Aphasiology

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: www.tandfonline.com/journals/paph20

Which words contribute most to the functional communication of people with post- stroke aphasia?

Pablo Jordi-Perea, Blanca Fuentes & Nereida Bueno-Guerra

To cite this article: Pablo Jordi-Perea, Blanca Fuentes & Nereida Bueno-Guerra (21 Jun 2026): Which words contribute most to the functional communication of people with post- stroke aphasia?, Aphasiology, DOI: 10.1080/02687038.2026.2690247

To link to this article: https://doi.org/10.1080/02687038.2026.2690247

APHASIOLOGY
https://doi.org/10.1080/02687038.2026.2690247

Which words contribute most to the functional communication of people with post- stroke aphasia?

Pablo Jordi-Perea[a,b], Blanca Fuentes[c*] and Nereida Bueno-Guerra[d,e*]

[a]Doctoral Program in Medical and Surgical Sciences Research, Complutense University of Madrid (UCM), Madrid, Spain; [b]Asociación Afasiactiva, Madrid, Spain; [c]Department of Neurology and Stroke Centre, Hospital La Paz Institute for Health Research-IdiPaz (La Paz University Hospital-Universidad Autónoma de Madrid), Madrid, Spain; [d]Department of Psychology, Universidad Pontificia Comillas, Madrid, Spain; [e]UNINPSI, Clinical Unit of Psychology, Madrid, Spain

ABSTRACT

Background: Functional vocabulary selection is a key component of post-stroke language rehabilitation, yet limited evidence exists regarding which words are perceived as most meaningful by people with aphasia themselves. This study examined which vocabulary items are prioritised by individuals with aphasia or their relatives, as well as by chronic aphasia patients enrolled in the DULCINEA clinical trial.

Methods and Procedures: This cross-sectional observational study comprised two phases. In Phase 1, 15 individuals previously affected by aphasia and their relatives completed an online survey designed to identify everyday words and short phrases considered essential for functional communication. After refinement and categorisation, a base list of 183 items was generated. In Phase 2, 20 participants with chronic post-stroke aphasia selected, with support from relatives and speech and language therapists, the words they wished totrain during the DULCINEA intervention. A total of 891 trained words were analysed. The analyses were conducted at two levels and on an exploratory basis. At the word level, generalized linear mixed models (GLMMs) with a negative binomial distribution were used to examine differences across semantic categories and selected subgroups (sex, age, and depressive symptoms). At the participant level, individual proportional distributions were compared between subgroups using Mann – Whitney U tests with false discovery rate (FDR) correction.

Results: Conversation was the most frequently selected category in both the baseline list and the training items, followed by Series, Foods and Drinks, Emotional and Physical States, and Verbs. Descriptive variations were observed in certain categories according to sex and, to a lesser extent, age. No differences were observed according to the status of depressive symptoms. However, these

ARTICLE HISTORY

Received 30 December 2025
Accepted 8 June 2026

KEYWORDS

Aphasia; functional communication; language therapy; patient preferences; stroke rehabilitation

CONTACT Blanca Fuentes blanca.fuentes@salud.madrid.org Department of Neurology and Stroke Centre, Hospital Universitario La Paz, Departamento de Medicina, Universidad Autónoma de Madrid, Paseo de la Castellana 261, Madrid 28046, Spain

*Blanca Fuentes and Nereida Bueno-Guerra share senior authorship.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02687038.2026.2690247

subgroup differences were not statistically significant after correction for multiple comparisons and were associated with small effect sizes.
Conclusions: Individuals with aphasia prioritise highly functional vocabulary relevant to everyday communication, particularly conversational terms. While exploratory variations were observed across subgroups, these findings should be interpreted with caution given their lack of statistical significance, the small sample size, and the exploratory nature of the analyses. Larger studies are needed to confirm these preliminary observations and to further inform patient-centered vocabulary selection approaches in aphasia rehabilitation.

Introduction

Stroke is one of the leading causes of disability worldwide, with a substantial and increasing global burden (Feigin et al., 2024). Communication plays a fundamental role in human interaction and has been identified as a key predictor of social reintegration after stroke (Beukelman et al., 2015; Gialanella et al., 2011; Wallace et al., 2017). It is estimated that around 4.5 million stroke survivors develop aphasia each year worldwide (RELEASE Collaborators, 2022) and approximately 15–42% experience some degree of language affecting production, comprehension, reading and writing (Flowers et al., 2016). Moreover, a considerable proportion of these individuals develop chronic aphasia, often requiring long-term rehabilitation and support (Brady et al., 2025; Gialanella et al., 2011).

Given its substantial impact on quality of life, addressing communication difficulties is a central objective in post-stroke rehabilitation (Brady et al., 2016; Bullier et al., 2020; Hebert et al., 2016; Pollock et al., 2014; Winstein et al., 2016).

Over the years, aphasia experts have become increasingly aware that current standard treatment approaches alone are insufficient to overcome language-related challenges, leading to questions about whether controlled tasks reliably represent or reflect the language processes or communication demands present in real-life situations (Barnes & Bloch, 2019; Carragher et al., 2012; Herbert et al., 2008; Hersh et al., 2012; Kagan et al., 2008). In practice, people with aphasia experience not only linguistic deficits but also disruptions in everyday communication that affect participation and social interaction. They must relearn how to exchange information, express ideas and feelings, and re-establish connections with listeners across multiple real-life contexts (Hammond et al., 2025). Consequently, communication should be understood not merely as a set of isolated linguistic abilities, but as a situated and interactive process.

In this context, the concept of functional communication has gained increasing relevance, understood as a contextualized, multimodal, and interactive process that occurs within a specific context, involves face-to-face interaction, uses multiple modalities, and is dynamically constructed during interaction (Doedens & Meteyard, 2022; Hammond et al., 2025; Wallace et al., 2017). Functional communication is therefore defined as the effective use of language to achieve meaningful goals in everyday life, such as expressing needs, participating in social interactions, or managing daily activities.

Consequently, aphasia treatment has evolved toward a more holistic perspective, in which approaches emphasise the importance of including functional and personally meaningful goals in therapy (Brady et al., 2025; Doedens & Meteyard, 2022; Palmer et al., 2017; Wallace et al., 2023; Worrall et al., 2011), aim to maximise patients' quality of life outcomes (Bullier et al., 2020; Kagan et al., 2008), and promote comprehensive rehabilitation addressing the different components of the ICF framework (medical-physical, social, and emotional) (Rose et al., 2013; Simmons-Mackie & Kagan, 2007; Wallace et al., 2017).

Within this framework of the International Classification of Functioning, Disability and Health (ICF) (Wallace et al., 2017; World Health Organization [WHO], 2001), communication difficulties resulting from language disorders such as aphasia are conceptualised in terms of activity limitations and participation restrictions, highlighting the barriers individuals face in maintaining meaningful communication. In line with this, approaches such as the Life Participation Approach to Aphasia (LPAA) emphasise the importance of orienting intervention toward personally meaningful goals, prioritising language use in real contexts and active patient involvement in the therapeutic process (Armour et al., 2019; Chapey et al., 2000). Together, these frameworks highlight that effective communication extends beyond linguistic accuracy to include successful participation in social interaction, functional language use in real contexts, and communicative purpose (Doedens & Meteyard, 2022; Hammond et al., 2025; Wallace et al., 2017).

Current intervention approaches increasingly incorporate multimodal strategies that address both linguistic and cognitive components of communication (Brady et al., 2025; Carragher et al., 2012; Rose et al., 2022; Salis et al., 2017), with the aim of promoting generalisation to everyday communicative contexts (Carragher et al., 2015; Doedens & Meteyard, 2022; Herbert et al., 2013). However, progress in word production and overall treatment outcomes are influenced by multiple factors, including treatment dose and intensity, mood, motivation, personal relevance, the therapist – patient relationship, and demographic, behavioural, cognitive, and neurological characteristics (Brady et al., 2025; Braun & Kiran, 2022; Doedens & Meteyard, 2022). In addition, psycholinguistic variables, such as lexical frequency, age of acquisition, semantic features, and phonological complexity, play a key role in word retrieval (Braun & Kiran, 2022; Law et al., 2015). Furthermore, task demands and the type of visual stimuli used can shape performance, with individuals with aphasia often benefiting from contextualised images that support comprehension and emotional connection (Beukelman et al., 2015; McKelvey et al., 2010).

Within this framework, increasing attention has been directed toward the ecological validity of therapeutic materials. Beyond improving isolated lexical processes, effective rehabilitation requires supporting language use in meaningful, interactive, and socially relevant contexts. Consequently, the selection of lexical material becomes a critical component of intervention, as certain words may better facilitate participation, initiate communicative exchanges, and sustain interaction in everyday life.

Although focusing on words that are personally meaningful to the individual with aphasia may maximise the usefulness and impact of intervention, few studies have examined which types of words should be included in therapy to ensure that training materials are truly relevant and functional. It has been suggested that, although individuals may prefer different words depending on factors such as age or gender,

commonalities can be identified among the words selected by people with aphasia, both at the level of specific lexical items and broader semantic categories (Beukelman et al., 2015; McKelvey et al., 2010; Palmer et al., 2017). Moreover, active involvement of the individual in selecting
therapeutic materials and stimuli may enhance the effectiveness and long-term impact of speech and language therapy (Vaezipour et al., 2020).

However, determining whether a word is important or relevant is not straightforward, as it depends on multiple dimensions, including individual preference, the subjective choice of the individual, functional relevance, the usefulness of a word in everyday life, and its potential to facilitate effective communication in real interactions or contribute to communicative success, understood as the extent to which the use of specific lexical items supports the successful achievement of communicative goals in interaction. Thus, not all words contribute equally to communicative success, as their usefulness depends on the contexts in which they are used and their function within everyday interactions (Doedens & Meteyard, 2022; Kagan et al., 2008; Palmer et al., 2017; Worrall et al., 2011).

The aim of the present study is to deepen the understanding of which words people with aphasia consider most important or relevant in the therapeutic context, recognising that such perceptions may reflect a combination of personal, functional, and communicative factors.

Methods

This study is framed within the DUbbing Language-therapy CINEma-based in Aphasia post-stroke (DULCINEA) randomised crossover feasibility clinical trial (NCT04289493) conducted at a Comprehensive Stroke Center from January 2021 until February 2023. We provided patients suffering from chronic post-stroke aphasia with software through which they could train words of their choice. Full details of the study design and its primary clinical outcomes have been published elsewhere (Bueno-Guerra et al., 2023; Fuentes et al., 2022). The study size for the present analysis was determined by the number of participants enrolled in the DULCINEA clinical trial and by the responses obtained from the survey conducted among individuals with previous aphasia and their relatives.

The present manuscript reports a secondary analysis of vocabulary data obtained during the DULCINEA trial, focusing on the words trained by participants rather than on the full set of words initially identified for training. This analysis differs from previous reports by examining patterns in the distribution of trained vocabulary across semantic categories and participants, rather than treatment outcomes.

It is important to note that the study was structured in two phases. The first phase consisted of selecting the words that would then be proposed to the clinical trial patients. To avoid therapist bias, the words were suggested by recovered individuals with aphasia who were currently able to communicate, with the assistance of their relatives (see details in Phase 1, below). None of them participated afterwards in the clinical trial. The second phase consisted of analysing which of the words proposed in Phase 1 were trained by the actual aphasia patients in the trial (see details in Phase 2, below).

This study was reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Vandenbroucke et al., 2007). The reporting of the online survey component followed the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) recommendations (Eysenbach, 2004). The corresponding checklists are provided as supplementary material.

Phase 1. words proposed by recovered patients and their relatives

The first phase consisted of an online survey designed for individuals who have experienced post-stroke aphasia, as well as their caregivers and/or family members. The objective was to identify which words and phrases they considered most relevant to include in speech and language therapy to promote functional communication, emotional engagement, and, consequently, patient adherence to therapy. In the present study, functional relevance is operationalised through the identification of words that participants consider important for real-life communication, reflecting not only their communicative intent but also the contextual usefulness of these lexical items in everyday situations. The survey collected information on participants' age, sex, and the period of time they had been living with aphasia. Participants were also asked which everyday words and phrases they believed were necessary to improve communication at home or during daily activities.

Specifically, participants were asked: "What words would you like yourself or your relative (with aphasia) to be able to say without difficulty?" (Supplementary Appendix A). The survey was created using Google Forms. For distribution, we collaborated with two patient associations: Afasia Activa and Asociación Ictus Madrid, who shared the survey digitally via their respective mailing lists. Therefore, participants were recruited using a non-probabilistic purposive sampling strategy combined with snowball recruitment. Given the exploratory nature of this phase and the relatively small sample size ($n = 15$), the findings should be interpreted with caution in terms of generalizability. Although communication partners were allowed to support individuals with aphasia during survey completion, when necessary, responses were recorded jointly and not attributed separately to patients or relatives. Accordingly, no formal comparison between responses provided by individuals with aphasia and those provided by caregivers was conducted. Therefore, the responses obtained should be interpreted as reflecting a shared perspective between individuals with aphasia and their communication partners, rather than an exclusive representation of patient priorities.

Since the survey was conducted in Spanish, once the responses were collected, the words were classified according to level of difficulty (based on word length in Spanish) and semantic category (e.g., Conversation, Emotional and Physical States, People, Body, Foods and Drinks, Verbs, House, Series (numbers, days, months), Places, Nature, Clothing, Objects). Most grammatical elements, such as articles and prepositions, were excluded from the initial set of candidate words, as this study focused on content words that could be meaningfully represented in audiovisual training materials. This decision was also based on the need to assess prosodic and speech timing variables that cannot be reliably captured through the isolated production of grammatical elements. However, these elements were not entirely excluded from the training, as they were integrated into short, multi-word phrases used during therapy. Words were initially grouped by syllable count: monosyllabic, disyllabic, more than two syllables, and phrases. Concretely, by

"phrases", we refer to language units consisting of a subject and a predicate that express a complete idea. These were simple phrases, composed of a maximum of two or three words (i.e., "I don't understand you - *No te entiendo*; I'm going to shower - *Voy a ducharme*"). They were further classified according to their communicative or pragmatic function, into the following types: declarative, imperative, interrogative, and courtesy/formulaic expressions. By formulaic expressions, we mean commonly used phrases, colloquialisms, or polite expressions that frequently appear in everyday communication (i.e., "Please; Good morning"). The final corpus of words and phrases collected consisted of 183 items.

When classifying words into semantic fields, some items could logically belong to more than one category. To maintain methodological consistency, each word was assigned to only one category. In cases of ambiguity, the criterion used was the most relevant functional context in everyday communication. To further explore the consistency of the categorisation system, a random 20% subset of the lexical items ($n = 37$ out of 183 total items) was selected using a computer-generated randomisation procedure in Microsoft Excel. Each item was assigned a random number using the RAND function, the list was sorted accordingly, and the first 37 items were selected for independent categorisation. Although the initial categorisation had been conducted by a single evaluator, the selected items were subsequently categorised independently by a language model based on artificial intelligence using the same predefined semantic categories and category definitions. Agreement between the original human categorisation and the AI-based categorisation was assessed using Cohen's Kappa coefficient (κ) in order to estimate inter-rater consistency. The coefficient was calculated using an online Cohen's Kappa Calculator (SocSciStatistics, n.d.). However, it is important to recognize a certain degree of subjectivity when classifying words into categories, since these categorization tasks inevitably involve lexical items that may belong to more than one category.

For example, terms such as "*rico*" (tasty) or "*caliente*" (hot) were included in the "Foods and Drinks" category because, in daily use, their meaning is typically interpreted in relation to food. Similarly, items such as "*cama*" (bed) or "*tele*" (TV) were classified under "House" rather than "Objects", given that their communicative use is primarily tied to the domestic environment. The "Objects" category was therefore reserved only for items that were not clearly associated with a more specific functional context. This approach prevented the proliferation of unnecessary categories and preserved the functional coherence of the lexical set.

Phase 2. trained words by patients within the DULCINEA clinical trial

Inclusion criteria for the DULCINEA Trial were as follows: a) Non-fluent aphasia due to ischemic stroke in the left hemisphere, with no evidence of lesions in the right hemisphere as shown by neuroimaging; b) Completion of a previously completed standard course of conventional speech and
language therapy (SLT); c) Severely restricted language, defined as poor repetition even for single words and moderately preserved language comprehension (i.e., below the 70th and 15th percentiles for repetition and auditory comprehension, respectively, as measured by the average score on the word repetition, sentence repetition, word comprehension, command comprehension, and complex ideational material subscales of the

Boston Diagnostic Aphasia Examination, BDAE). These criteria ensured that all participants presented with a predominantly motor/non-fluent aphasia profile, with articulatory difficulties and relatively intact comprehension.

Although a formal classification into classical aphasia subtypes was not undertaken (as the full diagnostic battery was not administered), the inclusion criteria applied are predominantly consistent with profiles characteristic of Broca's aphasia and, in some cases, with chronic and milder forms of global aphasia in which auditory comprehension has been partially recovered. Conversely, these criteria are not compatible with fluent aphasic variants such as anomic aphasia, conduction aphasia, Wernicke's aphasia, or transcortical sensory aphasia. Severity measures derived from BDAE subtests were used for descriptive purposes only and did not inform any analytic decisions.

The exclusion criteria were a) Any clinical condition or other characteristic that would hinder appropriate follow-up, and b) Simultaneous participation in any therapeutic trial evaluating post- stroke recovery.

Once participants had been selected to take part in the DULCINEA project and were about to begin their intervention sessions, both the participant and their family members were invited to an introductory session. In this session, the procedure and software were explained, and a printed version of the word inventory obtained in Phase 1 (Supplementary Appendix B) was presented. From this list, each participant selected a maximum of 50–60 preferred words intended to be trained in the DULCINEA intervention. Throughout this process, both the patient and the family member received guidance from the treating therapist to facilitate word selection.

Therapist involvement during this process was structured to support patient choice while minimising potential bias, taking into account each participant previously assessed linguistic and articulatory capacities. Although patient preference was prioritised, the therapist guided selection towards items that were considered feasible for training within the participant's functional abilities. Nevertheless, the degree of assistance was adapted to the individual needs of each participant, while allowing the patient the liberty to choose among the different words.

Importantly, for this secondary report, only the words that were actually trained were included in the analysis, rather than those initially selected. This meant that even though some participants chose 50 items, we only analysed the words that were actually trained. Table S3 lists all the words trained by the 20 DULCINEA participants, together with their repetition frequency and organised according to semantic category.

Although a maximum number of words was initially selected, no fixed number of words to be trained was established. The number of trained words therefore varied across participants and depended on each participant's individual abilities.

In some cases, participants' linguistic limitations prevented training on all selected items or reduced the number of words trained during the intervention period. This may have introduced a systematic bias toward simpler or more functionally accessible items and may also have affected the number of items recorded for each participant.

Therefore, the data analysed in this phase reflect the subset of words that were both selected and feasible for training, rather than the full range of words originally preferred by participants. This distinction is important, as it implies that the results may be partially influenced by feasibility constraints related to participants' linguistic abilities and

therapist guidance, and should not be interpreted as a direct representation of initial lexical preferences.

After analysing the responses or trained words of all participants in the DULCINEA study, the data were further examined according to sex, age, and the presence or absence of depression, to explore whether differences existed in the distribution of semantic categories within each group.

Given the relatively small sample size ($n = 20$), particularly for subgroup analyses, these comparisons should be considered exploratory in nature. The study may be underpowered to detect statistically significant differences between subgroups, and therefore any observed trends should be interpreted with caution.

Statistical analysis plan

A descriptive analysis was conducted to characterize the frequency and distribution of all trained words based on the word list defined after Phase 1 for use in the DULCINEA intervention. Each word was classified into one of the twelve predefined semantic categories. Both overall distributions and subgroup-specific distributions (sex, age group, and presence or absence of depressive symptoms) were calculated. Depressive symptoms were assessed using the Stroke Aphasia Depression Questionnaire (SADQ-10) (Sutcliffe & Lincoln, 1998). The presence of depressive symptoms was defined using a cut-off score of ≥ 14 indicating depressive symptomatology (Leeds et al., 2004). These analyses were performed at the word level and are presented for descriptive purposes only.

To obtain valid comparisons between groups and semantic categories without violating independence assumptions, and considering that individual trained words were nested within participants and that their total number varied across individuals and subgroups, two complementary levels of analysis were implemented. First, inferential modelling was conducted at the word level using generalized linear mixed models (GLMMs) with a negative binomial distribution to account for the nested structure of the data and formally test differences across semantic categories and subgroups. Second, participant-level proportional analyses were performed to characterise how individuals distributed their trained vocabulary across categories, providing an interpretable, participant-centred perspective.

In these models, the number of words produced within each semantic category was specified as the dependent variable. Semantic category and each grouping variable (sex, age group, and depressive status) were included as fixed effects, and interactions between semantic category and each grouping variable were evaluated. A random intercept for participant was included to account for the nested structure of word-level observations within individuals and to model between-participant variability. These models formally tested whether the observed differences across semantic categories and subgroups were statistically significant.

To complement word-level modeling and to characterize how a typical participant distributed their trained vocabulary across semantic categories, proportional measures were calculated at the participant level. For each participant, the proportion of trained words within each semantic category relative to the total number of words was computed. These proportions were summarized descriptively within each subgroup using medians and interquartile ranges (IQRs).

Between-group differences in the distribution of participant-level proportions were assessed using Mann – Whitney U tests. This nonparametric approach was chosen due to the small sample size and the bounded nature of proportional data. Moreover, conducting inferential comparisons at the participant level allowed for the evaluation of group differences without violating independence assumptions that would arise if individual trained words were treated as independent observations.

To correct for multiple comparisons across the twelve semantic categories, a false discovery rate (FDR) correction was applied using the Benjamini – Hochberg procedure separately for each grouping variable (sex, age group, and depressive status). Effect sizes for the Mann – Whitney U tests were estimated using rank-biserial correlation (r_b). Statistical significance was set at $p < .05$ after correction.

Generalized linear mixed models were estimated using jamovi (The jamovi Project, version 2.6.44). The significance of the GLMMs was evaluated using Wald χ^2 statistics and likelihood ratio tests. Descriptive statistics, participant-level proportion calculations, Mann – Whitney U tests, and FDR-adjusted *p*-values were computed using Microsoft Excel (Microsoft Corporation).

Data management and ethical considerations

Data were prospectively included in a study-specific database developed using REDCap software (REDCap 8.7.4 - 2021, Vanderbilt University) hosted on the IdiPAZ website. All data management procedures adhered to the principles of European biomedical research regulations, ensuring confidentiality. In accordance with European regulations and the Good Clinical Practice Guidelines of the International Conference on Harmonisation (ICH), both the investigator and the institution granted direct access to authorised representatives of the Ethics Committee to review the original medical records of the participants to verify study-related procedures and data. Monitoring was carried out by specialised staff from the Clinical Trials Unit at Hospital Universitario La Paz.

All patients, or their legal guardians or representatives, were asked to sign a written informed consent form after being given a detailed explanation of the study's nature and purpose, and prior to undergoing any procedures related to the clinical trial. An aphasia-friendly information sheet was provided, featuring large text and simplified language to ensure comprehension.

Results

Phase 1. words proposed by recovered patients

The survey link was disseminated through stroke associations and remained active for 30 days. During this period, 18 completed surveys were received. Three were excluded due to lack of informative content (two individuals did not understand the instructions and one survey was submitted blank). Therefore, data from 15 individuals (7 men and 8 women) were analysed. As shown in Table 1, the mean age of participants was 59.7 years (standard deviation [SD] = 15.1). A total of 303 items were collected, with a median of 13 words per participant (IQR = 5–26). The

Table 1. Sociodemographic characteristics and word contribution data of survey participants ($N = 15$).

Characteristic	Value
Age (years), mean (SD) (range)	59.7 (15.1) (31–84)
Women, n (%)	8 (53%)
Time elapsed since stroke (months), median (IQR)	12 (7; 48)
Words selected per participant, median (IQR)	13 (5; 26)

distribution of contributions was positively skewed, with most participants contributing between 5 and 26 words, whereas two individuals provided a larger number of items (38 and 99 words). Substantial overlap across respondents was observed, and the removal of duplicate entries markedly reduced the number of unique items. Consequently, the final word list included both words suggested by multiple respondents and items proposed by a single participant. A total of 42 repeated words were identified among the responses and are listed in Supplementary Appendix C.

After analysing these responses and excluding duplicates, a list of 218 words was compiled; 49 of them were excluded as they constituted articles, prepositions, or similar linguistic expressions (for example, two proposed items were "let's eat" and "I'm going to eat"). In cases where duplicates were identified, the simpler formulation was selected (e.g., "let's eat"). Additionally, when a word was documented but its antonym or gender-equivalent was absent, the missing form was incorporated into the list (for example, if "grandfather" was recorded but "grandmother" was not, or if "cold" was included but "hot" was missing, the absent word was added). In total, 14 additional words were incorporated.

The list of words recorded in the survey, organised by semantic category and level of difficulty, is presented in Table S1. It also includes the English translation of each item, as well as an indication of the words excluded from the final list and those incorporated after the selection process. In the end, the final Phase 1 word list consisted of 183 words.

Agreement between the original human categorisation and the artificial intelligence language model showed a very high level of inter-rater reliability. Out of the 37 lexical items included in the reliability analysis, both classifications coincided in 34 items, corresponding to an observed agreement of 91.89%. Cohen's Kappa coefficient indicated an almost perfect level of agreement ($\kappa = 0.90$), with an expected agreement by chance of 17.82%. These results suggest that the semantic category definitions were applied consistently across classifications.

As illustrated in Figure 1, most words in the final list were disyllabic ($n = 75$), followed by trisyllabic and polysyllabic words ($n = 51$), phrases ($n = 31$), and a smaller proportion of monosyllabic words ($n = 26$). This syllabic distribution reflects structural characteristics of the Spanish language and should therefore be interpreted within this linguistic context.

With regard to semantic categories, most words pertained to "Conversation" (57), followed by "Series" (27), "Foods and Drinks" (22), "House" (19), "Emotional and Physical States" (16), "Persons" (14), and "Verbs" (11), as shown in Figure 2.

Table S2 displays the distribution of the 183 words included in the final list by level of difficulty (syllable count or phrase type) and semantic category, offering a complementary descriptive overview of the data.

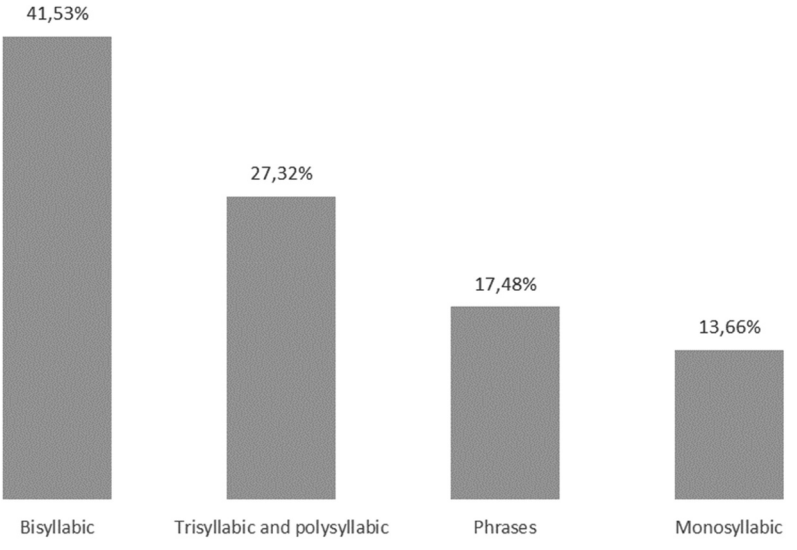

Figure 1. Distribution of the 183 words obtained in Phase 1 according to level of difficulty.

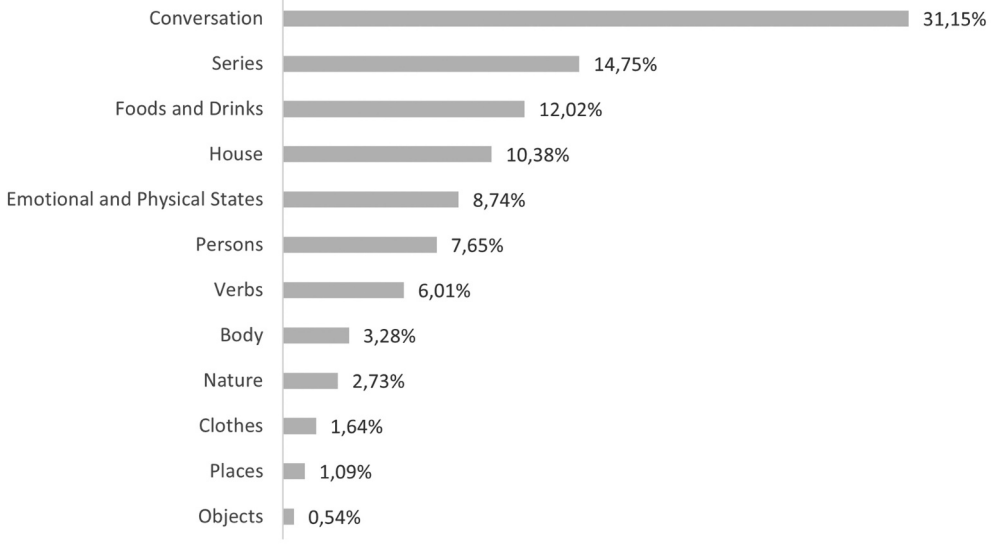

Figure 2. Distribution of the 183 words obtained in Phase 1 according to semantic category.

Phase 2. trained words by patients in the DULCINEA clinical trial

A total of 20 patients enrolled in the DULCINEA clinical trial participated in this phase. Table 2 presents their demographic and clinical characteristics.

Participants had a median age of 64.5 years, and nearly half had completed secondary education. Most were of European (Spanish) origin, with one participant of Latin American (Colombian) origin. The majority were in the chronic post-stroke phase, and

Table 2. Baseline sociodemographic and clinical characteristics of DULCINEA participants.

Sociodemographic and clinical characteristics of DULCINEA participants $N = 20$	
Age (years), median (RIC)	64.5 (54; 72)
Women, n (%)	8 (40%)
Ethnicity, n (%)	19 (95%) European (Spanish); 1 (5%) Latin American (Colombian)
Marital status, n (%)	
Married/common-law partner	50%
Single	20%
Widowed	25%
Divorced	5%
Living situation, n (%)	
Living with others	75%
Living alone	25%
Educational level	
No education, n (%)	1 (5%)
Primary education, n (%)	5 (25%)
Secondary education, n (%)	9 (45%)
Tertiary education, n (%)	5 (25%)
Mean time elapsed since stroke (months), median (RIC)	67 (48; 95)
BDAE comprehension, median (RIC)	26 (19; 30.5)
BDAE Repetition, median (RIC)	6 (4.5; 7)
Screen-positive on SADQ-10, n (%)	8 (40%)
SADQ-10, median (RIC)	12.4 (9.5;15.8)
Handedness, n (%)	19 (95%) Right-handed; 1 (5%) Ambidextrous
Comorbidity: dysarthria, n (%)	9 (45%)
Computer use	11 (55%)
Use of electronic tablets	10 (50%)
Cell phone use	16 (80%)

IQR = interquartile range; BDAE = Boston Diagnostic Aphasia Examination; SADQ = Stroke Aphasic Depression Questionnaire; (score ≥ 14 indicating depressive symptoms); DULCINEA = Dubbing Language-therapy CINEma-based in Aphasia post-stroke.

approximately 40% screened positive on the SADQ-10, indicating the presence of mood-related symptoms. Handedness was predominantly right-handed (19 right-handed, 1 ambidextrous).

Regarding aphasia-related comorbidity, dysarthria was the main associated motor speech disorder considered, affecting 9 participants (45%). No significant differences were observed in baseline BDAE scores across participants.

Regarding marital status, half of the participants were married or in a common-law relationship (50%), while 20% were single ($n = 4$), 25% were widowed ($n = 5$), and 5% were divorced ($n = 1$). Most participants were living with others ($n = 15$, 75%), whereas 25% were living alone ($n = 5$).

Word-level analysis

A total of 889 trained words were recorded across the 20 participants during the DULCINEA intervention. A descriptive analysis of these words was conducted by classifying them according to difficulty level and semantic category in order to examine their overall distribution.

The words analyzed in this section correspond to the subset of items that were trained during the DULCINEA trial and, therefore, do not fully represent participants' initial lexical preferences. Table S3 lists all the words trained (889) by the 20 DULCINEA participants, together with their repetition frequency, organised according to semantic category. The

individual response counts (per participant) for each subgroup (gender, age, and presence of depressive symptoms), are provided in Table S4.

A descriptive quantitative analysis was performed on the set of 889 words trained by the participants in the DULCINEA trial. Overall, participants obtained a median of 42 (IQR 37–51) words or short phrases during individual therapy, with individual contributions ranging from 13 to 82 trained items. As mentioned above, note that variability in the number of trained words contributed by each participant was accounted for in the inferential analyses through mixed-effects models including participant as a random intercept, thus ensuring that differences in individual word counts did not bias subgroup comparisons. As shown in Figure 3, the largest proportion of trained words consisted of words with three or more syllables ($n = 297$), followed by disyllabic words ($n = 287$) and short phrases ($n = 226$), whereas monosyllabic words ($n = 79$) were the least frequent.

Regarding semantic category, most trained words were related to "Conversation" ($n = 323$), followed by "Series" ($n = 123$), "Foods and Drinks" ($n = 89$), "Emotional and Physical States" ($n = 81$), "Verbs" ($n = 80$), and "House" ($n = 65$). The remaining categories comprised fewer than 50 trained words in the total sample, representing less than 5% of the trained words (Figure 4).

Word-level subgroup analysis

These data were also classified and analyzed by subgroups in order to examine the distribution of trained words according to sex, age, and the presence of depressive symptoms. These subgroup patterns may reflect differences in the distribution of trained lexical items and do not necessarily imply differences in communicative function. Table 3 presents the distribution of trained words by semantic category within each subgroup. Percentages were calculated relative to the total number of trained words within each subgroup (e.g., number of Conversation words trained by women divided by the total number of words trained by women).

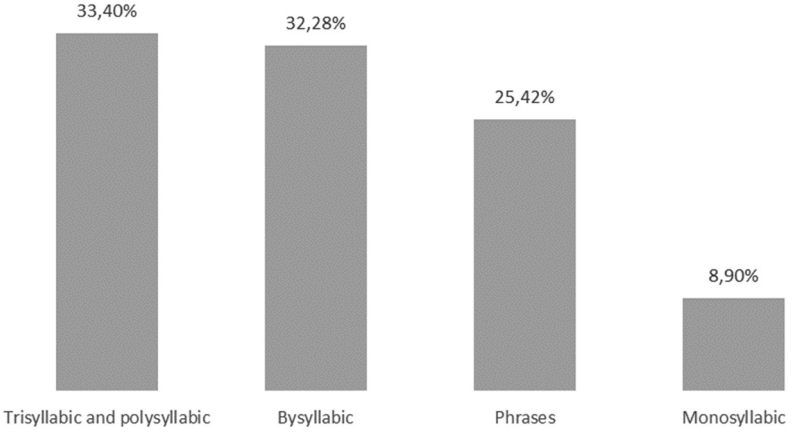

Figure 3. Distribution of the 889 trained words by DULCINEA participants according to level of difficulty.

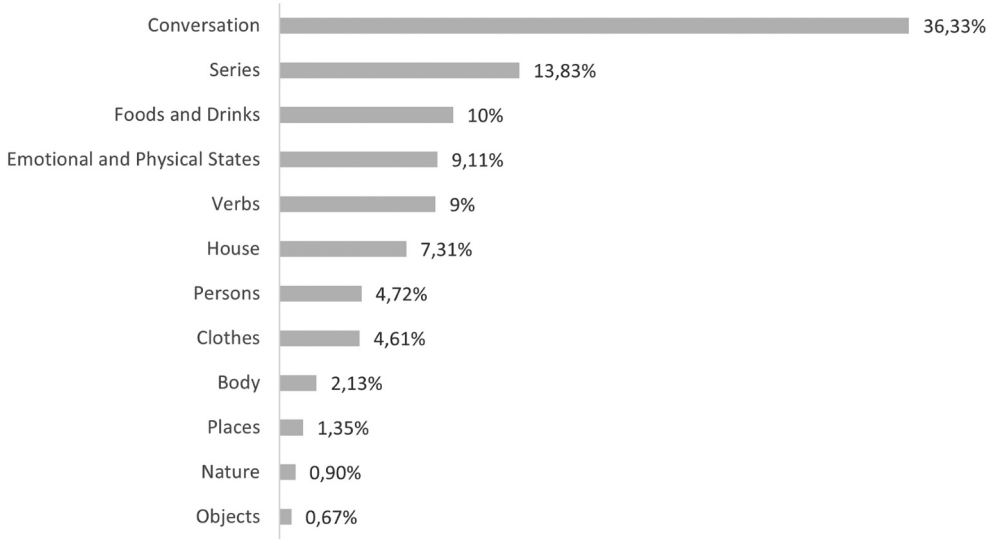

Figure 4. Distribution of the 889 trained words by DULCINEA participants according to semantic category.

Table 3. Distribution of trained words by semantic category across subgroups.

	Total trained words	Women	Men	>60 years	<60 years	With depressive symptoms	Without depressive symptoms
Conversation	323 (36.33%)	100 (34.01%)	223 (37.48%)	228 (35.13%)	95 (39.58%)	129 (38.27%)	194 (35.14%)
Emotional and Physical States	81 (9.11%)	22 (7.48%)	59 (9.92%)	56 (8.6%)	25 (10.42%)	36 (10.68%)	45 (8.15%)
Persons	42 (4.72%)	14 (4.76%)	28 (4.70%)	28 (4.31%)	14 (5.83%)	13 (3.86%)	29 (5.25%)
Body	19 (2.13%)	5 (1.70%)	14 (2.35%)	12 (1.84%)	7 (2.92%)	6 (1.78%)	13 (2.35%)
Foods and drinks	89 (10%)	24 (8.16%)	65 (10.92%)	71 (10.93%)	18 (7.5%)	36 (10.68%)	53 (9.60%)
Verbs	80 (9%)	43 (14.63%) *	37 (6.22%) *	55 (8.47%)	25 (10.41%)	34 (10.09%)	46 (8.33%)
House	65 (7.31%)	30 (10.20%) *	35 (5.88%) *	53 (8.16%)	12 (5%)	25 (7.41%)	40 (7.25%)
Series	123 (13.83%)	29 (9.86%) *	94 (15.8%) *	92 (14.17%)	31 (12.91%)	31 (9.19%) *	92 (16.68%) *
Places	12 (1.35%)	2 (0.68%)	10 (1.68%)	9 (1.38%)	3 (1.25%)	1 (0.29%) *	11 (1.99%) *
Nature	8 (0.9%)	5 (1.70%)	3 (0.50%)	7 (1.08%)	1 (0.42%)	6 (1.78%) *	2 (0.36%) *
Clothes	41 (4.61%)	18 (6.12%)	23 (3.86%)	34 (5.24%)	7 (2.92%)	17 (5.04%)	24 (4.35%)
Objects	6 (0.67%)	2 (0.68%)	4 (0.67%)	4 (0.61%)	2 (0.83%)	3 (0.89%)	3 (0.54%)
	889 (100%)	294 (100%)	595 (100%)	649 (100%)	240 (100%)	337 (100%)	552 (100%)

Cells marked with an asterisk (*) indicate categories in which the proportion of words in one subgroup was approximately twice that observed in the comparison subgroup. These differences are descriptive and did not reach statistical significance.

At the descriptive level, the "Conversation" category was the most frequently trained across all subgroups, with a relatively homogeneous pattern observed in the remaining major categories within each subgroup. Table S6 provides a complementary ranking of the most frequently trained semantic categories within each subgroup. Although it does not introduce additional statistical analyses, it offers a concise visual summary of the distribution of trained vocabulary across groups. In all subgroups, "Conversation" consistently emerged as the most frequently trained category. "Series", "Foods and Drinks" "Emotional and Physical States", and "Verbs" typically occupied the subsequent positions. Categories such as "House", "People", "Body", and "Clothing" tended to fall within an intermediate range (positions 5–9), whereas "Places", "Nature", and "Objects" generally appeared in the lowest positions.

Given the unequal subgroup sizes and for descriptive purposes only, the mean number of trained words per participant was initially estimated for each subgroup, based on the data presented in Table 3. Additional descriptive statistics on the distribution of trained words by subgroup and in total are provided in Table S5.

Women represented 40% of the sample (8 participants) and contributed 33.2% of the total trained words (294 words), whereas men represented 60% (12 participants) and contributed 66.8% (595 words). On average, this corresponds to 37 trained words per woman and 49.6 per man.

Participants older than 60 years represented 70% of the sample (14 participants) and contributed 73% of the total (649 words), whereas those younger than 60 years represented 30% (6 participants) and contributed 27% (240 words). On average, this corresponds to 46.5 words per participant in the older group and 40 in the younger group.

Participants with depressive symptoms represented 40% of the sample (8 participants) and contributed 38% of the total (337 words), whereas participants without depressive symptoms represented 60% (12 participants) and contributed 62% (552 words). On average, this corresponds to 42 trained words per participant with depressive symptoms and 46.3 per participant without depressive symptoms.

Although Table 4 shows descriptive variations between subgroups in the proportion of trained words within certain categories, these differences should be interpreted cautiously. Because percentages were calculated based on the total number of trained words within each subgroup, and given that both subgroup sizes and the number of

Table 4. Simple effects of sex, age, and depression within each semantic category.

Category	Sex χ^2	Sex p	Age χ^2	Age p	Depressive symptoms χ^2	Depressive symptoms p
House	0.637	.425	3.114	.078	0.018	.894
Foods and Drinks	4.222	.040*	2.594	.107	0.035	.852
Conversation	3.668	.055	0.029	.866	0.021	.884
Body	1.309	.252	0.309	.578	0.405	.525
Emotional and Physical States	3.867	.049*	0.003	.957	0.612	.434
Places	2.323	.127	0.158	.691	3.407	.065
Nature	1.481	.224	1.059	.303	3.350	.067
Objects	0.364	.546	0.012	.915	0.035	.851
Persons	0.605	.437	0.133	.716	0.892	.345
Clothes	0.176	.675	2.728	.099	0.064	.800
Series	8.384	.004*	0.991	.319	5.134	.023*
Verbs	3.915	.048*	0.013	.910	0.239	.625

All contrasts have df = 1. *p*-values marked with an asterisk (*) indicate statistical significance ($p < .05$).

trained words per participant were unequal, these variations reflect descriptive trends only and do not permit valid between-group comparisons. Treating each word as an independent observation would violate the independence assumptions required for valid inference. Accordingly, subgroup comparisons based on word-level frequencies are presented solely for descriptive purposes.

However, to analyze the effect of semantic category and each grouping variable (sex, age, and depression) on the number of words produced, and to determine whether the differences observed in the descriptive analyses were statistically significant, inferential analyses were conducted using generalized linear mixed models (GLMMs) with a negative binomial distribution. The estimated effects are presented below.

In all models, the number of words produced differed significantly across semantic categories (all $p < .001$), suggesting that some categories were associated with higher numbers of responses than others.

Subsequently, we examined whether this overall pattern varied as a function of the grouping variables (sex, age, and depression).

Sex

A significant interaction was identified between semantic category and sex (χ^2 (11) = 26.61, $p = .005$). Simple effects (Table 4) suggested possible differences differences between men and women in "Food" ($p = .040$), "Emotional and Physical States" ($p = .049$), "Series" ($p = .004$), and "Verbs" ($p = .048$). Specifically, in our sample, descriptively higher frequencies were observed in men compared to women in "Foods and Drinks", "Emotional and Physical States", and "Series", whereas higher frequencies were observed in women in "Verbs". No significant differences were observed in the remaining categories.

Age

No significant interaction was found between semantic category and age (χ^2 (11) = 8.92, $p = .629$), nor was there a main effect of age (χ^2 (1) = 1.97, $p = .161$). Simple effects confirmed the absence of significant differences between participants older than 60 years and those younger than 60 years across all categories (all $p > .05$).

Depressive symptoms

No significant interaction was found between semantic category and depressive symptoms status (χ^2 (11) = 15.36, $p = .167$), nor were there overall differences in total word production between participants with and without depressive symptoms (χ^2 (1) = 0.231, $p = .631$).

Simple effects showed no significant category-specific differences, except for "Series" ($p = .023$), where higher frequencies were observed in participants without depressive symptoms. However, given that the overall interaction was not significant and considering correction for multiple comparisons, this finding should be interpreted with caution. In this sample, non-significant trends were also observed in Places ($p = .065$) and Nature ($p = .067$).

Participant-Level analysis

To examine the relative distribution of semantic categories across participants, individual proportions were calculated for each category relative to the total number of words trained per participant. Descriptive statistics are presented as medians and interquartile ranges (IQRs) within each subgroup, as shown in Table 5.

In this context, the median of the proportions reflects the typical distribution of trained vocabulary for a representative participant within that subgroup, rather than an aggregated group-level proportion.

Participant-level subgroup analysis

Sex

Descriptively, the proportional distribution across categories was largely comparable between women and men. "Conversation" was the predominant category in both groups (women = 38% [26–42]; men = 39% [34–43]).

Regarding hierarchical order based on median proportions, in men the subsequent categories in descending order were "Series", "Foods and Drinks", and "Verbs", whereas in women, following Conversation, the most prominent categories were "Series", "Verbs", and "House". Overall, the hierarchical pattern appeared broadly similar between sexes, with minor variations in intermediate positions.

A descriptive trend toward a higher proportion of "Verbs" was observed in women compared to men (11% [8–17] vs. 6% [4–9]), which may represent an approximate five percentage point difference. Similarly, women showed a slightly higher proportion in "House" (9% [7–13] vs. 5%), with a difference of nearly four percentage points. Men

Table 5. Median individual proportions by semantic category and subgroup.

Category	Male	Female	>60 years	<60 years	Depressive symptoms	No depressive symptoms
House	0.05 [0.03–0.09]	0.09 [0.07–0.13]	0.12 [0.09–0.13]	0.08 [0.07–0.10]	0.09 [0.07–0.10]	0.06 [0.04–0.09]
F&D	0.11 [0.09–0.14]	0.08 [0.06–0.12]	0.08 [0.06–0.09]	0.07 [0.04–0.10]	0.06 [0.04–0.09]	0.07 [0.04–0.09]
Conversation	0.39 [0.34–0.43]	0.38 [0.26–0.42]	0.37 [0.28–0.40]	0.44 [0.37–0.48]	0.41 [0.35–0.43]	0.35 [0.30–0.41]
Body	0.02 [0.01–0.04]	0.05 [0.04–0.06]	0.02 [0–0.03]	0.03 [0.02–0.04]	0.02 [0–0.03]	0.02 [0.01–0.04]
E&P States	0.10 [0.08–0.12]	0.10 [0.04–0.12]	0.09 [0.05–0.11]	0.11 [0.06–0.13]	0.10 [0.08–0.13]	0.09 [0.05–0.11]
Places	0.02 [0–0.02]	0.01 [0–0.02]	0.01 [0–0.02]	0 [0–0.01]	0 [0–0]	0.02 [0–0.02]
Nature	0 [0–0]	0.01 [0–0.02]	0 [0–0]	0 [0–0]	0 [0–0]	0 [0–0]
Objects	0 [0–0.02]	0 [0–0]	0 [0–0.01]	0 [0–0.01]	0 [0–0]	0 [0–0]
Persons	0.04 [0.02–0.07]	0.05 [0.04–0.06]	0.05 [0.02–0.06]	0.06 [0.04–0.08]	0.04 [0.02–0.05]	0.06 [0.04–0.07]
Clothes	0.06 [0.04–0.09]	0.11 [0.08–0.17]	0.08 [0.04–0.11]	0.07 [0.03–0.07]	0.07 [0.02–0.10]	0.07 [0.04–0.09]
Series	0.14 [0.08–0.24]	0.13 [0.03–0.14]	0.14 [0.11–0.18]	0.06 [0.03–0.14]	0.12 [0.02–0.14]	0.14 [0.08–0.24]
Verbs	0.06 [0.04–0.09]	0.11 [0.08–0.17]	0.08 [0.05–0.13]	0.11 [0.06–0.13]	0.05 [0.03–0.07]	0.04 [0.02–0.06]

Values are expressed as the median percentage of the individual proportion [Q1–Q3]. Q1–Q3 = interquartile range.

exhibited a marginally higher proportion in "Foods and Drinks" (11% vs. 8%), although this difference was below five percentage points.

No substantial descriptive differences between sexes were observed in the remaining categories. "Nature" and "Objects" showed residual values in both groups.

Age

Median proportions were similar between participants older and younger than 60 years. "Conversation" was the most frequent category in both groups (>60 years = 37% [28–40]; <60 years = 44% [37–48]).

In this sample, in terms of hierarchical order, "Conversation" ranked first in both age groups. Among participants older than 60 years, "Series" clearly occupied the second position, followed by "Foods and Drinks" and "House". In participants younger than 60 years, "Series" also ranked second, with the remaining categories maintaining a similar order. Overall, the hierarchical structure appeared largely comparable across age groups.

The > 60 group showed a higher proportion in "Series" (14% [11–18] vs. 6% [3–14]), representing an approximate eight-percentage-point difference that may indicate a descriptive trend. In the remaining categories, including "Foods and Drinks" and "House", no relevant descriptive differences were observed between age groups. "Nature" and "Objects" again showed residual values in both groups.

Depressive symptoms

Median proportions were comparable between participants with and without depressive symptoms. "Conversation" was the predominant category in both groups (with symptoms = 41% [35–43]; without symptoms = 35% [30–41]).

In hierarchical terms, "Conversation" ranked first in both groups, followed by "Series". Intermediate positions of the remaining categories showed broadly similar patterns between participants with and without depressive symptoms, maintaining a similar proportional structure.

An approximate six-percentage-point difference was observed in "Conversation" (41% vs. 35%), which may be interpreted as a descriptive trend. In the remaining categories, including "Series", "Emotional and Physical States", "Foods and Drinks", and "House", no relevant descriptive differences were identified. "Nature" and "Objects" showed residual values in both cases.

To determine whether the observed descriptive differences were statistically significant, subgroup comparisons were conducted using Mann – Whitney U tests with false discovery rate (FDR) correction. No statistically significant differences were identified between subgroups in any of the analyzed categories. Overall, given the limited sample size and the lack of consistent statistical significance after correction for multiple comparisons, subgroup findings should be considered exploratory and interpreted with caution. Therefore, these variations are better understood as descriptive differences rather than as evidence of systematic patterns attributable to sex, age, or depressive symptoms. Complete U statistics as well as, p, p_FDR and *r* values are presented in Supplementary Table S7.

Discussion

This study explored the vocabulary most frequently represented among trained items during language rehabilitation. The data were analysed at both the overall sample and subgroup levels to explore possible variations in semantic category distribution according to sex, age group, and depressive symptoms. Subgroup analyses were conducted at both participant and word levels.

In the overall sample, the distribution of trained words showed a higher representation of certain semantic categories. In particular, "Conversation" clearly stood out as the most represented category. This was followed by categories such as "Series", "Foods and Drinks", "Emotional and Physical States", "Verbs", and "House", which also showed relatively high frequencies. In contrast, "People", "Body", and "Clothes" showed an intermediate presence, whereas "Places", "Nature", and "Objects" were the least represented categories within the analysed set.

Despite individual variations, most participants showed a similar pattern, with the six most frequently trained categories generally being "Conversation", followed by "Series", "Foods and drinks", "Verbs", "House", and "Emotional and Physical States". At the participant level, the same categories were generally most represented across subgroups, although their order varied slightly. Similarly, possible differences across subgroups in the distribution of semantic categories were also examined. At a descriptive level, higher frequencies were observed in men compared to women in categories such as "Foods and drinks", "Emotional and Physical States", and "Series". Conversely, higher frequencies were observed in women in the categories "Verbs" and "House". In the age groups, slightly greater representation was observed in participants over 60 years of age in the categories "House" and "Clothes". Regarding depressive symptoms, a larger share of words was observed in participants without depressive symptoms in the category "Series". Likewise, participant-level descriptive analyses showed a broadly similar pattern.

However, these descriptive differences were not statistically significant after correction for multiple comparisons, and effect sizes were small. Therefore, these variations should be interpreted cautiously as descriptive patterns rather than systematic subgroup differences. Similarly, previous research has shown interest in identifying relevant vocabulary for people with aphasia. For example, Palmer et al. (2017) proposed a large set of items distributed across 27 semantic categories. Notably, although their study included a considerably larger number of words and participants, 113 of the 183 words in our final basic list overlapped with theirs. In addition, the semantic categories used in our research closely align with those proposed by Palmer et al. This study included 27 semantic categories, many subdivided into multiple subgroups. Therefore, our study used a more simplified categorisation system. In addition, it is also worth noting that 48 of the non-overlapping words between both lists belonged to the category "Conversation". This was probably due to the nature of this category, which included terms typical of everyday communication, less linked to concrete lexical content or specific meanings and often consisting of simple phrases, which were not included in Palmer's study. However, some words from "Conversation" in our list could be associated with the main categories or subcategories used in their analysis. For example, "más" or "¿cómo?" could be included within "Grammatical/function words", and "hora" within the main category "Time".

In their study, Palmer reported differences in semantic categories with respect to gender and age, although these differ from those found in our work. For example, in their study, men selected more words related to "Nature and Gardening" and "Travel", whereas women showed higher percentages in categories such as "Clothes", "Money and Numbers", "Personal Care", "Shopping", and "Personal Items". Regarding age, participants under 65 years selected more words related to "Travel", "Actions", and "Time" compared to those over 65 years.

In this regard, the predominance of conversational vocabulary in our study may be consistent with observations from linguistic corpora such as the Corpus de Referencia del Español Actual (CREA) (Real Academia Española, 2024), where spontaneous spoken language is characterized by a high frequency of conversational elements, including discourse markers, interrogative particles, and expressions of politeness, which play an important role in everyday communication. This is also consistent with previous research on functional communication in aphasia recovery (Doedens & Meteyard, 2022). People with aphasia have identified participation in social, leisure, and work activities as a central element of their quality of life after stroke (Worrall et al., 2011). This notion may have been reflected to some extent in the responses collected in our study, as the most frequently represented categories across all subgroups ("Conversation", "Series", "Foods and Drinks", "Verbs", and "House") included words typically associated with everyday communication contexts.

Likewise, several contemporary approaches to language rehabilitation, based on the framework of the International Classification of Functioning, Disability and Health (ICF), emphasize the importance of considering not only linguistic deficits but also communicative activity and the person's participation in everyday life (Wallace et al., 2019). From this perspective, rehabilitation should not focus solely on the recovery of isolated words but also on restoring the functional use of communication in real-life contexts (Barnes & Bloch, 2019; Carragher et al., 2012; Herbert et al., 2008; Hersh et al., 2012). In this sense, identifying vocabulary that may be relevant for people with aphasia may provide a useful basis for informing therapeutic material selection and supporting more person-centred rehabilitation approaches (Doedens & Meteyard, 2022; Palmer et al., 2017). Involving people with aphasia in selecting training items may also increase engagement and support person-centred rehabilitation approaches (Vaezipour et al., 2020).

In this context, examining trained vocabulary that is associated with everyday communication contexts, while considering possible differences across sociodemographic variables such as sex or age, may represent a promising avenue for future research and clinical interventions.

Limitations

This study has several limitations that should be considered when interpreting the results. First, we included only patients with long-standing motor aphasia who had already completed standard speech therapy. Therefore, the results cannot be directly generalised to all people with aphasia, particularly those in the early stages of recovery or those with other aphasia profiles. In addition, working with a relatively specific sample may have influenced the nature of the responses collected, which limits the possibility of extrapolating these findings to broader populations.

One limitation of the present study concerns the fact that the semantic categorisation process was initially conducted by a single speech and language therapist, which may introduce a degree of subjectivity. To partially address this issue, an exploratory inter-rater reliability analysis was performed using an artificial intelligence-based language model, which showed high agreement with the original categorisation. Nevertheless, semantic categorisation inherently involves some degree of subjectivity, as certain lexical items may reasonably belong to more than one category. In addition, the final set of trained words may have been influenced by therapeutic guidance and participants' linguistic abilities.

Furthermore, the relatively small number of participants who contributed to the development of the word list and to the DULCINEA intervention means that the results should be interpreted as exploratory and may limit the generalisability of comparisons between subgroups. In this sense, the present study can be considered a preliminary step toward identifying vocabulary that may be relevant for aphasia rehabilitation. Future studies with larger and more diverse samples will be necessary to confirm and extend these findings.

Conclusions

Functional vocabulary plays an important role in language rehabilitation. In the present study, the words most frequently trained during therapy were mainly associated with the semantic categories "Conversation", "Series", "Foods and Drinks", "Emotional and Physical States", "Verbs", and "House". Examining the words most frequently trained by people with aphasia may provide a useful basis for guiding the selection of therapeutic materials and informing language therapy approaches. In this sense, these findings may offer preliminary insights for informing language therapy approaches. Although some descriptive differences between subgroups and categories were observed, further research with larger samples is needed to confirm and extend these findings.

Acknowledgements

The patients' association "*Afasia Activa*" was a partner of the research consortium. They contributed to the design of the research project; provided advice on an earlier draft of this manuscript and delivered a list of target words (together with two other Spanish associations for stroke patients, "*Asociación Ictus Madrid*" and "*Freno al Ictus*"). We acknowledge the collaboration of Cristian Sempere and the speech and language therapists Lidia de la Fuente and Celia Delgado (†). Also, we greatly appreciate the support of Morote Traducciones for editing assistance.

Disclosure statement

The authors declare that they have no competing interests. Some participants were recruited from Afasia Activa, where author Pablo Jordi works as a speech and language therapist. However, none of these participants were under his clinical care, and he remained blinded to their clinical background and progress throughout the study.

Funding

This study was promoted by Blanca Fuentes and the Fundación de Investigación Biomédica del Hospital Universitario La Paz, which hosted a research consortium joined by the Department of Neurology at La Paz University Hospital, the Department of Psychology at Comillas Pontifical University, and the patients' association, Afasia Activa. This project has received funding from "La Caixa" Foundation under the project code LCF/PR/HR19/52160009. The funder was not involved in any of the following processes: design of the trial, data collection, analysis, interpretation, or writing of the manuscript. BF is a member of the Spanish Stroke Research Network RICORS ICTUS [RD21/0006/0012; RD24/0009/0009] funded by the Instituto de Salud Carlos III, Ministerio de Ciencia, Innovación y Universidades, and confunded by the European Union.

ORCID

Blanca Fuentes http://orcid.org/0000-0002-0363-862X

Ethical approval

All versions of the DULCINEA protocol (version 1.0, dated 15 July, 2019 and version 2.0, dated 7 October, 2020) were approved by the Hospital Ethics Committee for Clinical Research. The protocol was registered with the ClinicalTrials.gov register: NCT04289493. All patients, or their guardian or legal representative, signed a written informed consent document after being given a detailed explanation of the nature and purpose of this study, and before undergoing any of the procedures related to the clinical trial. An aphasia-friendly information sheet containing large text and simplified language was provided.

References

Armour, M., Brady, S., Sayyad, A., & Krieger, R. (2019). Self-reported quality of life outcomes in aphasia using life participation approach values: 1-year outcomes. *Archives of Rehabilitation Research and Clinical Translation*, *1*(3), 100025. https://doi.org/10.1016/j.arrct.2019.100025

Barnes, S., & Bloch, S. (2019). Why is measuring communication difficult? A critical review of current speech pathology concepts and measures. *Clinical Linguistics and Phonetics*, *33*(3), 1–18.

Beukelman, D. R., Hux, K., Dietz, A. R., McKelvey, M. L., & Weissling, K. S. E. (2015). Using visual scene displays as communication support options for people with chronic, severe aphasia: A summary of AAC research and future research directions. *Special Education and Communication Disorders Faculty Publications*, *151*(3), 234–245. https://doi.org/10.3109/07434618.2015.1052152

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, *2016*(6), Article CD000425. https://doi.org/10.1002/14651858.CD000425.pub4

Brady, M. C., Mills, C., Prag Øra, H., Novaes, N., Becker, F., Constantinidou, F., Flöel, A., Sunnerhagen, K. S., Isaksen, J., Jagoe, C., Jesus, L. M., Marangolo, P., Meinzer, M., van der Meulen, I., Campbell, P., Ho, L., Hussain, S., & Hilari, K. (2025). European Stroke Organisation (ESO) guideline on aphasia rehabilitation. *European Stroke Journal*, *10*(4), 1189–1220. https://doi.org/10.1177/23969873241311025

Braun, E. J., & Kiran, S. (2022). Stimulus- and person-level variables influence word production and response to anomia treatment for individuals with chronic poststroke aphasia. *Journal of Speech, Language, & Hearing Research*, *65*(10), 3854–3872. https://doi.org/10.1044/2022_JSLHR-21-00527

Bueno-Guerra, N., Provencio, M., Tarifa-Rodríguez, A., Navarro, A., Sempere- Iborra, C., Jordi, P., de Celis-Ruiz, E., Alonso de Leciñana, M., Martín-Alonso, M., Rigual, R., Ruiz-Ares, G., Rodríguez-Pardo, J., Viués-Ortega, J., & Fuentes, B. (2023). Impact of post-stroke aphasia on functional

communication, quality of life, perception of health and depression: A case-control study. *European Journal of Neurology*, *31*(4), e16184. https://doi.org/10.1111/ene.16184

Bullier, B., Cassoudesalle, H., Villain, M., Cogné, M., Mollo, C., De Gabory, I., Dehail, P., Joseph, P. A., Sibon, I., & Glize, B. (2020). New factors that affect quality of life in patients with aphasia. *Annals of Physical and Rehabilitation Medicine*, *63*(1), 33–37. https://doi.org/10.1016/j.rehab.2019.06.015

Carragher, M., Conroy, P., Sage, K., & Wilkinson, R. (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions. *Aphasiology*, *26*(7), 895–916. https://doi.org/10.1080/02687038.2012.676164

Carragher, M., Sage, K., & Conroy, P. (2015). Outcomes of treatment targeting syntax production in people with Broca's-type aphasia: Evidence from psycholinguistic assessment tasks and everyday conversation. *International Journal of Language & Communication Disorders*, *50*(3), 322–336. https://doi.org/10.1111/1460-6984.12135

Chapey, R., Duchan, J. F., Elman, R. J., Garcia, L. J., Kagan, A., Lyon, J. G., & Simmons-Mackie, N. (2000). Life participation approach to aphasia: A statement of values for the future. *The ASHA Leader*, *5*(3), 4–6. https://doi.org/10.1044/leader.FTR.05032000.4

Doedens, W. J., & Meteyard, L. (2022). What is functional communication? A theoretical framework for real-world communication applied to aphasia rehabilitation. *Neuropsychology Review*, *32*(4), 937–973. https://doi.org/10.1007/s11065-021-09531-2

Eysenbach, G. (2004). Improving the quality of web surveys: The checklist for reporting results of internet e-surveys (CHERRIES). *Journal of Medical Internet Research*, *6*(3), e34. https://doi.org/10.2196/jmir.6.3.e34

Feigin, V. L., Stark, B. A., Johnson, C. O., Roth, G. A., Bisignano, C., Abady, G. G., Abdelkader, A., Abdelmasseh, M., Abd-Elsalam, S., Abdi, P., Abdollahi, A., Abdoun, M., Abd-Rabu, R., Abdulah, D. M., Abdullahi, A., Abebe, M., Abeldaño Zuñiga, R. A., Abhilash, E. S.Abiodun, O. O. . . . Vos, T. (2024). Global, regional, and national burden of stroke and its risk factors, 1990-2021. A systematic analysis for the global burden of disease study 2021. *Lancet Neurology*, *23*(10), 973–1003. https://doi.org/10.1016/S1474-4422(24)00225-1

Flowers, H. L., Skoretz, S. A., Silver, F. L., Rochon, E., Fang, J., Flamand-Roze, C., & Martino, R. (2016). Poststroke aphasia frequency, recovery, and outcomes: A systematic review and meta-analysis. *Archives of Physical Medicine & Rehabilitation*, *97*(12), 2188–2201.e8. https://doi.org/10.1016/j.apmr.2016.03.006

Fuentes, B., de la Fuente-Gómez, L., Sempere-Iborra, C., Delgado-Fernández, C., Tarifa-Rodríguez, A., Alonso de Leciñana, M., de Celis-Ruiz, E., Gutiérrez- Zúñiga, R., López-Tàppero, J., Martín Alonso, M., Pastor-Yborra, S., Rigual, R., Ruiz-Ares, G., Rodríguez-Pardo, J., Virués-Ortega, J., Borobia, A. M., Blanco, P., & Bueno-Guerra, N. (2022). Dubbing language-therapy cinema-based in aphasia post-stroke (DULCINEA): Study protocol for a randomized crossover pilot trial. *Trials*, *23*(1), 1–10. https://doi.org/10.1186/s13063-021-05956-5

Gialanella, B., Bertolinelli, M., Lissi, M., & Prometti, P. (2011). Predicting outcome after stroke: The role of aphasia. *Disability and Rehabilitation*, *33*(2), 122–129. https://doi.org/10.3109/09638288.2010.488712

Hammond, L., Christensen, T., Fridriksson, J., & den Ouden, D. B. (2025). Assessing functional communication in persons with aphasia: A scoping review of formal and informal measures. *International Journal of Language & Communication Disorders*, *60*(3), e70051. https://doi.org/10.1111/1460-6984.70051

Hebert, D., Lindsay, M. P., McIntyre, A., Kirton, A., Rumney, P. G., Bagg, S., Bayley, M., Dowlatshahi, D., Dukelow, S., Garnhum, M., Glasser, E., Halabi, M.-L., Kang, E., MacKay-Lyons, M., Martino, R., Rochette, A., Rowe, S., Salbach, N., Semenko, B., . . . Foley, N. (2016). Canadian stroke best practice recommendations: Stroke rehabilitation practice guidelines, update 2015. *International Journal of Stroke*, *11*(4), 459–484. https://doi.org/10.1177/1747493016643553

Herbert, R., Gregory, E., & Best, W. (2013). Syntactic versus lexical therapy for anomia in acquired aphasia: Differential effects on narrative and conversation. *International Journal of Language & Communication Disorders*, *49*(2), 162–173. https://doi.org/10.1111/1460-6984.12054

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, *22*(2), 184–203. https://doi.org/10.1080/02687030701262613

Hersh, D., Worrall, L., Howe, T., Sherratt, S., & Davidson, B. (2012). Smarter goal setting in aphasia rehabilitation. *Aphasiology*, *26*(2), 220–233. https://doi.org/10.1080/02687038.2011.640392

Kagan, A., Simmons-Mackie, N., Rowland, A., Huijbregts, M., Shumway, E., & McEwen, S., et al. (2008). Counting what counts: A framework for capturing real-life outcomes of aphasia intervention. *Aphasiology*, *22*(3), 258–280. https://doi.org/10.1080/02687030701282595

Law, S.-P., Kong, A. P.-H., Lai, L. W.-S., & Lai, C. (2015). Effects of context and word class on lexical retrieval in Chinese speakers with anomic aphasia. *Aphasiology*, *29*(1), 81–100. https://doi.org/10.1080/02687038.2014.951598

Leeds, L., Meara, R. J., & Hobson, J. P. (2004). The utility of the stroke aphasia depression questionnaire (SADQ) in a stroke rehabilitation unit. *Clinical Rehabilitation*, *18*(2), 228–231. https://doi.org/10.1191/0269215504cr685oa

McKelvey, M. L., Hux, K., Dietz, A., & Beukelman, D. R. (2010). Impact of personal relevance and contextualization on word-picture matching by people with aphasia. *American Journal of Speech-Language Pathology*, *19*(1), 22–33. https://doi.org/10.1044/1058-0360(2009/08-0021)

Palmer, R., Hughes, H., & Chater, T. (2017). What do people with aphasia want to be able to say? A content analysis of words identified as personally relevant by people with aphasia. *PLOS ONE*, *12*(3), e0174065. https://doi.org/10.1371/journal.pone.0174065

Pollock, A., St George, B., Fenton, M., & Firkins, L. (2014). Top 10 research priorities relating to life after stroke - consensus from stroke survivors, caregivers, and health professionals. *International Journal of Stroke*, *9*(3), 313–320. https://doi.org/10.1111/j.1747-4949.2012.00942.x

Reall Academia Española. (2024). *Banco de datos (CREA): Corpus de referencia del español actual*. https://www.rae.es

RELEASE Collaborators. (2022). Dosage, intensity, and frequency of language therapy for aphasia: A systematic review-based individual participant data network meta-analysis. *Stroke*, *53*(3), 956–967. https://doi.org/10.1161/STROKEAHA.121.035216

Rose, M., Cherney, L., & Worrall, L. (2013). Intensive comprehensive aphasia programs: An international survey of practice. *Topics in Stroke Rehabilitation*, *20*(5), 379–387. https://doi.org/10.1310/tsr2005-379

Rose, M. L., Nickels, L., Copland, D., Togher, L., Godecke, E., Meinzer, M., Rai, T., Cadilhac, D. A., Kim, J., Hurley, M., Foster, A., Carragher, M., Wilcox, C., Pierce, J. E., & Steel, G. (2022). Constraint-induced or multimodal personalised aphasia therapy compared with usual care for chronic post-stroke aphasia (COMPARE): A randomised controlled trial. *Journal of Neurology, Neurosurgery & Psychiatry*, *93*(6), 573–581. https://doi.org/10.1136/jnnp-2021-327897

Salis, C., Hwang, F., Howard, D., & Lallini, N. (2017). Short-term and working memory treatments for improving sentence comprehension in aphasia: A review and a replication study. *Seminars in Speech and Language*, *38*(1), 29–39. https://doi.org/10.1055/s-0036-1597262

Simmons-Mackie, N., & Kagan, A. (2007). Application of the ICF in aphasia. *Seminars in Speech and Language*, *28*(4), 244–253. https://doi.org/10.1055/s-2007-986521

SocSciStatistics. (n.d.). *Cohen's kappa calculator*. https://www.socscistatistics.com/tests/kappa/calculator/

Sutcliffe, L. M., & Lincoln, N. B. (1998). The assessment of depression in aphasic stroke patients: The development of the stroke aphasic depression questionnaire. *Clinical Rehabilitation*, *12*(6), 506–513. https://doi.org/10.1191/026921598672167702

Vaezipour, A., Campbell, J., Theodoros, D., & Russell, T. (2020). Mobile apps for speech-language therapy in adults with communication disorders: Review of content and quality. *JMIR mHealth and uHealth*, *8*(10). https://doi.org/10.2196/18858

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J., & Egger, M. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *PLOS Medicine*, *4*(10), e297. https://doi.org/10.1371/journal.pmed.0040297

Wallace, S. J., Worrall, L., Rose, T. A., Alyahya, R. S. W., Babbitt, E., Beeke, S., de Beer, C., Bose, A., Bowen, A., Brady, M. C., Breitenstein, C., Bruehl, S., Bryant, L., Cheng, B. B. Y., Cherney, L. R., Conroy, P., Copland, D. A., Croteau, C., Cruice, M., … Le Dorze, G. (2023). Measuring communication as a core outcome in aphasia trials: Results of the ROMA-2 international core outcome set development meeting. *International Journal of Language & Communication Disorders*, *58*(4), 1017–1028. https://doi.org/10.1111/1460-6984.12840

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A. C.Marshall, J. . . . Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, *14*(2), 180–185. https://doi.org/10.1177/1747493018806200

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Kong, A. P. H., Simmons-Mackie, N., Scarinci, N., & Gauvreau, C. A. (2017). What outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation*, *39*(14), 1364–1379. https://doi.org/10.1080/09638288.2016.1194899

Winstein, C. J., Stein, J., Arena, R., Bates, B., Cherney, L. R., Cramer, S. C., Deruyter, F., Eng, J. J., Fisher, B., Harvey, R. L., Lang, C. E., MacKay-Lyons, M., Ottenbacher, K. J., Pugh, S., Reeves, M. J., Richards, L. G., Stiers, W., & Zorowitz, R. D. (2016). Guidelines for adult stroke rehabilitation and recovery. *Stroke*, *47*(6), e98–e169. https://doi.org/10.1161/STR.0000000000000098

World Health Organization. (2001). *International classification of functioning, disability and health (ICF)*.

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology*, *25*(3), 309–322. https://doi.org/10.1080/02687038.2010.508530