



CONTRIBUTION OF PSYCHOLINGUISTIC PERSPECTIVE FOR SPEECH TECHNOLOGIES

Tatiana Slama-Cazacu

Laboratory of Psycholinguistics,
University of Bucharest, Romania

ABSTRACT

The need of a psycholinguistic (PL) perspective for Speech Technologies (STs) is argued in the paper, together with the importance of selecting an adequate PL model, as a first requirement in this concern. A realistic PL model, which might really be efficient for STs, is presented (the "contextual-dynamic model" - Slama-Cazacu). Research on the role of verbal stimuli (VS) in confront with the nonverbal ones (NVS) in automatic factories and in laboratory simulated conditions demonstrated the greater efficiency (more rapid reaction time and correct reactions) for VS, wherefrom application derives for computerized speech recognition and synthesis, and generally for a real communication situation in human activities involving computers (Cs). Other applications of the PL approach: to machine translation; for the identification of speakers; research on child language and education to be used in STs; PL analyses of "reading/writing" with C; PL research on the use of C as a "supermedium" in or for communication; the PL bases for an extension of the use of C with or for oral communication; decoding the new or future "hypertexts" on the C's screen; the PL based arguments for "multimedia". Scientific research on the effects on the human being of the new STs are much advocated for.

I. INTRODUCTION

0. As a participant in the two former ICSLP (1990, 1992), I advocated for including Psycholinguistics (PL) among the suggested areas proposed with specification of this name, for future Conferences. My contribution would be meant to underline the importance of having taken into consideration "PL" as such in the 3rd Conference, and to mention some research already done by the present author in this concern and to suggest new ways for future studies.

II. THE NEED OF AN ADEQUATE PL MODEL

1.0. Speech technologies (STs) mostly appealed, among the humanistic sciences, to Phonetics and Psychology; more rarely to "Linguistics" at large and, even, scarcely to PL.

1.1. The need of a PL perspective for STs is argued in the paper, together with the importance of selecting an adequate PL model, as a first requirement in this concern. A realistic PL model, which might really be efficient for STs, has to start from the reality of the act of human communication by natural language production/reception.

1.2. Such a model was proposed a longer time ago by the present author (Slama-Cazacu 1961, 1973, etc.), labelled "the contextual-dynamic model" (CD). It takes into consideration all the components of a concrete act of human communication, integrated in their *contexts* at various levels (linguistic, non-verbal, situational - explicit and implicit contexts), and in the dynamics of changes in the messages determined by the interaction between the interlocutors, by their psychological - cognitive communicative, attitudinal etc. - peculiarities, considered in a continuous *dynamics* (schemes 1 and 2 in Slama-Cazacu 1992, 686).

1.3. I would like here to help future demonstration, by introducing a *classification* within the large category of human "communication acts", of the *media of communication* (MC), having also in view the "mass media", various "instruments" and the special case of the *computer* (C): scheme 3 (here, end of our paper).

III. RESEARCH BASED ON (OR VALIDATING) THE CD MODEL, FOR USE IN STs

2.0. *Research* (Slama-Cazacu) [10], [11], [13], [14], etc. on the role of verbal stimuli (VS), in confront with the nonverbal ones (NVS) in automatic factories and in laboratory simulated conditions, demonstrated the greater efficiency (higher - i.e. more rapid - reaction time and correct reactions) for VS (words), wherefrom the application derive for computerized speech recognition and synthesis, and generally for a real communication situation in human activities involving computers (Cs). Messages which previously occurred mostly *visual* (produced on screens by the computer: illuminated written words

- as "Step up voltage!", "Alarm transformer!" -, or flickering of various electric bulbs, etc. - see Slama-Cazacu [12], [14:188-204]) may occur now as auditive "oral" ones, produced via Cs. This extrapolation of my results with visual messages may be of great interest for *speech synthesis* (production of V oral messages by electronic devices) and *speech recognition* (reception by the computer of the human oral messages). By that, not only is *communication facilitated* e.g. for asking for tourist *information*, but also for the direct control of automated installations, as *factories*. It was what I advocated from the very beginning of the 60's, hypothesizing ([11], [12], [13]), then demonstrating by laboratory experiments, namely that the human being is much more efficiently informed and more rapidly reacting when (s)he receives/sends V (*oral*) messages, than NV ones - perceiving images with a semiotic conventional value, or expressing by a reaction as pushing a button.

3.0. *Some other "new ways" for PL research* in close connection with Cs or undertaken for the benefit of computerized activities -for some, cf. also [20]: PL analyses of "reading/writing" on/with C; PL research on the use of the C as a "supermedium" in or for communication; the PL bases for an extension of the use of C with or for *oral* communication (including studies on the role of gestures, facial expression in human real communication); decoding the new or future "hypertexts" on the C's screen; the PL based arguments for "multimedia".

3.1. The C at large implies problems of *special production/emission of messages*. There is a need of learning of a new kind of reading - a special "literalization": Reinking [8] and of emitting messages. This is an important PL field of research.

3.2. The new concepts of *hypertexts* on the computer's screen require from the receiver a new training in reading such texts which are non-linear, non-sequential, for extracting information from banks of data composed of distinct textual units - Reinking 1992, 106; this requires, I would comment, another kind of reception than those envisaged by McLuhan, i.e. a complex "sensorial" one as named by him: it involves capacities for interpretive connections, and also much recourse to short-term memory, to "mnemonic storage" - Slama-Cazacu [18:133], even more than in reading "traditional" texts.

3.3. As there is a demand for *multimedia texts* which combine visual *and* auditive signs, one should see here the *need* for auditive supports in communication and maybe, first of all, for *orality* to complete the too "(verbally) silent computer".

Including on a screen "visual" (graphic images of words *and* drawings or pictures, etc. as well as other NV signs, iconic or gestures, etc. -see my concept of "mixed syntax: [15], [19]) *and* the "auditive" involves both the acoustic signs such as oral words but also music or other sounds that are parts of the communicated message. Under such conditions, of course, the reception of messages is complicated and different from the reading of a book or just hearing the radio: it involves the two major channels - and in each perceiving both the V and the NV signs -, decyphering the meaning of the signs transmitted via each of the channels simultaneously and integrating them into the same unit at every moment. The emission too is modified and the human had to re-adapt himself to it: specific (to what extend and in what manner, this is to be better shown by PL) for a TV production of oral blended with visual messages (in a more "formal" situation also? and how much is to be expressed visually vs auditive? etc.); or for a C operation when using it as a MC, etc. (a "hypertext" should not only be visually expanded into further information, but also *auditively* developed - maybe *orally* in the future?). What is the proportion of *visuality* and of *orality* in optimal V messages in TV e.g. or on the screen of a sophisticated C? (my hypothesis is that in "watching" TV one may feel more "frustrated" if there is *no* sound at all - i.e. image without any sound-, than if there is no image; but this may be mostly conditioned by the manner in which a TV message-production has been conceived, i.e. in laying stress on images, or on V sounds, such as in the messages transmitted by radio, where the occurrence of a message is "explained" *via* variation in voices, or by describing a situation, etc.; in any case, let us also think that mankind was no longer satisfied with "silent film" - so unilateral and not a "normal" way of communication - and invented-developed the "sound movie"). The following challenging hypothesis might be submitted to experimental and also to statistical check: in present-day communication, we are mostly submitted to a complexity of perceptions and production through the two major channels - visual and auditive -, but the *oral* components of communication are prevailing.

4.0. I can but mention *other topics* needing PL research: applications to machine translation (need of contrastive linguistics, based on a PL approach - also T.Slama-Cazacu, W.Nemser 1971, T.Slama-Cazacu 1970, 1975, etc.; or to the identification of speakers (T.Slama-Cazacu 1971) or PL research on child language and education to be used in STs, cortex activity and the C, etc. (my research in progress).

IV. SCIENTIFIC RESEARCH ON THE EFFECTS ON THE HUMAN BEING OF NEW STs

5.0. What is *the result, in communication*, of the C device (and maybe of other still waiting, in laboratories, to be born for all mankind, hopefully for its benefit)? What are the aspects which PL should study and should *be required* to make research upon?

5.1. A change in emission and reception of messages involves a *complexity* of sensorio-perceptual-manual, intellectual, volitional processes. It is not possible to speak about one kind of sensory aspect, and not even about just the sensorial multiplicity, which has been foreseen by McLuhan (however, mentioning it sporadically, without analyzing it and mostly raising doubts about the capacity of the human to adapt to "automation", which envelop us in new uncertainties: 1962, 275). When one think that this "electronic revolution" in communication occurred in less than one century, one may understand McLuhan's apprehension (and it had even been expressed thirty years ago, before many devices mentioned above became almost part of the usual life in the civilized world). But we have to take into account that these devices did not appear simultaneously: mankind became accustomed to the telephone, then to the radio, and only afterwards to the TV; and only recently, in the 80's especially, the C as a medium (producer/receiver, giving new form to human messages or transmitting the usual ones); and its use is expanding, little by little, under various forms. This hypothesis - of the rapid and "not dangerous" adaptation to the MC - may be submitted to validation in very interesting psycholinguistic research.

5.2. A corollary of such research would be the *kind of psycholinguistic* processes involved in the production and reception of messages *via* the C; these are not only sensorial processes, but complex *intellectual* ones and "highly verbal" (the *meaning*

and its *interpretive* production/reception almost neglected by McLuhan). The *visual* perception (of V messages but also of NV ones: images or gestures, etc.) is involved in TV, in communication *via/with* the computer, but I do not see many instances of visual messages completely isolated (as in a book) from *auditive* (V-oral) messages or components of messages. And this might be a consequence of what happens in usual life, in direct communication outside the MC and generally outside the instruments to which the modern individual has resorted for helping his/her communication.

5.3. One cannot ignore the major difference in communicating with a computer (also underlined, though incidentally, by Wenger 1987, 6), namely the *lack of a social relationship* (I would say: psycho-social) in the case of a pure computer-human dialogue (especially in the case of an "educational" dialogue, between a learner and the computer). However, when the computer is used as medium of communication, "behind" it there is *another* human. (This problem is quite different than "working with peers": see EJPE 1992).

A question in association with McLuhan again: will the *micro-computer*, in the miniature shape of a book, "less than 100 grs", lead again to an "isolation" of the individual as did the one described by McLuhan for the dawn of the Gutenberg age? Perhaps not, because such devices *serve for intercommunication* too, and not only for a use in solitude, for a game, etc.

6.0. Efforts to emphasize the links between PL and STs - as well as for stimulating research and generally interest in this respect - should have a bilateral direction, i.e. involving both sides: not only from PL to STs -a centrifugal trend-, if taking Ph as a reference point, but also a centripetal one, from STs toward PL.

Among the efforts and exhortations from PL side, I would like to mention on the one side the encouragement for "a multidisciplinary oriented PL", including relationship with STs, which comes from the ISAPL ("International Society of Applied Psycholinguistics") organizing its International Congress (June 1994) under this "flag". And, on the other hand, to stress upon the fact that the only scientific international magazine in PL, the "International Journal of Psycholinguistics" (published in Japan, ed.-in-chief T.Slama-Cazacu) stimulates such research, mentioning even in its cover-design: "Speech Technologies and Human Communication Models".

V. REFERENCES

- [1] *European journal of psychology of education* (Special issue: Interactional learning situation with computers, ed. A.Blave, P.Light, V.Rubtsov), 7, no.4, 1992.
- [2] A.Firth. Discourse at work: Negotiating by telex, fax and phone (Unpublished doctoral dissertation), Aalborg, University - Denmark (*apud* A.Firth, in *Newsletter discourse analysis* - "AILA Commission on Discourse Analysis", Spring, 1993), 1991.
- [3] *ICSLP* [International Conference on Spoken Language Processing] *Proceedings 90*. [Kobe], Japan. The Acoustical Society of Japan [2 vols] and [*idem*] 92. 1992 [Banff], Canada, University of Alberta.
- [4] "International journal of psycholinguistics" [Ed.-in-chief T.Slama-Cazacu; Publ.: Japan, 1993, [26], vol.9, no.1 and 2; 1994, [27], vol.10, no.1.
- [5] Marshall McLuhan. *The Gutenberg Galaxy. The making of typographic man*. Toronto, University of Toronto Press, 1962.
- [6] Seymour Papert. Etude comparée de l'intelligence chez l'enfant et chez le robot. In L.Apostel, J.B.Grize, S.Papert, J.Piaget. *La filiation des structures*. Paris, P.U.F. 1963, 131-183.
- [7] Seymour Papert. *Jaillissement de l'esprit. Ordinateurs et apprentissage*. Paris, Flammarion (in English: *Mindstorms, children, computers, and powerful ideas*, New York, Basic Books, 1980), 1981.
- [8] David Reinking. A nova alfabetização por (em) computador. *Letras de hoje* (Brasil), no.90, (Dec.1992), 99-112.
- [9] T.Slama-Cazacu. *Language et contexte*, The Hague, Mouton, 1961.
- [10] T.Slama-Cazacu. Remarques sur quelques particularités du message verbal employés dans l'automation. *Linguistics* (Holland), 1963, no.2, 60-84.
- [11] T. Slama-Cazacu. Comunicarea în procesul muncii ("Communication in the process of work"). București, Ed. Academiei, 1964a.
- [12] T.Slama-Cazacu. Problèmes psycholinguistiques posés par les messages verbaux employés dans l'automation. "*Revue de linguistique*", no.2, 119-130, 1964b.
- [13] T.Slama-Cazacu. The use of motor reaction time for the comparison of verbal and nonverbal stimuli, and verbalization in perception. *Revue roumaine des sciences sociales-Psychologie*, no.1, 81-88, 1966.
- [14] T.Slama-Cazacu. *Introduction to psycholinguistics*. The Hague, Mouton, 1973 (Romanian edition, 1968).
- [15] T.Slama-Cazacu. Nonverbal components in message sequence. In *Language and man*, ed. by W. McCormack, S.Wurm. The Hague-Paris, Mouton, 1976, 217-228.
- [16] T.Slama-Cazacu. Dialogue in children. The Hague-Paris, Mouton 1977 (Romanian edition 1961).
- [17] T.Slama-Cazacu. "Interpersonal relations and dialogue structuring: On "dialogic syntax". *International journal of psycholinguistics*, 1981, 8, no.4[24], 57-103.
- [18] T.Slama-Cazacu. "Niveau linguistique et décodage complexe en lecture". *Revue roumaine de linguistique/ Cahiers de linguistique théorique et appliquée*, 1989, 26, no.2, 123-140.
- [19] T.Slama-Cazacu. "Nonverbal components of communication and the field of applied psycholinguistics". (Plenary lecture -Kassel 1987). In *A case for psycholinguistic cases*, ed. by G.Appel, H.Deichert, Amsterdam-Philadelphia, Benjamins, 1991, 143-161.
- [20] T.Slama-Cazacu. "A new 'galaxy' in communication? Hypotheses for future studies and research on psycholinguistics", *International journal of psycholinguistics*, 1993, 9, no.2 [26], 215-230.
- [21] T.Slama-Cazacu. "Why a new framing for psycholinguistics: A multidisciplinary connected science?" (Key-note address "IVth International Congress of Applied Psycholinguistics" - Bologna 23-27 June 1994), in the *Proceedings*, 1994.
- [22] E.Wenger. *Artificial intelligence and tutoring systems. Computational and cognitive approaches to the communication of knowledge*. Los Altos, 1987, (*apud* EJPE 1992, 352).
- [23] N.Wiener. *Cybernétique et société. L'usage humain des êtres humains*. Paris, Union générale des éditions, 1971, (orig. ed. 1950).

Scheme 3

