“Moved deeply I am”
Autistic language in texts produced with FC

Arjuna Tuzzi¹, Marisa Cemin, Marco Castagna
¹Dept. of Historical and Political Studies – via del Santo, 28 – 35123 Padova – Italy
arjtuzzi@stat.unipd.it

Abstract
Using texts produced through Facilitated Communication (FC), this work is aimed at identifying the characteristic features of the language used by autistic subjects and understanding when these distinctive elements may distinguish it from the language of facilitators. Preliminary results show that autistic subjects actually use a special style of writing; this finding supports the hypothesis that texts are the fruit of individual production of autistic subjects, not inevitably influenced by facilitators. This first work, based on a restricted sample which is not necessarily representative, is important because it has permitted to better specify criteria by which subjects, texts, analyses to carry out and software to be employed will be chosen in future studies. Such a protocol may be later applied to a broader interdisciplinary project involving linguistics, statistics, computer science, pedagogy, neurology, psychology, social sciences and ethical-philosophy. This new project, named “EASIEST” (Autistic Expression: Interdisciplinary Study with Statistic-Textual data processing) is based on the idea that in the specific case of autism, each form of communication represents a resource to be evaluated even if this seems atypical and controversial.

Keywords: textual-approach, software Taltac, vocabulary comparison, characteristic textual units.

Acknowledgments:
We wish to thank people in charge of the institutes for Facilitated Communication training who have supplied us with data for this work: Patrizia Cadei, Vittoria Cristoferi and Sergio Vitali; and to thank also Lorenzo Bernardi, Sergio Bolasco and Patrizio Tressoldi for their useful insights. We are also very grateful to the (autistic) authors of the texts and their facilitators.

1. Introduction
Autism is a serious illness that affects the development of human neurological organisation. Nowadays it affects in its most serious form one children out of 2,000. As its origin is unknown, there is no resolutive remedy. The biggest problem for people affected by autism and their families is represented by difficulties to respond to other people and to interact with the surrounding world (autistics can be either totally dumb or speak with whimsical and strange or repetitive language expressions). It is therefore necessary to find effective ways of communication (including those alternative to spoken language) in order for people affected by autism to open up new spaces of exchange with their families, to start suitable educational programmes and to find possible solutions to relieve their troubles.

Facilitated Communication (FC) is a technique (Biklen, 1999; Crossley and McDonald, 1990) that aims at creating a situation to communicate between two (or more) people. After a specific training, FC can offer the possibility to interact through the use of a hardware (i.e. a personal computer) which helps the writing of messages and through a human support (the facilitator) who gives physical and emotional sustainment to the autistic subject. During a FC session the autistic subject and the facilitator are placed side by side in front of the PC and
chat using video and keyboard. The facilitator usually stimulate the conversation digitizing a question and giving the autistic subject the possibility to react (for instance by digitizing the answer). During the first sessions gestures of the autistic subject are helped by the facilitator’s hand placed on the subject’s wrist. The contact gradually moves to the elbow, arm, shoulder, back and knee (according to the specific case) in order to obtain the maximum autonomy of the subject. However there is no general agreement about FC reliability both as an educational method and a way of communication: in particular, those who affirm autistic cognitive incompetence are doubtful about the authenticity of their texts, considering them the fruit of the influence (and great sleight) of facilitators (Green, 1994; Jakobson et al., 1995). Moreover the validation method is still a matter of dispute among researchers all over the world. The idea of a survey on texts produced with FC from a statistic-linguistic point of view (Niemi and Kärnä-Lin, 2002; Zanobini and Scopesi, 2001) enters the above mentioned debate, since researchers try to find out useful methods to assess if (and how) texts written by autistic subjects are different in style and lexical choices from texts written by facilitators.

The present study uses texts written during FC sessions by subjects with clinical diagnosis of autistic disorder (according to DSM-IV). Three Italian institutes have provided us with texts: the “Center of Study on Facilitated Communication” (Chiavari, Genova), directed by Patrizia Cadei; the “Experimental Center for Development and Communication Disorders” (Padoua), directed by Vittoria Cristoferi Realdon and the Social Cooperative “Intervento” (Mestre, Venice), directed by Sergio Vitali. The written works of 11 people aged from 9 to 22 have been collected: everybody had reached the elbow facilitation level and had been writing on a PC for at least a year with expert/habitual facilitators. This implies that all the texts we have at disposal have been written by autistic subjects who have been writing for a long time, together with expert/habitual facilitators, meaning a high fluency and a high level of autonomy.

2. Texts organisation and coding

The texts written with the FC technique can be different: a daily dialogue concerning the activities done during the day; a test on a school lesson; the composition of a poem; an educational speech; a free composition; information about personal facts, etc.

Capital letters are used to designate the part of the text written by the facilitator, whereas lower-case letters are used by the autistic subject in order to distinguish them clearly. To give an example:

COME HAI IMPARATO A LEGGERE E SCRIVERE?
a scuola guardando gli altri leggere, la cf è servita a farlo capire
SEI STATO ALLA SCUOLA MATERNA? TI PIACEVA? COSA TI DISTURBAVA?
ma andiamo alla preistoria. per me è nebbia
C’È QUALCOSA CHE VORRESTI DIRE GLI INSEGNANTI E ALLE ALTRE PERSONE PRESENTI IN SALA?
l’autismo è guardare voi che vivete, amate, lavorate e intanto il cielo è sempre più blu, notte.

(P.C., anni 21)

HOW DID YOU LEARN TO READ AND WRITE?

at school looking at others reading, fc let it know

HAVE YOU BEEN AT PRIMARY SCHOOL? DO YOU LIKE IT? ANYTHING TROUBLED YOU?
we go back to prehistory. it’s fog for me

WOULD YOU LIKE TO TELL SOMETHING TO THE TEACHERS AND THE OTHER PEOPLE IN THIS ROOM?

autism is looking at you living, loving, working , meantime the sky is more and more blue, night

(P.C., 21 years old)

Two separate corpora have been therefore created in order to understand if there is a specific character of the autistic language and if there are perceptible differences with the language of
facilitators: 1) the first comprising the texts of the facilitators (questions and comments which have appeared during the sessions); 2) the second concerning the texts written by the facilitated autistic subjects (answers, comments, poems, written compositions, etc.). Since the two corpora were limited in dimension, no further distinction has been made (age, facilitation level, place of the session, argument etc.) as the subparts could be of little relevance for the analysis.

The two corpora have been studied using both simple word-types (graphic forms) and complex textual units coding procedures (fig.1). In the first phase only simple word-types have been chosen (Lebart et al., 1998). Step by step, a more complex recoding have been developed using an Italian common use multi-words list (phrase and polyrhematic recognition) and the extraction of repeated segments (multi-words and polyforms that make sense and other relevant sequences of words that gain or change meaning if considered as a block). All these operations have been done with the Taltac software (Bolasco et al., 2000) and applying linguistic resources to process Italian language.

3. Comparisons based on simple word-types vocabularies

3.1. Autistic subjects

The corpus of texts written by autistic subjects includes a total of \( N = 24,499 \) word-tokens and \( V(N) = 5,264 \) word-types. The Type-Token Ratio (T.T.R. = \( V(N)/N \)) and the hapax percentage (\( V_1/V(N) \)) point out the presence of a rich language; a result which can be partially explained through the limited dimension. The word-types vocabulary offers a very interesting description of the texts: at the top of the list which is ordered per decreasing frequency there are in fact some word-types showing a discourse which is mainly centered on self (‘mi / myself’, ‘io / I’, ‘sono / I am’, ‘ho / I have’, ‘me / me’, ‘voglio / I want’, ‘mio / my-male’, ‘mia / my-female’, ‘tu / you’, etc.). We can notice a massive use of negative forms (‘non / not’, ‘ma / but’, ‘no’) and of the word ‘molto / very’.

In order to read the vocabulary in a comparative perspective a reference lexicon taken from the Italian newspaper “La Repubblica” stored in the Taltac software has been used. It consists of a wide set of high frequency word-types (more than 50 occurrences in ten years of the newspaper publications) and is a good source of written Italian language. Such reference lexicon is useful to extract the key words from the corpus calculating over/under-usage with respect to the frequency of the lexicon assumed as reference language. If only word-types that autistic texts and reference lexicon have in common are taken into account (4,417 word-types of the intersection set out to 5,264, equal to 84.9%) and indexes of over/under-usage are computed (Bolasco, 1999) we can notice the over-usage of some topic words such as ‘autismo / autism’ and ‘autistico / autistic-male’ and of words used for answers and questions such as the names (of the facilitators and of the subjects) together with other simple word-types such as ‘ciao / hi’, ‘si / yes’, ‘ok’, ‘no / no’, etc. Moreover, there are elements that emphasize the per-

3.2. Facilitators

The facilitators’ corpus comprises a total of \( N = 23,386 \) word-tokens and \( V(N) = 3,714 \) word-types. The Type-Token Ratio (\( T.T.R. = \frac{V(N)}{N} = 15.9\% \)) and the hapax percentage (\( \frac{V_1}{V(N)} = 57.4\% \)) show that the language is in this case less rich than the one of the autistic subjects. This is a consideration based on a limited corpus that can nevertheless be compared with the other in terms of size and contents. The word-types vocabulary begins with ‘che / what / which’ followed by ‘di / of’ and ‘ti / you’: these are all devices that suggest the presence of a language based on questions (‘che cosa...? / what...?’, ‘ti...? / do you...?’, ‘perché...? why...?’). In the high and medium frequencies areas of the vocabulary there is also the presence of many colloquial words such as ‘vuoi / you want’, ‘fare / to do’, ‘oggi / today’, ‘anche / also / too’, ‘questo / this’, ‘quando / when’, ‘devi / you must’, ‘dire / say’, etc. The comparison with the “La Repubblica” reference lexicon allows to observe that in the part in common (3,196 word-types of the intersection set out of 3,714, equal to 86.0%) there are some word-types which are over-used by facilitators and that refer again to the question wording style: ‘ciao / hi’, ‘ti / you’, ‘hai / you have’, ‘vuoi / you want’, ‘scrivi / you write’, ‘vorresti / you would’, ‘stai / you are’, ‘ok’, ‘dimmi / tell me’, ‘senti / listen’. The result is not surprising because the occurrence of direct speech question/answer is not a characteristic feature of journalistic language, but on the other hand is very common in FC sessions. The analysis of word-types showing originality is probably more interesting (518 word-types of the complementary set out of 3,714, equal to 13.9% that occur in the texts of facilitators and never in the reference lexicon) because in this case the language of the educator is recognizable (FC is not only a way to communicate but also an educational method): ‘intendi / you mean’, ‘definisci / you define’, ‘aiutarti / help you’, ‘riferisci / you refer’, ‘controllarti / control yourself’, ‘capirti / understand you’, ‘descrivi / you describe’. Unlike the case of autistic subjects, here non-common use elements have not been found.

3.3. Autistic subjects versus facilitators

Following the same criterion as the case of the reference lexicon, we extract word-types showing originality directly comparing the word-type vocabulary of the facilitators with the one of autistic subjects. The two vocabularies have in common 1,950 word-types. This intersection set represents 52.5% of the facilitators’ vocabulary (1,950 word-types out of 3,714) and only 37.0% of the vocabulary of the autistic subjects (1,950 out of 5,264). Facilitators use 1,764 word-types never used by autistic subjects (equal to remaining 47.5% of the vocabulary), whereas autistic subjects put in their texts 3,315 word-types never used by facilitators (equal to 63.0%). This information (also represented in fig.2) suggests a greater wealth of words expressed by autistic subjects. However it is also interesting to take into consideration

It seems interesting to underline that each time the autistic subject uses an unknown (or not completely clear) for the facilitator word during a FC session, the facilitator immediately looks for an explanation about the use of the term (in the example given the names have been censured). Peculiar words of autistic subjects are therefore widely underestimated since during FC sessions some peculiar terms can be pronounced first by the facilitated autistic subject and then by the facilitator. This matter needs further investigation.

**Figure 2. Comparison of the word-types vocabularies of facilitators and autistic subjects**

4. Comparisons based on complex textual units vocabularies

4.1. Autistic subjects

As shown in text-figure 1, some intermediate phases for the analysis have been fixed in order to code the two corpora into new complex textual units. First of all we used a phrase and polyrhematic recognition based on an integrated Italian common multi-words list of the Tal-tac software; then we extracted repeated segments: multi-words and polyforms that are not contained in the Taltac lists and sequences of (less then seven) words repeated at least two times in the corpus. These non-hapax segments are 1,166 in total. This list shows the inci-
ence of segments comprising ‘non / not’ (148 segments) and ‘molto / very’ (33). There are a lot of emotional expressions addressed to their facilitators (‘mio buon nome_facilitatore / my dear facilitator_name’, ‘furba nome_facilitatrice / she-cunning facilitator_name). Moreover, there are some segments lacking grammatical elements (especially conjunctions and auxiliaries).

Grammatical elements may also be placed in positions that do not respect the correct structure of sentences in the Italian language. To give some examples: ‘penso proprio non lo so / I think definitely I don’t know’, ‘molto voglio dirvi grazie / a lot I want to say you thanks’, ‘sono solo molto pieno di pesante disperazione / I am only full very of heavy desperation’, ‘grande sono diventato / grown I have become’, ‘molto bravo sono con mia resistenza / clever very I am against my resistance’, ‘molto sono arrabbiato con te / angry very I am with you’, ‘molto sono emozionato / moved deeply I am’, ‘questo voglio fare non / this I want to do not’. These segments are undeniably interesting on linguistic grounds; it is also quite easy to recognize the qualitative specificity of these expressions. However, the evaluation of this hypothetical “autistic style” goes beyond this work as it ought to be studied by linguistics.

According to a statistic method, the Morrone’s IS index (Morrone, 1996; Bolasco et al., 2000) has been applied together with a manual selection in order to reduce the 1,166 segments to 555 (segments that make sense without redundancy). This index is able to recognize a meaningful sequence of word-types comparing the frequency of the sequence as a block with the frequency of its word-types components. This new list has been used for the lexicalization of the corpus in order to obtain a recoding of the corpus in new complex textual units (word-types, multi-words, polyforms, polyrhematic, idioms, and segments together in the same vocabulary).

4.2. Facilitators

The same criterion has been followed for the corpus containing the facilitators’ texts. Running the Taltac procedure of recognition of common use multi-word and extracting non-hapax segments of length below seven (a list of 1,405 multi-words reduced to 799 using Morrone’s IS index and a manual selection) we have obtained a list for lexicalization. By means of a comparison between the two lists (facilitators and autistic subjects) some differences have appeared. Among the segments of the facilitators with high absolute IS index (long sequences extraction) there are common use expressions like: ‘tu mi chiami / you call me’, ‘ti rendi conto / you realise’, ‘sicuramente non avrai problemi / surely you will not have any problem’, ‘hai lavorato bene / you have well worked’, ‘non hai risposto / you did not answer’. The list of segments of autistic subjects shows expressions like: ‘nidifico uova in vista di rivelarmi / I’m nesting eggs in order to reveal myself’, ‘abbandonato hai questa sterile vita / you left this sterile life’, ‘sforzo non potrà limitarmi / will not strain be able to limit me’, ‘opulenza di eccessi / opulence of excesses’. A technical language appears in the segments of the facilitators with high relative IS index (short sequences extraction): ‘comunicazione facilitata / facilitated communication’, ‘materie scolastiche / school subjects’, ‘chiesa romanica / Romanesque church’, ‘maestra nome_maestra / teacher teacher_name’. This language differs in qualitative terms from the corresponding expressions of the autistic subjects: ‘cultura umanistica / humanism’, ‘completa autonomia / complete autonomy’, ‘era nera / dark age’, ‘buon nome_facilitatore / dear facilitator_name’, ‘oserei dire / I dare say’.
4.3. Autistic subjects versus facilitators

The study of the two lists of selected segments suggests that the two languages are different as shown by the dimension of the intersection set of the two lists: only 62 common segments appear (fig.3).

If the two lists are merged together, a union set of 1,292 segments are obtained. This list has been used for the lexicalization of both corpora in order to obtain a partition comparable textual units. In fact, this operation creates two new vocabularies of complex textual units which can be compared. This comparison confirms what other analysis have already suggested (fig. 4): the facilitators’ vocabulary is only partly overlapping the autistic subjects’ vocabulary. In addition to this, the forms contained in the non common parts are different in quality.

In order to recognize the textual units present a great deal more (or a great deal less) in the facilitators’ corpus than in the autistic subjects’ one (elements useful to distinguish between the two groups) we can use the traditional ‘characteristic textual units’ method (Lebart et al., 1998). This simple tool is based on the hypergeometric model (Lafon, 1980) and by means of a probability of over-usage it can detect which elements are used frequently by a group (as well as which elements tend to be used rarely by means of a probability of under-usage). The hypergeometric model is not sufficient to draw general conclusions about differences between the two corpora because the researcher has no indication of the distribution of these ‘characteristic textual units’ within the corpus. In fact a textual unit can be identified as specific of the corpus even if it occurs only in the text of one or two facilitators (or autistic subjects) rather than being representative of the group as a whole. A further test on dispersion (Tuzzi and Tweedie, 2002; Tuzzi, 2003) has been used to select only ‘characteristic textual units’ that are not confined to a subset of the texts. The following list shows the ‘characteristic textual units’ (ranked in decreasing frequencies order) which can be therefore considered suitable to describe the specificity of the corpus written by autistic subjects versus the facilitator’s one.


5. Conclusions

Despite the fact that the validity of the FC technique has not been discussed in this paper, it is nevertheless impossible to ignore the striking nature of the language produced by autistic subjects during the sessions: they write in a very special way with a sparing use of words; unusual terms are used, short phrases are preferred and they talk about emotions, intentions and feelings.

According to this, many issues need further investigation and bigger and more relevant corpora. Which kind of language is the one which is never orally expressed? If an autistic language exists, how can the specific features that distinguish it from the one used by facilitators be explained? However, the main priority of people applying FC is obviously to increase the quantity and improve the quality of autistic subjects’ communication (Vidal et al., 2002).

The results obtained should never be used to further isolate people affected by autism, considering them like strange speaking individuals with an impenetrable and enigmatic inner world.

References

Bolasco S., Baiocchi F. and Morrone A. (2000). TALTAC versione 1.0. CISU.


