Towards an universal counting system for Intersteno competition texts

Contribution of Mr. Jean-Charles LE MASSON, Intersteno France,

To the scientific committee of Intersteno

July 2017

As you know, I have produced a report about the various ways to find a new universal system to count texts for Intersteno competitions.

In this report, which revealed the weaknesses of the current system, I evoked several ways based on the spelling, on the syllables, or on the steno strokes. But none of them presented guarantees stronger than the current system.

*I arrived at the conclusion that it was necessary to forget spelling or syllables, which generate a lot of questions on the way to define them, and that the solution could consist to be closer to **the initial signal emitted by the speaker**: the sound, **the phoneme**.*

*An **international phonetic alphabet** already exists since a long time, applicable to all languages. You can see these particular codes in all dictionaries, showing how to pronounce a word.*

My report was to promote this way which appeared to me –and remain today- as the most relevant way to setup in the future a text counting which will be perfectly fair.

With an automatic translator I found on Internet, it is very easy to compare the phonetic transcription of several languages, in order to determine for every language a "rate of phonemisation".

☞ **It is the rate between the phonemes quantity and the corresponding spelling characters quantity.**

i.e. :

human	hjuːmən
5 characaters	6 phonetic codes

Below, as an example without general ambition, an illustration realized from a sentence: the first article of the Universal Declaration of Human Rights:

Language	Text	Characters Count A	International Phonetic Alphabet	Phonem Count B	Rate A/B
CHINESE	人人生而自由,在尊严和权利上一律平等。他们赋有理性和良心,并应以兄弟关系的精神相对待。	43	zənzən ʂən ɑɻ tsɿjiɻʊ,tsaɪ tsuənjiɛn xɯʌ tɕʰuanli ʂaŋ jily pʰintəŋ. tʰamən fu jiɻʊliɕin xɯʌ liaŋɕin,piŋ jiŋ ji ɕiʊŋti kuanɕi tə tɕiŋʂən ɕiɑŋtu̠ɛɪ̠ taɪ̠.	144	0,298
ENGLISH	All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.	141	ˈɔl ˈhjuːmən ˈbiːɪŋz ɑr ˈbɔrn ˈfriː ənd ˈiːkwəl ˈɪn ˈdɪgnətiː ənd ˈraɪts. ˈðeɪ ɑr ɪnˈdæʊd ˈwɪð ˈriːzən ənd kənˈsaɪəns ənd ʃəd ˈækt ˈtoʊərdz ˈwən əˈnəðər ˈɪn ə ˈspɪrət əv ˈbrəðərˌhʊd.	153	0,921
FRENCH	Tous les êtres humains naissent libres et égaux en dignité et en droits. Ils sont doués de raison et de conscience et doivent agir les uns envers les autres dans un esprit de fraternité.	153	tu lez- ɛtʁ ymɛ̃ nɛs libʁ e ego ɑ̃ diɲite e ɑ̃ dʁwa. il sɔ̃ dwe də ʁɛzɔ̃ e də kɔ̃sjɑ̃s e dwav aʒiʁ lez- œ̃ ɑ̃vɛʁ lez- otʁ dɑ̃z- ɛn-ɛspʁi də fʁatɛʁnite.	120	1,275

In this board, we notice that the Chinese is a very compact language, because for 144 pronounced phonemes, are needed only 43 ideograms (a rate spelling vs phonetics of 0,298) at the same time, for French, 120 phonemes will be transcribed by 153 spelling characters (a rate spelling vs phonetics of 1,275).

By redoing these calculations, not on a single sentence, naturally, but on a corpus of text of several million words, we should be able to refine these rates until they are considered as exact rates for every studied language.

Then, we can realize a board which will reveal the relative rates between the languages, by taking naturally a reference basis (BASIS 100), for example the English language, and by placing all other languages with regard to English:

			Base 100
ENGLISH	0,900	0,900/0,900x100	100,00
CHINESE	0,298	0,298/0,900x100	33,11
FRENCH	1,275	1,275/0,900x100	141,67

Then, 1000 characters of my English basic text, will correspond to 331 Chinese ideograms only, and to 1416 French characters.

Three texts, of different spelling length, will count however the same number of phonemes.

IT WAS THE PURPOSE OF THIS STUDY …

Then, it will be easy to build **a new table** (columns), which will determine **for every minute** and **for every language,** the number of required **spelling** characters.

Here is what could look like a new table (datas are simple estimations, not definitive):

		ENGLISH			CHINESE		FRENCH			
		Nbr characters	Nbr phonems	Nbr characters	Nbr phoneme		Nbr caracters	Nbr words/min.	Nbr syllabes	Nbr phonems
Incrément par min.		36	=Nbr car /0,9	12	=Nbr car /0,298		51			=Nbr car /1,275
C	1	326	361,7	108	361,7		461	92	166	361,6
	2	361	401,5	120	401,5		512	102	184	401,3
	3	397	441,3	132	441,3		562	112	202	441,1
	4	433	481,1	143	481,1		613	123	221	480,9
	5	469	520,9	155	520,9		664	133	239	520,7
B	6	505	560,7	167	560,7		715	143	257	560,4
	7	540	600,5	179	600,5		765	153	275	600,2
	8	576	640,3	191	640,3		816	163	294	640,0
	9	612	680,1	203	680,0		867	173	312	679,7
	10	648	719,9	215	719,8		917	183	330	719,5
A	11	684	759,7	226	759,6		968	194	349	759,3
	12	719	799,4	238	799,4		1019	204	367	799,1
	13	755	839,2	250	839,2		1070	214	385	838,8
	14	791	879,0	262	879,0		1120	224	403	878,6
	15	827	918,8	274	918,8		1171	234	422	918,4

You can see that the phonems number is the same for all languages : the goal is reached

But all this is only **a very incomplete pre-study**.

We now have **to set up a task force** composed, for example, of 2 or 3 members of the Scientific Committee, plus some experts (linguists)

Carry out a more in-depth study, which:

1. Either **confirm the hypotheses** put forward by the pre-study and, in this case, must propose a concrete solution that can be implemented easily.

2. Either **the hypotheses** will be **invalidated,** and in this case:

    a) Will propose alternative tracks

    b) Or will abandon the project

In summary, the works which remain to make would be the following ones:

1. **Realize the translation in international phonetic characters** of a significative mass of text for every language susceptible to be used in competition

2. **Calculate, or refine, the rate spelling / phonemes** of each of these languages

3. From these rates spelling / phonemes, **establish a table of relativity between the spelling** volumes of every language

4. Establish the **table of characters per minute** for the various languages and the various minutes (in speech capturing and realtime)

5. Make all the necessary **discussions and tests** to validate and refine all the results

6. Realize **trainings** in real situation

7. Establish a **final report**

This study represents a **considerable workload** involving several people over a considerable period of time.

Therefore, *if it should lead to the validation of a new counting system*, **it is imperative that the Board undertakes to submit this system to the discretion and vote of the Jury and the Federal Council.**

I have no doubt that the honorable members of this Scientific Committee will understand this requirement and, moreover, I hope that some of them will be willing to participate in the task force to be created.

THANKS FOR YOUR ATTENTION