Three clusters of grammatical abilities in individuals with language deficits “split” the universal grammar hypothesis

Andrey Vyshedskiy (vysha@bu.edu)
Boston University https://orcid.org/0000-0002-1610-8716
Rohan Venkatesh
ImagiRation
Edward Khokhlovich
Independent researcher

Article

Keywords: recursive language, grammatical language, syntactic language, language uniqueness, first-language acquisition

Posted Date: November 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3181937/v1

Additional Declarations: (Not answered)

Abstract

In this study we assessed 15 language comprehension abilities in 55,558 individuals with language deficits 4 to 22 years of age using parent-generated reports. Data-driven cluster analysis identified three distinct clusters of co-expressed abilities. The first cluster, termed "command language," included knowing the name, responding to 'No' or 'Stop', responding to praise, and following simple commands. The second cluster, termed "modifier language," included understanding color and size modifiers, several modifiers in a sentence, size superlatives, and numbers. The third cluster, termed the "prepositional language," included understanding of spatial prepositions, verb tenses, flexible syntax, possessive pronouns, explanations about people and situations, simple stories, and elaborate fairy tales. Independently, we performed data-driven cluster analysis of participants. 19.8% of participants were limited to command language phenotype; 58% of participants were limited to modifier language phenotype; 22.2% of participants were clustered to the most-advanced prepositional phenotype. All neurotypical participants were clustered to the prepositional phenotype. While the Universal Grammar hypothesis attempts to preach a dichotomy between the presence of grammatical abilities in humans and their absence in animals, this study shows that there are additional distinct language comprehension phenotypes that require an explanation.

Introduction

The acquisition and use of language is a unique and defining characteristic of the human species. From a young age, children demonstrate an astounding ability to learn language, grasping complex grammatical structures and communicating effectively within a relatively short period. The question of how this remarkable feat is accomplished has intrigued scholars across various disciplines, leading to the development of numerous theories. One of the most influential theories in the field of linguistics is Noam Chomsky's theory of Universal Grammar [1]. Chomsky proposed that humans possess an innate mechanism that enables them to acquire and understand language effortlessly. Universal Grammar is characterized by a set of universal principles and parameters that define the fundamental properties of all human languages, while allowing for variation across different linguistic systems.

Some individuals, though, do not acquire full grammatical understanding. This can happen as a result of culture deprivation during early childhood or a genetic defect. An example of the former is illustrated by deaf linguistic isolates. About 90% of all congenitally deaf children are born to hearing parents [2]. In developed countries, these children typically are introduced to a formal sign language early and acquire full grammatical language. In countries without widespread testing for deafness and rigorous support network, however, deaf children living in remote villages may never be exposed to a formal sign language. To communicate, families usually spontaneously invent homesign (a.k.a. kitchensign), a system of iconic gestures that consists of simple signs. As a result, despite children's upbringing in a loving family, they are deprived of exposure to complex grammar. Invariably, these individuals do not acquire full grammatical language comprehension and no amount of postpubertal language therapy can reverse the effect of early childhood grammatical deprivation [3–8]. The simple act of withdrawing exposure to grammatical sentences can profoundly prevent the physical structure of the brain from typical development [9].

Genetic abnormalities can also preclude individuals from acquiring full grammatical comprehension. For example, the prevalence of this lifelong condition among individuals diagnosed with autism (that is widely considered to have genetic etiology [10–12]), is 30 to 40% [13] and can be as high as 60% among children enrolled into special ASD schools [14].

Accordingly, both aberrant nurture and nature can result in a similar impairment of grammatical language comprehension phenotype. However, the structure of this phenotype and its relationship to the Universal Grammar hypothesis are poorly understood. In 2014 we developed a language therapy app for children with language delay [15–19]. In order to receive the access to free structured lessons contained within this app, parents are asked to complete a 133-question survey once every quarter. Over ten years we accumulated more than 200,000 longitudinal assessments from children and adolescents with language deficits. This large database provides an opportunity to investigate grammatical abilities acquisition using cluster analysis. Accordingly, we extracted all 15 language-comprehension-related abilities polled by the 133-question parent survey (Table 1) and conducted data-driven cluster analysis of these abilities.

Table 1. Language comprehension items as they were posed to parents. The answers choices were: very true (0 points), somewhat true (1 point), and not true (2 points). Items 1 to 3 were assessed as part of the expressive language ATEC subscale 1; the rest of items were part of the MSEC subscale. A lower score indicates better language comprehension ability.

	Language comprehension items (verbatim)	Abbreviations used in dendrograms
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
5	Understands some simple modifiers (i.e., green apple vs. red apple or big apple vs. small apple)	Color or Size
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
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 fairy tales that are read aloud (i.e., stories describing FANTASY creatures)	Elab. Fairy tales
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

Methods

Study Participants

Participants were children and adolescents using a language therapy app that was made available gratis at all major app stores in September 2015 [15–19]. Once the app was downloaded, caregivers were asked to register and to provide demographic details, including the child’s diagnosis and age. Caregivers consented to anonymized data analysis and completed a 133-item questionnaire approximately every three months. Inclusion criteria for this study were as follows: absence of seizures (that commonly result in intermittent, unstable language comprehension deficits [20]), absence of serious and moderate sleep problems (that are also associated with intermittent, unstable language comprehension deficits [21]), age range of 4 to 21 years (the lower age cutoff was chosen to ensure that participants were exposed to all variety of items listed in Table 1 [22]; the upper age cutoff was chosen to avoid analysis of participants who may be linguistically declining due to aging). When caregivers have completed several evaluations, the last evaluation was used for analysis. Table 2 reports participants' diagnoses as reported by caregivers. A good reliability of such parent-reported diagnosis has been previously demonstrated [23]. As a control we have added 48 neurotypical children 4 to 8 years of age, whose data were previously collected for a different study [22] by approaching parents on a community online site and asking if they would be willing to complete a Google form. Thus, the study included the total of 55,558 participants, the average age was 6.4±2.7 years (range of 4 to 21 years), 76% 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 anonymized data analysis and publication of the results.

Statistics and Reproducibility

All fifteen available *language comprehension* items from the 133-item questionnaire were included in the cluster analysis (Table 1). Unsupervised Hierarchical Cluster Analysis (UHCA) was performed using Ward's agglomeration method with a Euclidean distance metric. Two-dimensional heatmap was generated using the "pheatmap" package of R, freely available language for statistical computing.

Table 2. Participants' diagnoses as reported by caregivers.

	Number of Participants	Percent of Total	Age, Mean(SD)	Percent Males
Neurotypical	48	0.1	6.3(1.3)	47.9
Not-diagnosed	8096	14.6	5.7(2.0)	68.2
Mild Language Delay	2211	4.0	5.3(1.6)	69.6
Apraxia	244	0.4	6.5(2.8)	68.0
Specific Language Impairment	2587	4.7	5.9(2.2)	70.7
Sensory Processing Disorder	416	0.7	6.5(2.6)	72.1
Social Communication Disorder	393	0.7	6.2(2.4)	73.3
Mild Autism Spectrum Disorder (ASD)	18144	32.7	6.2(2.5)	77.8
Moderate ASD	12841	23.1	6.8(2.8)	80.6
Severe ASD	8431	15.2	7.3(3.3)	80.7
Down Syndrome	864	1.6	8.3(3.5)	60.1
Other Genetic Disorder	569	1.0	7.9(3.6)	58.7
ADHD	714	1.3	6.4(2.4)	75.2
Total	**55558**	**100**	**6.4(2.7)**	**76.2**

Results

Clustering analysis of 15 language comprehension abilities

Parents and caregivers evaluated 15 language comprehension abilities (Table 1). To explore their co-occurrence, we used two common clustering methods: Unsupervised Hierarchical Cluster Analysis (UHCA) and Principal Component Analysis (PCA). Figure 1A depicts the dendrogram generated by UHCA. The height of the branches indicates the distance between clusters, which is an indicator of greater dissimilarity. Three clusters have inter-cluster distances that are significantly larger than the distances between subclusters. The right-most cluster includes knowing the name, responding to ‘No’ or ‘Stop’, responding to praise, and following some commands (items 1 to 4 in Table 1). This cluster was termed “command language.” The cluster in the middle includes understanding color and size modifiers, several modifiers in a sentence, size superlatives, and numbers (items 5 to 8 in Table 1). It was termed “modifier language.” The left-most cluster includes understanding of spatial prepositions, verb tenses, flexible syntax, possessive pronouns, explanations about people and situations, simple stories, and elaborate fairy tales (items 9 to 15 in Table 1). It was termed the “prepositional language.”

The PCA (Fig. 1B) also shows clear separation between the three clusters. The command items (knowing the name, responding to 'No' or 'Stop', responding to praise, and following some commands) are clustered in the top left corner. The modifier items (understanding color and size modifiers, several modifiers in a sentence, size superlatives, and numbers) are clustered in the lower middle. The prepositional items (understanding of spatial prepositions, verb tenses, flexible syntax, possessive pronouns, explanations about people and situations, simple stories, and elaborate fairy tales) are clustered in the top right corner.

The three-cluster solution was stable across multiple seeds as well as consistent across different evaluation methods (Ward.D2 method, Figure S1A; Ward.D method, Figure S1B; Average method, Figure S1C; Complete method, Figure S1D; Mcquitty method, Figure S1E), across different age groups (Figures S2 and S3, the dendrogram on the left), and across different time points (first evaluation, Figure S4; last evaluation, Fig. 1A).

Language comprehension phenotypes in participants

Independently from clustering 15 language comprehension abilities, we utilized UHCA to cluster all 55,558 participants, Fig. 2A. The distances between three clusters were significantly larger than those between subclusters, indicating that the three-cluster solution is the most favorite choice. The PCA also showed reasonable separation between the three participant clusters (Fig. 2B). This three-cluster solution was stable across different seeds, age groups (Figures S2 and S3), time points (first and last evaluations; Figures S4 and 2A, respectively), and evaluation methods (Euclidean, Manhattan, and Minkowski distance metrics; Figs. 3, S5, and S6, respectively).

A two-dimensional heatmap was used to relate participant clusters to language comprehension abilities clusters (Fig. 3). Columns represent the 55,558 participants and rows represent the 15 linguistic abilities. *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 lack of a linguistic ability (parent's response = *not true*). The three clusters of participants match the three clusters of linguistic abilities.

The cluster of participants termed "Prepositional Language Phenotype" shows the predominant blue color that indicated good skills across all three clusters of language comprehension abilities (22.2% of participants, Table 3). The cluster of participants marked as "Command Language Phenotype" shows the predominant blue color only among the command items and red colors across prepositional and modifier items (19.8%). The third cluster of participants marked "Modifier Language Cluster" shows the predominant blue color only across command and modifier items and white to red colors across prepositional items (58%).

Table 4 shows cluster assignment of participants by their diagnostic category. As expected, all neurotypical children were clustered into the prepositional cluster. Undiagnosed participants and participants with milder conditions have the highest proportion clustered to the prepositional language phenotype (46.6%) and the lowest proportion clustered to the command language phenotype (11.2%).

This observation is reversed among participants diagnosed with more severe conditions. For example, participants with severe autism have the lowest proportion clustered to the prepositional language phenotype (4.2%) and the highest proportion clustered to the command language phenotype (41.3%).

Table 3
Participant cluster statistics

	Prepositional Language Phenotype	Modifier Language Phenotype	Command Language Phenotype	Total
Number of participants	12363	32212	10983	55558
Percent of Total	22.2	58.0	19.8	100.0
Age, Mean(SD)	6.1(2.3)	6.5(2.8)	6.7(3.0)	6.4(2.7)
Percent Male	68.3	75.1	76.2	73.8

Table 4
Percentage of participants in the three language comprehension phenotypes.

	Prepositional Language Phenotype	Modifier Language Phenotype	Command Language Phenotype
Neurotypical	100	0	0
Not-diagnosed	46.6	42.2	11.2
Mild Language Delay	44.2	48.1	7.7
Apraxia	31.6	56.6	11.9
Specific Language Impairment	31.3	56.7	11.9
Sensory Processing Disorder	25.0	57.9	17.1
Social Communication Disorder	24.2	61.8	14.0
Mild ASD	24.5	62.8	12.7
Moderate ASD	9.4	65.1	25.5
Severe ASD	4.2	54.6	41.3
Down Syndrome	14.8	64.5	20.7
Other Genetic Disorder	19.2	61.0	19.9
ADHD	35.6	52.2	12.2

Discussion

The study analyzes 15 language comprehension abilities in 55,558 participants with language deficits. A three-cluster solution was consistent across a range of ages as well as parameters controlling the Unsupervised Hierarchical Cluster Analysis (UHCA). The first cluster included four abilities: comprehension of one's name, responding to 'No' or 'Stop,' responding to praise, and following simple commands (Fig. 1A); this cluster of concurrently acquired linguistic abilities was termed "command language." The second cluster included four abilities: comprehension of simple color/size modifiers, understanding of several modifiers in a sentence, understanding of size superlatives, and number comprehension; this cluster of simultaneously acquired abilities was termed "modifier language." The third cluster included the remaining 7 abilities: comprehension of spatial prepositions, verb tenses, flexible syntax, possessive pronouns, explanations, simple stories, and elaborate fairy tales; this cluster of related abilities was termed "prepositional language." Reassuringly, principal component analysis (Figs. 1B) corroborated the three-cluster solution.

Independently, UHCA assigned 55,558 participants to clusters. The three-cluster solution was consistent across a range of diagnosis, ages as well as parameters controlling the UHCA (Fig. 2). The two-dimensional heatmap analysis related participants' clusters (Fig. 3, the dendrogram shown on the top) to language comprehension abilities clusters (the dendrogram shown on the left). Participants in the prepositional language phenotype cluster acquired most language comprehension abilities tested by the 15 items (indicated by the predominant blue color across all items). Participants in the modifier language phenotype cluster acquired most command and modifier language abilities (indicated by the blue color) and did not acquire most prepositional language abilities (indicated by the red and white colors). Participants in the command language phenotype cluster acquired command language abilities (indicated by the blue color) and did not acquire prepositional and modifier language abilities (indicated by the predominant red color).

One explanation for differences in language comprehension could be differential exposure of participants to linguistic concepts. For example, if participants were never exposed to numbers, they would not understand the concept of numbers; if they were never exposed to spatial prepositions, they would not understand spatial prepositions. However, by four years of age (the lower cutoff in our study) participants were exposed to all variety of items listed in Table 1 [22]. Furthermore, limiting cluster analysis to older participants 6 to 21 years of age also demonstrated the same three language comprehension phenotype clusters (Figure S3). Therefore, differential exposure cannot explain differences in language comprehension. In the absence of exposure differences, differences in comprehension can only be explained by some common language mechanisms. For some reason, individuals in the modifier language phenotype were not able to acquire the mechanism underlying prepositional skills and individuals in the command language phenotype were not able to acquire the mechanism underlying modifier and prepositional skills. Taken together with cluster analysis of linguistic abilities (Fig. 1), these results suggest a single mechanism behind comprehension of spatial prepositions, verb tenses, flexible syntax, possessive pronouns, explanations, simple stories, and elaborate fairy tales; a different mechanism behind color, size, and number modifiers comprehension; and still a different mechanism behind commands comprehension. In other words, once a single mechanism is acquired, the whole cluster of linguistic abilities is manifested.

Limitations

Epidemiological studies leveraging app users as participants offer access to a substantial number of individuals, but they do have some evident drawbacks, such as reliance on parental reports. On one hand, parents may be prone to wishful thinking and may overestimate their children's abilities [24]. On the other hand, parents possess a deep understanding of their children, which can be particularly valuable for assessing grammatical language skills that can be challenging to evaluate in a clinical setting. Furthermore, several previous studies have indicated that parent reports of language abilities align closely with direct assessments conducted by clinicians [25,26]. Our own database studies also support the consistency and reliability of parent reports [21,23,27].

Implications for the universal grammar hypothesis

On the one hand, this study results strongly support the universal grammar hypothesis: when prepositional grammatical abilities are acquired, all of them tend to be acquired together as can be seen from both UHCA and PCA analysis (Fig. 1). It is really astounding that spatial prepositions, verb tenses, flexible syntax, possessive pronouns, explanations about people and situations, simple stories, and elaborate fairy tales are all co-expressed together. For example, at the onset of the study, we did not expect that “understanding of simple stories” and “understanding of elaborate fairy tales” would co-express together. These findings are consistent with a single innate faculty underlying all prepositional language abilities as predicted by the Universal Grammar hypothesis. Once this faculty has been acquired, all prepositional language abilities are manifested simultaneously.

On the other hand, our findings suggest that a more intricate view of the Universal Grammar hypothesis may be necessary: while the Universal Grammar hypothesis attempts to present grammatical abilities as a dichotomy between the presence of grammatical abilities in humans and their absence in animals, there is a clear split between modifier and prepositional grammatical abilities. Why some individuals’ innate language faculty limits their development to modifier language phenotype despite typical cultural exposure remains unknown.

This report aims to contribute to the ongoing debate surrounding Universal Grammar. By exploring organization of the innate faculty of language on a fine level, we hope to inspire further research and investigation into this intriguing topic. Understanding Universal Grammar is crucial not only for the advancement of linguistic theory but also for comprehending the nature of human cognition and the fascinating capacity for language acquisition.

Clinical implications

Communication level in individuals with language delay is commonly characterized in terms of their verbal level (*nonverbal*, *minimally verbal*, or *verbal*) [27–29] while their language comprehension ability is left undescribed. If there was an objective way to characterize an individual's language comprehension level, that would facilitate their language therapy and improve outcomes. This study presents evidence for a three-level classification of language comprehension abilities from basic to advanced: 1) command, 2) modifier, and 3) prepositional language comprehension phenotypes. Once a language comprehension phenotype has been assessed in a child, language therapy can be fine-tuned for his or her particular phenotype. Furthermore, using language comprehension phenotype *improvement* as a measure of success will help focus language therapy on the important aspect of language development.

Implications for the language evolution hypotheses

Humans split from chimpanzees six million years ago and developed a unique prepositional communication system [30]. The origin of this prepositional communication system remains a mystery. There is an opposition between saltationist and gradualist hypotheses of human language comprehension evolution. The saltationist hypothesis argues for an abrupt step-like transition from non-recursive to recursive language sometime around 100,000 years ago [31–34]. The gradualist scenario conjectures a much slower acquisition of grammatical and syntactic abilities over several million years [35–38]. The former hypothesis expects a uniquely-human neurological mechanism made possible by a single mutation that resulted in simultaneous acquisition of a whole range of uniquely-human linguistic abilities [39], while the latter believes that many different mutations enabled various uniquely human neurological mechanisms that underlie human linguistic abilities [35]. Therefore, the saltationist single-neurological-mechanism hypothesis predicts a two-cluster solution (one cluster of individuals who acquired that mechanism and another cluster of individuals who did not), while the gradualist scenario predicts the solution with many smaller clusters. This study results are partially consistent with both saltationist and gradualist hypotheses. The three clusters are more than anticipated by the saltationist hypothesis, but significantly fewer than predicted by the gradualist hypothesis. The ensuing discussion of evolutionary acquisition of the three language comprehension phenotypes is expected to be most interesting.

Implications for language terminology used in linguistics, philosophy, and psychology

The definition of language is ambiguous. Fitch, Hauser, & Chomsky [39] write that "It rapidly became clear ... that considerable confusion has resulted from the use of "language" to mean different things. We realized that positions that seemed absurd and incomprehensible, and chasms that seemed unbridgeable, were rendered quite manageable once the misunderstandings were cleared up. For many linguists, "language" delineates an abstract core of computational operations, central to language and probably unique to humans. For many biologists and psychologists, "language" has much more general and various meanings, roughly captured by “the communication system used by human beings.” Neither of these explananda are more correct or proper, but statements about one of them may be completely inapplicable to the other.” This report will facilitate interdisciplinary communication by simplifying terminology for describing different language comprehension phenotypes.

Declarations

Funding

This research received no external funding.

Acknowledgements

We wish to thank all participants’ caregivers who found time to complete children’s assessments. The authors are very grateful to Dr. Petr Ilyinskii for his scrupulous editing of this manuscript, Dr. Leonid Fridman for the advice on linguistics, as well as Dr. Natalya Markuzon, Dr. Konstantin Khrapko, and Michael Tselman for the advice on the study design and statistical analysis. The language therapy app used to collect the data presented in this manuscript was made possible by the contributions of Rita Dunn, Alexander Faisman, Jonah Elgart, Lisa Lokshina, and Yulia Dumov.

Author contributions

AV and EK designed the study. RV, EK, and AV analyzed the data. AV wrote the paper.

Competing Interests

Authors declare no competing interests.

Informed Consent

Caregivers have consented to anonymized data analysis and publication of the results. The study was conducted in compliance with the Declaration of Helsinki [40].

Compliance with Ethical Standards

Using the Department of Health and Human Services regulations found at 45 CFR 46.101(b)(4), the Biomedical Research Alliance of New York LLC Institutional Review Board (IRB) determined that this research project is exempt from IRB oversight.

Data Availability

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

Code availability statement

Code is available from the corresponding author upon reasonable request.

References

1. Chonsky, N. *Aspects of the Theory of Syntax*. (Cambridge, Mass.: MIT Press, 1965).
2. Mitchell, R. E. & Karchmer, M. A. Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. Sign Lang. Stud. 4, 138–163 (2004).
3. Curtiss, S. The case of Chelsea: The effects of late age at exposure to language on language performance and evidence for the modularity of language and mind. UCLA Work. Pap. Linguist. 18, 115–146 (2014).
4. Mayberry, R. I. & Eichen, E. B. The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. J. Mem. Lang. 30, 486–512 (1991).
5. Morford, J. P. Grammatical development in adolescent first-language learners. Linguistics 41, 681–722 (2003).
6. Newport, E. L. Maturational constraints on language learning. Cogn. Sci. 14, 11–28 (1990).
7. Ramírez, N. F., Lieberman, A. M. & Mayberry, R. I. The initial stages of first-language acquisition begun in adolescence: when late looks early. J. Child Lang. 40, 391–414 (2013).
8. Grimshaw, G. M., Adelstein, A., Bryden, M. P. & MacKinnon, G. E. First-language acquisition in adolescence: Evidence for a critical period for verbal language development. Brain Lang. 63, 237–255 (1998).
9. 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).
10. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237–241 (2012).
11. Ruzzo, E. K. *et al.* Inherited and de novo genetic risk for autism impacts shared networks. Cell 178, 850–866 (2019).
12. Satterstrom, F. K. *et al.* Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 180, 568–584 (2020).
13. Fombonne, E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. J. Autism Dev. Disord. 33, 365–382 (2003).

14. Vyshedskiy, A. *et al.* Novel Linguistic Evaluation of Prefrontal Synthesis (LEPS) test measures prefrontal synthesis acquisition in neurotypical children and predicts high-functioning versus low-functioning class assignment in individuals with autism. Appl. Neuropsychol. Child (2020) doi:https://doi.org/10.1080/21622965.2020.1758700.
15. 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).
16. Vyshedskiy, A. *et al.* Novel prefrontal synthesis intervention improves language in children with autism. Healthcare 8, 566 (2020).
17. 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).
18. 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, (2017).
19. 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, (2017).
20. 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. (2022) doi:10.1007/s10882-022-09851-y.
21. 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).
22. 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. (2022) doi:/10.1007/s10803-022-05769-8.
23. 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).
24. Scattone, D., Raggio, D. J. & May, W. Comparison of the vineland adaptive behavior scales, and the bayley scales of infant and toddler development. Psychol. Rep. 109, 626–634 (2011).
25. Miller, L. E., Perkins, K. A., Dai, Y. G. & Fein, D. A. Comparison of parent report and direct assessment of child skills in toddlers. Res. Autism Spectr. Disord. 41, 57–65 (2017).
26. Dale, P. S., Bates, E., Reznick, J. S. & Morisset, C. The validity of a parent report instrument of child language at twenty months. J. Child Lang. 16, 239–249 (1989).
27. 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. 9 423 (Multidisciplinary Digital Publishing Institute, 2021).
28. Chan, A. S., Cheung, J., Leung, W. W., Cheung, R. & Cheung, M. Verbal expression and comprehension deficits in young children with autism. Focus Autism Dev. Disabil. 20, 117–124 (2005).

29. Colle, L., Baron-Cohen, S. & Hill, J. Do children with autism have a theory of mind? A non-verbal test of autism vs. specific language impairment. J. Autism Dev. Disord. 37, 716–723 (2007).
30. Somel, M., Liu, X. & Khaitovich, P. Human brain evolution: transcripts, metabolites and their regulators. Nat. Rev. Neurosci. 14, 112–127 (2013).
31. Chomsky, N., Larson, R. K., Déprez, V. & Yamakido, H. *Chomsky, N. 'Some simple evo-devo theses: How true might they be for language' in 'The evolution of human language'*. (Cambridge University Press Cambridge, 2010).
32. Satık, D. The Strong Minimalist Thesis is too strong: syntax is more than just Merge. *Biolinguistics* 16, 1–28 (2022).
33. Progovac, L. & Progovac, L. Sudden (saltationist) approaches to language evolution. Crit. Introd. Lang. Evol. Curr. Controv. Future Prospects 15–29 (2019).
34. Berwick, R. C. & Chomsky, N. All or nothing: No half-Merge and the evolution of syntax. PLoS Biol. 17, e3000539 (2019).
35. Dediu, D. & Levinson, S. C. On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. Front. Psychol. 4, 397 (2013).
36. Progovac, L. & Progovac, L. Gradualist approaches to language evolution. Crit. Introd. Lang. Evol. Curr. Controv. Future Prospects 31–66 (2019).
37. Progovac, L. A gradualist scenario for language evolution: precise linguistic reconstruction of early human (and Neandertal) grammars. Front. Psychol. 7, 1714 (2016).
38. Martins, P. T. & Boeckx, C. Language evolution and complexity considerations: The no half-Merge fallacy. PLoS Biol. 17, e3000389 (2019).
39. Fitch, W. T., Hauser, M. D. & Chomsky, N. The evolution of the language faculty: clarifications and implications. Cognition 97, 179–210 (2005).
40. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 310, 2191–2194 (2013).

Figures

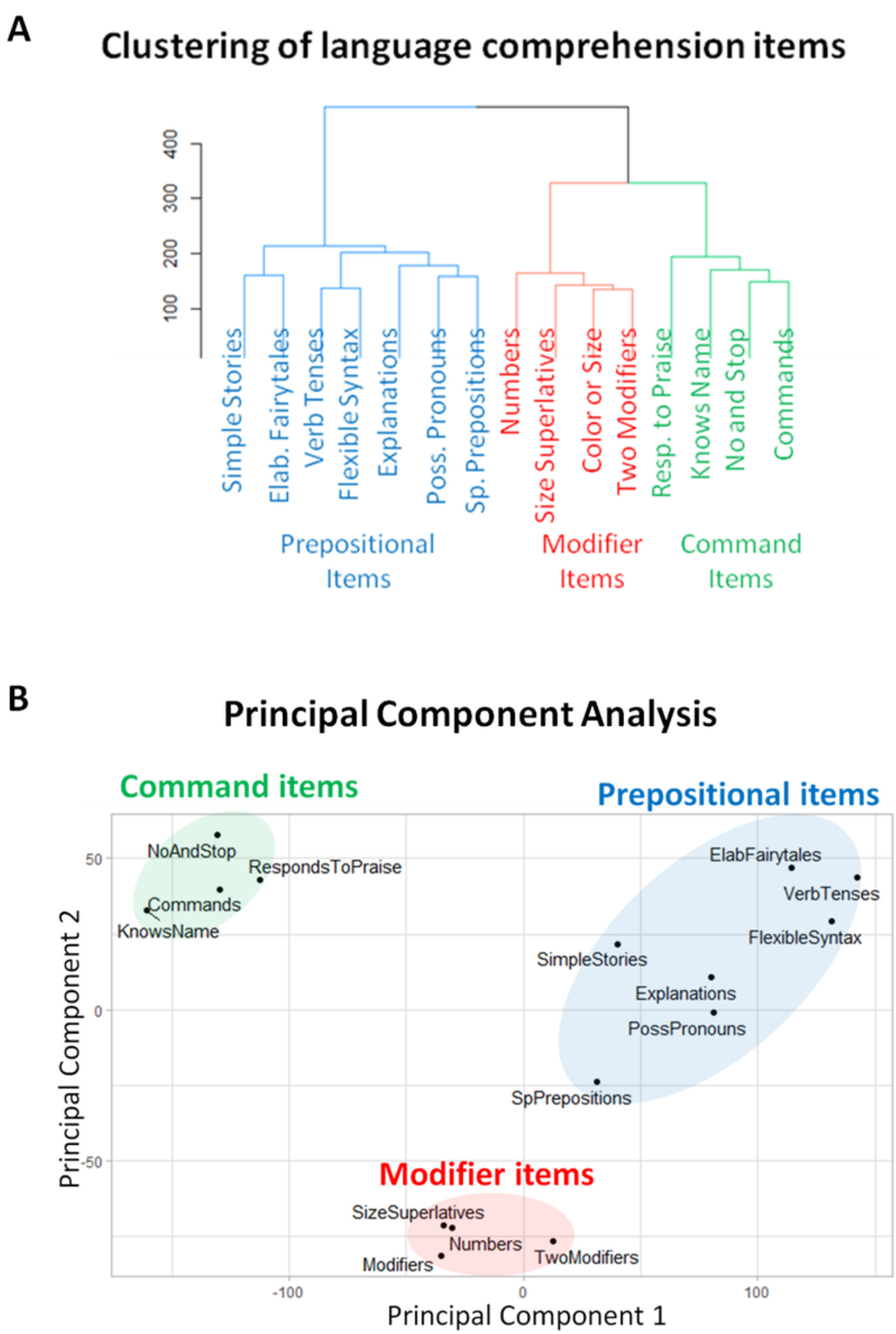

Figure 1

Clustering analysis of language comprehension items. (A) The dendrogram representing the hierarchical clustering of language comprehension abilities. (B) Principal component analysis of the 15 language comprehension abilities shows clear separation between command, modifier and prepositional items. Principal component 1 accounts for 42.5% of the variance in the data. Principal component 2 accounts for 11.0% of the variance in the data.

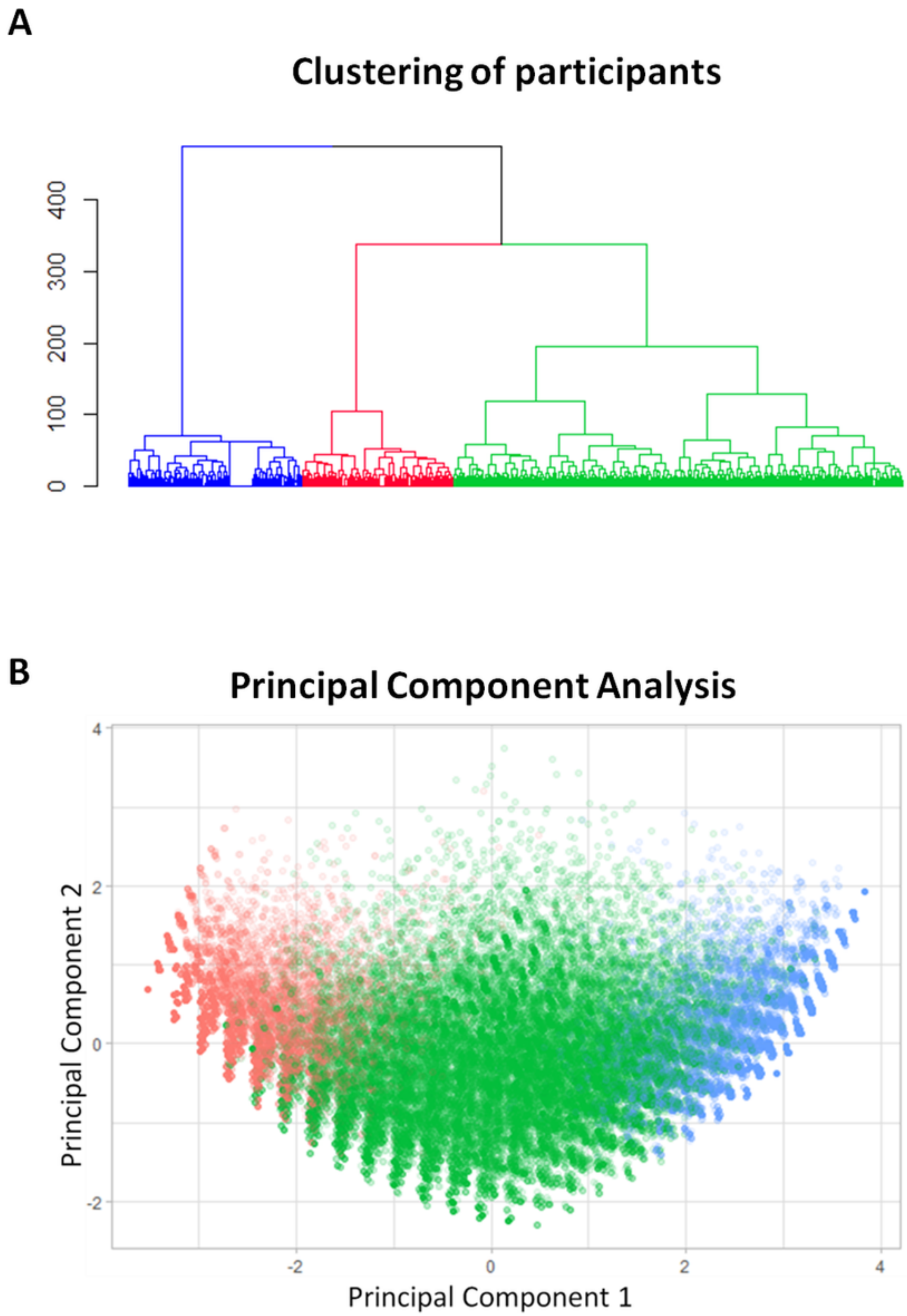

Figure 2

Clustering analysis of 55,558 participants. (A) The dendrogram representing the hierarchical clustering of participants. (B) Principal component analysis of participants shows reasonable separation between the three participant clusters. Principal component 1 accounts for 49.3% of the variance in the data. Principal component 2 accounts for 8.4% of the variance in the data.

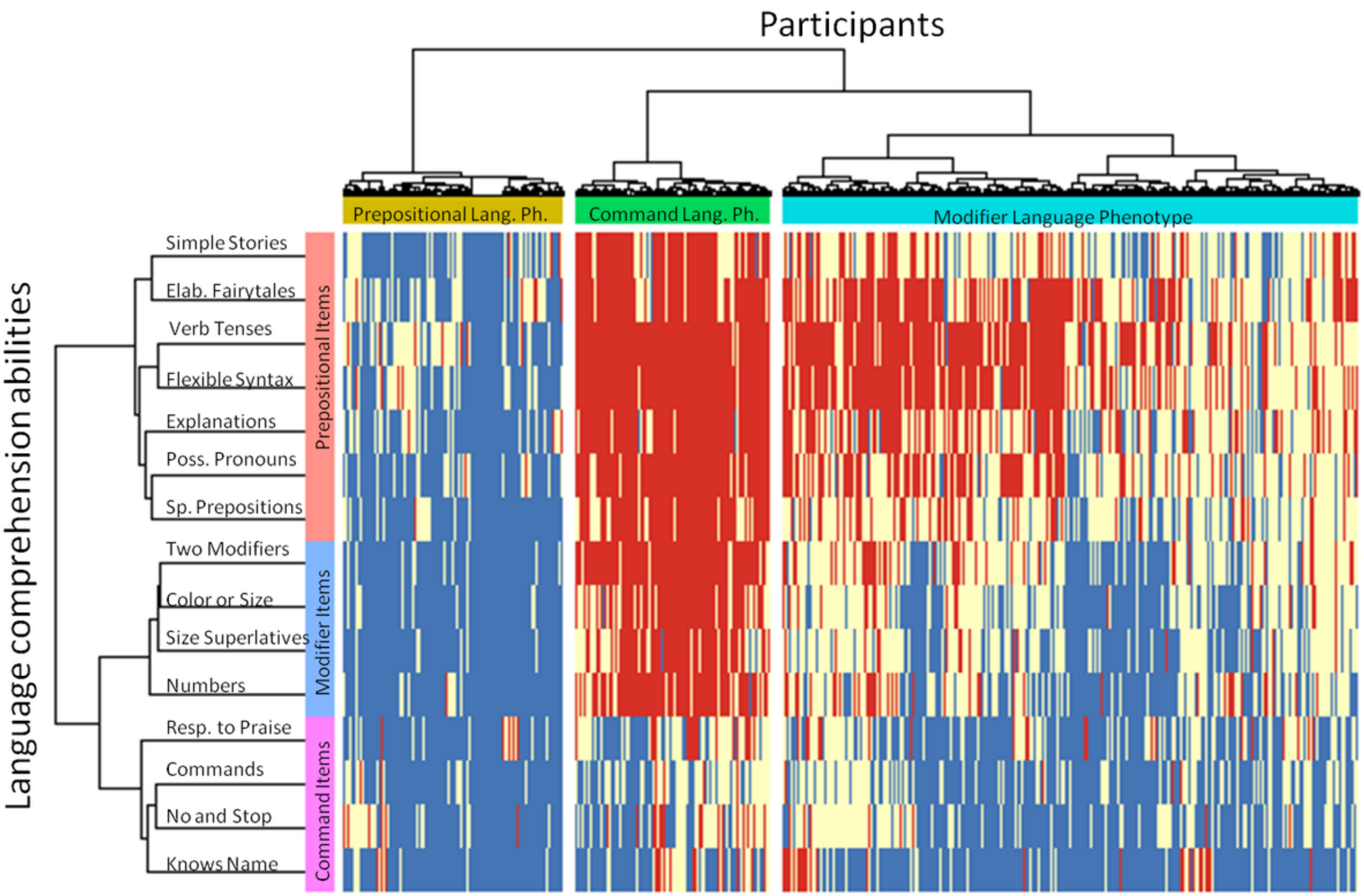

Figure 3

Two dimensional heatmap relating 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 55,558 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 indicates the "somewhat true" answer.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx