scientific reports

OPEN

Merge-based syntax is mediated by distinct neurocognitive mechanisms in 84,000 individuals with language deficits across nine languages

Elliot Murphy[1,2✉], Rohan Venkatesh[3], Edward Khokhlovich[3] & Andrey Vyshedskiy[4]

In the modern language sciences, the core computational operation of syntax, 'Merge', is defined as an operation that combines two linguistic units (e.g., 'brown', 'cat') to form a categorized constituent structure ('brown cat', a Noun Phrase). This structure can be further combined with additional linguistic units based on this categorial information, respecting non-associativity such that abstract grouping is preserved. Some linguists have embraced the view that Merge is an elementary, indivisible operation that emerged in a single evolutionary step. From a neurocognitive standpoint, different mental objects constructed by Merge may be supported by distinct tiers of processing demands: (1) simple command constructions (e.g., "eat apples"); (2) the merging of adjectives and nouns ("red boat"); and (3) the merging of nouns with spatial prepositions ("laptop behind the sofa"). Here, we systematically investigate participants' comprehension of sentences with increasing levels of syntactic complexity. Clustering analyses revealed behavioral evidence for three distinct structural types, which we discuss as potentially emerging at different developmental stages and subject to selective impairment. While a Merge-based syntax may have emerged suddenly in evolutionary time, responsible for the structured symbolic turn our species took, different processing tiers seem to underwrite the comprehension of various types of Merge-based objects.

In modern linguistics, Merge is a proprietary cognitive operation that combines two linguistic units (e.g., 'blue', 'cat') to form a categorized, labeled set ('blue cat', a Noun Phrase), which can then be further combined with additional linguistic units[1,2]. The field of linguistics has largely coalesced around the hypothesis that unbounded Merge is a uniquely human ability, contrasting with the limited linguistic abilities of modern large language models[3–6], and serving as the generative engine underlying the capacity to generate or infer an infinite number of expressions[7]. Some researchers have noted that due to the 'absolute' nature of Merge (one either has it, or does not have it), no partial or 'proto-Merge' is plausible: "The elementary operation of binary set formation (Merge) appeared in a single step"[8]. From a neurocognitive standpoint, it may be plausible to decompose distinct generic cognitive operations that subserve distinct components of a Merge-based syntax (e.g., the formation of hierarchical objects, the categorization of phrases, the short-term maintenance of categorial identity)[9,10]. From the perspective of linguistic formalism, the simplicity of the Merge operation is strikingly elegant. Its concise formulation echoes (intentionally) the parsimony of Hamilton's principle of least action and fundamental physical laws[11]. But from a neurocognitive standpoint, it is becoming increasingly plausible that different levels of Merge-based complexity may be supported by distinct nodes in the extended language network[12–19].

Studies involving large numbers of participants with language deficits offer a powerful approach to investigating the processing components of Merge-based syntactic knowledge. By including tens of thousands of individuals with a range of conditions associated with language impairments, researchers can observe the effect of a wide variety of genetic abnormalities linked to language deficits. Merge interfaces with various conceptual

[1]Department of Neurosurgery, UTHealth, Houston, TX, USA. [2]Texas Institute for Restorative Neurotechnologies, UTHealth, Houston, TX, USA. [3]Independent Researcher, Newton Centre, USA. [4]Boston University, Metropolitan College, Boston, MA, USA. ✉email: elliot.murphy@uth.tmc.edu

and performance systems, and if some of these exhibit digital/discrete representations but others exhibit analog/graded representations[20] we may expect to find genetic abnormalities that cause a gradual degradation of syntactic and semantic performance. Conversely, if Merge outputs to a single restricted and homogeneous domain-general interface, its loss should be catastrophic and complete.

Yet, previous studies have shown that neither of these two hypotheses is strongly supported. Two studies of tens of thousands of individuals identified three distinct general tiers of what a Merge-based syntax can output: Syntactic, Modifier, and Command[21,22]. The Syntactic Phenotype is exhibited by most adults and by children aged four years and older[23]. Individuals with this phenotype can comprehend sentences containing spatial prepositions, reversible word order, verb tenses, possessive pronouns, complex explanations, and elaborate fairytales demanding an array of these construction types (Table 1, *Syntactic Mechanism*). Approximately 2% of adults, as well as children between the ages of three and four, exhibit the Modifier Phenotype. Their ability is limited to combinations of nouns and adjectives (e.g., they can select 'a small green pencil' from a set of pencils, straws, and Lego pieces of different sizes and colors; Table 1, *Modifier Mechanism*). Finally, about 1% of adults, and children aged two to three years, exhibit the Command Phenotype, being restricted to simple commands (e.g., 'eat apple'; Table 1, *Command Mechanism*). Note that these categories delimit overt linguistic performance and major aspects of competence, but may not reflect the full extent of underlying competence.

These findings suggest that while Merge-based syntax can degrade in a seemingly catastrophic manner, it does not vanish entirely. For example, individuals functioning at the intermediate Modifier level lose the ability to comprehend fairytales, spatial prepositions, verb tense, and possessive pronouns, but they retain the capacity to integrate adjectives with nouns, form noun phrases with superlatives, and perform similar tasks over minimal compositional schemes. This dovetails with the observation that even in the most severe cases of lesion-based syntactic disruption, some capacity to execute basic hierarchical structure-formation is often preserved (in some format)[24].

Languages differ significantly in their grammatical structures, word order, word order flexibility, and morphological complexity. For instance, English typically follows a subject–verb–object word order, whereas languages like Japanese and Korean follow a subject–object–verb structure. In English, adjectives are generally pre-nominal—one would say "*the large cat*", not "*the cat large*". In contrast, Romance languages often place adjectives after the nouns they modify. Morphological complexity also varies widely: Russian, for example, features an extensive system of inflectional case endings, in contrast to the relatively simple morphology of English. This greater morphological richness in Russian contributes to its more flexible word order.

If the Command, Modifier, and Syntactic tiers differ across languages, we might be able to associate these grammatical differences with corresponding variations in them. Conversely, if no such differences are observed, it would suggest that these three tiers—Command, Modifier, and Syntactic—may be universal across languages.

Within this context of open questions, the goals of the present study were: (1) to relate the Command, Modifier, and Syntactic tiers to Merge-based syntax; (2) to analyze a larger cohort of participants than previous

		Language comprehension items (verbatim)	Abbreviations used in dendrograms
Command Mechanism	1	Knows own name	Knows Name
	2	Responds to 'No' or 'Stop'	No and Stop
	3	Responds to praise	Resp. to Praise
	4	Can follow some commands	Commands
Modifier Mechanism	5	Understands some simple modifiers (i.e., green apple vs. red apple or big apple vs. small apple)	Color or Size/Modifiers
	6	Understands several modifiers in a sentence (i.e., small green apple)	Two Modifiers
	7	Understands size (can select the largest/smallest object out of a collection of objects)	Size Superlatives
	8	Understands NUMBERS (i.e., two apples vs. three apples)	Numbers
Syntactic Mechanism	9	Understands spatial prepositions (i.e., put the apple ON TOP of the box vs. INSIDE the box vs. BEHIND the box)	Sp. Prepositions
	10	Understands verb tenses (i.e., I will eat an apple vs. I ate an apple)	Verb Tenses
	11	Understands simple stories that are read aloud	Simple Stories
	12	Understands elaborate fairytales that are read aloud (i.e., stories describing FANTASY creatures)	Elab. Fairytales
	13	Understands possessive pronouns (i.e., your apple vs. her apple)	Poss. Pronouns
	14	Understands the change in meaning when the order of words is changed (i.e., understands the difference between 'a cat ate a mouse' vs. 'a mouse ate a cat')	Flexible Syntax
	15	Understands explanations about people, objects or situations beyond the immediate surroundings (e.g., "Mom is walking the dog," "The snow has turned to water").	Explanations

Table 1. Three Language comprehension mechanisms—Syntactic, Modifier, and Command—have been identified in previous studies[21,22]. When one mechanism is acquired, the entire range of associated comprehension abilities is also gained. The Command-level abilities (Items 1 to 4) are acquired first. The Modifier-level abilities (Items 5 to 8) are attained next. The Syntactic-level abilities (Items 9 to 15) are acquired last. The Language comprehension items are presented exactly as surveyed with parents in both this and earlier studies. Response options were: very true, somewhat true, and not true. Items 1 to 3 were assessed as part of the expressive Language ATEC[30] subscale 1; the rest of items were a part of the MSEC subscale[31].

work; and (3) to confirm the dissociation among the three tiers within each of the nine languages analyzed individually.

Methods

Study participants

Participants were children and adolescents using a language therapy app that was made freely available at all major app stores (Apple App Store, Google Play Store, and Amazon App Store) in September 2015[25–29]. The app provides various structured language comprehension therapy exercises and is primarily used by caregivers of children with language impairments. Most of the caregivers are presumed to be parents. Once the app was downloaded, caregivers were asked to register and provide demographic details, including the child's diagnosis and age. Caregivers consented to pseudonymized data analysis and completed a 133-item questionnaire (77-item Autism Treatment Evaluation Checklist (ATEC)[30], Supplementary Tables 1–4; 20-item Mental Synthesis Evaluation Checklist (MSEC)[31], Supplementary Table 5; 10-item screen time checklist[32]; 25-item diet checklist[33]; and 1-item parent education survey) approximately every three months.

In what follows, we reproduce much of the Methods details from one of our previous publications[21].

All fifteen available *language comprehension* items from the 133-item questionnaire were included in the cluster analysis as in previously published articles[21,22] (Table 1). Answer choices were as follows: very true (0 points), somewhat true (1 point), and not true (2 points). A lower score indicates better language comprehension ability.

The inclusion criteria for this study remained consistent with those of previous studies[21,22]: absence of seizures (which commonly result in intermittent, unstable language comprehension deficits[34], absence of serious and moderate sleep problems (which are also associated with intermittent, unstable language comprehension deficits[35], age range of 4 to 22 years (the lower age cutoff was chosen to ensure that participants were exposed to complete sets of sentence structures listed in Table 1[36]; the upper age cutoff was chosen to avoid analysis of participants who may be linguistically declining due to aging). Previous studies were limited to individuals diagnosed with Autism Spectrum Disorder (ASD)[22] and individuals with fluid speech[21]. This study included all participants who submitted their assessments through the app, speaking one of the nine languages that the app is available in: English, Spanish, Portuguese, Italian, Russian, Chinese, French, German, and Korean. Table 2 reports participants' demographics in each language group as communicated by caregivers. Males outnumber females by approximately four to one, reflecting the predominance of autism among participants and its known male-to-female prevalence ratio.

Table 3 reports participants' diagnoses as communicated by caregivers. Autism level (mild/Level 1, moderate/Level 2, or severe/Level 3) was reported by caregivers. Pervasive Developmental Disorder and Asperger Syndrome were combined with mild autism for analysis as recommended by DSM-5[37]. A good reliability of such parent-reported diagnosis has been previously demonstrated[38]. We note that we do not have information on whether participants were diagnosed with or without language impairment.

When caregivers have completed several evaluations, the last evaluation was used for analysis as in previous studies[21,22]. Thus, the study included a total of 84,099 participants, the average age was 6.5 ± 2.7 years (range of 4 to 21.9 years), 74.7% participants were males. The education level of participants' parents was the following: 90.9% with at least a high school diploma, 68.6% with at least college education, 35.8% with at least a master's, and 5.6% with a doctorate. All caregivers consented to pseudonymized data analysis and publication of results. The study was conducted in compliance with the Declaration of Helsinki[39]. The study protocol was approved by the Biomedical Research Alliance of New York (BRANY) LLC Institutional Review Board (IRB). The data was accessed on July 9, 2025.

Statistics and reproducibility

Unsupervised Hierarchical Cluster Analysis (UHCA) is a data-driven approach that groups variables according to their similarity. This method generates tree-like structures, known as dendrograms, which illustrate the hierarchical relationships among clusters. For example, in cluster analysis of abilities, those that tend to co-occur are positioned in close proximity, whereas those that co-occur less frequently appear farther apart. UHCA was performed using Ward's agglomeration method with a Euclidean distance metric. The clustering analysis was

	Number of participants	Percent of total	Age, mean(SD)	Percent males
English	27,187	32.3	6.7(2.9)	75.8
Spanish	33,488	39.8	6.1(2.2)	68.3
Portuguese	7504	8.9	6.3(2.5)	76.3
Italian	6484	7.7	7.6 (3.2)	78.3
Russian	4778	5.7	6.8(2.5)	77.9
Chinese	2217	2.6	6.1(2.0)	79.2
French	1060	1.3	7.2(3.2)	72.5
German	927	1.1	7.2(3.2)	69.9
Korean	454	0.5	6.6(2.7)	74.2
Total	**84,099**	**100**	**6.5(2.7)**	**74.7**

Table 2. Participants' demographics.

	Number of participants	Percent of total	Age, mean(SD)	Percent males
Mild Autism Spectrum Disorder (ASD)	29,292	34.8	6.2(2.4)	75.8
Moderate ASD	19,094	22.7	6.9(2.8)	79.5
Severe ASD	10,839	12.9	7.6(3.3)	80.0
Not-diagnosed	9869	11.7	5.7(1.6)	54.2
Specific Language Impairment	4884	5.8	6.0(2.2)	69.2
Mild Language Delay	3973	4.7	5.4(1.7)	67.4
ADHD	1630	1.9	6.4(2.2)	72.8
Down Syndrome	1427	1.7	8.5(3.5)	60.4
Other Genetic Disorder	1104	1.3	8.0(3.5)	57.7
Social Communication Disorder	794	0.9	6.3(2.3)	69.1
Sensory Processing Disorder	727	0.9	6.8(2.7)	70.2
Apraxia	466	0.6	6.8(2.9)	64.6
Total	**84,099**	**100**	**6.5(2.7)**	**74.7**

Table 3. Participants' diagnoses.

data-driven without any design or hypothesis. A two-dimensional heatmap was generated using the "pheatmap" package of R, freely available language for statistical computing[40]. Code and data can be downloaded (https://doi.org/10.17605/OSF.IO/2QK5B).

Results
Clustering analysis of 15 Language comprehension abilities
Caregivers assessed 15 language comprehension abilities (Table 1). To examine patterns of co-occurrence among these abilities, we applied unsupervised hierarchical cluster analysis, a data-driven method that groups items based on their similarity. This technique produces dendrograms, which visually represent the hierarchical relationships between clusters of items. Abilities that frequently co-occur are positioned closely together, while those that co-occur less often appear farther apart.

Figure 1A depicts the dendrogram generated from the analysis of English-speaking participants. The height of the branches indicates the distance between clusters. A larger distance corresponds to greater dissimilarity between the clusters. Previous studies identified three clusters stable across different evaluation methods, age groups, time points, genders, and parental education[21,22]. The first cluster included *knowing the name*, *responding to 'No' or 'Stop'*, *responding to praise*, and *following some commands* (items 1 to 4 in Table 1) and was termed the Command Mechanism. The second cluster included *understanding color and size modifiers*, *several modifiers in a sentence*, *size superlatives*, and *numbers* (items 5 to 8 in Table 1) and was termed the Modifier Mechanism. The third cluster included *understanding of spatial prepositions*, *verb tenses*, *flexible syntax*, *possessive pronouns*, *explanations about people and situations*, *simple stories*, and *elaborate fairytales* (items 9 to 15 in Table 1) and was termed the Syntactic Mechanism. The analysis of English-speaking participants in Fig. 1A identified the same three clusters with inter-cluster distances that were significantly larger than the distances between subclusters. Principal component analysis (PCA) (Fig. 1B) also showed a clear separation between the three clusters: Command, Modifier and Syntactic.

As a control we calculated unsupervised hierarchical cluster analysis and PCA of the 15 comprehension abilities along with the "*hyperactivity*" (Figures S1), "*bed-wetting*" (Figure S2), and "*demands sameness*" (Figure S3) items. These items are not related to language and therefore should cluster into their own group. As expected, both unsupervised hierarchical cluster analysis and PCA clustered these items into their own group at a significant distance from the Command, Modifier, and Syntactic clusters, validating both clustering techniques.

Clustering analysis across spoken languages
Clustering analysis was conducted in all language groups with 400 or more participants. The three-cluster solution was consistent across all language groups explored: English, Spanish, Portuguese, Italian, Russian, Chinese, French, German, and Korean (Figs. 1, 2, 3, 4, 5, 6, 7, 8 and 9). In all spoken languages, unsupervised hierarchical cluster analysis sorted the 15 comprehension abilities into congruent three clusters (Command, Modifier, and Syntactic) and PCA showed a clear separation between the three clusters. Some language groups, such as Russian (Fig. 5B), demonstrated a greater separation between clusters in PCA.

These findings suggest that the three-cluster solution is not a result of differential cultural upbringing but rather a potentially general cognitive phenomenon constrained, consistent across spoken languages.

Language comprehension phenotypes in participants
Previous studies have employed unsupervised hierarchical cluster analysis to identify distinct language comprehension phenotypes of participants[21,22]. The principles underlying participant clustering are identical to those used for clustering abilities: participants with similar patterns of abilities are automatically organized into hierarchical dendrograms. Previous studies identified three distinct phenotypes: (1) Command Phenotype– participants who acquired only the Command Mechanism; (2) Modifier Phenotype–participants who acquired
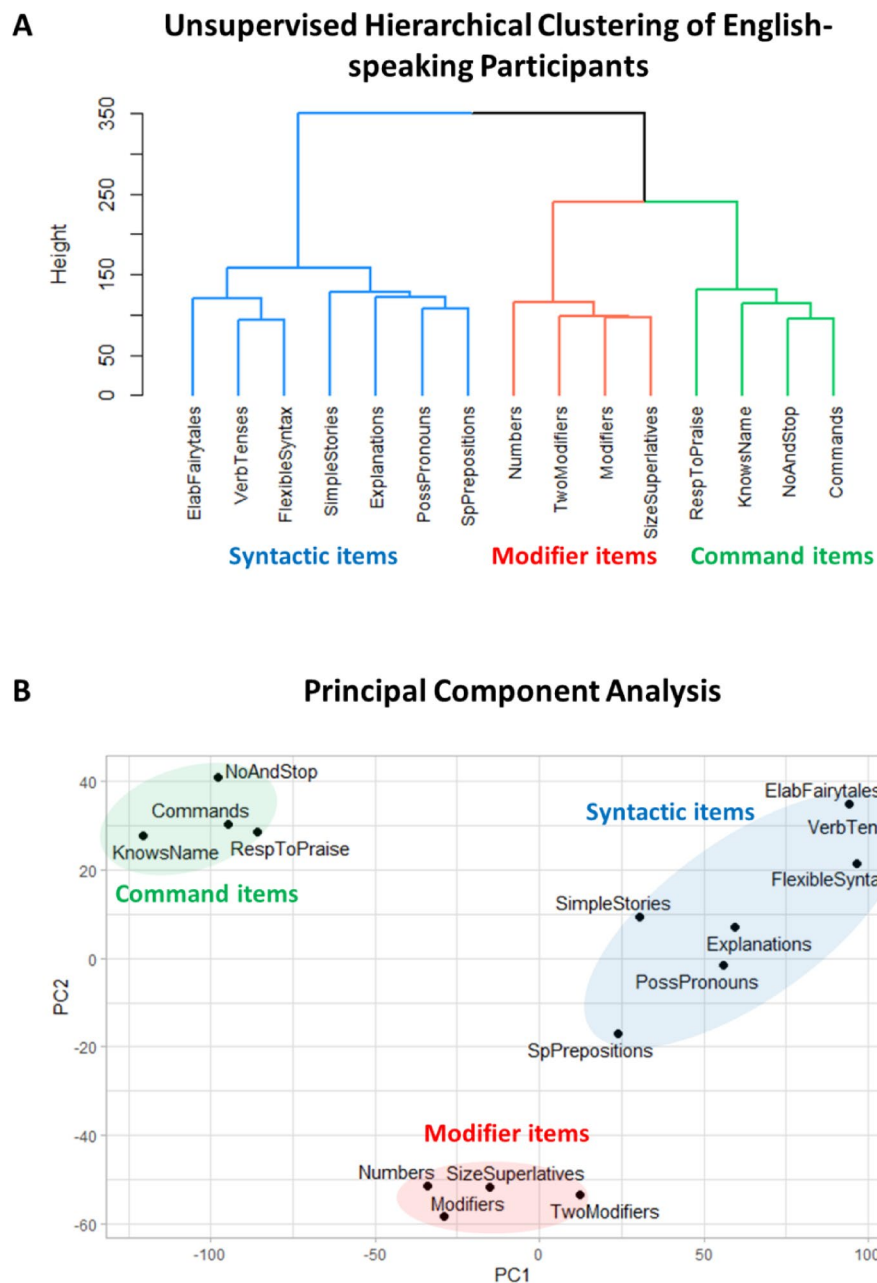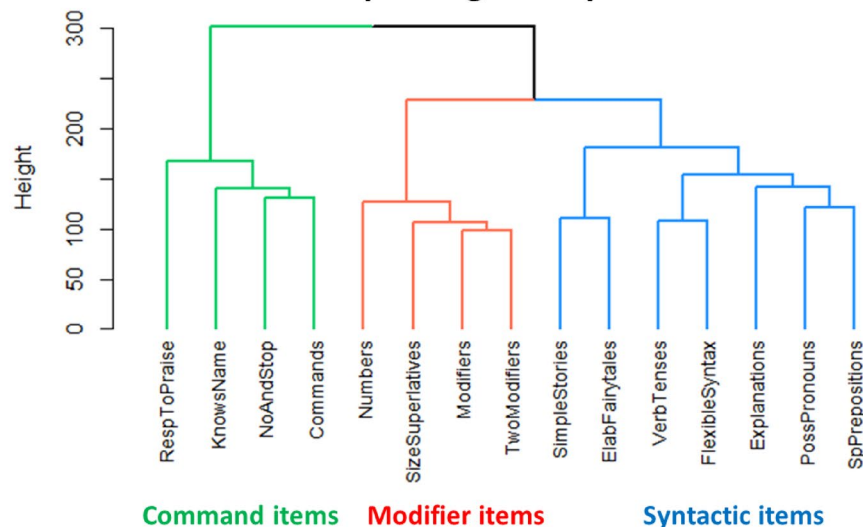
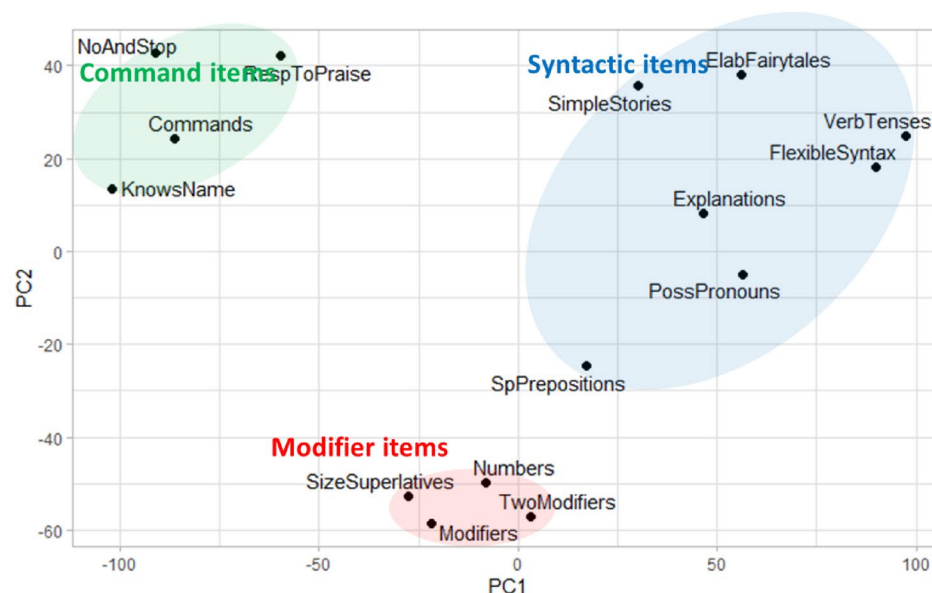**Fig. 1.** Clustering analysis of 15 comprehension items in English-speaking participants. **(A)** A dendrogram representing the unsupervised hierarchical clustering analysis (UHCA) of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 47% of the variance in the data. Principal component 2 accounts for 11.1% of the variance in the data.

both the Command and Modifier Mechanisms; and (3) Syntactic Phenotype–participants who acquired the Command, Modifier, and Syntactic Mechanisms.

The close correspondence between comprehension mechanisms and the resulting phenotypes is noteworthy. While various combinations of the three mechanisms are theoretically possible, such combinations were not observed empirically. For example, a hypothetical phenotype combining the Command and Syntactic Mechanisms (but not the Modifier Mechanism) could exist in theory. Another possibility would be a phenotype lacking all three mechanisms. However, these configurations did not emerge from the data. This absence suggests that the morphospace of comprehension phenotypes is constrained by cognitive faculties.

To investigate whether these constraints are consistent across different languages, this study conducted the unsupervised hierarchical cluster analysis of participants separately within each language group. The results are presented in Figs. 10, 11, 12, 13, 14, 15, 16, 17 and 18, which relate participant clusters to comprehension
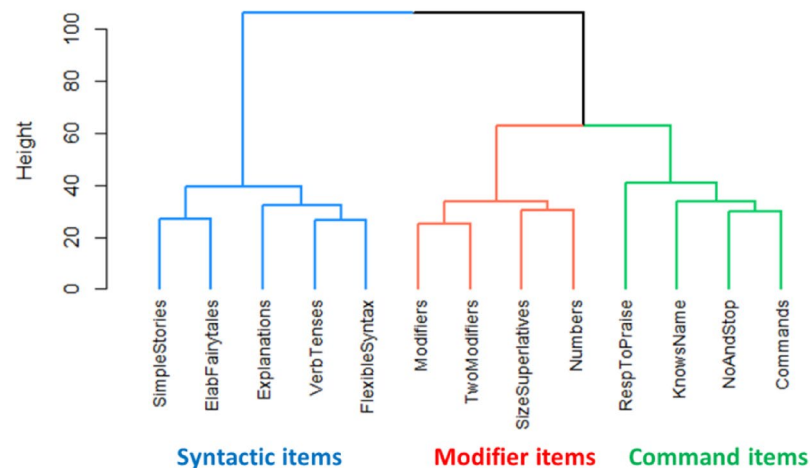
Fig. 2. Clustering analysis of 15 language comprehension items in Spanish-speaking participants. **(A)** A dendrogram representing the unsupervised hierarchical clustering analysis of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 32.4% of the variance in the data. Principal component 2 accounts for 11.5% of the variance in the data.

mechanisms. The three mechanism clusters (Command, Modifier, and Syntactic) are shown as rows (the dendrogram from Fig. 1A representing comprehension mechanisms is shown vertically on the left in Fig. 10) and the 27,187 English-speaking participants are shown as columns (the dendrogram representing participants is shown horizontally on the top). *Blue* indicates the presence of a linguistic ability (parent's response = *very true*); *white* indicates an intermittent presence of a linguistic ability (parent's response = *somewhat true*); and *red* indicates the complete lack of a linguistic ability (parent's response = *not true*).

In the heatmap of English-speakers (Fig. 10), the middle cluster of participants (marked "Syntactic Phenotype") shows the predominant blue color (representing good skills) across all abilities indicating that these
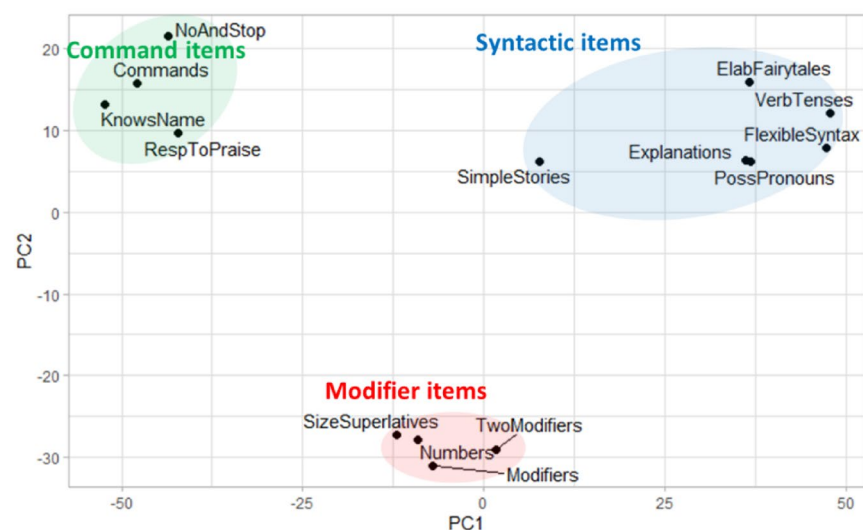
Fig. 3. Clustering analysis of 14 language comprehension items in Portuguese-speaking participants. One item ("spatial prepositions") was translated incorrectly and was therefore excluded from analysis. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 40.4% of the variance in the data. Principal component 2 accounts for 11.3% of the variance in the data.

participants acquired the Command, Modifier, and Syntactic Mechanisms. The leftmost cluster of participants (marked "Command Phenotype") shows the predominant blue color only among the Command Mechanism items and red colors across Syntactic and Modifier Mechanisms items, indicating that these individuals only acquired the Command Mechanism. The rightmost cluster of participants (marked "Modifier Phenotype") shows the predominant blue color only across Command and Modifier Mechanisms items and white to red colors across Syntactic Mechanism items, indicating that these individuals acquired the Command and Modifier Mechanisms.

This pattern was reproduced across all language groups (Figs. 10, 11, 12, 13, 14, 15, 16, 17 and 18). Participants acquired either: (1) the Command Mechanism alone (marked as the Command Phenotype), or (2) both the Command and Modifier Mechanisms (marked as the Modifier Phenotype), or (3) the Command, Modifier, and Syntactic Mechanisms (marked as the Syntactic Phenotype).
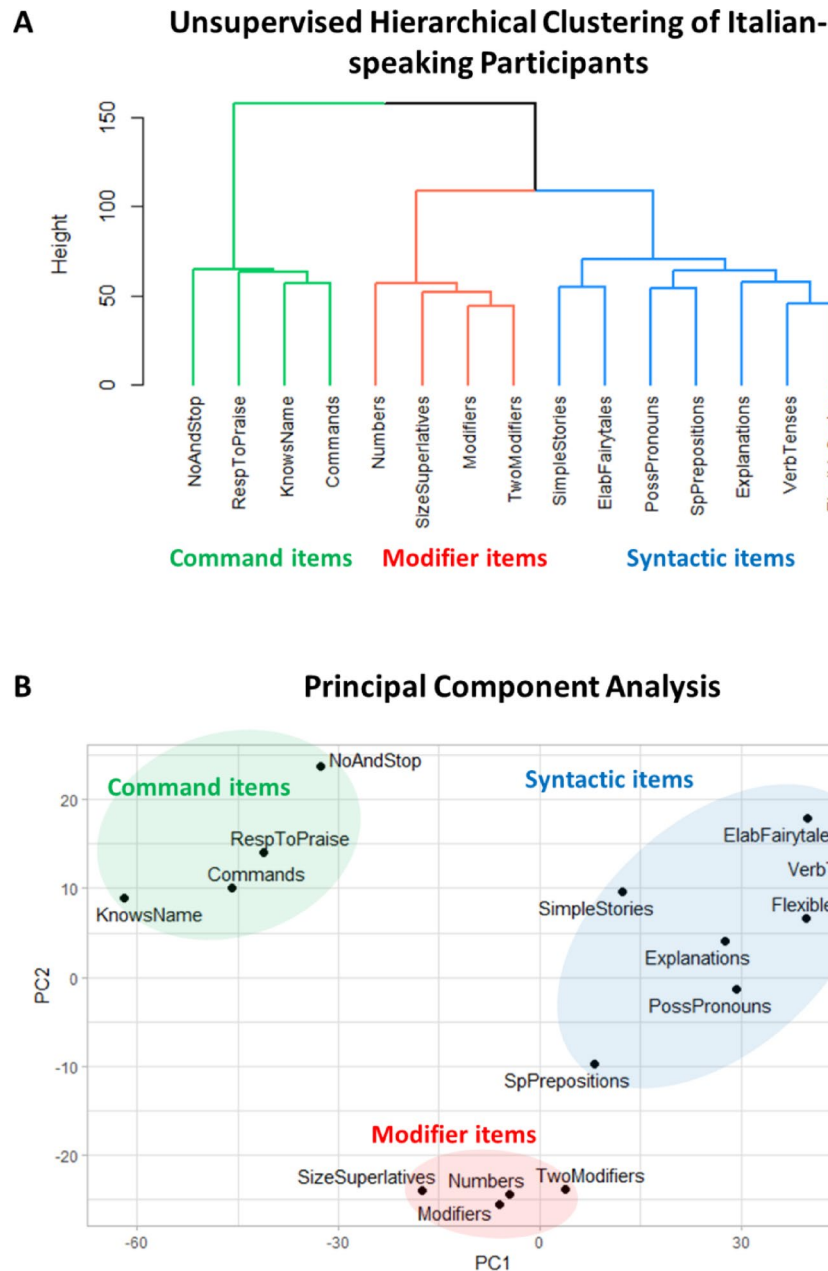
Fig. 4. Clustering analysis of 15 language comprehension items in Italian-speakers. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 43% of the variance in the data. Principal component 2 accounts for 10.8% of the variance in the data.

Discussion

We conducted a clustering analysis to examine the co-occurrence of fifteen language comprehension abilities in 84,099 individuals who spoke English, Spanish, Portuguese, Italian, Russian, Chinese, French, German, or Korean. The three identified clusters were identical between languages and congruent to those found in previous analyses[21,22]. Crucially, the clustering analysis in all studies was devoid of any design or hypothesis, as both unsupervised hierarchical clustering analysis and principal component analysis were entirely driven by the data. The outcome of our clustering analyses was a set of three coherent, discrete language ability clusters that appear to revolve around similar linguistic deficits concerning modifiers and complex syntactic operations—not a mixed amalgam of different patterns that would be expected if language abilities were mediated by many unrelated mechanisms.

The replication of the three-cluster solution across English, Spanish, Portuguese, Italian, Russian, Chinese, French, German, and Korean likely points to language-independent constraints. Notable, the languages
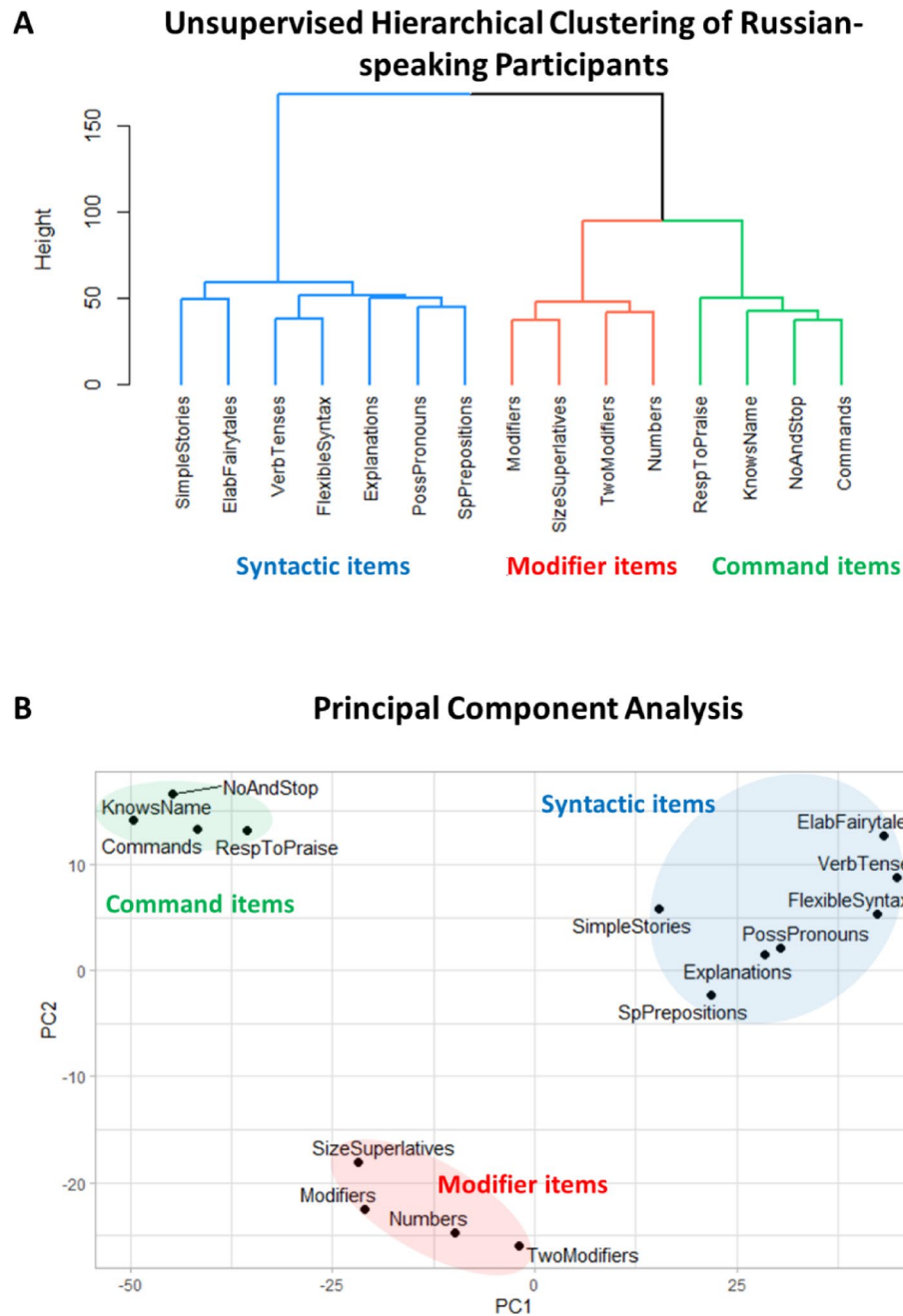
Fig. 5. Clustering analysis of 15 language comprehension items in Russian-speaking participants. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 52.7% of the variance in the data. Principal component 2 accounts for 10.3% of the variance in the data.

examined in this study vary widely in morphological complexity, word order and word order flexibility. For example, Russian has a much richer morphological system and greater word order flexibility than English; while Korean follows subject–object–verb structure, which differs from the subject–verb–object order typical of Indo-European languages. Despite these structural differences, all nine languages consistently revealed clear distinctions among the three tiers—Command, Modifier, and Syntactic—indicating that these distinctions may be universal.

Some may argue here that the three clusters reflect emergent probabilistic constructions rather than discrete cognitive mechanisms. However, the sharp boundaries we observe (especially the near-absence of Modifier-without-Command or Syntactic-without-Modifier profiles) are difficult to reconcile with a purely continuous competence model. Future work could operationalize the predictions of usage-based linguistic theories to more carefully explore these topics.
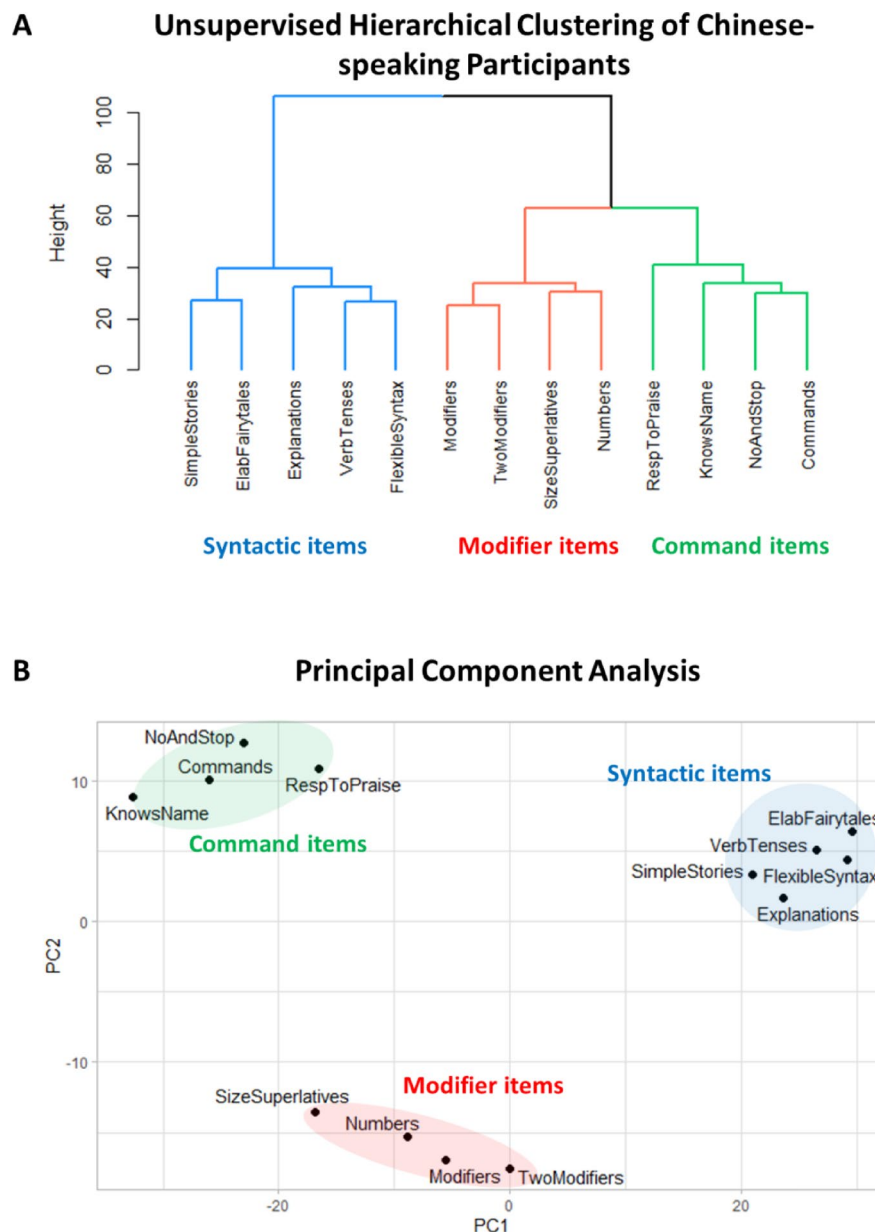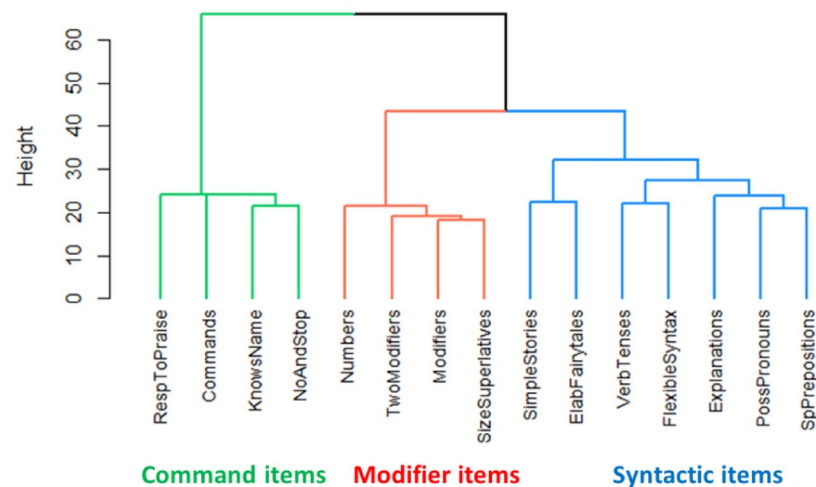
Fig. 6. Clustering analysis of 13 language comprehension items in Chinese-speaking participants. Two items ("spatial prepositions" and "possessive pronouns") were translated incorrectly and were therefore excluded from analysis. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 49.7% of the variance in the data. Principal component 2 accounts for 12.3% of the variance in the data.

The present behavioral gradient (Command → Modifier → Syntactic) is compatible with the hypothesis that a single explanatory construct, Merge, can be developmentally unpacked into sub-routines mastered at different developmental stages: the Command Mechanism by 1.6 years of age, the Modifier mechanism by 3.0 years of age, and the Syntactic Mechanism by 3.7 years of age[23]. Speculating for a moment, from an evolutionary perspective this aligns with a "saltation-plus-scaffolding" model: an initial binary structure-forming capacity may have arisen abruptly[8], but its efficient deployment in real-time cognition and communication required incremental recruitment of domain-general resources such as working memory, attention[14,41] and articulate speech[42]. It is possible that the layered neurocognitive architecture that supports modern human syntax and which breaks down in cases of language deficits[43–46] may provide some basis for the behavioral results we document here.

Our results reveal three distinct levels of language comprehension. When individuals with the Command Phenotype progress to the Modifier Phenotype, they consistently acquire the full range of Modifier-level abilities, such as combining size, color, number, and superlatives with nouns. Similarly, those advancing from the
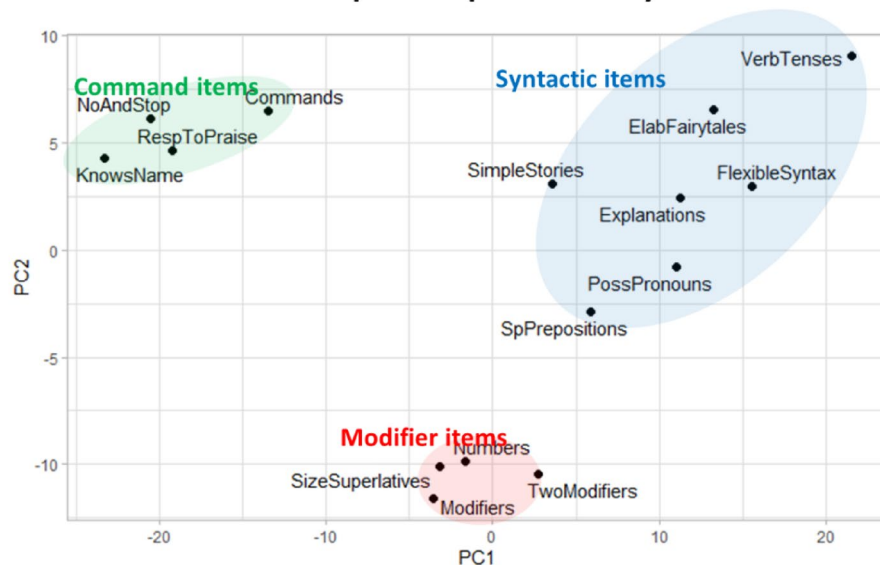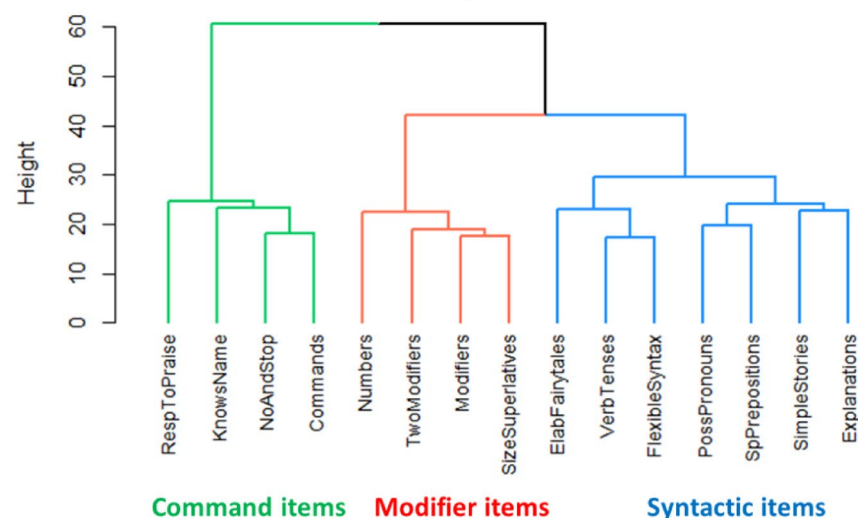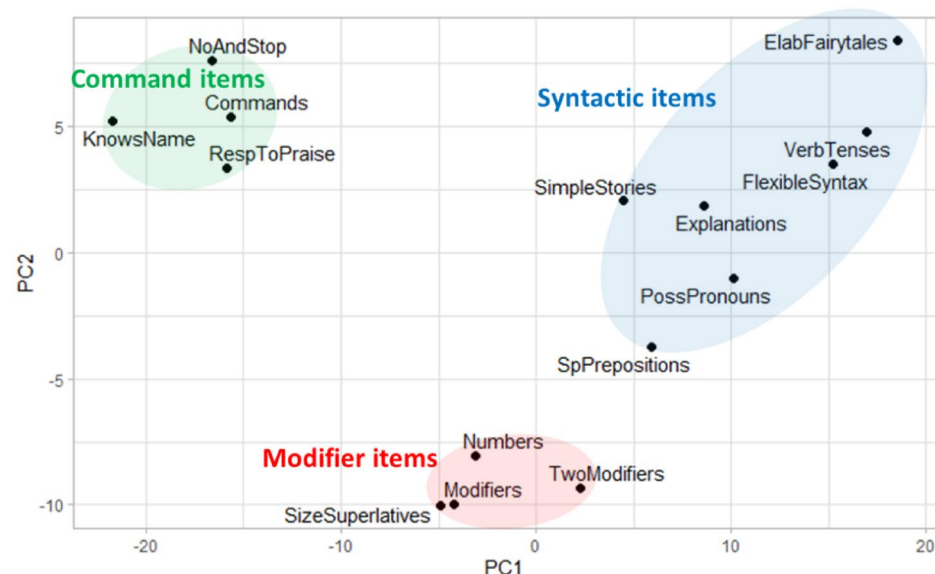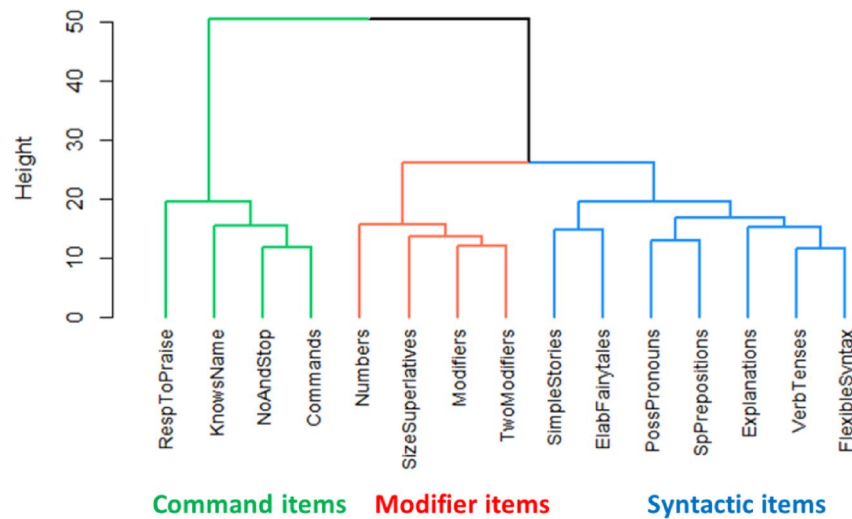
Fig. 7. Clustering analysis of 15 language comprehension items in French-speaking participants. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 42.6% of the variance in the data. Principal component 2 accounts for 11.3% of the variance in the data.

Modifier to the Syntactic Phenotype reliably acquire the complete set of Syntactic-level abilities, including spatial prepositions, possessive pronouns, understanding of verb tenses, and the capacity to comprehend stories and fairy tales. Crucially, no intermediate or "half-step" transitions have been observed. If such partial progressions were possible, we would expect to find clusters of individuals exhibiting distinct intermediate phenotypes. However, each transition appears to follow an all-or-nothing principle[8].

The three tiers we outline here are also to be thought of as developmental steps, or behavioral phenotypes of comprehension that operationalize specific algorithmic demands entailed by a Merge-based grammar. Our results reveal behavioral stratification of comprehension abilities that are entailed by a Merge-based grammar under algorithmic and implementational constraints. We also wish to note in this connection that while Merge-based syntax is a strictly computational-level formal framing, recent developments in neurolinguistics are increasingly pointing to a feasible set of neural mechanisms and causal structures that might enforce the algebraic properties of syntax[15], potentially allowing for a more precise assessment of our three linguistic mechanisms and
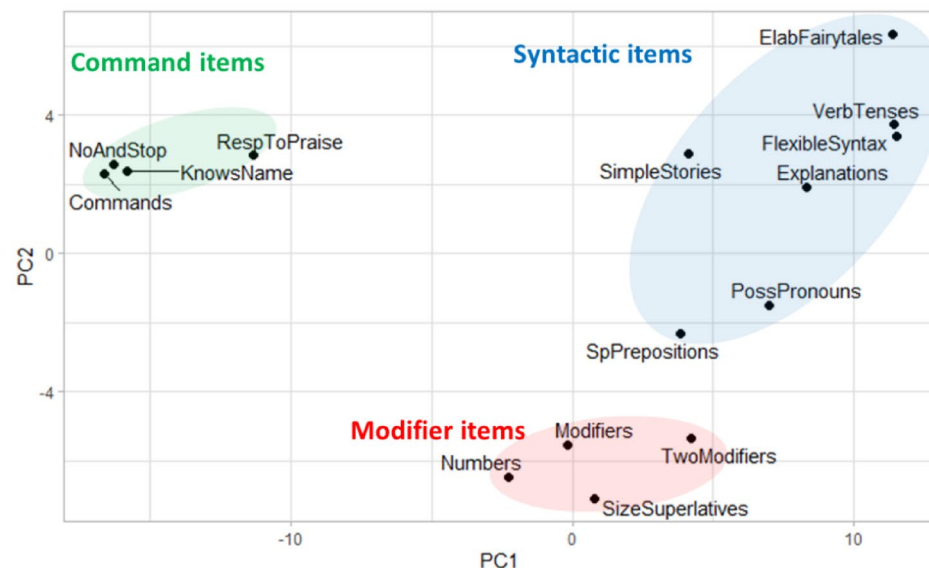
Fig. 8. Clustering analysis of 15 language comprehension items in German-speaking participants. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 42.4% of the variance in the data. Principal component 2 accounts for 10.7% of the variance in the data.

their neural bases. In other words, it is possible to construct predictions for the neural infrastructure of our Merge-based mechanisms based on the neural code outlined in these works[14,15,47]. Future work could explore these questions in more detail than what the current scope of this article allows.

Isolating behavioral dynamics of Command, Modifier, and Syntactic mechanisms might allow us to refine long-standing psycholinguistic debates about the grain-size of syntactic representations accessed during real-time comprehension. For example, classic garden-path effects show that comprehenders incrementally commit to local syntactic analyses that sometimes require costly reanalysis when later input forces revision. If the Modifier mechanism licenses phrase-internal operations such as adjective-noun union, while the Command mechanism governs clausal argument structure, then garden-path costs should be sharply magnified whenever disambiguation pivots on Command-level information (e.g., NP-attachment vs. VP-attachment ambiguities).

Fig. 9. Clustering analysis of 15 language comprehension items in Korean-speaking participants. **(A)** A dendrogram representing the hierarchical clustering of language comprehension abilities. **(B)** Principal component analysis (PCA) plot, where ovals highlight clusters identified by UHCA. The PCA reveals a distinct separation among Command, Modifier and Syntactic Mechanisms. Principal component 1 accounts for 49.5% of the variance in the data. Principal component 2 accounts for 8.7% of the variance in the data.

Conversely, ambiguities resolvable within the Modifier mechanism (e.g., prenominal adjective stacking) should yield milder slow-downs. The three comprehension types may also be grounded in varying levels of integrity and development of key frontoparietal and frontotemporal white matter tracts[48]. For example, individuals with limited to no abilities with the Syntactic tier often exhibit weaker fronto-parieto-temporal tracts[49,50].

We note here some limitations of our work: A 133-item survey completed repeatedly by motivated parents is invaluable, but still prone to optimism, fatigue, and socioeconomic skew. Future work could strengthen the validity of findings by quantifying inter-rater reliability (e.g., by having both parents independently complete the survey). While one study has demonstrated a strong correlation between the parent-reported survey used in this study and a clinicians-administered Language Phenotype Assessment ($r = 0.78$, $p < 0.0001$)[23], further validation could involve comparing a subsample of parent responses with results from standardized clinician-
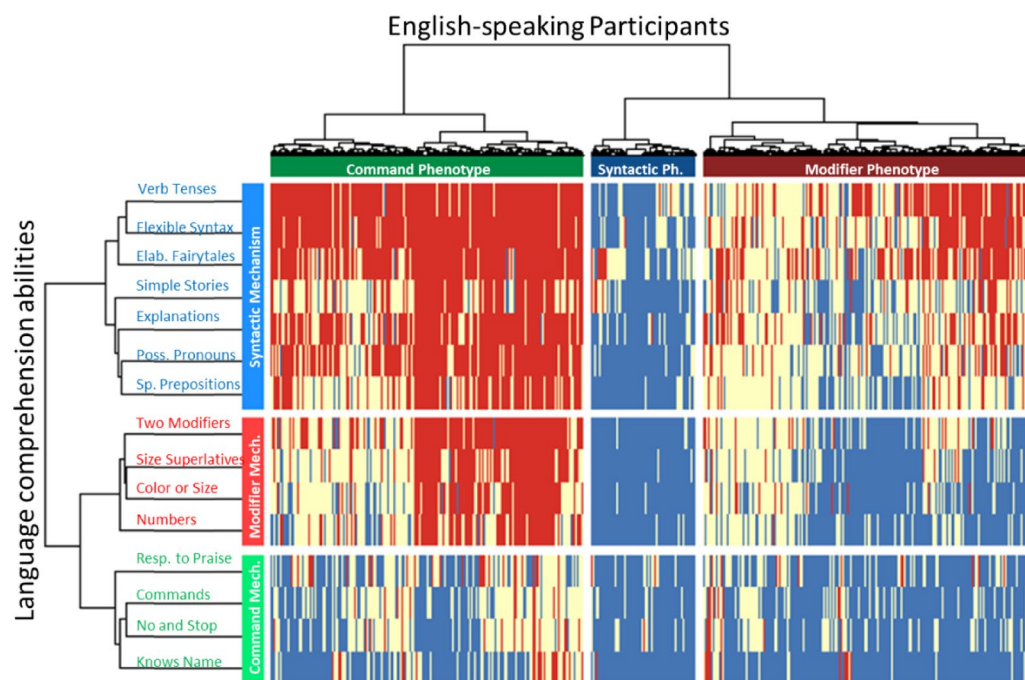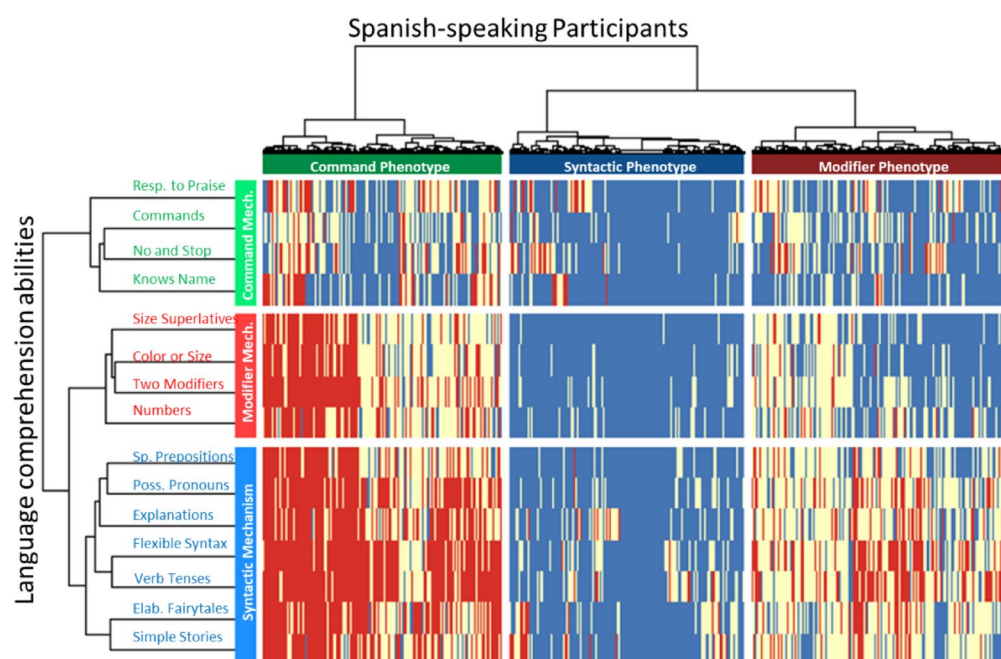
Fig. 10. Two-dimensional heatmap relating English-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 27,187 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.

Fig. 11. Two-dimensional heatmap relating Spanish-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 33,488 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.
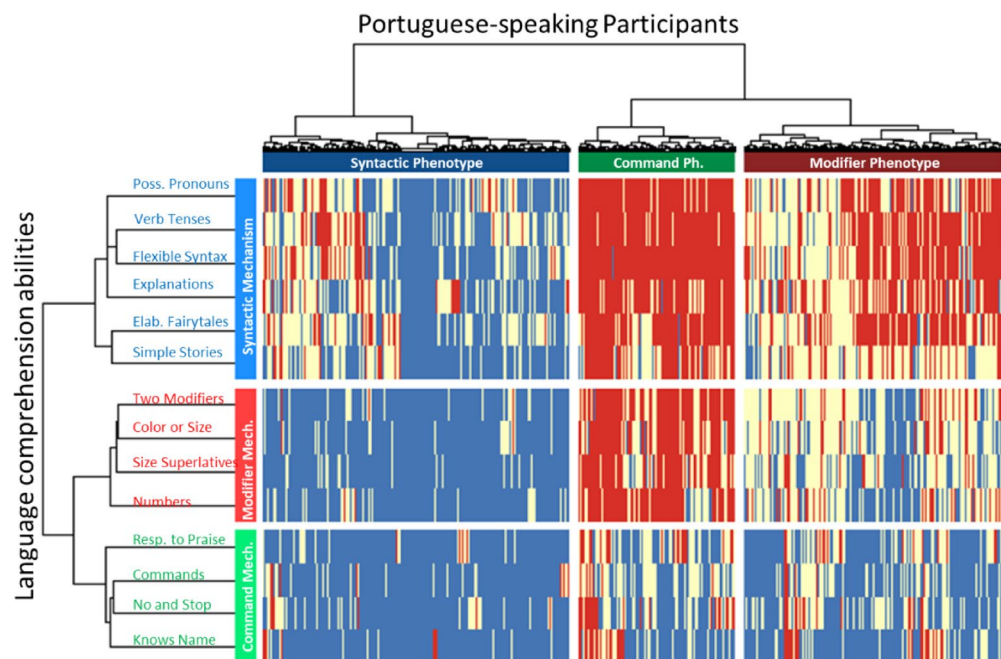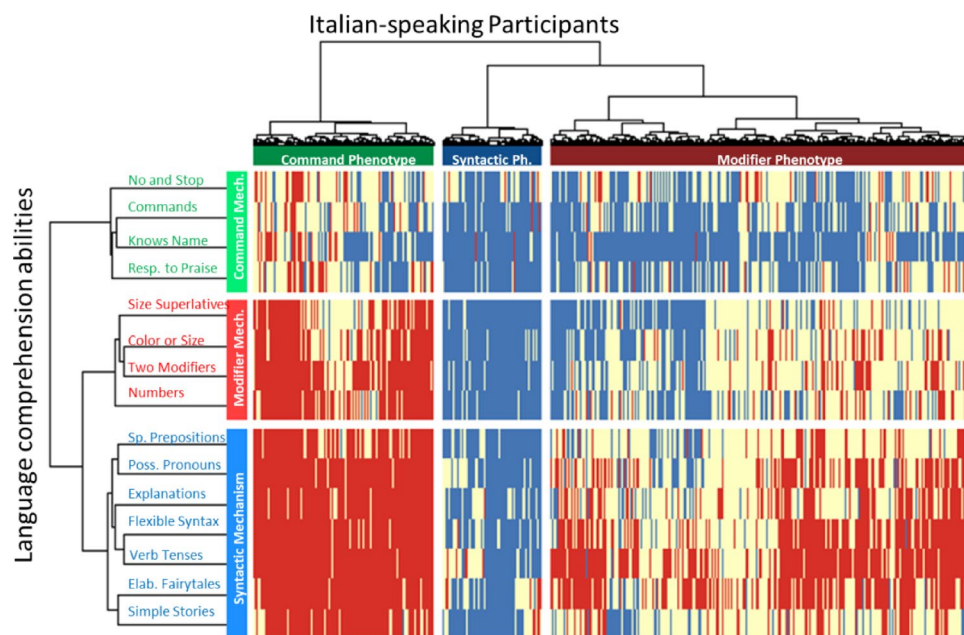
Fig. 12. Two-dimensional heatmap relating Portuguese-speaking participants to their language comprehension abilities. The 14 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 7,504 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.
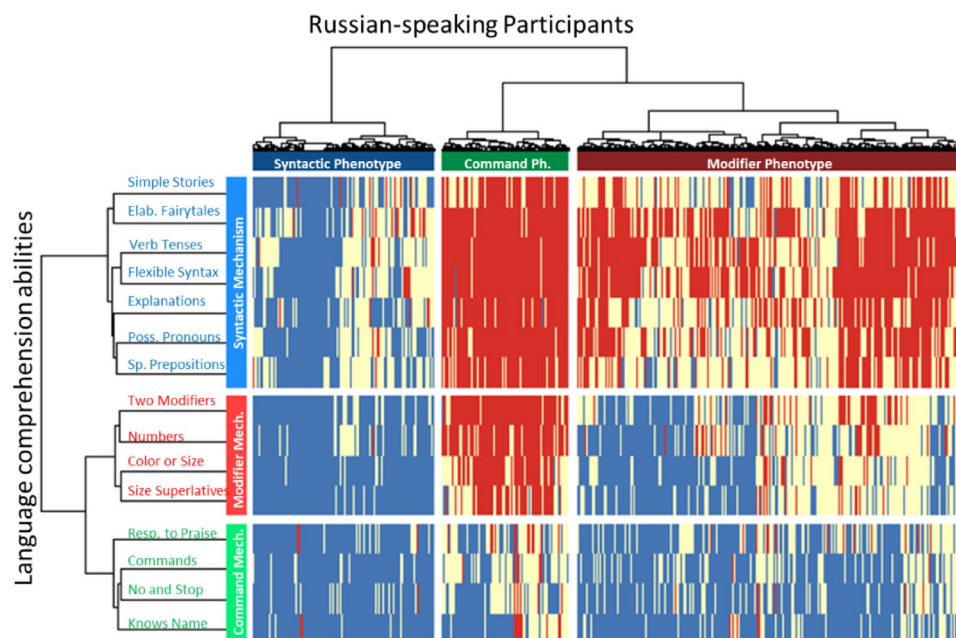

Fig. 13. Two-dimensional heatmap relating Italian-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 6,484 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.

Fig. 14. Two-dimensional heatmap relating Russian-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 4,778 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.
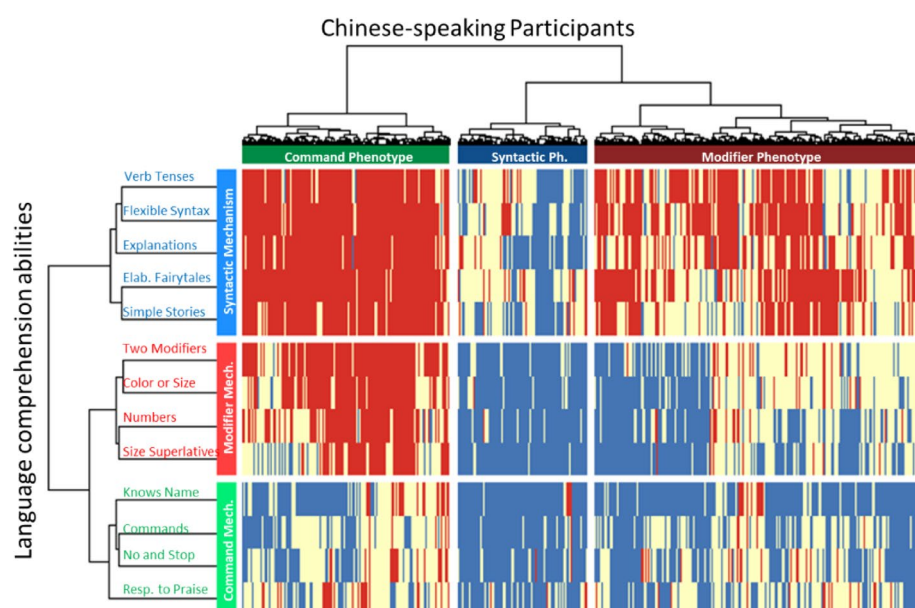

Fig. 15. Two-dimensional heatmap relating Chinese-speaking participants to their language comprehension abilities. The 13 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 2,217 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.
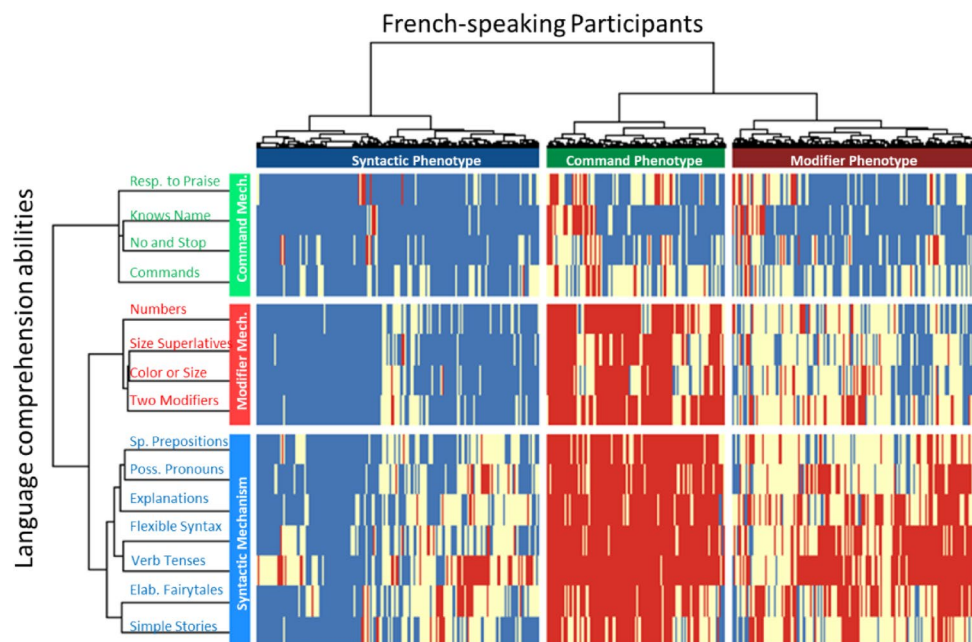
Fig. 16. Two-dimensional heatmap relating French-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 1,060 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.
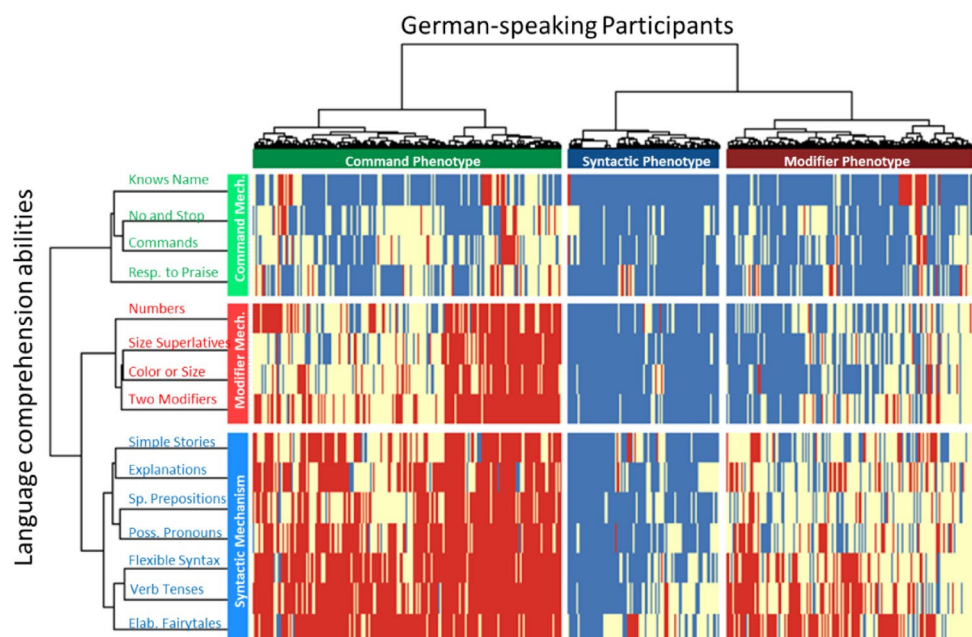

Fig. 17. Two-dimensional heatmap relating German-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 927 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.
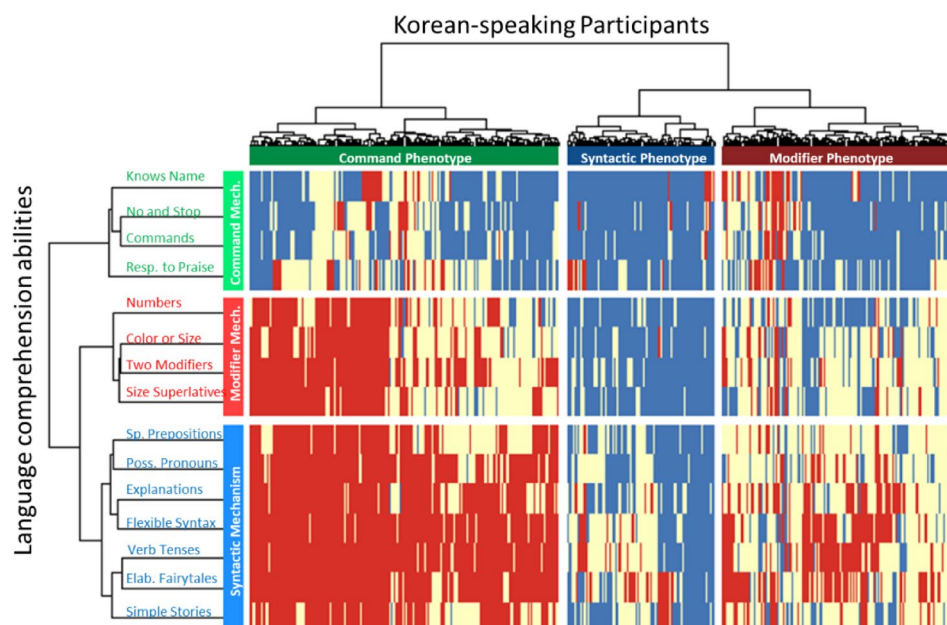
Fig. 18. Two-dimensional heatmap relating Korean-speaking participants to their language comprehension abilities. The 15 language comprehension abilities are shown as rows. The dendrogram representing language comprehension abilities is shown on the left. Participants are shown as 454 columns. The dendrogram representing participants is shown on the top. Blue color indicates the presence of a linguistic ability (the "very true" answer), red indicates the lack of a linguistic ability (the "not true" answer), and white-yellow indicates the "somewhat true" answer.

administered instruments (e.g., PLS-5[51], Token Test[52,53], CELF-5[54], TROG[55]. Our enrollment protocol also filters for relatively "tech-savvy", intervention-seeking parents and may under-represent low-SES households. A comparison of census-matched demographics would help establish external validity. In addition, the disorders reported in our study differ in etiology and linguistic phenotype; since the same caregiver questionnaire supplies both phenotype and explanatory variable, latent correlations may inflate cluster separability.

Overall, our results suggest that the conceptual and performance systems that access Merge-based syntax are fractionated behaviorally into distinct tiers. We hope that these results play some role in addressing a long-standing gap between formal linguistic theory and large-scale behavioral phenotyping. Our reported sample is an order of magnitude larger than typical language-impairment studies, improving power to detect stable substructures, but many further questions remain concerning the granularity of the documented sets of language deficits.

Data availability

De-identified raw data from this manuscript are available from the corresponding author upon reasonable request.

Code availability

Code is available from the corresponding author upon reasonable request.

References

1. Chomsky, N. *The Minimalist Program* (MIT Press. Camb. MA) (1995).
2. Chomsky, N. Minimalist inquiries: The framework. *Step Step Essays Minimalist Syntax Honor Howard Lasnik* 89–155 (2000).
3. Dentella, V., Günther, F., Murphy, E., Marcus, G. & Leivada, E. Testing AI on Language comprehension tasks reveals insensitivity to underlying meaning. *Sci. Rep.* **14**, 28083 (2024).
4. Leivada, E., Murphy, E. & Marcus, G. DALL·E 2 fails to reliably capture common syntactic processes. *Soc. Sci. Humanit. Open.* **8**, 100648 (2023).
5. Murphy, E., De Villiers, J. & Morales, S. L. A comparative investigation of compositional syntax and semantics in DALL·E and young children. *Soc. Sci. Humanit. Open.* **11**, 101332 (2025).
6. Murphy, E. & Leivada, E. A model for learning strings is not a model of language. *Proc. Natl. Acad. Sci.* **119**, e2201651119 (2022).
7. Tanaka, K. et al. Merge-generability as the key concept of human language: evidence from neuroscience. *Front. Psychol.* **10**, 2673 (2019).
8. Berwick, R. C. & Chomsky, N. All or nothing: no half-Merge and the evolution of syntax. *PLoS Biol.* **17**, e3000539 (2019).
9. Murphy, E. Labels, cognomes, and Cyclic computation: an ethological perspective. *Front. Psychol.* **6**, 715 (2015).

10. Yang, Q., Murphy, E., Yang, C., Liao, Y. & Hu, J. Neural mechanisms of structural inference: an EEG investigation of linguistic phrase structure categorization. *Preprint at.* https://doi.org/10.1101/2025.07.01.662085 (2025).
11. Murphy, E., Holmes, E. & Friston, K. Natural Language syntax complies with the free-energy principle. *Synthese* **203**, 154 (2024).
12. McCarty, M. J. et al. Intraoperative cortical localization of music and Language reveals signatures of structural complexity in posterior Temporal cortex. *iScience* **26**, 107223 (2023).
13. Murphy, E. *The Oscillatory Nature of Language* (Cambridge University Press, 2020). https://doi.org/10.1017/9781108864466
14. Murphy, E. ROSE: A neurocomputational architecture for syntax. *J. Neurolinguistics*. **70**, 101180 (2024).
15. Murphy, E. R. O. S. E. A universal neural grammar. *Cogn. Neurosci.* 1–32. https://doi.org/10.1080/17588928.2025.2523875 (2025).
16. Murphy, E. et al. Minimal phrase composition revealed by intracranial recordings. *J. Neurosci.* **42**, 3216–3227 (2022).
17. Murphy, E. et al. The Spatiotemporal dynamics of semantic integration in the human brain. *Nat Commun* **14**, 6336 (2023).
18. Murphy, E., Rollo, P. S., Segaert, K., Hagoort, P. & Tandon, N. Multiple dimensions of syntactic structure are resolved earliest in posterior Temporal cortex. *Prog Neurobiol.* **241**, 102669 (2024).
19. Murphy, E. & Woolnough, O. The Language network is topographically diverse and driven by rapid syntactic inferences. *Nat. Rev. Neurosci.* **25**, 705–705 (2024).
20. Jackendoff, R. S. & Erk, K. E. Toward a Deeper Lexical Semantics. *Top. Cogn. Sci.* tops.70013 (2025). https://doi.org/10.1111/tops.70013
21. Vyshedskiy, A., Venkatesh, R., Khokhlovich, E. & Satik, D. Three mechanisms of Language comprehension are revealed through cluster analysis of individuals with Language deficits. *Npj Sci. Learn.* **9**, 1–12 (2024).
22. Vyshedskiy, A., Venkatesh, R. & Khokhlovich, E. Are there distinct levels of Language comprehension in autistic individuals – cluster analysis. *Npj Ment Health Res* **3**, 19 (2024).
23. Vyshedskiy, A. et al. Language comprehension developmental milestones in typically developing children assessed by the new Language phenotype assessment (LPA). *Child. Basel Switz.* **12**, 793 (2025).
24. Dragoy, O., Akinina, Y. & Dronkers, N. Toward a functional neuroanatomy of semantic aphasia: A history and ten new cases. *Cortex* **97**, 164–182 (2017).
25. Vyshedskiy, A. & Dunn, R. Mental imagery therapy for autism (MITA)-An early intervention computerized brain training program for children with ASD. *Autism Open. Access.* **5**, 2 (2015).
26. Dunn, R. et al. Comparison of performance on verbal and nonverbal multiple-cue responding tasks in children with ASD. *Autism Open. Access.* **7**, 218 (2017).
27. Dunn, R. et al. Tablet-Based cognitive exercises as an early Parent-Administered intervention tool for toddlers with Autism - Evidence from a field study. *Clin Psychiatry* **3**, 1 (2017).
28. Dunn, R. et al. Children with autism appear to benefit from Parent-Administered computerized cognitive and Language exercises independent of the child's age or autism severity. *Autism Open. Access* **7**, 5 (2017).
29. Vyshedskiy, A. et al. Novel prefrontal synthesis intervention improves Language in children with autism. *Healthcare* **8**, 566 (2020).
30. Rimland, B. & Edelson, S. M. Autism treatment evaluation checklist (ATEC). *Autism Res. Inst. San Diego CA* (1999). http://www.autism.com
31. Braverman, J., Dunn, R. & Vyshedskiy, A. Development of the mental synthesis evaluation checklist (MSEC): A Parent-Report tool for mental synthesis ability assessment in children with Language delay. *Children* **5**, 62 (2018).
32. Fridberg, E., Khokhlovich, E. & Vyshedskiy, A. Watching Videos and Television Is Related to a Lower Development of Complex Language Comprehension in Young Children with Autism. in *Healthcare* vol. 9 423Multidisciplinary Digital Publishing Institute, (2021).
33. Acosta, A., Khokhlovich, E., Reis, H. & Vyshedskiy, A. Dietary factors impact developmental trajectories in young autistic children. *J. Autism Dev. Disord.* https://doi.org/10.1007/s10803-023-06074-8 (2023).
34. Forman, P., Khokhlovich, E. & Vyshedskiy, A. Longitudinal developmental trajectories in young autistic children presenting with Seizures, compared to those presenting without Seizures, gathered via Parent-report using a mobile application. *J. Dev. Phys. Disabil.* https://doi.org/10.1007/s10882-022-09851-y (2022).
35. Levin, J., Khokhlovich, E. & Vyshedskiy, A. Longitudinal developmental trajectories in young autistic children presenting with sleep problems, compared to those presenting without sleep problems, gathered via parent-report using a mobile application. *Res. Autism Spectr. Disord.* **97**, 102024 (2022).
36. Arnold, M. & Vyshedskiy, A. Combinatorial Language parent-report score differs significantly between typically developing children and those with autism spectrum disorders. *J. Autism Dev. Disord.* https://doi.org/10.1007/s10803-022-05769-8 (2022).
37. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5˚)* (American Psychiatric Pub, 2013).
38. Jagadeesan, P., Kabbani, A. & Vyshedskiy, A. Parent-reported assessment scores reflect ASD severity level in 2- to 7- year-old children. *Children* **9**, 701 (2022).
39. World Medical Association. World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191–2194 (2013).
40. R Foundation for Statistical Computing. R: A language and environment for statistical computing. (2021).
41. Murphy, E. No country for Oldowan men: emerging factors in Language evolution. *Front Psychol* **10**, 1448 (2019).
42. Vyshedskiy, A. Language evolution is not limited to speech acquisition: a large study of Language development in children with Language deficits highlights the importance of the voluntary imagination component of Language. *Res. Ideas Outcomes.* **8**, e86401 (2022).
43. Benítez-Burraco, A., Hoshi, K. & Murphy, E. Language deficits in GRIN2A mutations and Landau–Kleffner syndrome as neural dysrhythmias. *J. Neurolinguistics.* **67**, 101139 (2023).
44. Benítez-Burraco, A. & Murphy, E. The oscillopathic nature of Language deficits in autism: from genes to Language evolution. *Front Hum. Neurosci* **10**, 120 (2016).
45. Benítez-Burraco, A. & Murphy, E. Why brain oscillations are improving our Understanding of Language. *Front. Behav. Neurosci.* **13**, 190 (2019).
46. Murphy, E. & Benítez-Burraco, A. Language deficits in schizophrenia and autism as related oscillatory connectomopathies: an evolutionary account. *Neurosci. Biobehav Rev.* **83**, 742–764 (2017).
47. Van Der Burght, C. L. et al. Cleaning up the brickyard: how theory and methodology shape experiments in cognitive neuroscience of Language. *J. Cogn. Neurosci.* **35**, 2067–2088 (2023).
48. Friederici, A. D. & Becker, Y. The core Language network separated from other networks during primate evolution. *Nat. Rev. Neurosci.* **26**, 131–132 (2025).
49. Cheng, Q., Roth, A., Halgren, E. & Mayberry, R. I. Effects of early Language deprivation on brain connectivity: Language pathways in deaf native and late First-Language learners of American sign Language. *Front. Hum. Neurosci.* **13**, 320 (2019).
50. McFayden, T. C. et al. White matter development and Language abilities during infancy in autism spectrum disorder. *Mol. Psychiatry.* **29**, 2095–2104 (2024).
51. Zimmerman, I. L., Steiner, V. G. & Pond, R. E. PLS-5: preschool Language scale-5 [measurement instrument]. *San Antonio TX Psychol. Corp* (2011).
52. De Renzi, E. & Faglioni, P. Normative data and screening power of a shortened version of the token test. *Cortex* **14**, 41–49 (1978).
53. De Renzi, A. & Vignolo, L. A. Token test: A sensitive test to detect receptive disturbances in aphasics. *Brain J. Neurol* **85**, 665–678 (1962).

54. Wiig, E. H., Secord, W. A. & Semel, E. *Clinical Evaluation of Language Fundamentals: CELF-5* (Pearson, 2013).
55. Bishop, D. V. TROG 2: test for reception of grammar-version 2. *Ed Giunti OS Firenze* (2009).

Author contributions

AV designed the study. EK developed the statistical paradigm. RV wrote the statistical analysis software. AV analyzed the data. EM and AV wrote the paper.

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

The study was conducted in compliance with the Declaration of Helsinki. Informed consent was obtained from the caregivers of all participants. The study protocol was approved by the Biomedical Research Alliance of New York (BRANY) LLC Institutional Review Board (IRB).

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-28008-3.

Correspondence and requests for materials should be addressed to E.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.