**Carlo Aliprandi – Synthema - I**

# Advances in HLT: can Assistive Technology help professional reporting and text processing?

**Intersteno Congress - Prague July 2007**

## Carlo Aliprandi - Profile

He is a researcher and an industry expert in the area of Natural Language and Speech research, technology and applications. His research and development has pioneered a number of techniques to speed up human tasks like speech recognition for subtitling and reporting, text entry and fast typing.

From 1997 he is in charge of the Speech Recognition Research, leading the technical development of Voice SubTiling, the first speech subtitling system that was later adopted by RAI Radiotelevisione Italiana for live and deferred subtitling of sport and news events. He pioneered speech recognition techniques for report drafting of parliamentary proceedings, supporting the Italian Chamber of Deputies in adoption and deployment of CameraVox. CameraVox is a LVCSRS (Large Vocabulary Continuous Speech Recognition System) based on a language model of continuous speech inspired by parliamentary documentation.
From 2000 he is in chief of the Speech Solutions Labs and follows the strategic development of Voice Suite, the first distributed and multilingual speech reporting system. Voice Suite allowed the italian Fabrizio G. Verruso to become world champion of Voice Recognition at the Intersteno 2005 championships, setting the new world record of 174 words per minute.

**Advances in HLT: can Assistive Technology help professional reporting and text processing?**
**Carlo Aliprandi – Synthema**

It is for me a pleasure to attend this event and I'd like to thank all the organization committee, the Board, the Central Committee and the Scientific Committee of Intersteno. A thanks goes also to the Italian delegates, particularly to dr. Fausto Ramondelli and to Gian Paolo Trivulzio, who invited me to hold this lecture on what we are currently exploring in relation to typing and fast reporting.

Let me, first of all, to introduce my company and myself. Basically I am a technician, I work for an Information Technology company as a researcher in Natural Language and Speech technologies. In these years we have been pioneering a number of techniques to improve human tasks like speech subtitling and speech reporting, text entry and fast typing.

My company, Synthema, is an Italian Small to Medium Enterprise and has always been at the forefront of advancements in many applications related to Language Technologies.

But what are Language Technologies? I took the following definition from Wikipedia:

*"Language Technology is often called Human Language Technology (HLT) or Natural Language Processing (NLP) and consists of Computational Linguistics (CL) and Speech Technology as its core but includes also many application oriented aspects of them. Language Technology is closely connected to computer science and general linguistics."*

In the last 10 years we contributed to the development of *application oriented aspects* like Machine Translation, Automatic Speech Recognition and Live Subtitling, Language Understanding and Semantic Web.

In these years we have been working to bring to the market top technologies available in research labs, with a particular effort into making them "solutions for everyday life", for different kind of users, applications, and working conditions.

So it's an honour for me to speak to this community of professionals that in recent years has been looking at Information Technology giving special attention to its innovations.

In this lecture I will address recent advances in Human Language Technologies, particularly in relation to automatic Word Prediction methods and fast text entry.

Word Prediction is the task of guessing words that are likely to follow a given fragment of text. A Word Prediction software is a writing support: at each keystroke it suggests a list of meaningful predictions, amongst which the user can possibly identify the word he meant to type. By selecting a word from the list, the software will automatically complete the word being written, thus saving keystrokes.

Word prediction is facing a very ambitious task, as guessing and completing what word a user is willing to type is very complex.
It is complex first of all because natural language is complex. It is quite common to express in many different written forms the same concept or thought, in the same way as it is common – for example, in reporting- to produce different sentences upon the same speech: authors and reporters typically uses different words, styles, depending on their sensibility, experience, culture, education and so on.
Word prediction is complex due to usual ambiguities related to natural language that make it difficult to 'understand' what the user is willing to type. The inherent amounts of ambiguities (lexical, structural and semantic ambiguities but also pragmatic, cultural and even phonetic ambiguities for speech) that arise are complex problems to be solved by a computer.

For example, in the case of a lexical ambiguity, a word can be associated with more than one Part of Speech and the sentence can have more than one interpretation and meaning:

*the old man boats in the lake*

Can mean either "an old man travelling by boat" or "the boats of the old man are staying in the river".

When I speak about Word Prediction, the very first question that comes to mind if there are differences with respect to Mobile Phones typing predictors.
Before giving an answer, let's consider what functionalities are available or at least perceived as available from users in existing systems.
Letter prediction methods have become quite known as largely adopted in mobile phones and PDAs, where multitap is the input method. Multitap is the typing process that each user uses when, for example, composing a SMS. Multitap allows the user to write the word

*Hello*

By typing/tapping

44 33 555 555 666

on a 9 keys keyboard
Please note that a total a amount of 13 keystrokes is necessary to write a 5 chars word, without using any prediction method.
Commercial systems as Tegic Communications T9, Zi Corporation eZiText, and Motorola Lexicus iTAP are all successful systems that adopt a very simple method of prediction based on

dictionary disambiguation. At each user keystroke the system selects the letter between the ones associated with the key guessing it from a dictionary of words: hence they are commonly referred to as letter predictors.

Several measures and metrics exists for evaluating performances of prediction methods, a well-known metric is Keystroke Saving (KS):

$$KS = \frac{K_T - K_E}{K_T} \times 100$$

KS estimates the percentage of saved keystrokes, calculated by comparing the total number of keystrokes needed to type the text ($K_T$) and the effective number of keystrokes using word prediction ($K_E$).

Letter prediction brings a Keystroke Saving but it has been proven to be dependent from ambiguities that are more frequent for inflected languages. So it is not surprising that these methods had a great success for non inflected languages such as English.

But what are inflected languages? A language is inflected when it is possible to produce morphological forms from a root or lemma and a set of inflection rules. The degree of inflection of a language may vary from very high (e.g., Basque), to moderate (e.g., Spanish, Italian, French), to low (e.g., English).

Typical limitations of Letter Prediction especially when applied to inflected languages are that

- it doesn't have a sufficiently large dictionary to cover the user lexicon
- it doesn't "understand" what the user means to write as it doesn't understand the previous context.
- it doesn't complete the current word, it often limits to propose words for the typed keystrokes and the full typing of the word is often required.

So the answer to the question if Word Prediction is different from Mobile Phones letter predictors should be yes: Word Prediction has to be better, more usable and flexible.

Differently from Letter Predictors, Word Predictors typically use refined language resources in order to predict a full word instead of a single letter. Word prediction - in recent years - has received a growing interest and has been object of a flourishing research.

A very basic method for Word Prediction is, again, employing a dictionary of words combined with probabilities or frequencies for each word in the dictionary. Most frequent words in the dictionary are suggested. Suggestions may often appear inaccurate since many words may be

missing from the dictionary or words in the dictionary don't fit to the user lexicon and to its current domain. Word frequencies, moreover, do not take the sentence structure into account. Thus it is not surprising to find Word Predictors suggesting in the context:

*Tomorrow I w….*

A list of suggestions like this

1. *want*
2. *we*
3. *when*
4. *why*
5. *who*

Once more, as for letter predictors, most of research related to word prediction concerns non-inflected languages. A survey we conducted at the time of the start of our research showed that about 90% of available systems originates from English speaking countries, thus it is not surprising that when those technologies are transferred to other languages there is a loss in performances.

Therefore to offer the same level of performances for inflected languages, more sophisticated methods taking into deeper account the specific features of each language had to be designed, basically to enhance the predictive accuracy.

The central assumption of these methods is that it is useful to incorporate some syntactic and semantic information related to the dependencies between words. The assumption could be simplified by the assertion

*"contextual information affects the word to be predicted"*

So, in order to predict the most likely word it is necessary to have a high-order representation of the context: the task of word prediction can be modelled as the estimation of the probability to guess the nth word ($w_n$) given the current sequence of $n-1$ previous words ($w_1, w_2, \ldots, w_{n-1}$). This probability is denoted by:

$$\Pr(w_n \mid w_1, w_2, \ldots, w_{n-1})$$

Thus, to get back to the previous sample, the task of word prediction in the context

*Tomorrow I w….*

Becomes the task of finding the word that maximises the following probability:

$$P(w_n \mid I, Tomorrow)$$

Where $w_n$ wn is a word starting with the letter *w*

Simplifying, the Word Prediction task becomes the task of selecting the word $w_n$ among all the words with initial letter *'w'* that are more probable to appear after the word *"tomorrow"* and after the word *"I"*

With that in mind and with the necessary information available to a Word Prediction system, the most likely word that has to be predicted is

*will*

and not just the most frequent word in the list *want, we, when, why, who.*

So, after some studies on state of the art for non-inflected languages, in these last years we considered to further develop this innovative method. I'm going now to detail you its application to word prediction for inflected languages.

We developed FastType, a Word Prediction software specifically intended for disabled people. FastType is an innovative system for word and letter prediction based on combined statistical and lexical methods, relying on robust language resources.

FastType is the result of a collaborative RD project, supported by Synthema, the Department of Computer Science of the University of Pisa and the Service Unit for the Support and Integration of Disabled Students of the University of Pisa. FastType receives support and funding from Fondazione Cassa di Risparmio di Pisa.

It is typically proven that, for inflected languages reasonable KS is obtained limiting the number of words. In this project we outclassed this limitation showing that a significant KS above 40% can be offered and that, surprisingly, slightly better results have been proven for a large dictionary (of about 1.200.000 words) than for a limited dictionary (of 250.000 words).

Before detailing our results, let me introduce the specific intent FastType was conceived for, that is supporting people with disabilities.

Information technologies in modern society are getting more and more central in everyday life and it is commonly accepted that they are becoming so pervasive that they do influence vital factors like participation, communication, learning and interaction.

People with disabilities (physical, cognitive or even learning deficiencies) in many situations may have their communication and interaction ability reduced due to difficulties when using conventional communication modes. So one of the problems that these people have to overcome is their inability to communicate normally.

Information Technologies per se doesn't help to overcome this inability, on the contrary there is a common concern regarding the divide generated by Information Technologies for the wider community: the level of adoption of Information technology is still far from what happened for communications devices like telephones or mobile phones.

The gap between universal access and the real adoption of Information Technologies is referred to as Digital Divide and it represents a new form of exclusion and non-participation, namely, non-participation in the information society.

For disabled people, often, non-participation translates into non-communication. Considering that about 10% of European population is disabled, you can understand why the concern of reducing Digital Divide is become central to European Policies.

Assistive Technology (AT) has therefore become an essential support for people with disabilities. Assistive Technology refers to the use of technology to improve the abilities in performing functions that may be difficult or impossible to perform without help. Assistive technology enhances the rate of communication for people with physical and cognitive impairments.

Thanks to AT systems, people who cannot write or speak, obtain alternative means of communication. This is the case of a large group of people with speech and motor disabilities provided with alternative input systems in order to overcome their limitations.

Whereas in a normal conversation some 150/180 words per minute can be said, a disabled person can produce some 10/20 words per minute without the support of AT. An experienced typist can produce about 300 keystrokes per minute, a disabled person can produce some 20 keystrokes per minute.

Therefore Word Prediction has rapidly become a very used technique to enhance the rate of communication for physical and cognitive impaired people, proving to be vital for those persons who rely on typing to communicate.

We started our work with a group of disabled people that, before the introduction of an AT device, used to communicate pointing (typically using their nose) on a matrix of characters. In our case the matrix was hand-painted on a communication board (of glass, of plastic): the interlocutor had to get the pointed character and to build, character after character, a word. It's clear that, after a certain number of characters both the interlocutor and the disabled person had to interact in order to guess the word and then the sentence.

Later came the virtual keyboard, that is the 'software' translation of the communication board. A virtual keyboard is a software that acts as a virtual extension of a physical device with fewer buttons than a keyboard and a navigation or scanning method to select on-screen characters: a

virtual keyboard can be operated with a computer mouse or with a device like a joystick, a foot-button or an eye tracker.
Eventually Word Prediction has come as a natural complementary extension of virtual keyboards: Word Prediction integrated into a virtual keyboard becomes a real powerful tool that can substantially improve the communication rate. The final part of this contribute will motivate this assertion detailing results we got.

It is commonly proven that evaluating Word Prediction is difficult as many different metrics exist. Nevertheless, most of literature presents prediction results expressed in terms of the Keystroke Saving (KS) that, as already stated, estimates the percentage of saved keystrokes using Word Prediction.
There are two additional metrics we used to evaluate prediction results for FastType: Word Type Saving (WTS) and Keystrokes Until Completion (KUC).
Word Type Saving estimates the percentage of time saved by writing using Word Prediction. Keystrokes Until Completion estimates the average number of keystrokes before the desired word appears in the prediction list.
A parameter greatly influencing performance measurements is the length L of the prediction list, that is the number of suggestions presented to the user. FastType average results, for the Italian language, with values for L 5, 10, and 20 are shown in the following table:

| **L** | **KS** | **WTS** | **KUC** |
|---|---|---|---|
| **5** | 41,15% | 21,12% | 2,85 |
| **10** | 45,26% | 24% | 2,67 |
| **20** | 47,9% | 24,35% | 2,48 |

The average KS is among 41% and 48%, and the increase in KS, between 5 and 10 is way more relevant than the increase between 10 and 20. The same happens for WTS and KUC.
Performances are significantly good for WTS, meaning that –at a standard speed and without any added cognitive load- saving in time is average around 24%.
Particularly significant is also the KUC, meaning that the correct word is suggested after an average of 2.8 for L=5, 2.7 for L=10 and 2.5 for L=20.
The following figure presents a sample text:

| |
|---|
| Credeva di compiere un gesto di solidarietà ed ha invitato a cena due giovani vagabondi. Ma il gesto si è trasformato in un incubo: i soggetti le hanno somministrato del sonnifero e poi hanno derubato l'appartamento. |

Predicted keystrokes, yellow marked, are 103 out of a total of 218 keystrokes, thus producing, in this case, a KS of about 47%.

We met the Intersteno community in 2003 in Rome, presenting for the first time Synthema Voice Suite, a system for multilingual speech reporting. Some years later that software, became world champion at the Intersteno Congress of 2005, when Fabrizio Gaetano Verruso set the new Speech Recognition world record with 174 words per minute. That was a very interesting experience and represents a result of connections created between the Intersteno Congress and the world of research and industries.

The system was also adopted by A.S.FOR professional course "Multimedia Reporter expert in Subtitling", where it was employed by the teachers (Trivulzio and Crippa) to train speech reporters for live subtitling.

There is a common ground between Speech Recognition and Word Prediction that is of interest to take into consideration. One of the core motivations that pushed many companies (big companies like IBM or SCANSOFT/NUANCE) to develop Speech Recognition was its application for disabled people. That kind of users rapidly adopted it and worked as "catalyst" for further refinements and also for largely diffusing the benefits of Speech Recognition to other domains and users.

Similarly Word Prediction was born for disabled people, so we envision it can be extended to other domains and users in the same way.

For example Word Prediction can become a daily support for people in everyday activities, like writing a letter, writing an email or an SMS.

But also Word Prediction can be helpful for the so called "proficiency professionals", like attorneys and lawmen, translators and practitioners, that is all those professional people that make large use of text entry in their job.

And, finally, Word Prediction can be helpful for professional reporters, for whom text entry and text production "is" the job.

If is it true that among professional reporters almost 40 to 50 %[1] in reporting activity is currently performed with word processing and keyboarding, then we envisage a future where also most of professional reporters could get benefits out of this technologies.

But what is the awareness of this technologies and what is the level of integration between Word Prediction and, for example, Stenotyping?

We run some web searches in order to evaluate the level of interaction between these two worlds, finding an interesting "raw" result.

We run a first web search using the keyword "word prediction", finding lots of pages, namely 190.000. We run a second search with the keyword "stenotype", finding the same order of magnitude of citations (117.000). Then we run a third search with both keywords, i.e. "word prediction" and "stenotype", finding a very limited number of pages, only 16.

So the final conclusions are that we are at the very beginning of a new path, where an effort is necessary in order to move foreword the technology into a new domain, whose users we envision could take many benefits.

Nevertheless we are aware that most of the work is still to be done, especially in terms of ergonomics and of evaluation of practical benefits – for example it is true that a great saving in keystrokes doesn't necessary correspond to a saving in time- but we entrust, as happened for example for Speech Recognition that the contribution of all the players are needed: technicians, domain experts, professional users.

So, concluding, we all are here to put a seed for this future.

[1] G. Trivulzio, Accademia Aliprandi – ITC-IRST Congress: "Reporting: expertise, techniques, organizations". Povo, Trento Italy 12/02/2006