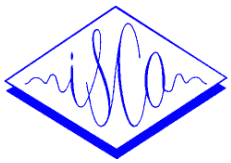

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

ORAL CULTURE IN THE 21st CENTURY: THE CASE OF SPEECH PROCESSING

Keynote speech at the 6th International Conference on Spoken Language Processing, Special Session on Speech Production Control, Bei Jing 16-20 October, 2000

Sven E. G. Ohman

Dept Linguistics, Uppsala university, Sweden

Abstract

In the 20th Century modern scientific research into speech and spoken language was launched by the pioneering efforts of men like Gunnar Fant, Ken Stevens, Al Liberman, Frank Cooper, and others. As this research progressed it had gradually to shake off the straight-jacket of traditional grammatical-linguistic conceptions, and to develop entirely fresh views. This paper will spell out some of the conceptual differences and discuss their probable consequences for the future in some detail.

10.21437/ICSLP.2000-9

The term "*oral* culture" in the title of this paper refers to the culture of *spoken* language, as contrasted with *written* language.

Of these two phenomena, speech, or spoken language, is primary both historically and conceptually. Since the implications hidden in the relationship between the two have kept me busy for a number of years in my job as responsible for the General Linguistics program at Uppsala university, I should like to take this opportunity to communicate some of my ideas on this rather fundamental matter.

WRITING

I think of the use of alphabetic script in Western writing systems as a *technique* applied in bringing an adequate writing tool (typically a "pencil") to work on some writing medium (typically a "piece of paper"), according to a code by which a mental representation of the sound of the speaker/writer's speech is interpreted as a linear string of *grámmata*, i. e. script signs. The encoding is made by the speaker/writer using her/his auditory sense and language competence, but without the aid of any external instruments besides the "pencil" and the "paper". The resulting text is to be readable by anyone mastering the same script code. This description seems to fit not only western, alphabetic writing systems but also others such as ancient and modern Chinese and Japanese scripts, ancient Egyptian hieroglyphics, Maya hieroglyphics, and many others. Here I focus on western alphabetic scripts, since they have been the source from which modern linguistics has picked up most of its ideas.

THEORY

This writing technique, script included, is informed by a system of specifically construed theoretical concepts which, together, constitute the *theory* of the script in question. Generally, since *mening is use*, a theory acquires its sense/meaning from its concrete applications. The theory of writing is, of old, also called *Grammar*

(from classical Greek *grámma* = script sign), In its modern version this theory has evolved into what is nowadays called *Linguistic Theory*.

A central point is that the vast majority of the most widely used grammatical-linguistic concepts of today are rooted in the gradual invention and practical employment of the *writing* technique. They are *ad hoc* conceptual constructions once worked out to facilitate the use of *letters* for purposes of linguistic communication.

I stress that grammatical-linguistic concepts such as the *letter (phoneme)*, the *word*, the *clause*, the *sentence*, the *paragraph*, and so on are NOT "structures already present in speech "by the very "physical nature of speech", The belief that they are, may be termed *realistic* in the technical sense in which this term is used to denote the confusion of concepts with things. The *bonus* of my approach is that we are under no obligation to conceptualize speech in the grammatical-linguistic way. On the contrary, as I try to show, we would do better to shake off the conceptual burden of contemporary linguistic theory! We may find other ways – *fresh* ways - that suit our purposes better. Digital coding is an excellent example with an out-and-out nonlinguistic theory.

LETTERS

In other words. Any representation of speech should be evaluated with respect to its performance in its practical applications! That is the true meaning of the concept of "adequacy" in linguistic theory.

Let me illustrate this point with two examples, the letter and the word.

The original use of a letter in script, say the letter *t* could roughly be said to be this: In writing down a speech utterance, write the letter *t* whenever you hear the speech sound we call *t*!

In formulating this description for print in the proceedings of this conference I obviously couldn't have replaced the phrase "the speech sound we call *t"* by the audible sound itself, since that sound is *not* a visible script sign which can be fixed on the page, but something that can only be heard. Of course that is the whole point of alphabetic script – to replace *audible* speech with *visible* script. Understanding script theory is understanding *this* among many other things. The letters of the alphabet were once *invented* by Semitic peoples in the Middle East, and this invention was at the same time a *construction* of certain useful script concepts, *e. g.* those we associate with our most common *speech sounds*. This construction created an *internal relation* between the visual letter shapes and the corresponding auditory sound impressions; "internal" in the sense that we cannot know one, e. g. a sound, without at the same time knowing the other, i. e. the letter. Clearly this construction could have been done in many different ways, but the one that has survived is motivated, i. e. it passes the adequacy test mainly by its success in implementing conventional writing and reading. This, of course, does not guarantee that it is adequate also for the speech processing purposes of our own digital-electronic age. We of course also have scientific notations for speech sounds available in phonetics and phonology, and again the formal stipulations of how these notations are to be used bring about new internal sound-sign relations. But these notations are still not adequate for electronic speech processing.

I would like to say that the description I have just given of the use of the letter *t*: "Write down *t* whenever you hear the sound *t!"* is typical of *all*

contemporary alphabetic signs, "Write down *p* whenever you hear the sound *p!"*, "Write down *k* whenever you hear the sound *k!"*,etc. - but that would obviously not be quite true. Capital letters have essentially the same uses as the corresponding lower case letters. But the difference in appearance is used for syntactic purposes and have no audible cues in the speech signal. Something similar may be said about script signs such as the *space*, the *comma*, the *period*, the *exclamation* and *question* marks, and some other signs as well. They are *script concepts* that we acquire in learning to play with these signs in writing practice. One might say that they are *non-phonetic* as against the *phonetic* letter uses such as that of the letter *t* just mentioned.

Note that what I have just termed "script concepts" are things you *do* in writing. To *have* such a concept is to have the *ability* to do the relevant thing. For instance, having the concept of the letter *t* is to be able to *write down* the letter *t* under the *appropriate auditory-phonetic circumstances*.

Moreover, the use of the letter *t* is not always elicited by the sound we call *t*. This letter may be written upon hearing other speech sounds also, as in the word *nation*, *n-a-t-i-o-n*, for instance. This is so in part for historical reasons, but also because in modern western script the letters are integral parts of lexical script words. Actually, in writing, speech is *not* coded primarily as a sequence of sounds associated to letters, but as a sequence of *words*, i. e. short sub-strings of letters forming characteristic visible units called *Bouma shapes* after the great Dutch psychologist of reading Herman Bouma. These shapes are visual units or "script Gestalts".

Before going into the matter of words we might note that a letter may have two or more phonetic uses, and that that which is commonly called a *phoneme* is usually *one* such use. A phoneme, as I understand it, and as usually talked about, is hence not necessarily anything *mental* but something thoroughly *practical*- a *use of a letter* in conventional writing. It is as misleading, as it is common to treat concepts as *being* mental images. Mental images may be helpful in guiding the *application* of a concept but the concept itself is not an image, but something more stable - a practical ability.

WORDS

Having at this point completed my account of what is involved in writing down the letter *t*, I now turn to the more adventurous task of explaining the writing of a *word.*

Erudite paleographers like *Paul Saenger* in his highly informative recent book *Space Between Words* (1) and *Malcolm Parkes* in his *Pause and Effect* (2) have shown in detail how the practice of dividing the string of letters into script words is the result of many centuries of experimentation starting relatively soon after the Phoenicians taught the Greeks to write, some six or seven centuries BC. The Greeks changed some of the Semitic consonant letters into vowel signs, a truly momentous step for Western Civilization, according to the American historian Eric Havelock(4). But they kept the Semitic practice of separating words by means of a special sign, the so-called *interpunctum*, a single dot. Soon the use of this sign was abandoned, however, and the Greeks started to employ so-called *scriptura continua* characterized by its paucity of punctuation - no word division, no sentence division, just a plain, long string of "samecase" letters in all texts! This was taken over by the Romans, and was the normal form for manuscripts well into the early Middle Ages. *Scriptura continua* was usually read aloud since

silent reading made parsing, and grasping the meaning of the text very difficult for the reader, since it forced the eye to jump back and forth over the line in whimsical saccades.

In the sixth Century AD the Roman grammarian Priscian gives a definition of the *word* (latin *dictio*) as the smallest meaningful part of a *sentence* (latin *oratio* - in everyday latin *oratio* meant *speech* – to "parse" means originally to detect the *parts* of speech. i. e. the words – *Priscian* does not tell us *which* the words of latin were however!). This definition suggests that by this time latin scribes had started to practice word division in preparing manuscripts, an assumption which is born out by Saenger's investigations. The conventions for what to include between the spaces that mark out a word was however a matter of experimentation for several centuries. As anyone even superficially familiar with modern acoustic phonetics will know there are no obvious acoustic cues in the acoustic speech signal to indicate any "natural" cuts corresponding to the word divisions of writing. The script units we call words have had to be stipulated conventionally, and we learn all words - all Bouma shapes along with learning to write and read. And when writing was first introduced into the various western so-called "vulgar languages"(i. e. *modern* languages) in the Renaissance and later the rules for word separation, for the collection of letters into individual words, had to be invented anew for each new language.

As Paul Saenger shows the string of letters making up a word normally fits into the foveal field of vision comprising about a 4° angle, and the next word or two on the line of writing are simultaneously noted in the so-called parafoveal (side vision) field of about a 15° angle. It is factors like these that have dictated the way words have been designed in our writing systems. Why, for instance can "today" and "maybe" be spelt as one word when "good morning" cannot? And why is French "ça y est!" *not* one word, but three? There are no "Bouma shapes" ready for our ears to pick up in the speech signal, so the word boundaries of writing must be stipulated *ad hoc* by tradition and education.

In summary the grammatical-linguistic concept of a *word*, though notorious in linguistics for its elusiveness is obviously among our deepest and oldest acquaintances in language. And we should beware of assuming that the structures imposed on speech in modern linguistics, e. g. morphemes, roots, junctures, and much more of the kind, though maybe linguistically motivated are any less arbitrary acoustically than the traditional ideas.

Since any representation of speech should be evaluated with respect to its performance in its intended practical applications, the conceptual construction of the word has proved able to increase the efficiency of mature readers' reading by orders of magnitude in comparison with reading *scriptura continua*.

It is worth noting, that the same type of mature reader reaches an even higher efficiency in reading Chinese script and Japanese Kan-ji.

A further advantage with standardized word division is that it makes *silent reading* easy, so easy in fact that practically any reader of modern text reads almost everything silently. Nowadays reading aloud *well* actually requires special practice.

An idea I try to elaborate elsewhere (5) is that the step from loud reading to silent reading as the normal practice in the West has forced the loud articulation of the text read to so to speak go

"underground". Reading has become a form of purely *mental speech.*

I believe that this fact carries a considerable historical responsibility for the contemporary mentalistic tendencies in linguistics. It would however lead too far to go into more detail with this matter here.

Let me repeat that any representation of speech should be evaluated with respect to its performance in its practical applications. Such an evaluation must give considerable praise to traditional alphabetic script. Our respect for its cultural prestige together with our tendency toward philosophical realism often tend to mislead us into taking for granted that the grammatical structures of linguistic theory are directly motivated by allegedly "corresponding structures" of the audible digital-acoustic speech signal. By the way, I invite all of you to listen carefully to the papers to be presented here to day, and to watch for confusions of the stuff of speech with the stuff of script! I believe you may detect many cases! Such confusions introduce metaphysics into speech research and actually flaws all attempts to measure the adequacy of the theory.

The theoretical superstructures of contemporary formalistic and nativistic phonology have their conceptual roots in the Prague school notions of Phonemes and Distinctive Features which, as I try to show elsewhere, are a kind of logical empiricist style "rational reconstruction" of the phonetic theory implicit in the alphabet.

In the form of formalistic phonology Grammar has exerted an influence on the thinking in speech research for about half a century, an influence that often has *not* been terribly helpful. Let me give two examples: *phoneme invariance* and *coarticulation*.

The considerable efforts that have been spent on finding the allegedly "acoustic invariants" presumed to somehow reside in the "physical matter" of speech sound have – it has to be admitted - largely failed, as could have been predicted from a consideration of the nature of the concepts that motivated the hypothesis of the physical existence of such "invariants". In all likelihood the motivating idea is the alphabetic concept of phoneme according to which a phoneme is *a phonetic use of a letter,* e. g. the use of the letter *t* when one intends the sound we call *t*. This use of the letter is invariant in the sense that it is always the same in writing the sound we call *t*. But from this fact the physical-acoustic invariance of the sound is a *non-sequitur,* and in fact a highly unlikely hypothesis from a phonetician's point of view!

Such facts have however in no way prevented researchers from pushing the hypothesizing upward into metaphysical spheres using the language of biology, and in particular that of genetics. Thus it is hypothesized that phoneme invariants nevertheless have their whereabouts in the higher levels of the auditory nervous system!

Here we have a nice example of how a confusion of languages belonging to quite different levels of analysis, that of practical linguistics and that of modern science, may trap us into metaphysical confusion.

A similar story may be told about so-called coarticulation theory. And here again I must count my own old self among the worst sinners!(7)

For, traditionally coarticulation simply means that two consecutive phonemes in the chain of speech are produced in such a way by the speech organ that one phoneme colors the other, or so that they overlap to some degree. This is in any case how literate phoneticians have experienced the

phenomenon subjectively.

My description implies that coarticulation presupposes phonemes, It is *phonemes* that are coarticulated. Without phonemes our theory has no object. Still it may have *some* kind of substance in the form of lots of carefully performed measurements that may one day become useful. The ensuing progress for speech processing endeavors such as text-to-speech, or speech-to-text systems is however probably rather slight.

FUTURE

If I were young today, I would concentrate my energies on contributing to the working out of methods for *writing in audible sound* directly based on the best contemporary digital-acoustic representation of spoken language, i. e. to develop a modern *acoustic* writing of *speech for the computer.* I wouldn't let abstruse linguistic theories rule over me too much, though I might not ignore them totally. In that way speech research could probably acquire a leading part in the spoken language race.

ACKNOWLEDGEMENTS

I should like to express my gratitude to Alejandro Engelmann of Uppsala university for his competent assistance with this paper.

REFERENCES

(1) Paul Saenger *Space Between Words. The Origins of Silent Reading,* Stanford UP 1997

(2) Malcolm Parkes *Pause and Effect. An Introduction to the History of Punctuation in the West,* Universitet. Calif. Opress 1993

(3) Walter Ong *Orality and Literacy. The Technologizing of the Word.* Methuen 1982

(4) Eric A. Havelock *Preface to Plato* Cambridge, Mass., cop. 1963, pr. 1982

(5) S. Ohman *Writing and Speech* (in Swedish) Ms Dept of Linguisics Uppsala university 1999

(6) S. E. G. Ohman *Coarticulation in VCV Utterances,* J. of the Acoustical Soc of America #39. 1966, and *A Numerical Model of Coarticulation,* J. of the Acoustical Soc of America #41,