gender. When gender-inclusive language, including pronouns, becomes a political position statement, the study of their effects on gender conceptualization also becomes compromised. This polarization is unlikely to recede over time. New research shows that younger men perceive gender equality as a threat to a larger extent than older men (Off et al., 2022). Moreover, the recent uprise of the misogynistic online milieu, referred to as the manosphere and its influencers, mainly addresses young men who are more susceptible to its anti-feminist messages than older men (Renström and Bäck, 2024).

Moreover, such misogynistic communities often intersect with alt-right communities, both of which promote traditional gender roles and nuclear families as societal ideals. Gender-inclusive language, which aims to increase the visibility of sexual minorities and reduce gender biases, directly challenges these ideals and is thus met with resistance. However, their implementation faces significant challenges if resistance persists and political polarization intensifies. Currently, there is very little research in this area, and a deeper understanding is needed of how the politicization of gender shapes attitudes toward, the use of, and the effects of gender-inclusive pronouns.

To date, there is also limited research on pronouns in other languages published in English. Hence, cross-cultural studies that explore gender-inclusive pronouns in different languages using similar study setups are desired. In such research, it is also desirable to include languages that vary in grammatical gender. As mentioned, gender-inclusive language may be more or less difficult to implement depending on grammatical structure. Regarding pronouns specifically, there is a lack of research on grammatical gendered languages. However, a recent Italian study showed that nonbinary individuals struggled to express their identity linguistically (Mirabella et al., 2024), which could be a consequence of the grammatical gender structure of Italian. Such research highlights the need for further scrutiny of pronouns in different languages.

Moreover, the present review is mainly concerned with Swedish and English, and while I call for more research on other Western languages with a variation of grammatical gender, a global perspective is also desirable. This would provide both a new linguistic and cultural/political perspective. For instance, there is no differentiation between feminine and masculine pronouns in spoken Chinese, but there is in written Chinese.

Practice implications

As should be evident, words and word choices matter. But how should this be translated into practice when the concept of gender is so contentious that use of the associated words has become political position statements? This question has no straightforward answer, and more research is required. What can be stated is that, despite the contentious nature of gender issues, gender-majority participants are not overwhelmingly negative toward gender-inclusive pronouns (e.g., Renström et al., 2022a, 2022b; Bradley, 2020). Furthermore, as shown by eye-tracking studies, gender-inclusive pronouns in texts are not particularly difficult to process (Vergoossen et al., 2020b).

In terms of gender-fair language, an important finding is that multi-gender strategies in equal opportunity statements within organizational descriptions did not reduce gender-majority participants' organizational attraction (Klysing et al., 2021).

However, in languages such as German, gender-inclusive language—particularly paired forms—was associated with reduced comprehensibility (Friedrich and Heise, 2019).

These results suggest that while some individuals may resist gender-inclusive language, its effects are not insurmountable. A practice recommendation, therefore, is to adopt gender-inclusive language whenever possible. This recommendation is further supported by research indicating that exposure to gender-fair language in languages such as Norwegian and German increases subsequent usage of such language (Kuhn, 2021).

Author contributions

ER: Conceptualization, Funding acquisition, Investigation, Writing – original draft, Writing—review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The present article was funded by the Swedish Research Council (grant no. 2023-00731).

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Altemeyer, B. (1981). Right-wing authoritarianism. Winnipeg, MA: University of Manitoba Press.

Bäck, E. A., Lindqvist, A., and Gustafsson Sendén, M. (2018). Hen. Bakgrund, attityder och användande. *Psykologiska Rapporter Från Lund* 8, 1–35.

Bailey, A. H., and LaFrance, M. (2017). Who counts as human? Antecedents to androcentric behavior. *Sex Roles* 76, 682–693. doi: 10.1007/s11199-016-0648-4

Balhorn, M. (2004). The rise of epicene they. *J. Eng. Ling.* 32, 79–104. doi: 10.1177/0075424204265824

Bem, S. L. (1993). The lenses of gender: Transforming the debate on sexual inequality. New Haven, CT: Yale University Press.

Bem, S. L., and Bem, D. J. (1973). Does sex-biased job advertising "aid and abet" sex discrimination? *J. Appl. Soc. Psychol.* 3, 6–18. doi: 10.1111/j.1559-1816.1973.tb01290.x

Beukeboom, C. J., and Burgers, C. (2019). How stereotypes are shared through language: a review and introduction of the social categories and stereotypes communication. *Rev. Commun. Res.* 7, 1–37. doi: 10.12840/issn.2255-4165.017

Bigler, R. S., and Leaper, C. (2015). Gendered language: psychological principles, evolving practices, and inclusive policies. *Policy Insights Behav. Brain Sci.* 2, 187–194. doi: 10.1177/2372732215600452

Blaubergs, M. S. (1980). An analysis of classic arguments against changing sexist language. *Women's Stu. Int. Q.* 3, 135–147. doi: 10.1016/S0148-0685(80)92071-0

Borghi, A. M., and Mazzuca, C. (2023). Grounded cognition, linguistic relativity, and abstract concepts. *Top. Cogn. Sci.* 15, 662–667. doi: 10.1111/tops.12663

Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). "Sex, syntax, and semantics" in Language in mind: Advances in the study of language and thought. eds. D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press).

Bradley, E. D. (2020). The influence of linguistic and social attitudes on grammaticality judgments of singular 'they'. *Lang. Sci.* 78:101272. doi: 10.1016/j.langsci.2020.101272

Bradley, E. D., Schmid, M. H., and Lombardo, H. (2019). Personality, prescriptivism and pronouns: factors influencing grammaticality judgements of gender-neutral language. *English Today* 35, 41–52. doi: 10.1017/S0266078419000063

Butler, J. (1988). Performative acts and gender constitution: an essay in phenomenology and feminist theory. *Theatr. J.* 40, 519–531. doi: 10.2307/3207893

Crawford, J. T., Brandt, M. J., Inbar, Y., and Mallinas, S. R. (2016). Right-wing authoritarianism predicts prejudice equally toward "gay men and lesbians" and "homosexuals.". *J. Pers. Soc. Psychol.* 111, e31–e45. doi: 10.1037/pspp0000070.supp

Decock, S., Van Hoof, S., Soens, E., and Verhaegen, H. (2024). The comprehensibility and appreciation of nonbinary pronouns in newspaper reporting: the case of hen and die in Dutch. *Appl. Linguis.* 45, 330–347. doi: 10.1093/applin/amad028

DeFranza, D., Mishra, H., and Mishra, A. (2020). How language shapes prejudice against women: an examination across 45 world languages. *J. Pers. Soc. Psychol.* 119, 7–22. doi: 10.1037/pspa0000188

Douglas, K. M., and Sutton, R. M. (2014). "A Giant leap for mankind" but what about women? The role of system-justifying ideologies in predicting attitudes toward sexist language. *J. Lang. Soc. Psychol.* 33, 667–680. doi: 10.1177/0261927X14538638

Duckitt, J. (2001). "A dual-process cognitive-motivational theory of ideology and prejudice" in Advances in experimental social psychology. ed. M. P. Zanna (New York, NY: Academic Press), 41–113.

Duckitt, J., and Sibley, C. G. (2010). Personality, ideology, prejudice and politics: A dual-process motivational model. *J. Pers.* 78, 1861–1894. doi: 10.1111/j.1467-6494.2010.00672.x

Duckitt, J., Wagner, C., du Plessis, I., and Birum, I. (2002). The psychological bases of ideology and prejudice: testing a dual process model. *J. Pers. Soc. Psychol.* 83, 75–93. doi: 10.1037/0022-3514.83.1.75

Eagly, A. H. (1987). Sex differences in social behavior: A social-role interpretation. New York, NY: Lawrence Erlbaum Associates, Inc.

Eagly, A. H., and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. doi: 10.1037/0033-295X.109.3.573

Eagly, A. H., and Kite, M. E. (1987). Are stereotypes of nationalities applied to both women and men? *J. Pers. Soc. Psychol.* 53, 451–462. doi: 10.1037/0022-3514.53.3.451

Elpers, N., Jensen, G., and Holmes, K. J. (2022). Does grammatical gender affect object concepts? Registered replication of Phillips and Boroditsky (2003). *J. Mem. Lang.* 127:104357. doi: 10.1016/j.jml.2022.104357

Fiedler, K. (2008). Language a toolbox for sharing and influencing social reality. *Perspect. Psychol. Sci.* 3, 38–47. doi: 10.1111/j.1745-6916.2008.00060.x

Flaherty, M. (2001). How a language gender system creeps into perception. *J. Cross-Cult. Psychol.* 32, 18–31. doi: 10.1177/0022022101032001005

Friedrich, M. C. G., and Heise, E. (2019). Does the use of gender-fair language influence the comprehensibility of texts? An experiment using an authentic contract. *Swiss J. Psychol.* 78, 51–60. doi: 10.1024/1421-0185/a000223

Gabriel, U., and Gygax, P. (2008). Can societal language amendments change gender representation? The case of Norway. *Scand. J. Psychol.* 49, 451–457. doi: 10.1111/j.1467-9450.2008.00650.x

Gabriel, U., Gygax, P. M., and Kuhn, E. A. (2018). Neutralising linguistic sexism: promising but cumbersome? *Group Process. Intergroup Relat.* 21, 844–858. doi: 10.1177/1368430218771742

Gaucher, D., Friesen, J., and Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *J. Pers. Soc. Psychol.* 101, 109–128. doi: 10.1037/a0022530

Gustafsson Sendén, M., Bäck, E. A., and Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Front. Psychol.* 6:893. doi: 10.3389/fpsyg.2015.00893

Gustafsson Sendén, M., Renström, E. A., and Lindqvist, A. (2021). Pronouns beyond the binary—the change of attitudes and use over time. *Gend. Soc.* 35, 588–615. doi: 10.1177/08912432211029226

Gygax, P. M., Elminger, D., Zufferey, S., Garnham, A., Sczesny, S., von Stockhausen, L., et al. (2019). A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Front. Psychol.* 10:1604. doi: 10.3389/fpsyg.2019.01604

Hegarty, P., and Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965-2004. *Rev. Gen. Psychol.* 10, 377–389. doi: 10.1037/1089-2680.10.4.377

Hegarty, P., Mollin, S., and Foels, R. (2016). "Binomial word order and social status" in Advances in intergroup communication. eds. H. Giles and A. Maass (London: Peter Lang Publishing), 119–135.

Hekanaho, L. (2020). Generic and nonbinary pronouns. Usage, acceptability and attitudes. Helsinki: Helsingfors University.

Hekanaho, L. (2022). A thematic analysis of attitudes towards English nonbinary pronouns. *J. Lang. Sexuality* 11, 190–216. doi: 10.1075/jls.21025.hek

Hellinger, M., and Bussman, H. (2002). Gender across languages. The linguistic representation of women and men, Amsterdam: John Benjamins Publishing Company.

Hjort Miltersen, E. (2020). Singular de and its referential use in talk-in-interaction. *Scand. Stu. Lang.* 11, 22–61. doi: 10.7146/sss.v11i2.123477

Hjorth-Nebel Miltersen, E., Clendenning-Jiménez, S., and Papadopoulos, B. (2022). Danish. Gender in Language Project. Available at: www.genderinlanguage.com/danish (Accessed April 22, 2024).

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., and van Anders, S. M. (2019). The future of sex and gender in psychology: five challenges to the gender binary. *Am. Psychol.* 74, 171–193. doi: 10.1037/amp0000307

Kesebir, S. (2017). Word order denotes relevance differences: the case of conjoined phrases with lexical gender. *J. Pers. Soc. Psychol.* 113, 262–279. doi: 10.1037/pspi0000094

Klysing, A. (2023). Who is woman and who is man? Normativity at intersections of gender and sexual orientation. Lund: Lund University.

Klysing, A., Renström, E. A., Gustafsson Sendén, M., and Lindqvist, A. (2021). Gender diversity in recruitment: influence of gender trouble on applicant attraction and evaluation. *J. Appl. Soc. Psychol.* 52, 781–802. doi: 10.1111/jasp.12809

Knisely, K. A. (2020). Le français non-binaire: linguistic forms used by nonbinary speakers of French. *Foreign Lang. Ann.* 53, 850–876. doi: 10.1111/flan.12500

Koch, S., and Lupyan, G. (2024). What is conceptual engineering good for? The argumetn from nameability. Providence, RI: Philosophy and Phenomenological Research.

Kuhn, E. A. (2021). Strategies that affect individuals' gender-fair language use. Doctoral thesis. Trondheim: Norwegian University of Science and Technology.

Ledin, P., and Lyngfelt, B. (2013). Om bruket av hen i bloggar, tidningstexter och studentuppsatser [the use of hen in blogs, newspaper articles and student work, in Swedish]. *Språk och Stil* 22, 141–174.

Lindqvist, A., Gustafsson Sendén, M., and Bäck, E. A. (2016). Vem tycker om hen? [Who likes hen? In Swedish]. *Språk Och Stil* 26, 101–129.

Lindqvist, A., Gustafsson Sendén, M., and Renström, E. A. (2020). What is gender, anyway: a review of the options for operationalising gender. *Psychol. Sexuality* 12, 332–344. doi: 10.1080/19419899.2020.1729844

Lindqvist, A., Renström, E. A., and Gustafsson Sendén, M. (2019). Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies. *Sex Roles* 81, 109–117. doi: 10.1007/s11199-018-0974-9

Liu, A. H., Shair-Rosenfield, S., Vance, L. R., and Csata, Z. (2018). Linguistic origins of gender equality and women's rights. *Gend. Soc.* 32, 82–108. doi: 10.1177/0891243217741428

Lucy, J. A. (1992). Language diversity and thought: A reformulation of the linguistic relativity hypothesis. Cambridge: Cambridge University Press.

Macrae, C. N., Milne, A. B., and Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *JJ. Pers. Soc. Psychol.* 66, 37.

Major, B., and Kaiser, C. R. (2017). Ideology and the maintenance of group inequality. *Group Process. Intergroup Relat.* 20, 582–592. doi: 10.1177/1368430217712051

Mazzuca, C., Borghi, A. M., van Putten, S., Lugli, L., Nicoletti, R., and Majid, A. (2024). Gender is conceptualized in different ways across cultures. *Lang. Cogn.* 16, 353–379. doi: 10.1017/langcog.2023.40

Mazzuca, C., Majid, A., Lugli, L., Nicoletti, R., and Borghi, A. M. (2020). Gender is a multifaceted concept: evidence that specific life experiences differentially shape the concept of gender. *Lang. Cogn.* 12, 649–678. doi: 10.1017/langcog.2020.15

Mazzuca, C., and Santarelli, M. (2023). Making it abstract, making it contestable: politicization at the intersection of political and cognitive science. *Rev. Philos. Psychol.* 14, 1257–1278. doi: 10.1007/s13164-022-00640-2

McConnell, A. R., and Fazio, R. H. (1996). Women as men and people: effects of gender-marked language. *Personal. Soc. Psychol. Bull.* 22, 1004–1013. doi: 10.1177/01461672962210003

Milles, K. (2013). En öppning i en sluten ordklass? Den nya användningen av pronomenet hen [an opening in a closed word class? The new use of the pronoun hen]. *Språk & Stil* 23, 107–140.

Mirabella, M., Mazzuca, C., De Livio, C., Di Giannantonio, B., Rosati, F., Lorusso, M. M., et al. (2024). The role of language in nonbinary identity construction: Gender words matter. *Psychol. Sexual Orientation Gender Div.* 16:729. doi: 10.1037/sgd0000729

Morgenroth, T., Gustafsson Sendén, M., Lindqvist, A., Renström, E. A., Ryan, M., and Morton, T. A. (2020). Defending the sex/gender binary: the role of gender identification and need for closure. *Soc. Psychol. Personal. Sci.* 12, 731–740. doi: 10.1177/1948550620937188

Morgenroth, T., and Ryan, M. K. (2018). Gender trouble in social psychology: how can Butler's work inform experimental social psychologists' conceptualization of gender? *Front. Psychol.* 9:1320. doi: 10.3389/fpsyg.2018.01320

Morgenroth, T., and Ryan, M. K. (2020). The effects of gender trouble: an integrative theoretical framework of the perpetuation and disruption of the gender/sex binary. *Perspect. Psychol. Sci.* 16, 1113–1142. doi: 10.1177/1745691620902442

Moulton, J., Robinson, G. M., and Elias, C. (1978). Psychology in action - sex bias in language use—neutral pronouns that arent. *Am. Psychol.* 33, 1032–1036. doi: 10.1037/0003-066X.33.11.1032

Off, G., Charron, N., and Alexander, A. (2022). Who perceives womens rights as threatening to men and boys? Explaining modern sexism among young men in Europe. *Front. Polit. Sci.* 4:909811. doi: 10.3389/fpos.2022.909811

Parks, J. B., and Roberton, M. A. (1998). Influence of age, gender, and context on attitudes toward sexist/nonsexist language: is sport a special case? *Sex Roles* 38, 477–494. doi: 10.1023/A:1018766023667

Parks, J. B., and Roberton, M. A. (2005). Explaining age and gender effects on attitudes toward sexist language. *J. Lang. Soc. Psychol.* 24, 401–411. doi: 10.1177/0261927X05281427

Peresman, A., Carroll, R., and Bäck, H. (2021). Authoritarianism and immigration attitudes in the UK. *Political Stu.* 71, 616–633. doi: 10.1177/00323217211032438

Perez-Arche, H., and Miller, D. J. (2021). What predicts attitudes toward transgender and nonbinary people? An exploration of gender, authoritarianism, social dominance, and gender ideology. *Sex Roles J. Res.* 85, 172–189. doi: 10.1007/s11199-020-01212-3

Peterson, B. E., and Zurbriggen, E. L. (2010). Gender, sexuality, and the authoritarian personality. *J. Pers.* 78, 1801–1826. doi: 10.1111/j.1467-6494.2010.00670.x

Poteat, V. P., and Mereish, E. H. (2012). Ideology, prejudice, and attitudes toward sexual minority social policies and organizations. *Psychology* 33, 211–224. doi: 10.1111/j

Pratto, F., Sidanius, J., Stallworth, L. M., and Malle, B. F. (1994). Social dominance orientation: a personality variable predicting social and political attitudes. *J. Pers. Soc. Psychol.* 67, 741–763. doi: 10.1037/0022-3514.67.4.741

Prentice, D. A., and Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: the contents of prescriptive gender stereotypes. *Psychol. Women Q.* 26, 269–281. doi: 10.1111/1471-6402.t01-1-00066

Prewitt-Freilino, J. L., Caswell, A. T., and Laakso, E. K. (2012). The gendering of language: a comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles* 66, 268–281. doi: 10.1007/s11199-011-0083-5

Renström, E. A., Bäck, H., and Carroll, R. (2022b). Protecting the ingroup? Authoritarianism, immigration attitudes, and affective polarization. *Front. Psychol.* 4:919236. doi: 10.3389/fpos.2022.919236

Renström, E. A., and Klysing, A. (2024). Ideological origins of resistance against gender-inclusive language reforms: singular they as a de-gendering or multi-gendering strategy. *Political Psychol.* 5, 1–19. doi: 10.1111/pops.13058

Renström, E. A., Lindqvist, A., Akbas, G., Hekanaho, L., and Gustafsson Sendén, M. (2023b). Are gender-neutral pronouns really neutral? Testing a male bias in the grammatical genderless languages Turkish and Finnish. *J. Lang. Soc. Psychol.* 42, 476–487. doi: 10.1177/0261927X221146229

Renström, E. A., Lindqvist, A., and Gustafsson Sendén, M. (2022a). The multiple meanings of the gender-inclusive pronoun hen: predicting attitudes and use. *Eur. J. Soc. Psychol.* 52, 71–90. doi: 10.1002/ejsp.2816

Renström, E. A., and Bäck, H. (2024). Manfluencers and Young Men's Misogynistic Attitudes: The Role of Perceived Threats to Men's Status. *Sex Roles*.

Renström, E. A., Lindqvist, A., Klysing, A., and Gustafsson Sendén, M. (2023a). Personal pronouns and person perception - do paired and nonbinary pronouns evoke a normative gender bias? *Br. J. Psychol.* 115, 253–274. doi: 10.1111/bjop.12686

Riggio, R. E., and Friedman, H. S. (1986). Impression formation: the role of expressive behavior. *J. Pers. Soc. Psychol.* 50, 421–427. doi: 10.1037/0022-3514.50.2.421

Rosch, E. H. (1973). Natural categories. *Cogn. Psychol.* 4, 328–350. doi: 10.1016/0010-0285(73)90017-0

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., and Nauts, S. (2012). Status incongruity and backlash effects: defending the gender hierarchy motivates prejudice against female leaders. *J. Exp. Soc. Psychol.* 48, 165–179. doi: 10.1016/j.jesp.2011.10.008

Samuel, S., Cole, G., and Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychon. Bull. Rev.* 22, 1–20. doi: 10.3758/s13423-019-01652-3

Santacreu-Vasut, E., Shenkar, O., and Shoham, A. (2014). Linguistic gender marking and its international business ramifications. *J. Int. Bus. Stud.* 45, 1170–1178. doi: 10.1057/jibs.2014.5

SAOL (2015). Svenska Akademins Ordlista 14 [the dictionary of the Swedish academy]. Stockholm: Nordstedts.

Sczesny, S., Formanowicz, M., and Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Front. Psychol.* 7:25. doi: 10.3389/fpsyg.2016.00025

Shoham, A., and Lee, S. M. (2018). The causal impact of grammatical gender marking on gender wage inequality and country income inequality. *Bus. Soc.* 57, 1216–1251. doi: 10.1177/0007650317696231

Sidanius, J., and Pratto, F. (1999). Social dominance: An intergroup theory of social hierarchy and oppression. Cambridge: Cambridge University Press.

Sidanius, J., Pratto, F., Van Laar, C., and Levin, S. (2004). Social dominance theory: its agenda and method. *Polit. Psychol.* 25, 845–880. doi: 10.1111/j.1467-9221.2004.00401.x

Siewierska, A. (2013). "Gender distinctions in independent personal pronouns" in The world atlas of language structures online. eds. M. S. Dryer and M. H. Leipzig (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Skewes, L., Fine, C., and Haslam, N. (2018). Beyond Mars and Venus: the role of gender essentialism in support for gender inequality and backlash. *PLoS One* 13:e0200921. doi: 10.1371/journal.pone.0200921

Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S. (2007). "Representation of the sexes in language" in Social communication. ed. K. Fiedler (London: Psychology Press), 163–187.

Strunk, W., and White, E. B. (1972). The elements of style. *2nd* Edn. New York, NY: Plain Label Books, 55–56.

Swan, T. (1992). All about eve: women in Norwegian newspapers in the 20th century. *Working Papers Lang. Gender Sexism* 2, 37–54.

Tee, N., and Hegarty, P. (2006). Predicting opposition to the civil rights of trans persons in the United Kingdom. *J. Community Appl. Soc. Psychol.* 16, 70–80. doi: 10.1002/casp.851

Thoma, B. C., Eckstrand, K. L., Montano, G. T., Rezeppa, T. L., and Marshal, M. P. (2021). Gender nonconformity and minority stress among lesbian, gay, and bisexual individuals: a meta-analytic review. *Perspect. Psychol. Sci.* 16, 1165–1183. doi: 10.1177/1745691620968766

Van Assche, J., Koç, Y., and Roets, A. (2019). Religiosity or ideology? On the ideological predictors of sexism. *Personal. Individ. Differ.* 139, 191–197.

Vergoossen, H. P., Pärnamets, P., Renström, E. A., and Gustafsson Sendén, M. (2020b). Are new gender-neutral pronouns difficult to process in Reading? The case of hen in SWEDISH. *Front. Psychol.* 11:574356. doi: 10.3389/fpsyg.2020.574356

Vergoossen, H. P., Renström, E. A., Lindqvist, A., and Gustafsson Sendén, M. (2020a). Four dimensions of criticism against gender-fair language. *Sex Roles* 83, 328–337. doi: 10.1007/s11199-019-01108-x

Wagener, A. (2021). No need to iel: why France is so angry about a gender-neutral pronoun. *The* Conversation.

Whorf, B. L. (1956). Language, thought, and reality: Selected writings. Cambridge, Cambridge, Mass: Technology Press of Massachusetts Institute of Technology.

Williams, A., Cotterell, R., Wolf-Sonkin, L., Blasi, D., and Wallach, H. (2021). *Trans. Assoc. Comput. Ling.* 9, 139–159. doi: 10.1162/tacl_a_00355

World Economic Forum (2021). Global Gender Gap Report. Available at: https://www3.weforum.org/docs/WEF_GGGR_2021.pdf (Accessed February 18, 2024).

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 19 March 2025
DOI 10.3389/fpsyg.2025.1517311

OPEN ACCESS

EDITED BY
Antonio Bova,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Daniele Panizza,
University of Messina, Italy
Sarah V. Di Pietro,
University of Zurich, Switzerland

*CORRESPONDENCE
Rafiah Ansari
✉ rafiah.ansari.2@city.ac.uk

RECEIVED 25 October 2024
ACCEPTED 07 January 2025
PUBLISHED 19 March 2025

CITATION
Ansari R, Chiat S, Cartwright M and Herman R (2025) Vocabulary interventions for children with developmental language disorder: a systematic review.
Front. Psychol. 16:1517311.
doi: 10.3389/fpsyg.2025.1517311

Vocabulary interventions for children with developmental language disorder: a systematic review

Rafiah Ansari [1]*, Shula Chiat[1], Martin Cartwright[2] and Ros Herman[1]

[1]Department of Language and Communication Science, School of Health and Psychological Sciences, City St George's, University of London, London, United Kingdom, [2]Department of Health Services Research and Management, School of Health and Psychological Sciences, City St George's, University of London, London, United Kingdom

Introduction: Developmental language disorder (DLD) is a neurodevelopmental condition often characterised by vocabulary difficulties that lead to academic and social challenges. The acquisition of vocabulary is a complex, dynamic process of mapping word sound (phonology) to meaning (semantics) supported by contextual cues; a complexity that vocabulary interventions need to address. To understand the key features and impact of such interventions, a systematic review of word-learning studies involving children aged 5–11 with DLD was conducted.

Method: A structured search covered seven electronic databases for the period 1990–2023. In addition, the reference lists of identified studies were searched manually. Studies were appraised for quality and data was extracted relating to word-learning effectiveness and intervention characteristics. Findings were reported as written summaries and quantitative data ranges.

Results: Sixteen relevant studies were identified with most appraised as medium quality. Interventions tended to be delivered individually in school by speech and language therapists. The most common outcome measure was expressive target-word tests, such as picture naming and word definitions. Interventions explicitly targeting phonological and semantic word features had the most high-quality studies reporting significant vocabulary gain. The inclusion of stories to provide context implicitly during phonological and semantic interventions was beneficial, though stories alone were less effective. Specificity in learning was noted across studies. Gains did not generally transfer to non-targeted words and showed depreciation following therapy. Intervention responses were influenced by children's language profiles. For example, children with more severe language difficulties were less responsive to contextual cues during story reading and were more distracted by extraneous music during multimedia-supported word learning.

Discussion: Whilst the available studies have limitations in range and quality, they do suggest some benefits of combining explicit and implicit vocabulary strategies and considering children's presenting profiles. Implications for practitioners supporting the individual needs of children with DLD are discussed. This includes addressing issues with the generalization and maintenance of vocabulary gains by targeting the most relevant words and encouraging recall and self-management strategies. Further research should explore the influence of home-school carryover.

10.3389/fpsyg.2025.1517311

Systematic review registration: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022327345, PROSPERO, Reg: CRD42022327345.

KEYWORDS

developmental language disorder, child language acquisition, word learning, vocabulary interventions, systematic review

1 Introduction

Developmental language disorder (DLD) is a condition characterized by significant challenges in daily communication that are unlikely to resolve without specialist intervention (Bishop et al., 2017). It is estimated to affect 7% of children of primary school age, i.e., 5–11-year-olds (Norbury et al., 2016), and can present in isolation, or alongside other neurodevelopmental disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD) and Dyslexia (Bishop et al., 2017).

The term DLD was advocated by Bishop et al. (2017) following a multinational, multidisciplinary consensus study which aimed to align criteria and terminology for children with language difficulties. Contributors to the consensus included speech and language therapists (SLTs), educational psychologists, psychiatrists, pediatricians and specialist teachers. DLD was proposed as a descriptive label focusing on the sustained social and educational impact of language difficulties (Bishop et al., 2017). It served to replace Specific Language Impairment, a term that had been commonly used by clinicians and academics to identify language impairment based on discrepancies between language and intelligence scores (Stark and Tallal, 1981; Tomblin et al., 1997), but which had been criticized for variations in diagnostic threshold and criteria when applied in research and practice (Aram et al., 1993).

Whilst DLD can affect many aspects of language, a restricted vocabulary size and range is among the most common. Almost half of primary-school-aged children with DLD may struggle with vocabulary skills (Rice and Hoffman, 2015). Unsurprisingly, the identification of needs is most common in the primary school years (Lindsay and Strand, 2016), given that this period marks an expected estimated vocabulary growth from 3,000 to 8,000 words (Anglin et al., 1993; Biemiller and Slonim, 2001). This also means that, without intervention, children with DLD are at high risk of falling behind their peers.

Poor vocabulary rarely occurs on its own and is associated with wider language difficulties including issues with grammar and narration (Justice et al., 2018; Khan et al., 2021), whilst also predicting progression in areas of learning such as reading and numeracy, and overall academic attainment (Bleses et al., 2016; Matte-Landry et al., 2020). In addition, vocabulary development trajectories are strongly associated with subsequent behavioral and emotional needs (Westrupp et al., 2020).

1.1 Theoretical models

Language acquisition theory and research commonly use a connectionist model of language processing to explain how word learning is a dynamic process of identifying, connecting and mapping spoken sounds (phonology) to their correct meaning (semantics) (Plaut, 1999; Trueswell et al., 2013). It is suggested that children with DLD may struggle with their phonological processing, their semantic processing, and/or connecting the two during word learning, leading to issues with understanding (receptive) and/or use (expressive) of words (Best et al., 2015; Chiat, 2001).

Context plays a crucial role in word learning, providing social, perceptual, cognitive and linguistic signals to help connect sounds to meaning (Monaghan, 2017; Pomper, 2020). A cross-situational learning model has been proposed to describe how learners use contextual information from multiple situations to track the co-occurrence of word sounds and their meanings and to resolve ambiguity in sound-meaning associations (Roembke et al., 2023; Hartley et al., 2020). There are indications of restricted capacity in using this contextual inferencing to support language processing in children with DLD (Broedelet et al., 2023; McGregor et al., 2022).

1.2 Intervention approaches

Interventions for children with DLD are predominantly led by SLTs, who deliver evidence-based therapy in conjunction with parents, educators, and partner professionals such as educational psychologists. Therapy tends to occur in the home or clinic when children are younger and then usually moves to mainstream or specialist school settings. Language goals can vary but often include those relating to the child's interest, school curriculum and family routine (Dennis et al., 2017).

Empirical coverage of vocabulary interventions for children with DLD tend to focus on strategies that elaborate and connect the sound and meaning components of words (Steele and Mills, 2011). The goal is to increase children's accuracy in the understanding and use of word sound features (e.g., initial sounds, syllables, and rhymes) and word meaning features (e.g., function, location, category, attributes). Activities are explicit, meaning the child is actively taught word features through tasks involving imitation, repetition, feedback, and recall; written words and pictures often serve as prompts (for example activities see Parsons et al., 2005). The rationale is that by directly targeting the sound and meaning of words, children with DLD are supported to undertake the phonological and semantic integration required for word learning, a method that is in line with the connectionist model of language processing (Plaut, 1999; Trueswell et al., 2013).

An alternative approach to vocabulary interventions for children with DLD is the use of implicit, incidental strategies to provide contextual cues to support word learning, a method more aligned to a cross-situational learning model (Broedelet et al., 2023). Presenting target words in a narrative using story-based activities

is an example of a context-based intervention for vocabulary enrichment (see Nash and Donaldson, 2005). The content of the stories can facilitate vocabulary learning by providing information regarding word definitions as well as examples of how the word can be used outside of therapy (Marks and Stokes, 2010). In addition, the grammatical structures surrounding the target word when presented in a narrative can provide important clues around word meaning. This is referred to as syntactic bootstrapping, a process where the syntactic frame and morphological markers associated with a novel word help to determine the meaning of the word (Rice et al., 2000).

Whilst explicit semantic-phonological interventions and more indirect contextual interventions have differing theoretical basis and strategies, they can be considered complementary approaches to supporting vocabulary skills in children with DLD. A survey of SLT vocabulary-intervention practice for school-aged children (Steele, 2020, US survey with 357 respondents) identified that therapists most frequently used direct explicit strategies (endorsed by 91.3% of respondents) with the next most common approach being context-based strategies (endorsed by 79.7% of respondents). It is worth noting that therapy decisions were predominantly driven by professional experience rather than a consideration of research findings, a pattern that has emerged in other studies of SLT vocabulary-intervention practice (Justice et al., 2014; Marante and Hall-Mills, 2024). The reason appears to be difficulty applying research into practice given the heterogeneity of the DLD population. The nature, extent, and implications of language difficulties may vary not only between children but also within a child due to differential influence of internal physiological and psychological and external social and environmental factors (Law et al., 2022). This highlights the need for research that helps practitioners understand not only the outcomes of vocabulary-intervention studies, but also the key characteristics of the interventions to help incorporate evidence-based strategies into individualized support for children with DLD.

1.3 Intervention research

Whilst there are several systematic reviews that have synthesized the evidence base for vocabulary interventions for children with DLD, only two have covered the primary school period, i.e., 5–11 years. Cirrin and Gillam (2008) analyzed 21 peer-reviewed studies of language interventions for children aged 5–18 years. Of the studies reviewed by Cirrin and Gillam (2008) only one reported the significance and size of vocabulary intervention effects for children in the primary school years (Wing, 1990, $n = 10$, age 5; 11–7; 01). This was a non-randomized matched-group comparison of a specialist-school-based SLT-delivered intervention targeting word phonology (picture cards to support the understanding and use of initial sounds, syllables, and rhymes for target words) versus a semantic approach (picture cards to support understanding and use of word category, function, and attributes). Only the phonological approach led to a significant vocabulary gain with a moderate effect size ($p < 0.05$, $d = 0.7$) as measured using a standardized expressive vocabulary test.

The second systematic review (Rinaldi et al., 2021) focused on randomized controlled language intervention trials for children aged 3–8 years. Only one vocabulary intervention paper was identified (Smeets et al., 2014), comparing target words implicitly presented in picture or video animation narrated e-stories. The children, who were based in specialist schools, were encouraged to view the e-stories independently. Two studies were conducted, both adopting a randomized alternating treatment crossover design. The first study ($n = 28$, age 5; 0–6; 8) found significantly greater gain on an expressive sentence completion test for words exposed through e-stories compared to non-targeted words ($p < 0.001$, $d = 1.54$, large effect size) with greater gain for picture e-stories without background music than animated e-stories with background music ($p < 0.01$, $d = 0.48$, medium effect). The second study found that, in the absence of background music, vocabulary gain did not differ between e-story types ($n = 21$, age 5; 0–7; 6).

The above studies are informative, however, having only two vocabulary-intervention papers for primary-school aged children with DLD that have been reviewed as empirically robust, limits the evidence-base for practitioners, researchers and academics. Given that one review of studies is over a decade old, an update is warranted to cover subsequent relevant studies. While the other review, by considering only randomized control trials and having an upper age cut-off of 8 years, may have excluded studies of relevance. There is therefore a need for a systematic review that explores the fundamental features and impact of vocabulary interventions in studies that spans the primary school years and have sufficient methodological rigor to be of empirical value. In keeping with this, the current review addresses the following research question:

What are the key characteristics of vocabulary interventions for primary-school-aged children with DLD and their influence on word-learning outcomes?

2 Method

A systematic review of word-learning studies was undertaken to identify the core components of vocabulary interventions, the intervention effects, and the study design that generated the outcomes. The review was completed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, Page et al., 2021). The review is registered in PROSPERO (Patel et al., 2022, Reg: CRD42022327345; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022327345).

2.1 Eligibility

Eligible studies included children aged 5; 0–11; 11 with diagnoses of DLD or equivalent according to the criteria in Bishop et al. (2017) as described below:

- Children presenting with difficulty producing or understanding language that affected everyday functioning (everyday social interactions or educational progress).
- If the child was multilingual, then the difficulties presented in all languages.

- The presentation had to be suggestive of poor prognosis with difficulties emerging in the course of development. This is based on research evidence indicating that language problems apparent from preschool years that are still evident at 5 years and over are likely to persist (Stothard et al., 1998).
- The language difficulties could not be acquired or associated with a known biomedical cause. Children with a language difficulty secondary to a biomedical condition, where language needs occur as part of more complex impairments patterns, were excluded on the basis of requiring specialized intervention. Differentiating conditions include brain injury, neurodegenerative conditions, cerebral palsy, sensori-neural hearing loss, and genetic conditions such as Down syndrome. In line with recommendations from Bishop et al. (2017), autism and intellectual disability were considered differentiating conditions in this review as they are commonly linked to genetic or neurological causes.
- Children with a language and co-occurring cognitive, sensori-motor or behavioral disorders, which may affect pattern of language impairment and intervention response but where causal relation is unclear, were considered to meet the criteria for DLD. These co-occurring disorders include attentional problems, motor problems, literacy problems, speech problems, limitations of behavior and emotions.

While DLD is the empirically-advocated term for the population in this review (Royal College of Speech Language Therapists: RCSLT, 2021; McGregor et al., 2020), the search strategy also accounted for overlapping academic and clinical diagnostic labels including (specific) language impairment, language delay and language difficulties (Green, 2020; Georgan and Hogan, 2019).

The study designs considered were randomized control trials (RCTs), non-randomized controlled studies and pre-post comparisons. Single-subject designs were only considered if outcomes were measured at multiple timepoints. Studies had to have reported on or had sufficient data to calculate the significance of the change with or without effect size values.

Any intervention aimed at improving vocabulary (with or without standard treatment) was considered for the review as well as any comparator. Studies were required to have a vocabulary measure as a primary outcome. Secondary measures, such as grammar and literacy outcomes, were recorded but were not a key focus as it would be difficult to infer direct causal effects.

Only English language publications were reviewed, as time and resources were not available for reliable translation, though this is acknowledged as a limitation. Only studies reported on or after 1990 were considered based on the date of the earliest vocabulary intervention study identified in previous comparable systematic reviews (Cirrin and Gillam, 2008).

2.2 Search strategy

To identify relevant studies, searches were conducted for published trials between 1990 and 2023 in PubMed, CINAHL, PsycINFO, the Cochrane Library, and ERIC. In addition, unpublished literature between 1990 and 2023 was searched using SCOPUS and Open Dissertations.

For each database a search strategy was developed by considering MESH and free terms which covered the following: (Language AND (disorder* OR impair* OR delay* OR difficult*)) AND (child* OR infant* OR P?ediatric*) AND (vocabulary OR word*) AND (therap* OR intervention* OR instruction* OR treatment* OR teaching OR learning OR support*). In addition, manual searches of reference lists of identified studies were conducted.

2.3 Manual data management

Study data were transferred for refinement and coding using EPPI-Reviewer systematic review software (V4 https://eppi.ioe.ac.uk/cms/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4).

2.4 Data selection

The primary author reviewed selected papers, screened titles and abstracts to remove ineligible studies, then conducted a full text review of the remaining articles to identify eligible studies. A third of selected titles and abstracts, as well as all selected full texts, were reviewed for consensus by a secondary rater. To quantify the level of agreement between the two raters, the Cohen's Kappa measure of inter-rater reliability was used. An agreement value of 0.82 was achieved for the selection of titles and abstracts and then again for the selection of full texts. This was acceptable as a value >0.8 is considered satisfactory (Pérez et al., 2020).

2.5 Data extraction

Data extraction was guided by the Cochrane Data Extraction Form for RCTs and non-RCTs (Wilson, 2016) and the Template for Intervention Description and Replication (TIDieR; Hoffmann et al., 2014). Collectively, this provided information on the studies investigating the interventions (author, year, country, design, participants, sample sizes, target words, outcome measures, follow up) and the characteristics of the interventions (type, dosage, provider, mode, location).

2.6 Quality appraisal

The selected studies were appraised to assess their methodological quality and the extent to which each study had addressed the possibility of bias in its design, conduct and analysis. The Joanna Briggs Institute critical appraisal checklists (Tufanaru et al., 2020) were used to assess quality as they can be applied to multiple study designs and have precedent for use with DLD populations (Alduais et al., 2022; Wanicharoen and Boonrod, 2024; Zupan et al., 2022). The primary author and a

second rater independently appraised each included study and reached full consensus.

The checklists assessed areas relating to selection bias, study design, confounders, and data collection methods with a choice of yes/no/unclear/not applicable responses. Full details of the checklist and results are available in Appendix A. Raw scores were calculated for each selected study by dividing the number of positive responses by the total number of applicable statements in the critical checklists, these were then converted to percentage scores. Studies with percentage scores of <49% were classed as low quality with high risk of bias, studies between 50% and 79% as medium quality, and studies ≥80% as high in quality with low risk of bias. This classification has been used in prior systematic reviews (e.g., Zupan et al., 2022).

As per recommended guidelines (Tufanaru et al., 2020), studies of low-quality were included in the qualitative reporting to provide a complete view of the evidence available to inform the review question. However, findings from low-quality studies were omitted from the quantitative synthesis to minimize the impact of study biases when pooling numerical data.

3 Results

This section details the word-learning studies identified in the review and provides a narrative summary of intervention characteristics. This is followed by a quantitative synthesis of studies with comparable design in order to synthesize intervention outcomes by prominent intervention characteristics.

3.1 Identification of studies

The search was completed on 1st October 2023 and database alerts were set up to signpost any new studies that were subsequently published (none were identified). In total, the search yielded 16 studies meeting the eligibility criteria, 12 with interventions delivered in English (UK, 7 studies; US, 4 studies; New Zealand, 1 study), two in Dutch (set in the Netherlands), one in French (set in Switzerland), and one in German (set in Germany), providing collective data for 288 participants aged 5; 0–11; 11 (167 males, 78 females, 43 unknown). A PRISMA flow diagram of the article screening process is presented in Figure 1.

3.2 Study data

A summary of data from the selected studies is presented in Tables 1–4. Studies in this review were aligned to four intervention approach types to enable better synthesis and analysis of findings: interventions that explicitly focused on phonological and semantic word features to support learning (Table 1), interventions utilizing

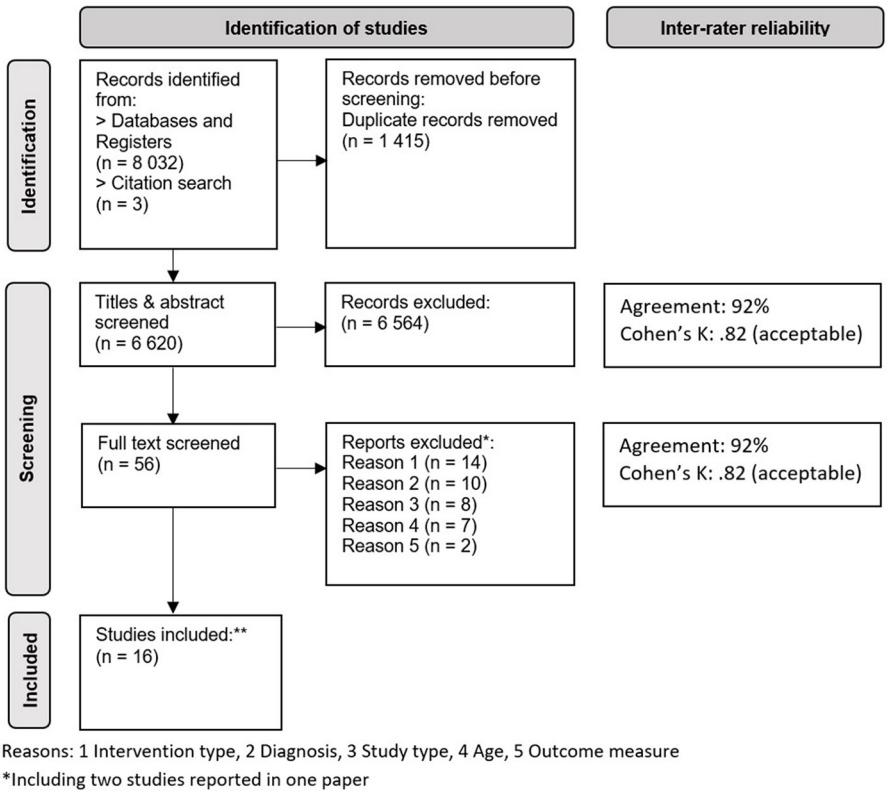


FIGURE 1
Flow diagram of the article screening process (based on PRISMA guidelines, Page et al., 2021).

TABLE 1 Summary of reviewed studies – phonological and/or semantic vocabulary intervention approaches.

Authors, year, country	Design	Participants	Intervention	Dosage	Delivery	Outcome measure	Results	Follow-up	JBI quality rating*
Wing (1990), US	Non-randomized matched group comparison: 5 children per group	8 males & 2 females with existing diagnosis of LI (age 5; 11–7; 1)	Phonological versus semantic approach	30 × 25 min sessions over 2.5 months per group Total: 750 min/12.5 h Number of target words unclear	**Provider**: School SLT **Mode**: Face-to-face group sessions in English **Location**: Room in specialist school	Within group, pre-post comparison using the standardized Test of Word Finding (German, 1986), a picture-naming test of expressive vocabulary	• Significant gain with phonological therapy ($p < 0.05$, $d = 0.7$, moderate effect) • Gain with semantic therapy was not significant	Not measured	Low quality
Wright (1993), UK	Matched no-treatment control group: 2 males & 2 females with SLI (aged 7; 9–8; 5)	2 males & 2 females with existing diagnosis of SLI (age 7; 6–8; 8)	Combined phonological & semantic approach	18 × 20 min sessions over 4 weeks Total: 360 min/6 h 90 target words per child, each presented once	**Provider**: School SLT **Mode**: Face-to-face group sessions in English **Location**: Room in specialist school	Within group, pre-post comparison using a researcher-created picture-naming test of target words to assess expressive vocabulary	• Significant gain for target words ($p < 0.01$) & untargeted control words ($p < 0.05$) • No significant change for control group on any measure Effect sizes not reported	Loss in gains at 1-month follow up	Medium quality
Parsons et al. (2005), UK	Pre-post comparison	2 males with existing diagnosis of SLI (age 8; 10–9; 5)	Combined phonological & semantic approach	18 × 25–35 min sessions. 2–3 sessions a week over 8 weeks Total: 450-630 min/7.5–10.5 h 18 target words, single presentation in therapy with home-school reinforcement	**Provider**: Healthcare SLT with carer follow-up at home & teaching staff follow-up in class **Mode**: Face-to-face individual sessions in English **Location**: Room in mainstream school	1. Individual pre-post comparison using a researcher-created test which matched targeted words with pictures/synonyms to assess receptive vocabulary 2. Individual pre-post comparison using standardized vocabulary measures: British Picture Vocabulary Scale (Dunn et al., 1982) to assess receptive vocabulary & Test of Word Finding (German, 1989) to assess expressive vocabulary	1. Target word tests • Child A. Significantly greater gain for target words than control words ($p < 0.01$) • Child B. Significantly greater gain for target words than control words ($p < 0.001$) Effect sizes not reported 2. Standardized tests. No change for Child A or B	Not measured	Medium quality
Zens et al. (2009), New Zealand	Randomized alternating treatment crossover design, no washout time in-between	19 children with existing diagnosis of SLI (age 6; 2–8; 3, gender unknown)	Phonological/ semantic/ combined approach + Ongoing specialist support for 10 children (specifics unclear)	12 h of one intervention over 6 weeks (2 × 1 h weekly) Followed by 12 h alternate intervention over 6 weeks (2 × 1 h weekly) Total: 720 min/12 h 27 target words, multiple presentations	**Provider**: University-affiliated SLT **Mode**: Face-to-face group sessions in English **Location**: Room in mainstream school	1. Within group, pre-post-test comparison using the standardized Test of Language Development 3rd Ed (Newcomer and Hammil, 1997) & a researcher-created categories test. Raw scores were combined to provide an expressive vocabulary score 2. Within group, pre-post-test comparison using the non-standardized Phonological Awareness Probes (Stahl and Murray, 1994): phoneme blending, isolation, segmentation & deletion	1. Expressive vocabulary tests • Significant gain with phonological therapy ($p = 0.001$, f = 0.65, large effect) • Significant gain with phonological + semantic therapy ($p = 0.001$, $f = 0.65$, large effect) • Significant gains with semantic therapy ($p = 0.004$, $f = 0.62$, large effect) • Significant gain with semantic + phonological therapy ($p = 0.004$, $f = 0.62$, large effect) 2. Phonological awareness tests • Significant gain with phonological therapy ($p < 0.001$, $f = 1.06$, large effect) • Significant gain with phonological + semantic therapy ($p < 0.001$, $f = 1.06$, large effect)	Not measured	Medium quality

(Continued)

10.3389/fpsyg.2025.1517311

TABLE 1 (Continued)

Authors, year, country	Design	Participants	Intervention	Dosage	Delivery	Outcome measure	Results	Follow-up	JBI quality rating*
							• Significant gain with semantic + phonological therapy ($p < 0.001, f = 1$, large effect) • Gain with semantic therapy was not significant		
Motsch and Marks (2015), Germany	Randomized control trial. Control: 55 males & 24 females with SLI (mean age 9;6, SD 0.27)	53 males & 25 females with existing diagnosis of SLI (mean age 9;6, SD 0.16) German-speaking. 38 children received group therapy (2 children per group). 40 received individual therapy	Combined phonological & semantic approach + Variety of ongoing individual & group SLT/teacher language support (specifics unclear, however statistical significance of intervention gains maintained when this additional-support cohort removed)	20 sessions once per week over 5 months. 45 min group sessions & 30 min individual sessions Total: Individual therapy – 600 min/10 h Group therapy – 900 min/15 h Number of target words unclear	**Provider**: SLTs with home-school follow-up **Mode**: Face-to-face individual versus group sessions in German **Location**: Room in special schools	Within and between group pre-test to 4-month follow-up comparison using a range of standardized language tests in German (post-test scores not reported): 1. WWT 6–10 (Glück, 2011). A picture- naming test of expressive vocabulary 2. P-ITPA - Vocabulary subtest (Esser et al., 2010). A sentence-completion test to assess expressive vocabulary 3. P-ITPA – Analogies subtest (Esser et al., 2010). An analogy-generation test to assess expressive vocabulary 4. SET 5–10 (Petermann, 2010). A sentence comprehension to assess receptive syntax	Within group 1. Picture naming test of expressive vocabulary • Significant gain with group therapy ($p < 0.001, d = 0.73$, large effect) • Significant gain with individual therapy ($p < 0.001, d = 0.54$, medium effect) 2. Sentence completion test of expressive vocabulary • Significant gain with group therapy ($p = 0.004, d = 0.38$, medium effect) • Gains did not reach significance with individual therapy 3. Analogies test of expressive vocabulary. No gains reached significance 4. Sentence comprehension test • Significant gain with group therapy ($p = 0.02, d = 0.41$, medium effect) • Significant gain with individual therapy ($p < 0.001, d = 0.57$, medium effect) Between group 1. Picture naming test. Gain for group therapy significantly greater than control group gain ($p = 0.039$) 2. Sentence completion test. No significant between-group difference 3. Analogies test. Gain for individual therapy significantly greater than control group gain ($p = 0.01$) 4. Sentence comprehension test. Gain for individual therapy significantly greater than control group gain ($p = 0.039$)	Change from post-test to follow-up not reported, only pre-test to 4-month follow-up	High quality
Best et al. (2018), UK	Randomized control trial Control: 5 males & 4 females with DLD (aged 6; 3–8; 7)	6 males & 5 females diagnosed with DLD as part of the study (age 6; 0–7; 8)	Word webs for combined phonological & semantic approach No other intervention accessed	Weekly 30 min sessions for 6 weeks Total: 180 min/3 h 25 target words, multiple presentation	**Provider**: University SLT **Mode**: Face-to-face individual sessions in English **Location**: Mostly room in mainstream school	Within and between group pre-post comparison using a researcher-created picture-naming test of target words to assess expressive vocabulary	Between group • Significantly greater gain for therapy group than control group on target-words ($p < 0.0001, d = 2.30$, large effect), no significant difference on control words	Not measured	High quality

(Continued)

10.3389/fpsyg.2025.1517311

TABLE 1 (Continued)

Authors, year, country	Design	Participants	Intervention	Dosage	Delivery	Outcome measure	Results	Follow-up	JBI quality rating*
Best et al. (2021), UK	Randomized alternating treatment crossover design with 6-week washout	12 males & 8 females diagnosed with DLD as part of the study (age 6; 4–8; 8)	Word webs for phonological versus semantic approach No other intervention accessed	Weekly 30 min sessions for 6 weeks per approach. Total: 180 min/3h 50 target words, 25 per approach, multiple presentation	**Provider**: University SLT **Mode**: Face-to-face individual sessions in English **Location**: Mostly room in mainstream school	1. Within group pre-post comparison using a researcher-created picture-naming test of target words to assess expressive vocabulary 2. Outcomes according to language profile	1. Within group • Significantly greater target-word gain for semantic therapy than phonological therapy ($p = 0.014$, $d = 0.489$, medium effect) • No significant order effects or change in control words 2. Outcomes according to language profile • Children with semantic & phonological needs ($n = 11$): 3 children showed significant gain from semantic intervention only, 2 children from phonological intervention only, 5 children from both interventions, 1 child showed no significant gain from either intervention • Semantic needs ($n = 6$): 4 children showed significant gain from semantic intervention only, 2 children showed no significant gain from either intervention • Phonological needs ($n = 3$): 2 children showed significant gain from phonological intervention only, 1 child showed no significant gain from either intervention	Loss in gains at 6-week follow up	High quality
Ardanouy et al. (2023), Switzerland	Pre-post comparison	8 French-speaking children with existing diagnosis of DLD (age 6–10 years, gender unknown)	Combined phonological & semantic approach with context cues No other intervention accessed	5 months of 45 min session per week. 4 sessions per theme covering 4 themes (sports, animals, vegetables, & school materials) Total: 840 min/14 h 60 target words, 15 per category, multiple presentation	**Provider**: University SLT supported by Educational Psychologists **Mode**: Face-to-face group sessions in French **Location**: Specialist clinic	Within group and individual pre-post comparison using a researcher-created picture-naming test of target words to assess expressive vocabulary. Target words were grouped by category	Within group • Veg: Significantly greater gain for target words than control words ($p = 0.01$, $r = 0.89$, large effect) • Animals: Significantly greater gain for target words than control words ($p = 0.01$, $r = 0.89$, large effect) • Sports: Significantly greater gain for target words than control words ($p = 0.01$, $r = 0.89$, large effect) • School: Significantly greater gain for target words than control words ($p = 0.02$, $r = 0.84$, large effect) Individual level • Veg: Significant gain for 6 out of 8 children ($p < 0.05$) • Animals: Significant gain for all children ($p < 0.05$) • Sports: Significant gain for 7 out of 8 children ($p < 0.05$) • School theme: Significant gain for 4 out of 8 children ($p < 0.05$) • Control words: no significant change	Veg: no change at 1.5-month follow-up Animals: no change at 3-month follow-up Sports: loss in gains at 4.5-month follow-up	Medium quality

*Based on JBI quality appraisal rating (see Appendix A).

LI, Language Impairment; SLI, Specific Language Impairment; DLD, Developmental Language Disorder.

10.3389/fpsyg.2025.1517311

TABLE 2 Summary of reviewed studies – story-based vocabulary intervention approaches.

Authors, year, country	Design	Participants	Intervention	Dosage	Delivery	Outcome measure	Results	Follow-up	JBI quality rating*
Nash and Donaldson (2005), UK	Non-randomized alternating treatment crossover design with 1-week washout	13 males & 3 females with existing diagnosis of SLI (age 5; 5–9; 0)	Exposure to spoken stories + Corresponding picture books Versus Explicit semantic approach	Two 20–30 min sessions over 2 consecutive days for each learning context Total: 40–60 min 8 target words per child, 4 per approach, multiple presentations	**Provider:** University SLT **Mode:** Explicit teaching: Face-to-face individual sessions with SLT Incidental: Playing of pre-recorded story by SLT **Location:** Mostly room in mainstream school	Within group comparison of gain between first and second therapy session, and between group comparison of final scores, using a range of researcher-created target-word tests to assess expressive & receptive vocabulary. 1. Picture naming test of target words (expressive vocabulary) 2. Target word definition test (expressive vocabulary) 3. Spoken word to picture matching – match target word to 1 of 4 pictures (receptive vocabulary) 4. Spoken word recognition test – correct pronunciation of target word from choice of 4 (receptive vocabulary) 5. Meaning recognition test – Y/N allocation to given category & attribute (receptive vocabulary)	Within group 1. Picture naming test • Significant gain with story exposure ($p < 0.01$) • Significant gain with semantic therapy ($p < 0.01$) 2. Word definition test • Significant gain with story exposure ($p < 0.05$) • Significant gain with semantic therapy ($p < 0.01$) 3. Spoken word to picture matching • No significant change with story exposure • Significant gain with semantic therapy ($p < 0.01$) 4. Spoken word recognition. • Significant gain with story exposure ($p < 0.001$) • Significant gain with semantic therapy ($p < 0.01$) 5. Meaning recognition • Significant gain with story exposure ($p < 0.05$) • No significant change with semantic therapy Between group Overall gain from semantic therapy significantly greater than story exposure for the Word Definition test ($p < 0.05$) and the Meaning Recognition test ($p < 0.05$). All other between group comparisons were non-significant Effect sizes not reported. No order effect analysis	Not measured	Medium quality
Smeets et al. (2014), Study 1, Netherlands	Randomized alternating treatment crossover design with no washout	24 males & 5 females with existing diagnosis of SLI (age 5; 0–6; 8)	Exposure to narrated e-stories with pictures (no background sound) versus videos (with background sounds)	Each approach presented across 8 sessions over 4 weeks in random order. Session lengths unknown 28 target words, 14 per approach, multiple presentations	**Provider:** Academics in psychology **Mode:** Individual sessions in Dutch using headphones & television screen **Location:** Room in specialist school	Between group comparison of pre-post gains using a researcher-created sentence completion test of target words to assess expressive vocabulary	• Target-word gain significantly greater for picture e-books than with video e-books ($p < 0.01$, $d = 0.48$, medium effect) • Target-word gain significantly greater for e-books (picture & video condition combined) than control words ($p < 0.001$, $d = 1.54$, large effect)	Not measured	Medium quality
Smeets et al. (2014), Study 2, Netherlands	Randomized alternating treatment crossover design with no washout	13 males & 10 females with existing diagnosis of SLI (age 5; 0–7; 6)	Exposure to narrated e-storybooks with picture/video with/without background sounds	Each approach presented across 16 sessions over 8 weeks in random order. Session lengths unknown 72 target words, 18 per approach, multiple presentations	**Provider:** Academics in psychology **Mode:** Individual sessions in Dutch using headphones & television screen **Location:** Room in specialist school	Between group comparison of pre-post gains using a researcher-created sentence completion test of target words to assess expressive vocabulary	• No significant differences in target-word scores found between the intervention groups • A significant correlation was found between increased language severity and negative influence of background sound ($p < 0.05$, $d = 0.43$, medium effect)	Not measured	Medium quality

*Based on JBI quality appraisal rating (see Appendix A).
SLI, Specific Language Impairment.

10.3389/fpsyg.2025.1517311

TABLE 3 Summary of reviewed studies – story-based with semantic and/or phonological vocabulary intervention approaches.

Authors, year, country	Design	Participants	Intervention	Dosage	Delivery	Outcome measures	Results	Follow-up	JBI quality rating*
Marks and Stokes (2010), UK	Pre-post comparison	1 male with LI (aged 8; 1)	Story read to child + semantic approach No other intervention accessed	8 × 50–60 min sessions, over 3 weeks. Total: 400–480 min/6 h 40 min–8 h 30 target words across 4 stories, multiple presentations	**Provider**: Healthcare SLT **Mode**: Face-to-face individual sessions in English **Location**: Room in mainstream school	Within group, pre-post comparison using two researcher-created tests of target words to assess vocabulary: 1. Picture naming test of target words (expressive vocabulary) 2. Spoken word to picture matching (receptive vocabulary)	1. Picture naming test • Significant gain with target words ($p < 0.001$) • No significant changes with control words 2. Spoken word to picture matching • Significant gain with target words ($p = 0.016$) • No significant changes with control words Effect sizes not reported	Loss in gains at 8-month follow up	Medium quality
Steele et al. (2013), US	Randomized alternating treatment crossover design with no washout in-between	10 males & 2 females with existing diagnosis of LI (Mean age 10; 3, SD: 9.32 months)	Child-read story + Phonological and/or semantic approach	One session per condition, length and frequency unknown 15 target words per child, 5 per condition, multiple presentations	**Provider**: University SLT & SLT students **Mode**: Face-to-face individual sessions in English **Location**: Room in mainstream school	Within group, pre-post comparison using a researcher-created target-word definition test to assess expressive vocabulary	• Significantly greater target-word gain for story + phonological + semantic therapy (combined) compared to the control condition ($p = 0.028$) • Significantly greater target-word gain for story + semantic therapy compared to the control condition ($p = 0.002$) • No significant difference between story + phonological therapy compared to the control condition Effect sizes not reported. No order effect analysis	Not measured	Medium quality
Lowman and Dressler (2016), US	Randomized alternating treatment crossover design with no washout time in-between	18 children with existing diagnosis of SLI (age 10; 0–11; 11, gender unknown)	Child-read storybooks + Phonological, semantic & syntactic word cues via an iPod Versus Story reading Ongoing language support continued for both conditions (specifics unclear)	Eight 15-min video viewing sessions, over 4 weeks (2 sessions a week) Total: 120 min/2 h Plus, reading time (not measured) 24 target words, 12 per condition, multiple presentations	**Provider:** Self-directed viewing of iPods/book reading with University-affiliated SLT/SLT students supporting as required **Mode**: Technology vs book **Location**: Room in mainstream school	Between group comparison of pre-post gains using a range of researcher-created target-word tests of expressive & receptive vocabulary 1. Word definition test (expressive vocabulary) 2. Semantic recognition test (receptive vocabulary) 3. Definition selection test (receptive vocabulary) 4. Sentence generation test (expressive syntax) 5. Syntactic recognition test (receptive syntax)	1. Word definition test Significantly greater target-word gain for combined therapy compared to story-reading alone ($p < 0.05$, $f^2 = 0.13$, small effect) 2. Semantic recognition test Significantly greater target-word gain for combined therapy compared to story-reading alone ($p < 0.01$, $f^2 = 0.11$, small effect) 3. Definition selection test Significantly greater target-word gain for combined therapy compared to story-reading alone ($p < 0.01$, $f^2 = 0.46$, large effect) 4. Sentence generation test Significantly greater target-word gain for combined therapy compared to story-reading alone ($p < 0.05$, $f^2 = 0.17$, medium effect) 5. Syntactic recognition test Significantly greater target-word gain for combined therapy compared to story-reading alone ($p < 0.05$, $f^2 = 0.20$, medium effect) No significant order effects	Not measured	Medium quality
Storkel et al. (2019), US	Randomized alternating treatment crossover design with 2–3 weeks in between to measure learning retention	21 males & 13 females diagnosed with DLD as part of the study (age: 5; 0–6 ;2)	Story read to child + Explicit semantic approach	Dosage to achieve 36 target-word exposures varied by no. of therapy sessions per word (4, 6, 9) & word exposures per session (9, 6, 4). Total: 160–299 min 60 target words, 10 per approach, multiple presentations	**Provider**: University SLT & SLT students **Mode**: Face-to-face individual sessions in English **Location**: Mostly room in mainstream school	Within and between group, pre-post comparison using a researcher-created target-word definition test to assess expressive vocabulary	Within group Significant target-word gain from all dosage-delivery variations: • Variation 1: $p = 0.002$ • Variation 2: $p < 0.0001$ Between group No significant difference in scores between dosage delivery variations Effect sizes not reported	Approx. 60% loss in gains at 1-week follow up rising to 70% at 3-week follow up	High quality

*Based on JBI quality appraisal rating (see Appendix A).

LI, Language Impairment; SLI, Specific Language Impairment; DLD, Developmental Language Disorder.

10.3389/fpsyg.2025.1517311

TABLE 4 Summary of studies included in review – orthography-based vocabulary intervention approach.

Study	Participants	Intervention	Design	Dosage	Delivery	Outcome measures	Results	Follow-up	JBI quality rating*
Best (2005), UK	2 males and 3 females, with pre-identified mixed profiles of language/learning needs (age 6;10-10;7)	Child introduced to digital aid to independently sound out initial letter of target words + Regular specialized language/learning input (specifics unclear)	Pre-post comparison	6 × 60 min sessions once a week for 6 weeks Total duration: 360 min/6 h 27–30 target words, multiple presentations	**Provider**: University SLT **Mode**: Combined digital & face-to-face individual sessions. Regular updates to child's SLT & teaching staff **Location**: Setting unknown	1. Individual pre-post comparison using a researcher-created picture-naming test of target words to assess expressive vocabulary 2. Individual pre-post comparison using the standardized Word Finding Vocabulary Test (Renfrew, 1995) to test expressive vocabulary of single words and Test of Word Finding in Discourse (German, 1991) to assess expressive vocabulary in sentences 3. Individual pre-post comparison using the standardized British Picture Vocabulary Scales (BPVS, Dunn et al., 1982) to assess receptive vocabulary 4. Pre-post comparison of group average percentage score for literacy (reading) & numeracy using child, parent, teacher & therapist completed visual scale "views questionnaire"	1. Picture naming • Significant gain for target words ($p <$ 0.025 for all children) • No significant changes for control words 2. Standardized test of expressive vocabulary. Significant gain for 2 out of 5 children 3. Standardized test of receptive vocabulary. No significant change 4. Literacy (reading) & numeracy. No significant change in numeracy score from pre-test (51%) to post-test (51%) or reading score from pre-test (56%) to post-test (63%) Effect sizes not reported	Loss in gains at 6-month follow up	Low quality

*Based on JBI quality appraisal rating (see Appendix A).

10.3389/fpsyg.2025.1517311

written and/or verbal stories to target word learning (Table 2), use of stories and phonological/semantic approaches combined (Table 3), and interventions targeting orthography (spellings) to support word learning (Table 4).

3.3 Quality appraisal

Full details of the appraisal scores are available in Appendix A, and the overall quality rating for each study is included in the summary of studies tables (Tables 1–4, Column 3). Two studies were appraised as low quality and were omitted from quantitative synthesis (Best, 2005; Wing, 1990). Factors limiting the quality of these two studies included lack of control groups, small sample sizes, and ambiguity regarding access to ongoing specialist language support in addition to the experimental intervention.

3.4 Study properties

3.4.1 Study design and sample sizes

The review identified one pre-post comparison case-study (Marks and Stokes, 2010, $n = 1$), two case-series reporting on pre-post comparisons of multiple individual case-studies (Best, 2005, $n = 5$; Parsons et al., 2005, $n = 2$), and one pre-post comparison group study (Ardanouy et al., 2023, $n = 8$). All the case-studies used control words as comparators in place of control groups.

Five controlled group studies were identified (Best et al., 2018, $n = 11$; Wright, 1993, $n = 4$; Motsch and Marks, 2015, $n = 78$; Nash and Donaldson, 2005, $n = 16$; Wing, 1990, $n = 10$). Two of the controlled studies were randomly assigned (Best et al., 2018; Motsch and Marks, 2015), while the remainder were matched by language profiles with the justification that population heterogeneity meant there was difficulty recruiting sufficient children with comparable characteristics.

Seven studies adopted a randomized crossover design to mitigate population heterogeneity. Interventions were delivered at different time periods, and three of the studies included a no-intervention period between interventions (Break between interventions: Best et al., 2021, $n = 20$; Steele et al., 2013, $n = 12$; Storkel et al., 2019, $n = 34$; No break: Lowman and Dressler, 2016, $n = 12$; Smeets et al., 2014, Study 1 $n = 28$ & Study 2 $n = 21$; Zens, 2009, $n = 19$). One of the crossover studies analyzed both individual and group effects (Best et al., 2021).

A pattern of chronology emerged in that over half of the studies (11) built on work covered in earlier studies in the review, enabling preliminary findings to be replicated, extended or generalized to a larger sample.

3.4.2 Participant information

Half of the studies (8) identified participants as presenting with an existing diagnosis of Specific Language Impairment (SLI). The criteria for SLI as defined by Stark and Tallal (1981) is based on a discrepancy profile, with language ability significantly below the child's cognitive or chronological age; non-verbal ability within the normal range on standard scores; and typical hearing, sensory and socioemotional development. Whilst participants appeared to align to these diagnostic criteria for SLI within the studies reviewed, wide variation in assessments used and thresholds for diagnosis were noted. Although this weakens the strength of the synthesis, this diagnostic variability for SLI is typical both academically and clinically (Aram et al., 1993).

Three studies described their samples as children with an existing diagnosis of Language Impairment (Marks and Stokes, 2010; Steele et al., 2013; Wing, 1990). The criteria matched that of SLI as described in the above paragraph.

Four studies referred to their sample as presenting with DLD (Ardanouy et al., 2023; Best et al., 2018, 2021; Storkel et al., 2019). DLD deploys more descriptive diagnostic criteria than SLI, focusing on the sustained social and educational impact of pervasive language difficulties and allowing for certain co-morbidities (Bishop et al., 2017). Ardanouy et al. (2023) recruited children with an existing DLD diagnosis whilst the others diagnosed as part of their studies. Overall, the characteristics of participants aligned with that of SLI, apart from one child in the study by Ardanouy et al. (2023) who also presented with attention deficit hyperactivity disorder (ADHD).

One study defined their sample using descriptive profiles of existing language needs (Best, 2005). Characteristics aligned with that of SLI, apart from one child who also presented with ADHD and moderate learning difficulties.

All children in the reviewed studies were identified by investigators as presenting with vocabulary difficulties. The assessments used, the threshold for identification, and the nature of the difficulty (e.g., receptive and/or expressive) varied widely between the studies.

Only two studies provided socio-economic information. Steele et al. (2013) reported that most participants were African American (96%) and received subsidized lunches (85%). Parent education was also reported by Steele et al. (2013) with parents typically having a college degree, partial college accreditation, or high school diploma. Storkel et al. (2019) referred to most children in their study as being White, non-Hispanic (79%), and with married parents (53%). They also reported on parental education with most parents holding a college degree, partial college accreditation, or high school diploma.

Where reported, children across the studies were identified as monolingual, with the exception of one child studied by Best (2005) who spoke English and Spanish (replicating parents) and presented with difficulties in both languages. This child received intervention in English only.

3.4.3 Target words

Lowman and Dressler (2016) used an established tiered framework (Beck et al., 2002) to select words that were described as unlikely to appear in everyday conversation but that would benefit from teaching as could be used across multiple contexts. Selection consensus was reached between the investigating SLT, SLT student and school SLT. Storkel et al. (2019) used words from a previously published study (Justice et al., 2005) which had adopted a similar target-word selection method as Lowman and Dressler (2016). Parsons et al. (2005) selected words from the UK national school curriculum.

For six studies, the target words were those considered by the research team to be relevant and meaningful (Ardanouy et al., 2023; Smeets et al., 2014: Study 1 & 2; Motsch and Marks, 2015; Wing, 1990; Zens et al., 2009). One of these six studies (Ardanouy et al., 2023) analyzed intervention effect according to target word categories (sports, animal, vegetable, school). Whilst another, Zens et al. (2009), ensured balance of word properties: meaning (category) and sound (syllable length, phonotactic probability).

For the remaining seven studies, target words were selected from pre-published vocabulary lists standardized for factors such as word properties, age of acquisition, frequency of occurrence, and familiarity. Of these seven, three studies (Best, 2005; Best et al., 2018, 2021) also targeted words of personal relevancy to the children, but the type and number varied for each child and outcomes were not included in their statistical analysis.

3.4.4 Outcome measures

For ease of reading, Table 5 provides a summary of the measures used. Most studies (13) used researcher-created expressive target-word tests such as picture naming and word definitions to measure outcomes. Other vocabulary measures included researcher-created tests to assess receptive knowledge of target words and standardized expressive and/or receptive vocabulary tests.

Generally, studies focused on vocabulary outcomes only, however, some did extend to standardized and non-standardized measures of phonological awareness (Zens et al., 2009), syntax (Lowman and Dressler, 2016; Motsch and Marks, 2015), literacy, and numeracy (Best, 2005). The range of study designs and small data sets meant it was not appropriate to synthesize or make conclusions around intervention effects beyond vocabulary outcomes.

3.4.5 Follow up (maintenance of outcomes effects)

Of the six studies that measured changes in vocabulary scores from post-therapy to follow-up, five reported loss (Best, 2005: at 6 months follow-up; Best et al., 2021: at 6 weeks; Motsch and Marks, 2015, at 4 months; Storkel et al., 2019: at 1 week and 2 weeks; Wright, 1993: at 1 month) and one reported either no change or loss depending on time measured (Ardanouy et al., 2023: vegetable vocabulary–no change at 1.5 months; animals–no change at 3 months; sports–loss at 4.5 months).

No studies reported losses that marked a return to pre-therapy scores.

3.5 Intervention characteristics

3.5.1 Intervention type

Studies in this review were grouped into four intervention approach types: phonological and/or semantic interventions (8 studies), story-based interventions (3 studies), story-based with semantic and/or phonological interventions (4 studies), orthography-based interventions (1 study).

Studies exploring phonological and semantic approaches focused on both teaching the child phonological and semantic word features and cuing the child to self-generate information about word sounds and meaning. Wing (1990) compared phonological and semantic approaches, whereas Wright (1993), Motsch and Marks (2015), and Parsons et al. (2005) investigated combined strategies. Zens et al. (2009) studied the influence of ordering phonological and semantic approaches. Best et al. (2018) utilized word webs, which are visual diagrams to cue, map, and record phonological and semantic word features. Best et al. (2021) also used word webs but considered differential responses to intervention based on presenting profiles. Ardanouy et al. (2023) added contextualization to phonological and semantic word-learning strategies by connecting to personal experiences and considering words in multiple environments with a range of communication partners. All the studies in this category used objects and/or pictures to support word learning, and all bar Wing (1990) and Wright (1993) also used the written form of the target word.

The second intervention type involved exposure to stories to provide contextual cues for vocabulary learning. Nash and Donaldson (2005) compared incidental word learning through exposure to pre-recorded narrated stories and corresponding picture books against explicit exposure to semantic word features. Smeets et al. (2014) conducted two studies: one comparing the child being exposed to narrated electronic story books with either static images or videos with background and music audio, and the other comparing static picture vs. video narrated electronic written story books either with or without background noise and music.

The third intervention approach combined stories to contextualize target words with explicit exposure to phonological and/or semantic word features. In Marks and Stokes (2010) an adult introduced a spoken story with corresponding pictures,

TABLE 5 Summary of outcome measures.

	Standardized	Non-standardized
Expressive vocabulary	Best, 2005; Motsch and Marks, 2015; Parsons et al., 2005; Wing, 1990	Ardanouy et al., 2023; Best, 2005; Best et al., 2018, 2021; Lowman and Dressler, 2016; Marks and Stokes, 2010; Nash and Donaldson, 2005; Smeets et al., 2014 (Study 1 & 2); Steele et al., 2013; Storkel et al., 2019; Wright, 1993; Zens et al., 2009
Receptive vocabulary	Best, 2005; Parsons et al., 2005	Lowman and Dressler, 2016; Marks and Stokes, 2010; Nash and Donaldson, 2005; Parsons et al., 2005
Phonological awareness		Zens et al., 2009
Syntax	Motsch and Marks, 2015	Lowman and Dressler, 2016
Literacy & numeracy		Best, 2005

followed by explicit teaching of semantic features and supporting of story generation. In Steele et al. (2013), children were supported by an adult to read story text containing the target word as well as learning and generating specific phonological and/or semantic word features. Lowman and Dressler (2016) exposed children to pre-recorded phonological, semantic, and syntactic word features via an iPod, in addition to the children independently reading storybooks containing the target words, versus independent storybook reading alone. Storkel et al. (2019) introduced target words through shared adult-child reading of written story picture books in addition to explicitly teaching semantic word features.

The final intervention reviewed used alphabetic letters to support vocabulary skills. Best (2005) introduced a digital aid that children could use to sound out the initial letters of target words presented in picture form, SLT support was provided as required.

Four studies (Best, 2005; Lowman and Dressler, 2016; Motsch and Marks, 2015; Zens et al., 2009) mentioned the continuation of ongoing specialist language support from the children's SLT and/or school in addition to the experimental interventions. Details were brief making the influence of the additional support difficult to ascertain, with the exception of Motsch and Marks (2015) where only a sub-sample received additional support and where intervention effects were maintained even once data for the additional-support cohort was removed. Other studies explicitly stated that no other specialist support was accessed (Best et al., 2018, 2021; Ardanouy et al., 2023), whilst the remaining studies did not comment either way.

3.5.2 Intervention dosage

Interventions ranged from 40 min spread over 2 days (Nash and Donaldson, 2005) to 900 min/15 h spread over 5 months (Motsch and Marks, 2015).

Most studies (9) delivered intervention weekly or biweekly lasting 30–60 min over 4–6 weeks. Three studies provided interventions three or more times a week (Wright, 1993; Parsons et al., 2005; Wing, 1990) lasting 20–60 min over 3–10 weeks. Three studies provided insufficient detail to calculate intervention duration (Smeets et al., 2014: Study 1 & 2; Steele et al., 2013).

One study (Storkel et al., 2019), building on previous work that identified 36 exposures per target word as the optimal for vocabulary interventions (Storkel et al., 2017), compared intervention frequencies for optimal exposure. Dosage delivery ranged between 10 and 23 sessions lasting 13–16 min over 5–12 weeks; no differential effect of dosage on intervention outcomes was found.

The number of target words varied between studies ranging from 4 to 90 words per child per intervention. There was a general correlation between intervention dosage and number of words targeted. However, as number of exposures to target words was inconsistently reported across the studies, findings could not be pooled.

3.5.3 Intervention provider

In most studies (9), intervention was provided by university-affiliated SLTs who were the primary investigators (ranging from professors to doctoral students).

The exceptions were four studies where the SLT was affiliated to the child's school or health provider (Marks and Stokes, 2010; Parsons et al., 2005; Wing, 1990; Wright, 1993), one study where the affiliation of the SLTs delivering the intervention was unclear (Motsch and Marks, 2015), and two studies where intervention was delivered by psychology students who were the primary investigators (Masters and PhD: Smeets et al., 2014).

In three studies the SLT was supported in delivering the intervention by SLT students (Lowman and Dressler, 2016; Steele et al., 2013; Storkel et al., 2019) and in one study by educational psychologists (Ardanouy et al., 2023).

Two studies mentioned that the intervention details were shared with caregivers and teaching staff for follow-up (Motsch and Marks, 2015; Parsons et al., 2005) and one mentioned follow-up by the child's SLT, caregivers and teaching staff (Best, 2005); none of these studies quantified the follow-up support which limits replicability.

3.5.4 Intervention mode

In most studies (9), intervention was delivered as face-to-face individual sessions with the intervention provider. Four further studies (Ardanouy et al., 2023; Wing, 1990; Wright, 1993; Zens et al., 2009) delivered intervention through face-to-face group sessions. One study (Motsch and Marks, 2015) compared face-to-face individual sessions and face-to-face group sessions, and found greater expressive vocabulary gain with group therapy and greater syntax gain (expressive and receptive) with individual therapy.

Three studies used multimedia interventions (Smeets et al., 2014: Study 1 & 2–television; Lowman and Dressler, 2016–iPod; Best, 2005–digital aid sounding out alphabet letters). In Lowman and Dressler (2016), children viewed videos containing text, pictures and animation related to target words in addition to reading storybooks embedded with target words, and reported greater gains than with reading alone. Smeets et al. (2014) investigated the differential effect of multimedia by comparing word learning using narrated e-books with static pictures versus narrated e-books with animation, both conditions were presented either with or without background music/sound. Although no mode effect was found at the group level, a significant correlation emerged between increased language severity and negative influence of background sounds. Finally, Best (2005) adopted a hybrid model with the digital aid introduced during face-to-face individual sessions with a SLT, and found greater gains with the aid than without.

3.5.5 Intervention location

One study was conducted in a specialist clinic (Ardanouy et al., 2023), five studies were conducted in a quiet room in a specialist school (Motsch and Marks, 2015; Smeets et al., 2014: Studies 1 & 2; Wing, 1990; Wright, 1993), with the remaining studies undertaken in mainstream schools, bar one study which did not confirm location (Best, 2005).

3.6 Quantitative synthesis of intervention outcomes

Meta-analysis was precluded by variations in the collection, analysis and reporting of quantitative data compounded by the heterogeneity of study design, clinical populations, diagnostic criteria, and intervention characteristics. Instead, outcomes were summarized according to the Synthesis Without Meta-analysis (SwiM) guidelines (Campbell et al., 2020) with *p*-value and effect size ranges reported.

Across the review, intervention type was the only intervention characteristic consistently identified as a study variable, and vocabulary gain was the only consistently reported intervention effect measure. As a result, quantitative findings have been synthesized and compared according to the effect of intervention type on vocabulary outcomes. Table 6 provides a summary of this quantitative synthesis with further detail.

3.6.1 Phonological and/or semantic interventions (8 studies: 3 high design quality, 5 medium)

Six studies found significant vocabulary gain when phonological and semantic interventions were combined. Of these, three studies showed gain on expressive target-word scores (Ardanouy et al., 2023; Best et al., 2018; Wright, 1993), one showed gain on both target-word and non-target word tests of expressive vocabulary Zens et al., 2009, one showed gain on standardized expressive vocabulary tests (Motsch and Marks, 2015) and one showed gain on receptive target-words but not on standardized tests of receptive or expressive vocabulary (Parsons et al., 2005). With semantic intervention alone, three studies found significant gain on expressive vocabulary (target-word scores: Best et al., 2021; Nash and Donaldson, 2005, target and non-target word scores: Zens et al., 2009). With phonological intervention alone, one study found significant gain on expressive vocabulary (target and non-target word scores: Zens et al., 2009).

Two studies evaluated the effects of ordering phonological and semantic interventions (Best et al., 2021; Zens et al., 2009). Neither found an intervention order effect. However, Best et al. (2021) reported greater target-word gain during the semantic than the phonological intervention phase at the group level. In contrast, Zens (2009), found significantly greater target-word naming gain after day one with the phonological intervention but not with the semantic intervention (though the significance was not maintained over the course of the phonological intervention).

Best et al. (2021) also explored the extent to which individual children's responses to semantic and/or phonological intervention were influenced by their profile of need at baseline. Ninety-one percent of children (10/11) with semantic and phonological needs showed expressive target-word gain from either phonological, semantic, or combined intervention; two-thirds of the children (4/6) with relatively more semantic difficulties benefited from semantic but not phonological intervention and two-thirds with relatively more phonological difficulties benefited from

TABLE 6 Summary of quantitative synthesis.

		Intervention		
Outcome		**Phonological (P) and/or Semantic (S) - Eight studies**	**Story - Three studies**	**Story + P and/or S - Three studies**
	Expressive vocabulary	**Phonological and Semantic combined (Five studies)** Significance range: $p < 0.0001$ to 0.01 *Source:* Ardanouy et al., 2023 (n = 8); Best et al., 2018 (n = 11); Wright, 1993 (n = 4); *Motsch and Marks, 2015 (n = 78); Zens et al., 2009 (n = 19)* Effect size: Large *Source:* Ardanouy et al., 2023; Motsch and Marks, 2015; Zens et al., 2009	**Story (Two studies)** Significance: $p < 0.001$–0.05 *Source: Nash and Donaldson, 2005 (n = 16); Smeets et al. (2014) Study 1 (n = 28)* Effect size: Large. *Source: Smeets et al., 2014, Study 1 (n = 28)*	**Story + P&S (Two studies)** Significance range: $p = 0.028$–0.05 *Source: Lowman and Dressler, 2016; (n = 18); Steele et al., 2013 (n = 12)* Effect size: Small *Source: Lowman and Dressler, 2016*
		Semantic only (Three studies) Significance range: $p = 0.004$ to <0.01 *Source: Best et al., 2021 (n = 20); Nash and Donaldson, 2005 (n = 16); Zens et al., 2009 (n = 19)* Effect size range: Medium to large *Source: Best et al., 2021; Zens et al., 2009*		**Story + S (Two studies)** Significance range: $p =$ <0.0001 to 0.002 *Source: Steele et al., 2013 (n = 12); Storkel et al., 2019 (n = 34)* Effect size not reported.
		Phonological only (One study) Significance: $p = 0.001$ Effect size range: Large. *Source: Zens et al., 2009 (n = 19)*		
	Receptive vocabulary	**Phonological and Semantic combined (One study)** Significance range: $p < 0.001$ to <0.01 *Source: Parsons et al., 2005 (n = 2)* Effect size not reported	**Story (One study)** Significance range: $p < 0.001$ to <0.05 (Spoken word recognition & Meaning recognition) *Source: Nash and Donaldson, 2005 (n = 16)* Effect size not reported	**Story + P&S (One studies)** Significance: $p < 0.01$ *Source: Lowman and Dressler, 2016 (n = 12)* Effect size range: Small to large *Source: Lowman and Dressler, 2016 (n = 12)*
		Semantic only (One study) Significance: $p < 0.01$ *Source: Nash and Donaldson, 2005 (n = 16)* Effect size not reported		

phonological but not semantic intervention (2/3). Altogether, Best et al. (2021) identified that 90% (18/20) of children in their study showed expressive target-word gain after accessing the semantic and/or phonological components of word web intervention.

3.6.2 Story-based interventions (3 studies: all with medium design quality)

One study (Nash and Donaldson, 2005) investigated incidental word learning through exposure to narrated picture stories and reported significant target-word gain on expressive and receptive vocabulary measures. When compared to an explicit semantic intervention, greater gain was reported from the semantic intervention on both expressive and receptive target-word measures.

Two studies explored the effect of exposure to narrated electronic story books on target words. One reported greater gain for targeted words introduced as a picture or animated e-story compared to control words (Smeets et al., 2014, Study 1), and the other found a significant correlation between increased language severity and the negative influence of background sound in e-books (Smeets et al., 2014, Study 2).

3.6.3 Stories + phonological and/or semantic interventions (3 studies: 1 high quality, 2 medium)

Two studies reported significant gains on expressive and receptive target-word measures when stories were combined with phonological and semantic interventions: in one study (Steele et al., 2013) therapy was provided in person, while in the other (Lowman and Dressler, 2016) multimedia was used. Two studies found significant gain on expressive measures when stories were combined with semantic interventions (Steele et al., 2013; Storkel et al., 2019). It is worth noting that Storkel et al. (2019) reported a significant correlation between higher language scores and greater word-learning gain.

4 Discussion

This review considered available studies targeting vocabulary skills in children with DLD in order to identify and describe key intervention features. What follows is a discussion of the characteristics of the vocabulary interventions that have been reviewed and the effects on word learning, consideration is also given to the design of the studies that generated the reported findings.

4.1 The benefits of an integrated therapeutic model

The use of combined phonological and semantic word-learning strategies was the most frequently evaluated intervention, with most of the reviewed studies having a high-quality design and demonstrating statistically significant expressive vocabulary gain with a large effect size. Studies investigating semantic or phonological approaches alone also demonstrated gain but less consistently. These findings align with the connectionist model of language processing and the proposal that vocabulary difficulties in children with DLD may stem from both the processing of sound and meaning, and the interactions between them (Chiat, 2001).

There were some indications of the benefit of explicit sound-letter cueing to support phonological processing during word learning, however, as findings were restricted to one study which was appraised as having a low design quality, more research is needed for firm conclusions to be drawn.

There were stronger indications that integrating contextual story narratives with explicit phonological and semantic cues strengthened vocabulary gains, as demonstrated in four studies of medium to high quality. This is in keeping with the cross-situational model of language acquisition which proposes that children's vocabulary is enhanced when information is presented in multiple contexts because patterns of co-occurrence strengthen word learning (Roembke et al., 2023). However, exposure to stories alone was not as effective as explicit cuing in improving word-learning. This indicates that the interactive, conversational aspect of narratives may be key in supporting children with DLD, as highlighted in the work by Law et al. (2022) who explored the influence of social context for children with DLD.

Taken together, these results align with current thinking around the benefits of integrating intervention approaches to improve outcomes for children with DLD. For example, Baron and Arbel (2022a) discuss how differential learning systems are engaged when explicit and implicit intervention approaches are combined. According to Baron and Arbel (2022a), explicit learning, where the child is systematically taught facts, can address underlying processing difficulties in children with DLD. However, there is a risk of dependency on rote-learning and memorisation which incidental learning can mitigate through inferencing and generalization opportunities. Applying this theory to word learning, by combining explicit phonological-semantic vocabulary strategies with social narratives, children can use explicit and implicit cues to acquire, apply, and refine sound-meaning associations.

Effective integration of explicit and incidental word-learning strategies is not limited to exposing DLD children to targeted words within narratives during therapy. Research into word recall has demonstrated that word retention in children with DLD is influenced not only by the robustness of the initial learning phase but also meaningful retrieval and practice opportunities (Kueser et al., 2021; Leonard et al., 2020). It is important to consider that most studies in this review that reported on maintenance of word-learning gains observed depreciation at follow-up relative to the gains made from pre- to post-intervention. This was regardless of the intervention approach used, indicating that there are other factors at play. One such factor could be the selection of target words as most studies used pre-set word lists, however, words from the child's personal life, interests, or school curriculum would have provided more incidental reinforcement opportunities. Studies that utilized curriculum and personal interest words did not include follow-up measures; therefore, implications for retention require further exploration.

The need to target functional words for meaningful and sustained word learning is further reinforced by indications in this review that on the whole therapy gains did not generalize beyond the directly targeted words. This was demonstrated by the minimal transfer of gains to control words or standardized vocabulary assessments.

Another strategy for integrating explicit and implicit word-learning opportunities that was evident in the reviewed interventions was the use of self-cuing. Most studies embedded self-cueing strategies where children were encouraged to independently apply word-learning techniques once they had been modeled and practiced with an adult. However, there was only one example where this strategy continued to be studied post-therapy and even then, it was informally rather than through standardized observations and evaluation (Best et al., 2021). Given that autonomous self-cueing has been identified as an area of implicit learning that is compromised in children with DLD due to poor executive functioning and language planning (Baron and Arbel, 2022b; Senter, 2022), further research is warranted.

Combining individual and group therapy sessions can also provide children with DLD the opportunity to benefit from explicit and implicit word-learning opportunities. Most studies in this review adopted either individual or group therapy, however, one study directly compared the two approaches and found greater vocabulary gain at word level with group therapy and at sentence level with individual therapy (Motsch and Marks, 2015). This differential response aligns with findings for language therapy more generally, and the proposed activation of separate learning mechanisms. Group therapy is considered to have the advantage of targeting words in peer discourse with multiple communication partners, whereas individual therapy enables more tailored support (Watt and White, 2018). Further exploration of the effects of individual versus group vocabulary therapies is required.

4.2 Individualized approaches to provide intervention that aligns with needs

A key finding that emerged from the reviewed vocabulary intervention studies was the importance of considering children's presenting profiles. This was most evident in the study by Best et al. (2021) which completed both group- and individual-level analyses. As a group, children with DLD showed significantly greater word-learning gain from explicit semantic interventions compared with explicit phonological therapy. However, case-series analysis indicated differential patterns of response depending on presentation. Children presenting with both phonological and semantic difficulties responded to both or either intervention. However, children presenting with greater semantic than phonological difficulties tended to benefit from semantic but not phonological therapy. The converse was also true with phonologically-impaired children tending to respond to phonological but not semantic therapy. Crucially, some children did not respond to either intervention, indicating that word-learning strategies for children with DLD cannot be chosen solely on presenting profiles.

Threshold of intervention gain is another feature that may have differential implications depending on the language profile of a child with DLD. One study (Storkel et al., 2019), identified that children with more severe language difficulties were less responsive to combined explicit and implicit word-learning strategies. This built on previous work by Storkel et al. (2017) exploring dosage, which identified a plateau in word-learning gains in children with DLD in response to increased intervention exposure. The ceiling appeared related to language proficiency, with children who had more severe needs reaching ceiling of benefit sooner. Taken together, this indicates the need for an individualized approach for vocabulary interventions and careful monitoring of progress when considering the optimal dosage and level of response.

Whilst the use of multimedia during word learning was found to be of benefit for children with DLD (Best, 2005; Lowman and Dressler, 2016; Smeets et al., 2014), this was another area where the nature of responses differed according to the profile of language needs. More specifically, Smeets et al. (2014) demonstrated that children with more severe language difficulties benefited from digital images and videos to support word-learning only when extraneous background music and sounds were removed. Given that remote working measures introduced during COVID-19 have led to a subsequent rise in digital practice to support children with DLD (McCartney and Forbes, 2023; Ansari et al., 2022), further research is needed to determine how technology may support individual needs.

5 Limitations

5.1 Limitation of reviewed intervention studies

Intervention quality was generally found lacking in the reviewed interventions. Of the 16 studies reviewed, only four were appraised as being of high quality; methodological weaknesses included small sample sizes, weak or lack of control measures, and omission of effect size calculations. It is promising that the later-dated studies were the ones of high quality (2015 onwards), and worth noting that they generally built on work from the earlier more exploratory studies indicating the value of the earlier work. A key point, however, is that one of the high-quality RCTs compared intervention effect at the group level with the individual level and reported discrepancies (Best et al., 2021). This is inevitable given the heterogeneity of the DLD population; however, as advised by numerous researchers in the field, it can be addressed by adopting multi-method case-series and controlled group trial approaches. This would maintain research rigor whilst providing findings that can be applied to a heterogeneous population as is required by SLTs delivering interventions in practice (Best et al., 2019; Forsythe et al., 2022).

The heterogeneity of the population was further compounded between studies by differences in diagnostic terminology. The use of Language Impairment or Specific Language Impairment as diagnostic labels was evident in most studies published before 2018, whereas all studies from 2018 and after adopted DLD. This mirrors a broader diagnostic shift by clinicians and academics

away from identifying language impairment based on discrepancies between language and intelligence scores (Stark and Tallal, 1981; Tomblin et al., 1997) to a more descriptive approach focusing on the sustained social and educational impact of language difficulties (Bishop et al., 2017). While the rationale is well-documented (Royal College of Speech Language Therapists: RCSLT, 2021; McGregor et al., 2020), it introduces further heterogeneity among already varied study samples, which in turn impacts the comparability and synthesis of findings.

Poor consideration of the impact of the intervention provider and setting was a further limitation of the reviewed studies. Only three studies extended intervention to the home and school environment (Best, 2005; Motsch and Marks, 2015; Parsons et al., 2005), and none quantified this additional support. Wider research involving children with a range of communication needs has demonstrated the benefits of vocabulary learning when intervention strategies are embedded within the home and classroom (Throneburg et al., 2000; Ebbels et al., 2019; Roberts and Kaiser, 2011). Theoretically, this aligns with the cross-situational model of language acquisition (Roembke et al., 2023), which emphasizes the value to children of repeated vocabulary exposure across physical and social environments. Given the shortage of SLTs to support children with DLD (Christopulos and Redmond, 2023; Gibbons, 2021), and the findings from this review that children experience deterioration in word-learning gains once intervention is withdrawn, there is a need for research that robustly explores long-term home-school reinforcement of vocabulary interventions.

Another limitation across most of the reviewed studies was the poor reporting of children's socioeconomic status. This has implications for research and practice given that low SES is considered a risk factor for DLD and may influence access to and response to treatment (Bishop et al., 2017; McGregor, 2020).

5.2 Limitations of systematic review

Variation in study design meant the Synthesis Without Meta-analysis (SWiM, Campbell et al., 2020) guidelines were applied and quantitative synthesis was limited to reporting *p*-value and effect size range. Unlike meta-analysis, this approach is limited in analyzing the extent of variation in *p*-values and effect sizes among studies or revealing causes of heterogeneity. However, it provides greater insight into the measurable impact of an intervention on outcomes and influential factors than a qualitative summary alone (Siedler, 2022).

To assist with the synthesis of the findings, this review was limited to studies reported in English only. It is acknowledged that this excludes a body of work that may be relevant.

6 Recommendations

6.1 Research implications

Based on the systematic review findings, the following recommendations are made for future research on vocabulary interventions for children with DLD:

1. Combine case-study and controlled group methods to generating findings relevant to a heterogeneous population whilst maintaining research rigor.
2. Use controlled variable studies to explore the impact of explicit sound-meaning word-learning strategies reinforced implicitly over time across providers (teachers, caregivers) and settings (school, home).
3. Select target words that are of relevance for children with DLD and that they are likely to be re-exposed to outside of the therapy environment.
4. Recruit more representative samples with diverse socioeconomic status and ethnicity, with detailed reporting of participant characteristics.

6.2 Clinical implications

The following recommendations from this systematic review are relevant for professionals supporting children with DLD:

1. As word-learning gains are typically restricted to directly targeted words, practitioners, educators, and caregivers should work in partnership with children who have DLD to identify the most relevant target words. These should relate to the children's interests, curriculum, and routines, so that the words are useful across multiple environments and with various communication partners.
2. Children with DLD may show greater response to the phonological, semantic and phonological-semantic cues that match the areas they struggle with. However, most children will respond to phonological-semantic cues combined, and this should form the start of therapy which can then be refined depending on how the child responds. Both Best et al. (2018) and Parsons et al. (2005) provide detailed protocols for phonological-semantic interventions (see also Branagan and Parsons, 2021, Word Aware 3).
3. For maximum benefit, explicit phonological-semantic word-learning strategies should be targeted both individually and in groups and then reinforced by providing frequent opportunities for generalization. Generalization strategies should include exposure to words in conversational narratives, encouragement of self-cuing, and on-going opportunities to hear and use the words at home and school.
4. When planning interventions, take into consideration that children with DLD require repeated target-word exposures during combined explicit and implicit word-learning to make significant vocabulary gains. The maximum beneficial exposure will vary between children depending on language severity, therefore, joint SLT-school-home planning for intervention delivery and monitoring is important.
5. The use of multimedia to support vocabulary skills, i.e., audio, images, video, is advocated. However, multimedia word-learning should be planned carefully and monitored closely, as each child is likely to respond differently depending on their language and learning profiles, as well as individual preferences.

7 Conclusions

This review considered the available research on vocabulary interventions to understand the intervention features leading to vocabulary gain in children with DLD. The range and quality of studies are currently restricted but enabled some conclusions to be drawn.

The results suggest word-learning benefits from integrating explicit phonological-semantic prompts with implicit contextual cues. Factors that may influence outcome include the child's language and learning profile, vocabulary relevance, reinforcement opportunities, use of self-cuing, and peer-modeling.

More research is needed across all aspects of vocabulary interventions for children with DLD, with gaps most evident in relation to the influence of socioeconomic status and generalization to the home and school environment.

This paper is important for professionals working with children who have DLD, as it offers practical, evidence-based guidance for applying research to meet children's individual word-learning needs. The findings also have bearing for researchers and academics as they guide the direction of future research.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

RA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft. SC: Methodology, Supervision, Validation, Writing – review & editing. MC: Methodology, Resources, Software, Supervision, Validation, Writing – review & editing. RH: Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The primary author, RA (Clinical Doctoral Research Fellow, ICA-CDRF-2018-04-ST2-029), was funded by the Health Education England (HEE)/National Institute for Health Research (NIHR) for this research project.

Acknowledgments

The authors would like to thank Hilary Lowe, Department of Language and Communication Science, City University London, UK for support with methodology and Rhiannon Sells, NIHR Intern and Pediatric Speech and Language Therapist for inter-rater reliability testing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1517311/full#supplementary-material

References

Alduais, A., Majorano, M., Andrés-Roqueta, C., Hamaguchi, P., Persici, V., and Qasem, F. (2022). Conceptualizing, defining, and assessing pragmatic language impairment in clinical settings: a scoping review. *Infant Child Dev.* 31:e2368. doi: 10.1002/icd.2368

Anglin, J. M., Miller, G. A., and Wakefield, P. C. (1993). Vocabulary development: a morphological analysis. *Monograp. Soc. Res. Child Dev.* 1993:186. doi: 10.2307/1166112

Ansari, R., Loraine, E., and Gréaux, M. (2022). Impact of COVID-19 on digital practice in UK paediatric speech and language therapy and implications for the future: a national survey. *Int. J. Lang. Commun. Disord.* 57, 1112–1129. doi: 10.1111/1460-6984.12750

Aram, D. M., Morris, R., and Hall, N. E. (1993). Clinical and research congruence in identifying children with specific language impairment. *J. Speech Lang. Hear. Res.* 36, 580–591. doi: 10.1044/jshr.3603.580

Ardanouy, E., Delage, H., and Zesiger, P. (2023). Effectiveness of a group intervention for lexical enrichment in 6-to-10-year-old children with developmental language disorder. *Child Lang. Teach. Ther.* 39, 218–233. doi: 10.1177/02656590231188523

Baron, L. S., and Arbel, Y. (2022a). An implicit–explicit framework for intervention methods in developmental language disorder. *Am. J. Speech-Lang. Pathol.* 31, 1557–1573. doi: 10.1044/2022_AJSLP-21-00172

Baron, L. S., and Arbel, Y. (2022b). Inner speech and executive function in children with developmental language disorder: implications for assessment and intervention. *Perspect. ASHA SIG* 7, 1645–1659. doi: 10.1044/2022_PERSP-22-00042

Beck, I. L., McKeown, M. G., and Kucan, L. (2002). *Bringing Words to Life*. New York, NY: Guilford Press.

Best, W. (2005). Investigation of a new intervention for children with word-finding problems. *Int. J. Lang. Commun. Disord.* 40, 279–318. doi: 10.1080/13682820410001734154

Best, W., Fedor, A., and Thomas, M. S. (2015). Intervening to alleviate word-finding difficulties in children: case series data and a computational modelling foundation. *Cogn. Neuropsychol.* 32, 133–168. doi: 10.1080/02643294.2014.1003204

Best, W., Hughes, L., Masterson, J., Thomas, M. S., Howard, D., Kapikian, A., et al. (2021). Understanding differing outcomes from semantic and phonological interventions with children with word-finding difficulties: a group and case series study. *Cortex* 134, 145–161. doi: 10.1016/j.cortex.2020.09.030

Best, W., Hughes, L. M., Masterson, J., Thomas, M., Fedor, A., Roncoli, S., et al. (2018). Intervention for children with word-finding difficulties: a parallel group randomised control trial. *Int. J. Speech Lang. Pathol.* 20, 708–719. doi: 10.1080/17549507.2017.1348541

Best, W., Ping Sze, W., Edmundson, A., and Nickels, L. (2019). What counts as evidence? Swimming against the tide: valuing both clinically informed experimentally controlled case series and RCTs in intervention research. *Evid. Based Commun. Assess. Interv.* 13, 107–135. doi: 10.1080/17489539.2019.1597444

Biemiller, A., and Slonim, N. (2001). Estimating root word vocabulary growth in normative & advantaged populations: evidence for a common sequence of vocabulary acquisition. *J. Educ. Psychol.* 93:498. doi: 10.1037/0022-0663.93.3.498

Bishop, D. V., Snowling, M. J., Thompson, P. A., and Greenhalgh, T. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *J. Child Psychol. Psychiatry* 58, 1068–1080 doi: 10.1111/jcpp.12721

Bleses, D., Makransky, G., Dale, P. S., Højen, A., and Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Appl. Psycholinguist.* 37, 1461–1476. doi: 10.1017/S0142716416000060

Branagan, A., and Parsons, S. (2021). *Word Aware 3: Teaching Vocabulary in Small Groups for Ages 6 to 11*. London: Routledge. doi: 10.4324/9781003159377

Broedelet, I., Boersma, P., and Rispens, J. (2023). Implicit cross-situational word learning in children with & without DLD and its relation to lexical-semantic knowledge. *Front. Commun.* 8:1021654. doi: 10.3389/fcomm.2023.1021654

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., et al. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 368:l6890. doi: 10.1136/bmj.l6890

Chiat, S. (2001). Mapping theories of developmental language impairment: premises, predictions and evidence. *Lang. Cogn. Process.* 16, 113–142. doi: 10.1080/01690960042000012

Christopulos, T. T., and Redmond, S. M. (2023). Factors impacting implementation of universal screening of developmental language disorder in public schools. *Lang. Speech Hear. Serv. Sch.* 54, 1080–1102. doi: 10.1044/2023_LSHSS-22-00169

Cirrin, F. M., and Gillam, R. B. (2008). Language intervention practices for school-age children with spoken Language Disorders: a systematic review. *Lang. Speech Hear. Serv. Schools* 39:1. doi: 10.1044/0161-1461(2008/012)

Dennis, J., Law, J., and Charlton, J. (2017). Speech and language therapy interventions for children with primary speech and/or Language Disorders: (Protocol). *Cochrane Database System. Rev.* 2017, 1–21. doi: 10.1002/14651858.CD012490

Dunn, L. M., Dunn, L. M., Whetton, C., and Pintillie, D. (1982). *The British Picture Vocabulary Scale*. NFER-Nelson: Windsor.

Ebbels, S. H., McCartney, E., Slonims, V., Dockrell, J. E., and Norbury, C. F. (2019). Evidence-based pathways to intervention for children with Language Disorders. *Int. J. Lang. Commun. Disord.* 54, 3–19. doi: 10.1111/1460-6984.12387

Esser, G., Wyschkon, A., Ballaschk, K., and Hänsch, S. (2010). *Potsdam–Illinois Test für Psycholinguistische Fähigkeiten (P-ITPA) [Potsdam–Illinois test for psycholinguistic abilities, P-ITPA]*. Göttingen: Hogrefe

Forsythe, R., Murphy, C. A., Tulip, J., and Law, J. (2022). Why clinicians choose their language intervention approach: an international perspective on intervention for children with DLD. *Folia Phoniatr. Logopaedica* 73,537–551. doi: 10.1159/000513242

Georgan, W. C., and Hogan, T. P. (2019). *The many terms used for DLD. DLD and Me*. Available at: https://dldandme.org/terminology/ (accessed January 26 2022).

German, D. J. (1986). *Test of Word Finding*. Austin, TX: Developmental Learning Materials.

German, D. J. (1989). *Test of Word Finding*. Dallas, TX: DLM Teaching Resources.

German, D. J. (1991). *Test of Word Finding in Discourse*. Leicester: Taskmaster.

Gibbons, M. (2021). *Development and delivery of an educational intervention that increased teachers' awareness, knowledge, and actions related to Developmental Language Disorder (DLD)* (M.Sc. thesis). Institute of Technology, Sligo. Available at: https://research.thea.ie/handle/20.500.12065/4518 (accessed January 26, 2024).

Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für Sechs- bis Zehnjährige (WWT 6–10) [Test of vocabulary size and word-finding abilities for children between six and ten, WWT 6–10]*. München: Elsevier.

Green, L. (2020). The specific language impairment/developmental Language Disorders forum: fostering a discussion of terminology. *Perspect. ASHA Spec. Interest Groups* 5, 3–5. doi: 10.1044/2019_PERSP-19-00184

Hartley, C., Bird, L. A., and Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition* 200:104265. doi: 10.1016/j.cognition.2020.104265

Hoffmann, T., Glasziou, P., Boutron, I., Milne, R., Perera, R., Moher, D., et al. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 348:1687. doi: 10.1136/bmj.g1687

Justice, L., Cain, K., Jiang, H., Logan, J., Jia, R., Jiang, H., et al. (2018). Modeling the nature of grammar and vocabulary trajectories from prekindergarten to third grade. *J. Speech Lang. Hear. Res.* 61, 910–923. doi: 10.1044/2018_JSLHR-L-17-0090

Justice, L. M., Meier, J., and Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Lang. Speech Hear. Serv. Sch.* 36, 17–32. doi: 10.1044/0161-1461(2005/003)

Justice, L. M., Schmitt, M. B., Murphy, K. A., Pratt, A., and Biancone, T. (2014). The 'robustness' of vocabulary intervention in the public schools: targets and techniques employed in speech–language therapy. *Int. J. Lang. Commun. Disord.* 49, 288–303. doi: 10.1111/1460-6984.12072

Khan, K. S., Logan, J., Justice, L. M., Bowles, R. P., and Piasta, S. B. (2021). The contribution of vocabulary, grammar, and phonological awareness across a continuum of narrative ability levels in young children. *J. Speech Lang. Hear. Res.* 64, 3489–3503. doi: 10.1044/2021_JSLHR-20-00403

Kueser, J. B., Leonard, L. B., Deevy, P., Haebig, E., and Karpicke, J. D. (2021). Word-learning trajectories influence long-term recall in children with developmental Language Disorder and typical development. *J. Commun. Disord.* 94:106160. doi: 10.1016/j.jcomdis.2021.106160

Law, J., Reilly, S., and McKean, C., editors. (2022). *Language Development: Individual Differences in a Social Context*. Cambridge: Cambridge University Press. doi: 10.1017/9781108643719

Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S. L., and Kueser, J. B. (2020). After initial retrieval practice, more retrieval produces better retention than more study in the word learning of children with developmental language disorder. *J. Speech Lang. Hear. Res.* 63:2763. doi: 10.1044/2020_JSLHR-20-00105

Lindsay, G., and Strand, S. (2016). Children with language impairment: prevalence, associated difficulties, and ethnic disproportionality in an English Population. *Front. Educ.* 1:2. doi: 10.3389/feduc.2016.00002

Lowman, J. J., and Dressler, E. V. (2016). Effects of explicit vocabulary videos delivered through iPods on students with language impairments. *J. Spec. Educ. Technol.* 31, 195–206. doi: 10.1177/0162643416673914

Marante, L., and Hall-Mills, S. (2024). Exploring speech-language pathologists' perception of and individualized education program goals for vocabulary intervention with school-age children with language disorders. *Lang. Speech Hear. Serv. Sch.* 55, 368–380. doi: 10.1044/2023_LSHSS-23-00078

Marks, I., and Stokes, S. F. (2010). Narrative-based intervention for word-finding difficulties: a case study. *Int. J. Lang. Commun. Disord.* 45, 586–599. doi: 10.3109/13682820903277951

Matte-Landry, A., Boivin, M., Tanguay-Garneau, L., Mimeau, C., Brendgen, M., Vitaro, F., et al. (2020). Children with persistent versus transient early language delay: language, academic, and psychosocial outcomes in elementary school. *J. Speech Lang. Hear. Res.* 63, 3760–3774. doi: 10.1044/2020_JSLHR-20-00230

McCartney, E., and Forbes, J. (2023). SLPs and education: social capital relations during Covid-19 disruption. *Scott. Edu. Rev.* 1, 1–22. doi: 10.1163/27730840-20231004

McGregor, K. K. (2020). How we fail children with developmental language disorder. *Lang. Speech Hear. Serv. Sch.* 51, 981–992. doi: 10.1044/2020_LSHSS-20-00003

McGregor, K. K., Goffman, L., Van Horne, A. O., Hogan, T. P., and Finestack, L. H. (2020). Developmental language disorder: applications for advocacy, research, and clinical service. *Perspect. ASHA Spec. Interest Groups* 5, 38–46. doi: 10.1044/2019_PERSP-19-00083

McGregor, K. K., Smolak, E., Jones, M., Oleson, J., Eden, N., Arbisi-Kelm, T., et al. (2022). What children with developmental Language Disorder teach us about cross-situational word learning. *Cogn. Sci.* 46:13094. doi: 10.1111/cogs.13094

Monaghan, P. (2017). Canalization of language structure from environmental constraints: a computational model of word learning from multiple cues. *Top. Cogn. Sci.* 9, 21–34. doi: 10.1111/tops.12239

Motsch, H. J., and Marks, D. K. (2015). Efficacy of the Lexicon Pirate strategy therapy for improving lexical learning in school-age children: a randomized controlled trial. *Child Lang. Teach. Ther.* 31, 237–255. doi: 10.1177/0265659014564678

Nash, M., and Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *J. Speech Lang. Hear. Res.* 48, 439–458. doi: 10.1044/1092-4388(2005/030)

Newcomer, P. L., and Hammil, D. D. (1997). *Test of Language Development - Primary: Third Edition*. Austin, TX: Pro-Ed.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., et al. (2016). The impact of nonverbal ability on prevalence and clinical presentation of Language Disorder: evidence from a population study. *J. Child Psychol. Psychiatry* 57, 1247–1257. doi: 10.1111/jcpp.12573

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br. Med. J.* 372:n71. doi: 10.1136/bmj.n71

Parsons, S., Law, J., and Gascoigne, M. (2005). Teaching receptive vocabulary to children with specific language impairment: a curriculum-based approach. *Child Lang. Teach. Ther.* 21, 39–59. doi: 10.1191/0265659005ct280oa

Patel, R., Chiat, S., Cartwright, M., Herman, R., and Sells, R. (2022). *Efficacy of vocabulary interventions for children of primary-school-age (5-11 years) with Language Disorder: protocol for a systematic review*. PROSPERO 2022 CRD42022327345 Available at: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022327345 (accessed May 17, 2022).

Pérez, J., Díaz, J., Garcia-Martin, J., and Tabuenca, B. (2020). Systematic literature reviews in software engineering. Enhancement of the study selection process using Cohen's K. *J. Syst. Softw.* 168:110657. doi: 10.1016/j.jss.2020.110657

Petermann, F. (2010). *Sprachstandserhebungstest für Fünf- bis Zehnjährige (SET 5–10) [Test of language development for children between five and ten, SET 5–10]*. Göttingen: Hogrefe.

Plaut, D. C. (1999). "Connectionist approaches to word processing," in *Handbook of Psycholinguistics* (ed) M. Gernsbacher (London: Academic Press Ltd), 877–905.

Pomper, R. (2020). *Combining Cues in Novel Word Learning*. The University of Wisconsin-Madison (Doctoral dissertation). Available at: Combining Cues in Novel Word Learning-ProQuest.

Renfrew, C. (1995). *Word Finding Vocabulary Test (4th ed.)*. Bicester: Speechmark Publishing.

Rice, M. L., Cleave, P. L., and Oetting, J. B. (2000). The use of syntactic cues in lexical acquisition by children with SLI. *J. Speech Lang. Hear. Res.* 43, 582–594. doi: 10.1044/jslhr.4303.582

Rice, M. L., and Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment:alongitudinal study from 2; 6 to 21 years of age. *JSLHR* 58, 345–359. doi: 10.1044/2015_JSLHR-L-14-0150

Rinaldi, S., Caselli, M. C., Cofelice, V., D'Amico, S., De Cagno, A. G., Della Corte, G., et al. (2021). Efficacy of the treatment of developmental Language Disorder: a systematic review. *Brain Sci.* 11:407. doi: 10.3390/brainsci11030407

Roberts, M. Y., and Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: a meta-analysis. *Am. J. Speech-Lang. Pathol.* 20, 180–199 doi: 10.1044/1058-0360(2011/10-0055)

Roembke, T. C., Simonetti, M. E., Koch, I., and Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. *Front. Psychol.* 14:1175272. doi: 10.3389/fpsyg.2023.1175272

Royal College of Speech and Language Therapists: RCSLT (2021). *Giving Voice to People with DLD*. Available at: https://www.rcslt.org/wp-content/uploads/2021/10/rcslt-dld-factsheet.pdf (accessed 20 May 2023).

Senter, R. (2022). *Speech-Language Pathologists' Services for Children with Co-Occurring Language and Executive Function Deficits* Doctoral dissertation, University of Maryland. Available at: ProQuest Dissertations & Theses Global: www.proquest.com/dissertations-theses/speech-language-pathologists-services-children/docview/2688138491/se-2?accountid=10406 (accessed October 22, 2022).

Siedler, M. (2022). *Sink or SWiM? When and How to Use Narrative Synthesis in Lieu of Meta-Analysis*. Available at: https://usblog.gradeworkinggroup.org/2020/05/ (accessed October 22, 2023).

Smeets, D. J., Van Dijken, M. J., and Bus, A. G. (2014). Using electronic storybooks to support word learning in children with severe language impairments. *J. Learn. Disabil.* 47, 435–449 doi: 10.1177/0022219412467069

Stahl, S. A., and Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *J. Educ. Psychol.* 86, 221–234. doi: 10.1037//0022-0663.86.2.221

Stark, R. E., and Tallal, P. (1981). Selection of children with specific language deficits. *J. Speech Hear. Disord.* 46, 114–122. doi: 10.1044/jshd.4602.114

Steele, S. C. (2020). Vocabulary intervention: a national survey of school-based speech–language pathologists. *Commun. Disord. Q.* 41, 151–161. doi: 10.1177/1525740119827008

Steele, S. C., and Mills, M. T. (2011). Vocabulary intervention for school-age children with language impairment: a review of evidence and good practice. *Child Lang. Teach. Ther.* 27, 354–370. doi: 10.1177/0265659011412247

Steele, S. C., Willoughby, L. M., and Mills, M. T. (2013). Learning word meanings during reading: effects of phonological and semantic cues on children with language impairment. *Int. J. Speech Lang. Pathol.* 15, 184–197. doi: 10.3109/17549507.2012.700322

Storkel, H. L., Komesidou, R., Fleming, K. K., and Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with SLI: identifying adequate progress and successful learning patterns. *Lang. Speech Hear. Serv. Schools* 48, 108–124. doi: 10.1044/2017_LSHSS-16-0058

Storkel, H. L., Komesidou, R., Pezold, M. J., Pitt, A. R., Fleming, K. K., and Romine, R. S. (2019). The impact of dose and dose frequency on word learning by kindergarten children with developmental Language Disorder during interactive book reading. *Lang. Speech Hear. Serv. Sch.* 50, 518–539. doi: 10.1044/2019_LSHSS-VOIA-18-0131

Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., and Kaplan, C. A. (1998). Language-impaired preschoolers: a follow-up into adolescence. *J. Speech Lang. Hear. Res.* 41, 407–418. doi: 10.1044/jslhr.4102.407

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. A., and Paul, P. J. (2000). A comparison of service delivery models: effects on curricular vocabulary skills in the school setting. *Am. J. Speech-Lang. Pathol.* 9, 10–20. doi: 10.1044/1058-0360.0901.10

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *JSLHR* 40, 1245–1260. doi: 10.1044/jslhr.4006.1245

Trueswell, J. C., Medina, T. N., Hafri, A., and Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cogn. Psychol.* 66, 126–156. doi: 10.1016/j.cogpsych.2012.10.001

Tufanaru, C., Munn, Z., Aromataris, E., Campbell, J., and Hopp, L. (2020). "Systematic reviews of effectiveness," in: *JBI Manual for Evidence Synthesis* (eds) E. Aromataris, C. Lockwood, K. Porritt, B. Pilla, and Z. Jordan (JBI). Available at: https://synthesismanual.jbi.global (accessed March 1, 2022).

Wanicharoen, N., and Boonrod, V. (2024). Effect of music therapy on language skills in children with specific language impairment: a systematic review. *J. Assoc. Med. Sci.* 57, 96–103. doi: 10.12982/JAMS.2024.011

Watt, A., and White, S. (2018). Efficacy of group versus individual therapy for advancing receptive and expressive language development for children aged 6–12 years within community settings: a critically appraised topic. *Evid. Based Commun. Assess. Interv.* 12, 54–71. doi: 10.1080/17489539.2018.1444914

Westrupp, E. M., Reilly, S., McKean, C., Law, J., Mensah, F., and Nicholson, J. M. (2020). Vocabulary development and trajectories of behavioral and emotional difficulties via academic ability and peer problems. *Child Dev.* 91, e365–e382. doi: 10.1111/cdev.13219

Wilson, J. (2016). *Data Collection Form for Intervention Reviews for RCTs and Non-RCTs - Template*. The Cochrane Collaboration. Available at:https://dplp.cochrane.org/data-extraction-forms (accessed March 1, 2022).

Wing, C. S. (1990). A preliminary investigation of generalization to untrained words following two treatments of children's word-finding problems. *Lang. Speech Hear. Serv. Sch.* 21, 151–156. doi: 10.1044/0161-1461.2103.151

Wright, S. H. (1993). Teaching word—finding strategies to severely language—impaired children. *Eur. J. Disord. Commun.* 28, 165–175. doi: 10.3109/13682829309041464

Zens, N. K. (2009). *Doctoral Thesis: Facilitating word-learning abilities in children with specific language impairment*. University of Canterbury. Available at: Microsoft Word - Naomi's Thesis examination editing.doc (canterbury.ac.nz).

Zens, N. K., Gillon, G. T., and Moran, C. (2009). Effects of phonological awareness and semantic intervention on word-learning in children with SLI. *Int. J. Speech-Lang. Pathol.* 11, 509–524.

Zupan, B., Hutchings, S. M., Everitt, L. E., and Gupta, C. (2022). Language disorder and internalizing mental health problems in youth offenders: a systematic review. *Int. J. Lang. Commun. Disord.* 57, 1207–1212 doi: 10.1111/1460-6984.12759

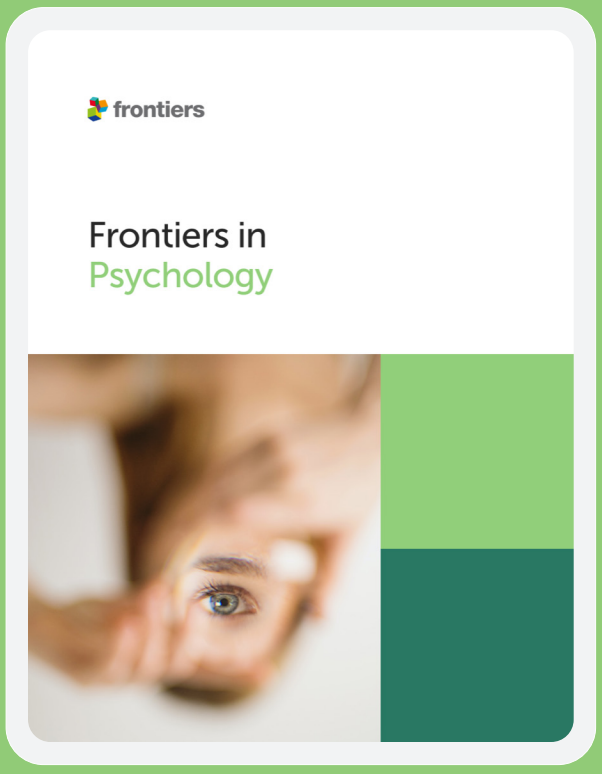
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact