

Transcription and Annotation of Spoken Corpora

Claudio Bendazzoli
claudio.bendazzoli@unito.it

University of Torino (Italia)
School of Management and Economics
Department of Economic and Social Studies

Outline

- Transcription
 - What
 - How
- Annotation
- Representativeness
 - DIRSI-C (*ethnographic approach*)

TRANSCRIPTION in CL / CIS

- No international standard
- Spoken language features
- Machine-readable vs. User-friendly
- Alignment

Verba volant, scripta manent

- G. Jefferson's system
- Conversation Analysis
- Dialogue, prosodic features

: vowel / consonant lengthening
= latching
(.) pause
° xxx ° lower volume

people have been warm and friendly the the food has been wonderful the city is beautiful and I think my children have tasted every single flavour of gelato that there is ok so I am going to talk about cystic fibrosis-related diabetes



Making transcription...

Machine readable

User friendly

Transcriber / Annotator friendly!

Replications
Data-sharing

transcription [should] be limited to features to be subsequently analyzed, [which] in turn are determined [...] by the specific purpose of the research project [...]: only what is to contribute systematically to data analysis should be transcribed, and only what makes the data analysis intelligible should be presented

(O'Connell and Kowall 1994, as quoted in Setton 2011)

TEI
www.tei-c.org

■ Text Encoding Initiative:

- international **standard?**
- attributes (XML tags)
- header (contextual info) + ‘text’

CIS: Cencini 2000

elementi inseriti come attributi portatori di informazioni di varia natura (Cencini 2000, pp. 193 e ss.):

```

<?xml version="1.0"?>
<?xml-stylesheet href="TIC.css" type="text/css"?>
<!-- The TIC.dtd was compiled using the website
www.hcu.ox.ac.uk/TEI/newpizza.html;
the tagsets chosen are
base: transcribed speech
additional tagsets: linking, corpus and names and dates.
--&gt;
&lt;!DOCTYPE TEI.2 SYSTEM "TIC.dtd"&gt;
&lt;TEI.2&gt;

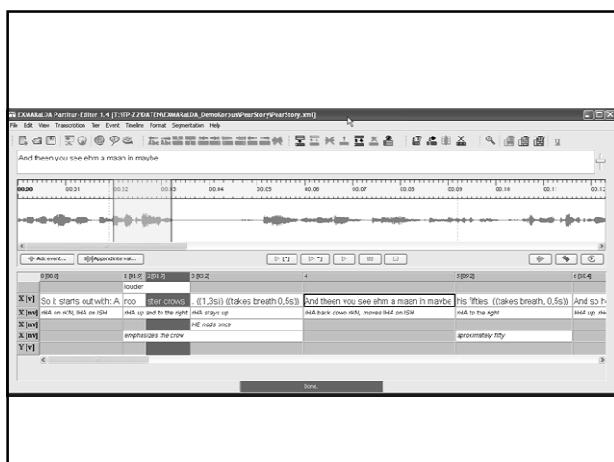
&lt;teiHeader&gt;
    &lt;fileDesc&gt;
        &lt;titleStmt&gt;&lt;title&gt; The Verona Initiative - an electronic
transcription&lt;/title&gt;
        &lt;respStmt&gt;&lt;resp&gt;transcribed and encoded:&lt;/resp&gt;&lt;name&gt;Marco
Cencini&lt;/name&gt;&lt;/respStmt&gt;&lt;/titleStmt&gt;
        &lt;extent&gt;words: 1869 kb: 14&lt;/extent&gt;
        &lt;publicationStmt&gt;&lt;authority&gt;release authority:
SSLIMIT&lt;/authority&gt;&lt;availability status="free"&gt;&lt;p&gt;Available for
purposes of academic research and teaching
only&lt;/p&gt;&lt;/availability&gt;&lt;/publicationStmt&gt;
        &lt;sourceDesc&gt;&lt;recordingStmt&gt;&lt;recording&gt; &lt;equipment&gt;&lt;p&gt; Audio
recorded during the conference "The Verona
Initiative".&lt;/p&gt;&lt;/equipment&gt;&lt;/recording&gt; &lt;/recordingStmt&gt;
        &lt;/sourceDesc&gt;
    &lt;/fileDesc&gt;&lt;encodingDesc&gt;
        &lt;classDecl&gt;&lt;taxonomy&gt;
            &lt;category id="mod1"&gt;
                &lt;catDesc&gt;consecutive&lt;/catDesc&gt;&lt;/category&gt;
            &lt;category id="mod2"&gt;
                &lt;catDesc&gt;simultaneous&lt;/catDesc&gt;&lt;/category&gt;
            &lt;category id="mod3"&gt;
                &lt;catDesc&gt;chuchotage&lt;/catDesc&gt;&lt;/category&gt;
            &lt;category id="pos1"&gt;
                &lt;catDesc&gt;on-screen interpreter&lt;/catDesc&gt;
            &lt;/category&gt;
            &lt;category id="pos2"&gt;
                &lt;catDesc&gt;off-screen interpreter&lt;/catDesc&gt;
            &lt;/category&gt;
        &lt;/taxonomy&gt;&lt;/classDecl&gt;&lt;/encodingDesc&gt;
    &lt;profileDesc&gt;&lt;creation&gt;&lt;date&gt;
        24 Sept 2000&lt;/date&gt;
    &lt;/creation&gt;&lt;langUsage&gt;&lt;language id="eng"&gt;
        english&lt;/language&gt;
        &lt;language id="ita"&gt;
            italian&lt;/language&gt;
    &lt;/langUsage&gt;&lt;/profileDesc&gt;
    &lt;particDesc&gt;
        &lt;person id="bCR" sex="f" role="speaker"&gt;
            &lt;persName&gt; Cathy Reed&lt;/persName&gt;
            &lt;firstLang&gt; English&lt;/firstLang&gt;&lt;/person&gt;
        &lt;person id="iPS001" sex="f"
role="interpreter"&gt;
            &lt;persName&gt; Cristina Mazza &lt;/persName&gt;
            &lt;firstLang&gt;Italian&lt;/firstLang&gt;&lt;/person&gt;
        &lt;/particDesc&gt;
    &lt;settingDesc&gt;&lt;setting&gt;
        &lt;p&gt; The conference was held in Verona
        on 7-8 July 2000.&lt;/p&gt;
        &lt;p&gt; This transcription refers to the first speech
        of the open session of the concluding day
        of the conference.&lt;/p&gt;
        &lt;p&gt; The speaker was reading from her notes.
        &lt;/p&gt;
        &lt;p&gt; Interpreters were not given the texts. &lt;/p&gt;
        &lt;p&gt; They were working in groups of two&lt;/p&gt;
        &lt;p&gt;The interpretation was delivered through the
        simultaneous mode.&lt;/p&gt;
    &lt;/setting&gt;&lt;/settingDesc&gt;
    &lt;textClass&gt;&lt;catRef target="mod2"/&gt;
    &lt;/textClass&gt;&lt;/profileDesc&gt;
&lt;/teiHeader&gt;
&lt;text&gt;
    &lt;body&gt;
        &lt;div1&gt;
        &lt;gap desc="material missing"/&gt;
    &lt;u who="bCR" id="u1" lang="eng" corresp="u2"&gt;
        thank you &lt;aanchor id="s1" synch="s3"/&gt;&lt;pause/&gt;
        my name is Cathy Reed and I'm a director of the Board of the North West
        England regional development Agency &lt;pause/&gt;
        with ehm lead responsibility for health and regeneration &lt;pause/&gt;
        I chair the North West Health Partnership &lt;pause/&gt;
        and my background is political leader in local government &lt;pause/&gt;
    &lt;/div1&gt;
&lt;/body&gt;
&lt;/text&gt;
</pre>

```

Tools

- **Transcriber** <http://trans.sourceforge.net/>
 - **Transana** www.transana.org
 - **SpeechIndexer** (Szakos & Glavitsch, 2004a, 2004b, 2007)
 - **Winpitch** <http://www.winpitch.com/>
 - **Exmaralda** www.exmaralda.org

10



TRANSCRIBING (WHAT)

transcription is a *selective process*

annotation of **basic features**

machine-readable >><< user-friendly

linguistic: orthographic transcription,
no punctuation signs

paralinguistic: disfluencies + pauses
(unfinished words, mispronounced words)

extralinguistic: header
(file, speech, speaker)

Annotation and encoding

■ **Time-tagging:** Transana (2.12) www.transana.org

■ **POS-tagging:** TreeTagger (Schmid 1994)

■ * **Corpus WorkBench** CWB-CQP (Christ 1994)

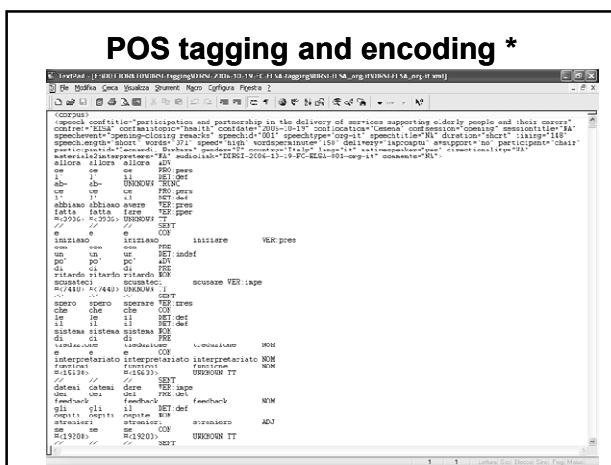
■ ** **C-ORAL-ROM** (Cresti & Moneglia 2005)

Spanish corpus (Moreno & Guirao 2006)

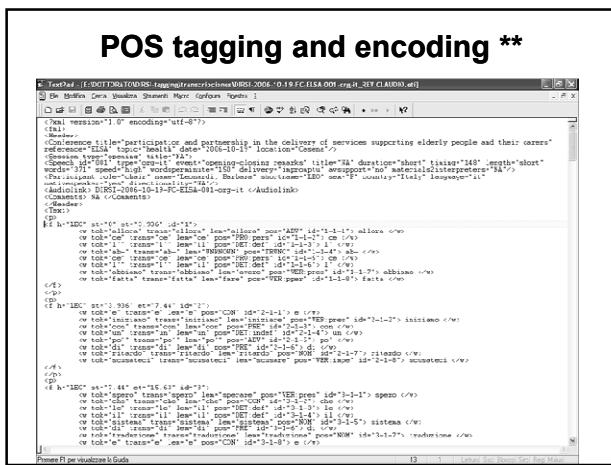
CHIEDE (Garrote 2008)

Time tagging (text-sound alignment)

DIRSI-2006-05-20-VR-CFF4-001- org-it [header]	DIRSI-2006-05-20-VR-CFF4-001- int-it-EN [header]
Gabriella può chiudere le porte là in cima per favore n<4999> // incominciamo // buongiorno e benvenuti a tutti a questo quarto seminario detto di primavera che è un appuntamento annuale n<20756> // diciamo dedicato a a a mettere a fuoco alcuni temi centrali della fibrosi cistica n<29415> temi di ricerca temi di cura di assistenza e di prospettive sul da farsi n<35724> //	could you close the doors at the top of the room please n<4046> // good morning and welcome to everybody n<10428> // welcome to this fourth spring seminar which is an annual meeting n<18122> which is dedicated to the focus on some central issues within cystic fibrosis be they research-based or care-based issues and also the prospects for future development n<36510> //



#<409826>	#<409826>	UNKNOWN	TT
//	//	//	SENT
then	then	then	RB
we	we	we	PP
also	also	also	RB
have	have	have	VHP
a	a	a	DT
third	third	third	JJ
of	of	of	IN
the	the	the	DT
professionals	/opresprofessionals/	professional	NNS
who	who	who	WP
think	think	think	VVP
that	that	that	IN
it	it	it	PP
's	's	be	VBZ
useful	useful	useful	JJ
to	to	to	TO
use	use	use	VV



ALIGNMENT

Sound-Text

(Transana, Anvil, Winpitch, SpeechIndexer)

buongiorno a tutti // cominciamo in maniera
rapida i lavori di questo working group che
ha uno scopo eminentemente pratico

good afternoon everybody // so we would like
now to get our working groups' proceedings off
to a rapid start // we've got practical aims

ALIGNMENT

Parallel texts

Content-based

(source | target)

buongiorno a tutti //	good afternoon everybody //
cominciamo in maniera rapida i lavori di questo working group che ha uno scopo eminentemente pratico	so we would like now to get our working groups' proceedings off to a rapid start // we've got practical aims

ALIGNMENT

Parallel texts

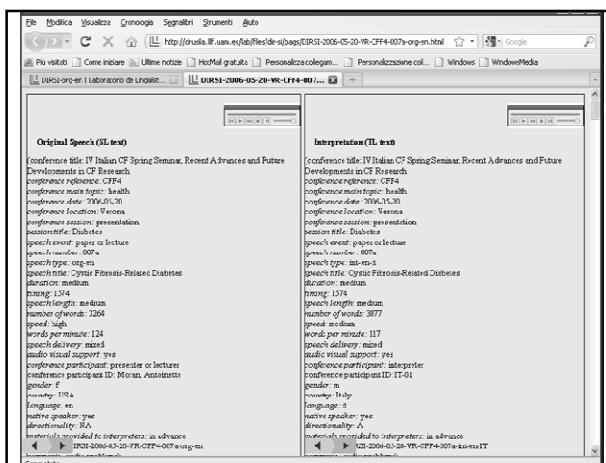
Décalage-based (e.g. Exmaralda)

buongiorno a tutti // cominciamo
in maniera rapida i lavori di
good afternoon everybody // so we would like now

DIRSI ALIGNMENT

Semi-automatic alignment:

- Web page with 2 column table
- New line when //
- Manual adjustment (Kompozer 0.7.10)

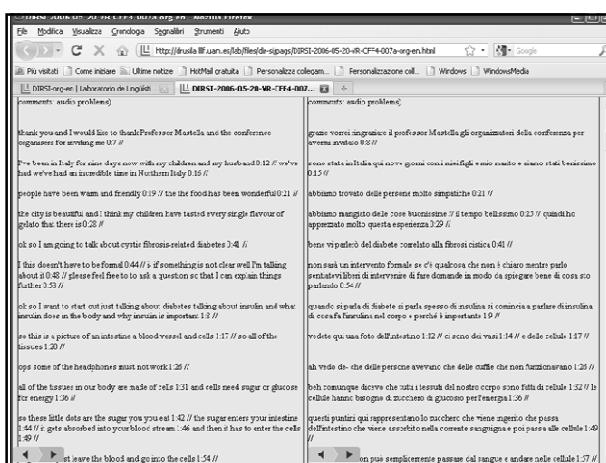


Original Spec's (SL text)

conference title: 17th Italian CF Spring Seminar, Recent Advances and Future Developments in CF Research
conference report code: CFP4
conference main topic: health
conference date: 2006-05-20
conference location: Pescara
conference session: presentation
participants: Diabetes
speech event: lecture or lecture
place: online
speech type: org-ns
speech title: Cystic Fibrosis-Related Diabetes
speech duration: 13'45
speech length: medium
number of words: 3264
speed: 117
words per minute: 124
speech delivery: mixed
audio visual support: yes
participants: participant or lecturer
conference participant ID: Mavan_Antoniotti
gender: f
country: ITA
language: en
native speaker: yes
diamonality: it
interpretation code: interpretation in advance
interpretation date: 2006-05-20-PR-CFP4-007a-org-ns

Interpretation (IL text)

Developments in CF Research
conference reference: CFP4
conference main topic: health
conference date: 2006-05-20
conference location: Pescara
conference session: presentation
session title: Diabetes
speech event: paper or lecture
place: online
speech type: int-ns
speech title: Cystic Fibrosis-Related Diabetes
speech duration: 13'45
speech length: medium
number of words: 3077
speed: 117
words per minute: 117
speech delivery: mixed
audio visual support: yes
participants: participant or lecturer
conference participant ID: Mavan_Antoniotti
gender: m
country: ITA
language: it
native speaker: yes
diamonality: it
interpretation code: interpretation in advance
interpretation date: 2006-05-20-PR-CFP4-007a-org-ns



Comments audio problems

Thank you and I would like to thank Professor Martella and the conference organizers for inviting me 0.0//

I've been in Italy for over days now with my children and my husband 0.12 // we've had a wonderful time in Northern Italy 0.15 //

people have been warm and friendly 0.19 // the food has been wonderful 0.21 //

the city is beautiful and I think my children have tasted every single flavour of gelato that there is 0.28 //

oh so I want to start out not talking about diabetes talking about insulin and what insulin does in the body and why insulin is important 1.2 //

so that is a picture of an insulin a blood vessel and cells 1.17 // so all of the tissues 1.20 //

so some of the beta cells must not work 1.26 //

all of the tissues in our body are made of cells 1.31 and cells need sugar or glucose for energy 1.36 //

so these little dots are the sugar you eat 1.42 // the sugar enters your intestine 1.44 // it gets absorbed into your blood stream 1.46 and then it has to enter the cells 1.49 //

it leaves the blood and goes into the cells 1.54 //

Comments audio problems

grazie vorrei ringraziare il professor Martella gli organizzatori della conferenza per avermi invitato 0.51//

vorrei anche ringraziare gli ospiti italiani che sono stati molto simpatici 0.15//

abbiamo trovato delle persone molto simpatiche 0.21//

abbiamo mangiato delle cose succulente // il tempo bellissimo 0.23 // quindi approfittando molto questa esperienza 3.29//

bene ripetete del diabete correlate alla fibrosi cistica 0.41//

non so se un intervento finale se c'è qualcosa che non è chiaro potete rivolgersi direttamente a me per fare domande in modo da spiegare bene di cosa sto parlando 1.54//

grande sigaretta di sigaretta si prende spesso di solito si comincia a fumare di solito di circa 15 minuti nel corpo e poi è importante 1.59//

vedete qua una foto dell'adiposita 1.12 // ci sono dei vaselli 1.14 // e delle cellule 1.17//

ah vede da destra delle persone avevano che due cellule che non funzionavano 1.26//

beh comunque devono che tutta l'isola del nostro corpo sono fatte di cellule 1.32// le cellule hanno bisogno di zuccheri di glucosio per funzionare 1.36//

questi puntini qui rappresentano lo zucchero che viene inghiottito passa attraverso la parete della vena e poi passa alle cellule 1.39//

// on può semplicemente passare dal sangue e andare nelle cellule 1.57//

As soon as you start transcribing:

From transcription...

...to data classification

EPIC >> [verbatim reports](#) (web)

DIRSI-C >> ?

EP debate:

Economic crisis, growth and employment

February 15, 2012

Mario Monti

■ Verbatim report

- (official/revised [version](#))
- (spoken [version](#))

■ Recordings: [EN](#) [PT](#) [ES](#)



on which this Parliament voted by an overwhelming majority, are a tool that I have encouraged elsewhere without ever seeing them as a possible source of fiscal indiscipline. They are a tool to ensure greater integration of financial markets which, if properly constructed along the lines of the Commission's Green Paper and the report adopted today by the European Parliament, may in fact help to make the actions of the financial markets into a disciplining factor for public budgets. It is just one example of tools that can be implemented, while respecting the various sensitivities: if a major country believes that stability bonds should come a little bit later rather than earlier, again, this sensitivity should be respected, as long as no one claims that exploring them is dangerous. I believe that one day, the time will come – but not for a while – to reconsider whether the way we currently treat the act of private investment, but also of public investment, is today as appropriate as it could be.

I am happy that in our constitutions now, following the fiscal compact, we are introducing the rule of a structurally correct balanced budget; I am a little less happy that we have called it the 'golden rule'. This is an important discipline but once, even in the German economic culture, the golden rule stated that there could be state debt but only to the limit of public investment; so we can finance with debt what is essentially the formation of capital.

(Applause)

One day, once we have completed the great transformation which – thanks especially to Germany – we have made in recent years towards a culture of discipline, Europe will be able to afford, without avoiding the issue or watering it down, to look more coolly and calmly even at tools for growth such as these.

The final point, Mr President, which I have little time to develop: I beg to refer to an article published this morning in the French, Italian and German press, and which bears my signature along with that of Ms Sylvie Goulard. We need to reconcile – and for now we are not managing to do this so well – integration, which we need in ever increasing doses, and democracy, which we certainly do not wish to give up. How do you reconcile democracy and integration? Well ...

(Murmurs of dissent)

I think it is certainly possible to reconcile democracy and integration.

(Applause)

Of course, only a deeply superficial, insular culture...

(Applause)

... might naively believe that integration means a super-state.

(Call of 'Who elected you?')

A vast majority in the Italian Parliament.

(IT) Mr President, please forgive me if I am overtaken by my passion for European integration, which does not mean a superstate at all, but rather a continuous operation of the principle of subsidiarity and needs to be reconciled with democracy. It is clearly evident in the last few days what might happen if citizens, in certain Member States, get the impression that integration occurs to the detriment of democracy.

The final point, Mr President, which I have little time to develop: I beg to refer to an article published this morning in the French, Italian and German press, and which bears my signature along with that of Ms Sylvie Goulard. We need to reconcile – and for now we are not managing to do this so well – integration, which we need in ever increasing doses, and democracy, which we certainly do not wish to give up. How do you reconcile democracy and integration? Well ...

(*Murmurs of dissent*)

I think it is certainly possible to reconcile democracy and integration.

(*Applause*)

Of course, only a deeply superficial, insular culture...

(*Applause*)

... might naively believe that integration means a super-state.

(*Call of 'Who elected you?'*)

A vast majority in the Italian Parliament.

(*IT*) Mr President, please forgive me if I am overtaken by my passion for European integration, which does not mean a superstate at all, but rather a continuous operation of the principle of subsidiarity and needs to be reconciled with democracy. It is clearly evident in the last few days what might happen if citizens, in certain Member States, get the impression that integration occurs to the detriment of democracy.

Instead of trying to re-invent the wheel, as we say in Italian, I believe that we need to work on the existing institutions. There is in Europe – I believe you have heard of it – an institution directly elected through universal suffrage by citizens, known as the European Parliament.

L'ultima considerazione, signor Presidente, che ho poco tempo di sviluppare: mi permetto di fare riferimento a un articolo apparso stamani sulla stampa francese, italiana e tedesca e che reca la mia firma unitamente a quella dell'onorevole Sylvie Goulard. Noi abbiamo bisogno di conciliare – cosa che per ora non ci riesce tanto bene – integrazione, di cui abbiamo bisogno in dosi sempre maggiori, e democrazia, alla quale non vogliamo certo rinunciare. Come si fa a conciliare democrazia e integrazione? Ebbene...

I think it is deeply possible to reconcile democracy and integration.

(*Applause*)

Of course only a deeply superficial, insular culture...

(*Applause*)

... might naively believe that integration means a super-state.

(*Call of 'Who elected you?'*)

A vast majority in the Italian Parliament.

Signor Presidente, la prego di scusarmi se mi faccio prendere dalla passione per l'integrazione europea, la quale non significa affatto un *superstate* ma piuttosto un continuo operare del principio della sussidiarietà e deve essere riconciliata con la democrazia. È ben evidente in questi giorni cosa possa accadere se i cittadini, in taluni Stati membri, hanno l'impressione che l'integrazione avvenga a scapito della democrazia.

Anziché pretendere di inventare la ruota, come si suol dire in italiano, ritengo che occorra lavorare sulle istituzioni esistenti. Esiste in Europa – credo ne abbiate sentito parlare – un'istituzione eletta direttamente dai cittadini a suffragio universale denominata Parlamento europeo. Questa Istituzione condiziona molto incisivamente la vita comunitaria: io ho fatto parte di una

Un giorno un giorno l'Europa, un giorno l'Europa completata la grande trasformazione che – per merito soprattutto della Germania – abbiamo compiuto in questi anni verso la cultura della disciplina, potrà permettersi, senza eluderla senzae attenuarla, di considerare in modo più freddo e pacato anche strumenti per la crescita come questi.

L'ultima considerazione, signor Presidente, che ho poco tempo di sviluppare: ma mi permetto di fare riferimento a un articolo che questa mattina è apparso-uscito stamani-nesulla stampa francese, italiana e tedesca e che reca la miaho firmato unitamente a quella delcon l'onorevole Sylvie Goulard. Noi abbiamo bisogno di conciliare – cosa chee per ora non ci riesce-riusciamo tanto bene – integrazione, di cui abbiamo bisogno in dosi sempre maggiori, e democrazia, alla quale non vogliamo certo rinunciare. Come si fa a conciliare[interruption?] a conciliare democrazia e integrazione? Ebbene...

INT: penso che sia possibilissimo riconciliare la democrazia e l'integrazione [applause] naturalmente solamente una cultura insulare molto superficiale può ritenere ... può ritenere ... ingenuamente che l'integrazione significa un super stato ... una grande maggioranza del Parlamento italiano risponde il presidente Monti

I think it is deeply possible to reconcile democracy and integration. Of course integration

(Applause)

Of course of course only a deeply superficial ehm- insular culture might consider might consider that that...

(Applause)

... might might might naively might naively believe that integration means a super-state.

(Call of 'Who elected you?')

on on Aa vast majority in the Italian-European Parliament sorry in the Italian Parliament.

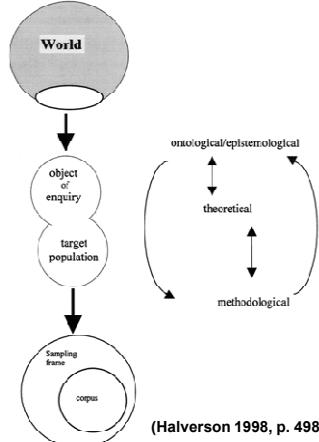
Signor Presidente, la prego di scusarmiscusi se mi faccio prendere dalla passione per l'integrazione europea che sono sicuro [applause] integrazione [applause] integrazione europea che, la quale non significa affatto un-a superstate ma piuttosto-significa un continuo operare del principio della sussidiarietà ed- deve essere riconciliata con la democrazia. È ben evidente in questi giorni Vediamo in questo momento cosa possa accadere se i cittadini le persone, in taluni certi Stati membri, hanno l'impressione che l'integrazione avvenga a scapito della democrazia.

e io credo che aAnziché pretendere di inventare la ruota, come si suol dire in italiano, ritengo che occorra lavorare sulle istituzioni esistenti che esistono. Esiste in Europa— credo che voi ne abbiate sentito parlare – in Europa un'istituzione eletta direttamente dai cittadini a suffragio universale denominata si chiama Parlamento europeo. Questa Istituzione condiziona molto incisivamente la vita comunitaria: io ho fatto parte

representativeness

«a sample which is maximally representative of the variety under examination»

(McEnery & Wilson 2001, p. 30).



(Halverson 1998, p. 498 – based on Biber 1993)

«The specification of a corpus [...] is hardly a job for linguists at all, but more appropriate to the sociology of culture». Sinclair (1991: 13)

- Sociolinguistics
- Linguistic Anthropology
- Ethnography of Speaking
- Conversation Analysis
- Discourse Analysis
- (non-mediated) Communication Studies

Conference as communicative event

stage activity
platform arrangement
(Goffman 1981)

social discourse event
(Ventola 2002: 27)

social institution
(Shalom 2002: 51)

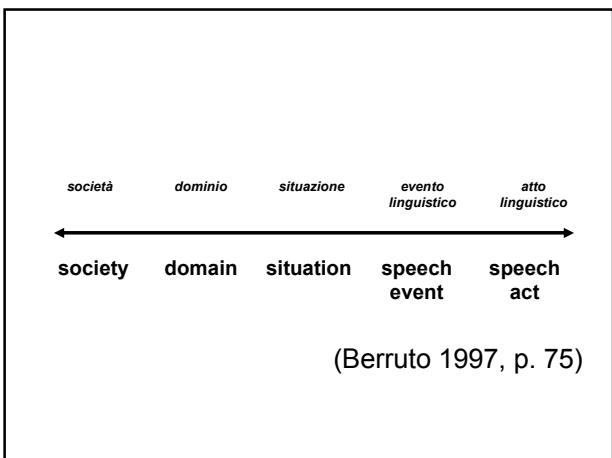
S	SITUATION	(Hymes, 1980)
P	PARTICIPANTS	
E	ENDS	
A	ACT SEQUENCES	
K	KEY	
I	INSTRUMENTALITIES	
N	NORMS	
G	GENRES	

Communicative event /
Speech **situation**

VS.

Speech **event**

(Duranti 2005: 258;
Hymes 2003: 38-39)



«overall ‘map’ of the interaction in terms of its typical ‘phases’ or ‘sections’»
(Heritage 1997: 166)

episodes
(Levinson 1992)

external + internal point of view

CORPUS DESIGN: representativeness

Communicative situation:

- **structure** (main phases or sections)
- **participants** (roles)
- **‘texts’** (speech events)

Macro-structure

PRE-EVENT

call for papers -> submission of abstracts -> evaluation of abstracts -> drawing up of programme

THE CONFERENCE

opening plenary -> sessions and social programme -> closing plenary

POST EVENT

submission of papers -> evaluation of papers -> conference proceedings/publication

(Shalom 2002: 53)

Conference sessions

Ventola (2002)	Shalom (2002)	Russo (1999)
<i>panel</i>	<i>paper presentation session</i>	sessione d'apertura
<i>roundtable</i>	<i>poster session</i>	sessioni di lavoro
<i>section</i>	<i>panel/round table</i>	sessione di chiusura
<i>workshop</i>	<i>plenary or keynote lecture</i>	
	<i>workshop</i>	
<i>plenary</i>		scambio di domande e risposta
<i>paper</i>	<i>presentation</i>	presentazione di relazioni
<i>poster</i>	<i>discussion</i>	
<i>discussion</i>		

conference session:	opening	
	presentation	
	paper	
	plenary	
	discussion	
	round table or panel	
	poster	
closing		

conference title:	(full name)
conference reference:	CFF4 CFF5 ELSA
conference main topic:	health
conference date:	year-month-day
conference location:	Verona Cesena
conference session:	opening presentation discussion closing
session title:	(see official programme)

Micro-structure

Ratified speech events:

- 'One speaker at a time'
- Use of microphone
- From **turn** to **floor** allocation

Floor / 'facoltà di parola'

opening remarks	<i>intervento di apertura</i>
paper presentation	<i>relazione</i>
lecture, plenary presentation	<i>conferenza, lezione magistrale, plenaria</i>
floor allocation	<i>assegnazione della facoltà di parola</i>
procedure	<i>intervento procedurale</i>
housekeeping announcements	<i>avvisi</i>
question	<i>domanda</i>
answer	<i>risposta</i>
comment	<i>commento</i>
closing remarks	<i>intervento di chiusura</i>

speech event:	opening-closing remarks paper or lecture floor allocation procedure or housekeeping announcements question answer comment
speech number:	000
speech type:	org-it org-en int-it-en int-en-it
speech title:	(see official programme and check power point presentation)

duration:	timing:	short medium long	< 900 900-1800 > 1800
speech length:	number of words:	short medium long	< 1650 1650-3300 > 3300
speed:	words per minute:	low medium high	< 100 100-120 > 120
speech delivery:		impromptu read mixed	
audio visual support:		yes no	

Participants	
initiator	<i>iniziatore</i>
organizer	<i>organizzatore</i>
sponsor	<i>patrocinatore o sponsor</i>
chair	<i>presidente di sessione o moderatore</i>
discussant or respondent	<i>discussant o commentatore</i>
presenter	<i>relatore</i>
lecturer	<i>conferenziere</i>
audience	<i>pubblico</i>
interpreter	<i>interprete</i>

conference participant:	organizer sponsor chair discussant presenter or lecturer audience interpreter
conference participant ID:	Surname, Name IT-01 IT-02 IT-03 IT-04 UK-01
gender:	m f
country:	(specify country of origin)
language:	it en
native speaker:	yes no

directionality:	A B
materials provided to interpreters:	in advance on the spot none
audio link	(full name of relevant audio file)
comments:	

References

- Berruto, Gaetano (1997) *Fondamenti di Sociolinguistica*. Roma: Laterza.
- Biber, Douglas (1993) Representativeness in corpus design. *Literary and Linguistic Computing* 8/4, pp. 243-257.
- Cencini, Marco (2000) Il Television Interpreting Corpus (TIC). *Proposta di codifica conforme alle norme TEI per trascrizioni di eventi di interpretazione in televisione*. Unpublished MA thesis, Università di Bologna – Sede di Forlì, Scuola Superiore di Lingue Moderne per Interpreti e Traduttori.
- Christ, Oli (1994) A Modular and Flexible Architecture for an Integrated Corpus Query System, *COMPLEX '94, Budapest*. Online (Corpus Work Bench info page): <http://www3.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CorpusWorkbench.html>
- Cresti, Emanuela & Massimo, Moneglia (eds.) (2005) *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins.
- Duranti, Alessandro (1997) *Linguistic Anthropology*. New York: Cambridge University Press.
- Duranti, Alessandro (2005) *Antropologia del Linguaggio*. Roma: Meltemi editore.
- Garrote Salazar, Marta (2008) *CHIEDE. Corpus de Habla Infantil Espontánea del Español*. PhD thesis, Universidad Autónoma de Madrid.
- Goffman, Erving (1981) *Forms of Talk*. Oxford: Blackwell.
- Halverson, Sandra (1998) Translation Studies and representative corpora: Establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta* 43/4, pp. 494-514/1-22. Online: <http://www.erudit.org/revue/meta/1998/v43/n4/003000ar.pdf>
- Heritage, John (1997) Conversation analysis and institutional talk: analyzing data. In Silverman, David (ed.) *Qualitative Research: Theory, Method and Practice*. London: SAGE Publications, pp. 161-182.
- Hymes, Dell (1980) *Fondamenti di sociolinguistica. Un approccio etnografico*. Translation into Italian by Filippo Beghelli. Revision by Gaetano Berruto. Bologna: Zanichelli.
- Hymes, Dell (2003) Models of the interaction of language and social life. In Bratt Paulson, C. & G.R. Tucker (eds.) *Sociolinguistics. The Essential Readings*. Malden/Oxford: Blackwell Publishing, pp. 30-47.
- Levinson, C. Stephen (1992) Activity types and language. In Drew, P. & J. Heritage (eds.) (1992a), pp. 66-100.
- McEnery, Tony & Andrew, Wilson (2001) *Corpus Linguistics. An Introduction*. Edinburgh: Edinburgh University Press.
- Moreno Sandoval, A. & Guirao, José María (2006). Morphosyntactic Tagging of the Spanish C-ORAL-ROM Corpus: Methodology, Tools and Evaluation. In Kawaguchi, Yuji; Zaima, Susumu & Toshihiro, Takagaki (eds.) *Spoken Language Corpus and Linguistic Informatics*. Amsterdam/Philadelphia: John Benjamins, pp. 199-218.
- O'Connell, C. Daniel & Sabine, Kowal (1994) Some current transcription systems for spoken discourse: a critical analysis. *Pragmatics* 4/1, pp. 81-107.
- Russo, Mariachiara (1999) La conferenza come evento comunicativo. In Falbo, Caterina; Russo, Mariachiara & Francesco, Straniero Sergio (eds.) *Interpretazione simultanea e consecutiva. Problemi teorici e metodologie didattiche*. Milano: HOEPLI, pp. 87-102.
- Schmid, Helmut (1994) Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing. September 1994*.
- Shalom, N. Celia (2002) The academic conference: a forum for enacting genre knowledge. In E. Ventola, Shalom, Celia & Susan Thompson (eds.) *The Language of Conferencing*. Frankfurt am Main: Peter Lang GmbH, pp. 51-68.
- Sinclair, John (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Szakos J. and U. Glavitsch (2004a) "Portability, modularity and seamless speech-corpus indexing and retrieval: a new software for documenting (not only) the endangered Formosan Aboriginal languages", in *Proceedings of the EMELD Language Digitization Project Conference, Workshop on Linguistic Databases and Best Practice, Detroit, USA, July 15 - 18, 2004*, online: <http://emeld.org/workshop/2004/szakos-paper.html>
- Szakos J. and U. Glavitsch (2004b) "Seamless speech indexing and retrieval: developing a new technology for the documentation and retrieving of endangered Formosan languages". Paper presented at the Sixth International Conference on Teaching and Language Corpora (TALC6). University of Granada (Spain) 4-7 July 2004.
- Szakos J. and U. Glavitsch (2007) "SpeechIndexer in action: managing endangered Formosan languages", in *Proceedings of Interspeech*. Paper presented at ISCA 2007, August 27 - 31, Antwerp, Belgium.
- Text Encoding Initiative: www.tei-c.org
- Ventola, Eija (2002) Why and what kind of focus on conference presentations? In E. Ventola, Shalom, Celia & Susan Thompson (eds.) *The Language of Conferencing*. Frankfurt am Main: Peter Lang GmbH, pp. 15-50.

You will find more references in the bibliography of the following books (they are written in Italian but they can be dowloaded for free and have many extra references):

- Bendazzoli, C. (2010a) *Testi e contesti dell'interpretazione di conferenza: uno studio etnografico*. Bologna: Asterisco. [Open access: <http://amsacta.unibo.it/2905/>]
- Bendazzoli, C. (2010b) *Corpora e interpretazione simultanea*. Bologna: Asterisco. [Open access: <http://amsacta.unibo.it/2897/>]