Using Support Vector Machines to identify determinants of pronoun difficulty in aphasia: a preliminary critical review and meta-analysis of individual data

Cecilia Devers, Silvia Martínez Ferreiro, Seçkin Arslan

19th International Science of Aphasia Conference - Venice

Using Support Vector Machines to identify determinants of pronoun difficulty in aphasia: a preliminary critical review and meta-analysis of individual data

Cecilia Devers[1], Silvia Martínez Ferreiro[2], Seçkin Arslan[3]

[1] *Leeds Beckett University, UK* [2]*University of Copenhagen, Denmark* [3] *Laboratoire Bases, Corpus, Language (BCL) Université Côte d'Azur, CNRS, BCL, France*

Introduction

The literature has strongly evidenced that people with aphasia (PWA) demonstrate a difficulty processing pronouns (Choy & Thompson, 2010; Vasić, Avrutin, & Ruigendijk, 2006). This difficulty is not well understood due to variability in sample sizes, conditions tested, and linguistic constraints relevant to their processing in different languages. Attempts have been made to theorise the underpinning elements contributing to this difficulty, however, with little resulting agreement across authors.

Differential performance patterns have been documented under different on-line and off-line experimental paradigms (de Roo, 2003; Piñango & Burkhardt, 2001); and, with relation to differentiating variables relating to pronoun use. These include different modalities (comprehension/production) (Abuom, Shah, & Bastiaanse, 2013; Fabbro & Frau, 2001; Goral, Levy, & Kastl, 2010), aphasia type, pronoun position and type (e.g. subject, personal, clitic; (Baauw, Ruigendijk, Cuetos, & Avrutin, 2011; Luzzatti, Toraldo, Ghirardi, Lorenzi, & Guarnaschelli, 2001), simplified versus complex structures (e.g. passive, wh-movement; Arslan, Gür, & Felser, 2017; Kljajevic & Murasugi, 2010).

A collection of theoretical explanations has been developed to explain the difficulty underlying pronoun processing in PWA. Plausible explanations straddle two primary conceptual constructs: representational and processing capacity accounts. Theoretical explanations range from a general syntactic processing impairment that encompasses computations necessary to resolve referential relationships (Edwards & Varlokosta, 2007; Kohn, Cragnolino, & Pustejovsky, 1997; Love, Swinney, & Zurif, 2001; Ruigendijk & Avrutin, 2003) to limitations of processing capacity necessary to execute syntactic operations during pronominal resolution (Burkhardt, Avrutin, Piñango, & Ruigendijk, 2008; Grodzinsky, Wexler, Chien, Marakovitz, & Solomon, 1993) and to a possible lexical integration and/or discourse-linking impairment (Bos, Dragoy, Avrutin, Iskra, & Bastiaanse, 2014; Choy & Thompson, 2010).

Given the uncertainty in understanding these phenomena, the present meta-analysis was designed to systematically analyse the current literature corpus relevant to studies that have investigated pronoun processing in PWA with a view to draw on common themes that may clarify why these referential linguistic items are problematic. The findings from this study will advance the understanding of pronoun processing in PWA.

Methods

Data

An exhaustive search in PubMed/MEDLINE and Web of Science was conducted with the following key word combination: "aphasia AND [pronoun OR clitic]". A total of 105 papers were retrieved. Those reporting individual accuracy results from PWA were included (n=30). Papers that merely report group means data (n=24) were noted and their inclusion postponed pending access to individual results. The remaining 51 papers were excluded due to a) reports of individuals with right hemisphere lesions (n=4); b) reports of individuals with non-aphasic symptoms primarily (e.g., dementia, echolalia, SLI; n=20) or of healthy individuals (n=7), and c) review articles (n=6) and articles that report no accuracy data for pronouns (n = 14). The original search was complemented through a subsequent search in Google Scholar and 22 additional papers and dissertations were added, increasing the total amount of studies included so far to 52.

These studies reported pronoun data across 21 languages comprising a total of 476 PWA (ranging from 1-50 per study). The languages reported in our meta-analysis include Cantonese (N=10), Catalan (N=6), Croatian (N=6), Czech (N=2), Danish (N=4), Dutch (N=54), English (N=126), French (N=35), Friulian (N=4), Galician (N=22), German (N=43), Greek (N=5), Hebrew (N=1), Italian (N=81), Québec French (N=5), Russian (N=20), Scledense (N=2), Spanish (N=59), Swahili (N=11), Swedish (N=2), Turkish (N=13), and Venetian (N=1).

Data analysis

The data were analysed using the Support Vector Machines regression model (SVM; Scholkopf & Smola, 2001) as this machine learning algorithm is well-suited for continuous data with large number of predicting variables and is also able to calculate variable importance. The meta-data was assessed using variable importance (see Figure 1a), variables that are important were further included in our analyses. For the time being, the present study has focused on experimental data reporting individual percent accuracy.

Results

Figure 1a. illustrates the variable importance measures, which showed that linguistic factors such as reflexives, relative pronouns, *wh*-movement (applicable in interrogatives), and passives ranked as important variables. Our trained SVM regression model predicted individual pronoun performances with a moderate success rate (RMSE=25.5), see Figure 1b for an illustration. It is conceivable that the model accuracy will increase as more data and features are added into our meta-dataset. A detailed look into the data showed that reflexive elements are found to be better preserved (80%) than non-reflexive elements (62.3%). PWA seemed to perform slightly better when pronouns are placed in subject positions (65%) than in object positions (45.7%) in production tasks. However, this asymmetry did not seem to hold in comprehension tasks (subjects=67.6%, objects=65.5%). An informative factor was the presence of pronouns in passive sentences, which seems to render pronoun comprehension harder (59.4%) than when used in non-passive sentences (67%). Crucially, individual languages were also important factors. Whether

pronouns appear in interrogative or declarative sentences (65.4% vs. 64.4%), and whether pronouns are expressed as clitics or not did not seem to matter (66.6% vs. 61.2%) to an important extent.

Discussion

Findings from this preliminary meta-analysis have led us to three general lines of conclusion so far. First, pronoun difficulty in aphasia persists across languages notwithstanding whether pronouns are expressed as clitics or full morphemes. Second, our findings suggest that pronoun difficulty increases as the structures encompassing pronouns also increase in complexity (e.g. passives), providing support to the limited processing capacity account in aphasia, as opposed to a structural deficit account. Further, reflexive elements being better preserved in aphasia than non-reflexives seems to be in line with this line of reasoning. At present, we are unable to contemplate on each specific hypothesis, as we are currently seeking access to more data. This meta-analysis has provided us with insights into how pronoun processing is impaired in aphasia varies across languages. Nonetheless, this cross-linguistic variability may be linked to diverse profiles of individuals and to the different nature of tasks used in examining pronoun resolution.

References

Abuom, T., Shah, E., & Bastiaanse, R. (2013). Sentence comprehension in Swahili-English bilingual agrammatic speakers. *Clinical Linguistics & Phonetics, 27*(5), 355 - 370.

Arslan, S., Gür, E., & Felser, C. (2017). Predicting the sources of impaired wh-question comprehension in non-fluent aphasia: A cross-linguistic machine learning study on Turkish and German. *Cognitive Neuropsychology, 34*(5), 312-331.

Baauw, S., Ruigendijk, E., Cuetos, F., & Avrutin, S. (2011). The interpretation of stressed and non-stressed pronouns in Spanish language breakdown. *Aphasiology, 25*(3), 386-408.

Bos, L., Dragoy, O., Avrutin, S., Iskra, E., & Bastiaanse, R. (2014). Understanding discourse-linked elements in aphasia: Athreefold study in Russian. *Neuropsychologia, 57*, 20-28.

Burkhardt, P., Avrutin, S., Piñango, M., & Ruigendijk, E. (2008). Slower-than-normal syntactic processing in agrammatic Broca's aphasia: Evidence from Dutch. *Journal of Neurolinguistics, 21*, 120-137.

Choy, J., & Thompson, C. (2010). Binding in agrammatic aphasia: Processing to comprehension. *Aphasiology, 24*(5), 551-579.

de Roo, E. (2003). Null subject pronouns in Broca's speech production. *Aphasiology, 17*(11), 1057-1072.

Edwards, S., & Varlokosta, S. (2007). Pronominal and anaphoric reference in agrammatism. *Journal of Neurolinguistics, 20*, 423-444.

Fabbro, F., & Frau, G. (2001). Manifestations of aphasia in Friulian. *Journal of Neurolinguistics, 14*(2-4), 255 - 279.

Goral, M., Levy, E., & Kastl, R. (2010). Cross-language treatment generalisation: A case of trilingual aphasia. *Aphasiology, 103*(1-2), 203 - 204.

Grodzinsky, Y., Wexler, K., Chien, Y., Marakovitz, S., & Solomon, J. (1993). The breakdown of binding relations. *Brain and Language, 45*, 396-422.

Kljajevic, V., & Murasugi, K. (2010). The role of morphology in the comprehension of wh-dependencies in Croatian aphasic speakers. *Aphasiology, 24*(11), 1354-1376.

Kohn, S., Cragnolino, A., & Pustejovsky, J. (1997). Pronoun production in agrammatic speakers: Patterns of use and avoidance in context neutral sentences. *Aphasiology, 11*(2), 157-175.

Love, T., Swinney, D., & Zurif, E. (2001). Aphasia and the time-course of processing long distance dependencies. *Brain and Language, 79*, 169-171.

Luzzatti, C., Toraldo, A., Ghirardi, C., Lorenzi, L., & Guarnaschelli, C. (2001). Comprehension of reversible active and passive sentences in agrammatism. *Aphasiology, 15*, 419-441.

Piñango, M., & Burkhardt, P. (2001). Pronominals in Broca's aphasia comprehension: The consequences of syntactic delay. *Brain and Language, 79*(1), 167-168.

Ruigendijk, E., & Avrutin, S. (2003). The comprehension of pronouns and reflexives in agrammatic and Wernicke's aphasia. *Brain and Language, 87*, 17-18.

Scholkopf, B., & Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press.

Vasić, N., Avrutin, S., & Ruigendijk, E. (2006). Interpretation of pronouns in VP-ellipsis constructions in Dutch Broca's and Wernicke's aphasia. *Brain and Language, 96*(2), 191-206.