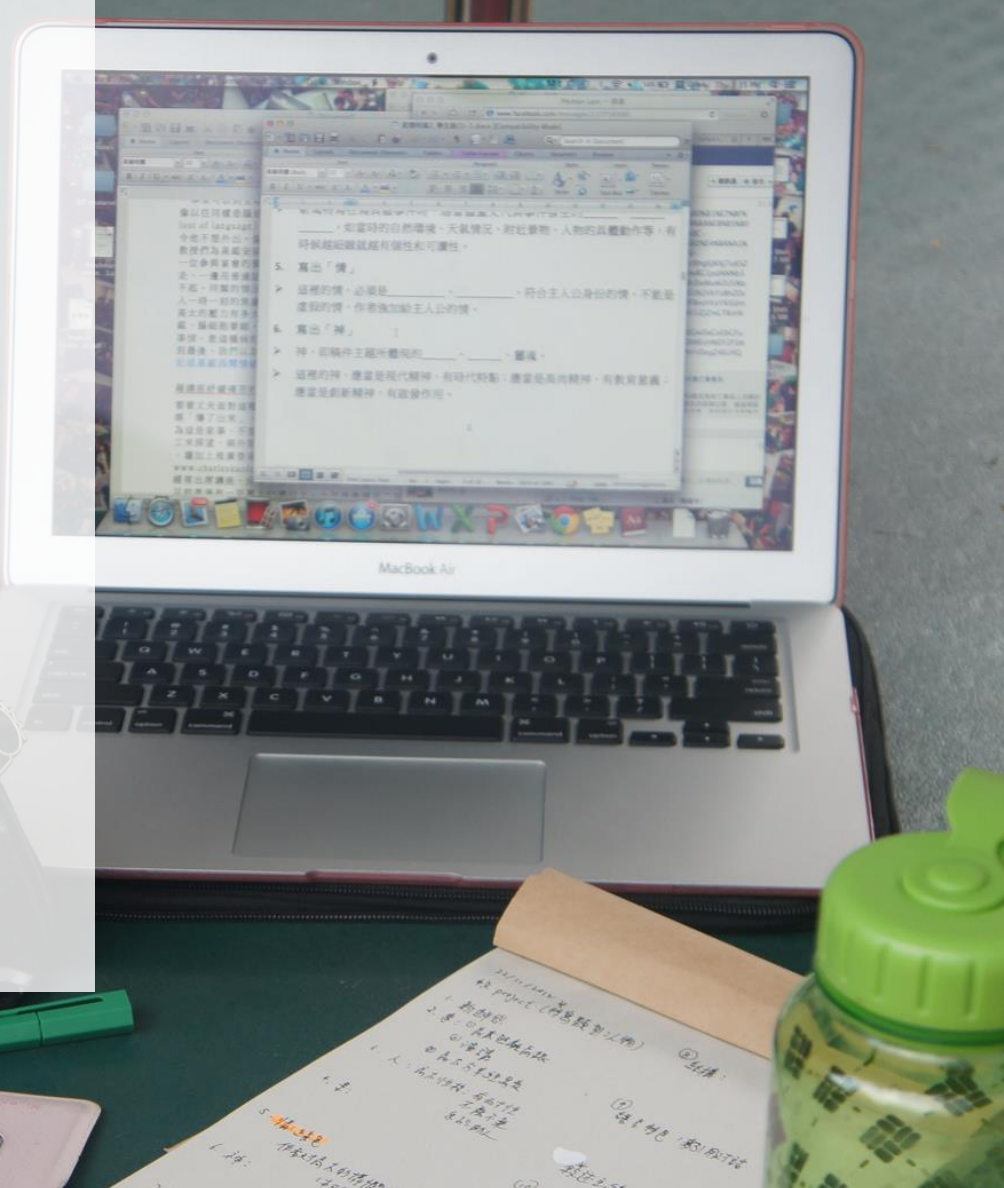


Modulo A

Valutare i sistemi automatici di interazione linguistica

7. Lingua scritta e lingua parlata

26 ottobre 2016



Linguistica italiana II
Mirko Tavosanis
A. a. 2016-2017

Questioni pratiche

- Il 31 ottobre non ci sarà il ricevimento
- I ricevimenti successivi saranno regolari

Oggi

Preliminari per la costituzione di un corpus:

- Scritto e parlato
- Le caratteristiche della lingua parlata in generale
- Le caratteristiche dell'italiano parlato, oggi
- Riferimento: la voce di Miriam Voghera sulla «lingua parlata»

Il modello della lingua scritta

Sembra che spesso le persone che hanno seguito un percorso scolastico si rappresentino la lingua parlata più o meno come la lingua scritta

- Discorsi ordinati
- Battute precise come in un romanzo o in un copione per il teatro

La realtà è molto diversa e crea molti problemi ai sistemi di riconoscimento automatici (come mostrano i WER molto diversi)

Scritto e parlato: diversità

Esempio di parlato in trasmissione radiofonica (dal corpus LIP):

... dopo una bella parentesi musicale in quanto_ questa sera c'è un po' di di di pigrizia nel farsi_ nel farsi largo nelle nel comunicare vi ricordo che siete all'ascolto della trasmissione Radio Incontri quindi pertanto gli incontri se ci mancano gli antagonisti non si possono incontrare allora in mancanza di antagonisti noi andiamo avanti con la nostra bellissima serata musicale una serata eh sulla sulle ali del romanticismo musica scelta con molta acutezza dal nostro regista Pino non prima di ricordarvi i telefoni che sono sei otto zero tre due zero sei otto zero tre tre tre quindi stasera abbiamo sentito il Polpo e Immacolata che si sono incontrati in in prima serata stranamente perché di solito le le vostre comunicazioni arrivano in un seconda tempo ma adesso invece che la serata che la notte va verso la la chiusura no manca ancora poco più di un'ora circa un'ora scarsa cioè un'ora abbondante però ah c'è una persona che telefona pronto...

Scritto e parlato: diversità

Bazzanella (1994) ha messo a confronto una relazione scritta di argomento linguistico e la sua presentazione a voce.

- Versione scritta:

sintagmi come sort of, kind of, I mean to say sono stati definiti da Quirk (1955) come «segnali di familiarità» e da Abercrombie (1967) come «riempitivi».

- Versione parlata:

le prime definizioni che si sono avute a proposito ehm della conversazione sono quelle di Quirk e di Abercrombie (-) uhm i quali hanno eh (--) ehm definite espressioni come (-) uhm sort of (-) del tipo kind of oppure you know è una cosa (-) un'espressione che si usa spesso in inglese /,/ [Intervento del docente: «qui sentiremo x più avanti»] sì eh (-) sono appunto espressioni che sono definite da Quirk come eh (-) signals / intimacy signals (-) cioè segnali di familiarità eh che indicano appunto una condizione molto più informale e spontanea eh sì spontanea! mentre eh sono stati definiti da Abercrombie come eh silence fillers (-) cioè riempitivi /,/.

Dal corpus CLIPS (DGtdBo1L, da 5')

p1#143: i piedi ? <sp> cosa vuoi dai piedi ? #<p2#144> <laugh>#

p2#144: #<p1#143> <tongue-click> che so dimmi dimmi com'è# l'ombra se no <P>
<NOISE> <P> l'ombra del

#<p1#145> piede c'è ?#

il primo "ombra" è pronunciato con un volume di voce più alto rispetto ai termini circostanti

p1#145: #<p2#144> l'ombra sta<aa>#

sì dietro poco poco

p2#146: sì

<sp>

p1#147: <inspiration> a<aa>+ ci sta il piede destro

p2#148: <mh>

p1#149: che si vede soltanto l'alluce

<sp>

p2#150: sì

p1#151: invece quello sinistro si intravedono poco

poco anche gli #<p2#152> altri#

p2#152: #<p1#151> sì#



Riferimento

- Per riferimento useremo qui la voce «Lingua parlata» di Miriam Voghera (*Enciclopedia dell'italiano*, Roma, Treccani, 2011), che penso di inserire anche nella bibliografia d'esame
- La voce è anche liberamente leggibile in line sul sito Treccani
http://www.treccani.it/enciclopedia/lingua-parlata_%28Enciclopedia_dell'Italiano%29/
- È un'ottima sintesi delle caratteristiche del parlato italiano e vale la pena discuterla in dettaglio qui
- Soprattutto, è una sintesi *aggiornata* (ed equilibrata) di ciò che si pensa adesso del parlato, anche per il caso italiano

Sintesi generale

«L'espressione *lingua parlata* identifica un insieme di caratteristiche strutturali e funzionali che si manifestano primariamente, ma non in modo esclusivo, quando si usa la lingua attraverso il canale fonico-uditivo in condizioni naturali e spontanee. La lingua parlata è quindi costituita da un insieme di usi linguistici prodotti dalle specifiche condizioni enunciative del parlare, non (o solo parzialmente) osservabili in altre modalità di trasmissione (per es., nella lingua scritta).»

Una precisazione importante

- «... spesso l'espressione *lingua parlata* è usata per identificare la lingua colloquiale o informale o varietà diastratiche basse. L'identificazione di lingua parlata e registri informali o varietà basse induce fraintendimenti e confusioni tra piani di variazione diversi. È bene dunque distinguere le proprietà generali della lingua parlata da quelle che derivano da fattori di variazione diafasica o diastratica.»
- Un conto sono insomma i fenomeni necessariamente collegati al **canale** e un conto quelli collegati agli **usi sociali** (anche se il rapporto c'è)
- Questa distinzione è basilare, ma non sempre riconosciuta: si può avere scritto informale e parlato formale

Grammatica del parlato

- «Le differenze registrate tra parlato e scritto non dipendono (...) da ‘grammatiche’ diverse, come nel caso, per es., della differenza tra enunciati italiani e tedeschi, ma dal fatto che, data la diversità delle condizioni in cui generalmente si parla e si scrive, si scelgono le strutture che sono meglio compatibili con le diverse situazioni enunciative. Ciò significa che non si può propriamente parlare di una grammatica del parlato opposta a una grammatica dello scritto, ma che è meglio parlare di usi linguistici tipici e preferiti nell’una o nell’altra modalità di comunicazione.»
- Io spesso parlo di «vincoli pragmatici»: l’idea è che siano importanti questi, non una «variazione diamesica» che non esiste in quanto tale

Fattori di condizionamento

- Il sistema fonico-uditivo
- Il processo produttivo e ricettivo
- La relazione tra elementi verbali e non verbali
- La struttura dell'interazione verbale
- Le condizioni di coerenza e coesione
- La gestione e il controllo degli elementi enunciativi

L'ultimo fattore dice in sostanza significa solo che i fattori precedentemente indicati possono combinarsi in vario modo, e che questa combinazione può essere conseguenza di scelte

Il sistema fonico-uditivo

- a) l'organizzazione temporale del segnale;
- b) la non-ripetibilità e permanenza del segnale;
- c) la contemporaneità tra produzione e ricezione del segnale.

In sostanza, tempo invece di spazio...

Il processo produttivo e ricettivo

- Il problema di base: si interagisce in tempo reale
- Chiaramente, c'è chi è più svelto... e chi no

La relazione tra elementi verbali e non verbali

- Il ruolo degli elementi non verbali è talmente importante che la comunicazione a faccia a faccia viene definita anche *multimodale*
- In molti casi il risparmio di tempo e di energia prodotto dalla comunicazione multimodale è tale da rendere preferibile il vedersi di persona
- Tuttavia, con il telefono ci siamo abituati senza troppi problemi a una comunicazione verbale in cui la voce è l'unico punto di contatto

La struttura dell'interazione verbale

- Il modello di base è il **dialogo**, non il monologo
- «Questa dimensione interattiva coinvolge la testualità e le strutture del parlato in modo pervasivo, tanto che secondo alcuni autori è da ritenersi un elemento distintivo primario del parlato rispetto allo scritto (Koch 1988).»
- Per lo scritto, la comunicazione elettronica smentisce in buona parte questa idea: per esempio, nelle chat la dimensione interattiva è fondamentale – giustifica l'esistenza stessa dello strumento
- In ogni caso, gli assistenti vocali si basano molto sull'idea di dialogo

Le condizioni di coerenza e coesione

- Nel parlato «sono per definizione in fieri» e «sono il risultato di un processo aperto anche al ricevente»
- D'accordo...

Sì, ma quali sono i tratti linguistici della lingua parlata?

Le caratteristiche possono essere

- a) foniche
- b) testuali
- c) lessicali e sintattiche

Caratteristiche foniche

- Bassa specificazione del segnale
Fusione, elisione, mutamenti di timbro... Non ce ne accorgiamo, perché il cervello di chi ascolta ci mette del suo
 - Aggiungo io: fonicamente, il parlato degli esseri umani è molto diverso tra maschi e femmine, adulti e bambini, individui diversi e così via – anche qui, non ce ne accorgiamo perché il cervello ci fa credere che sia tutto uguale (lo vedremo meglio la prossima settimana)
 - Questo è un grosso ostacolo per i computer!
- Disfluenza
Interruzione del flusso del parlato dovuta alla necessità di organizzare il discorso

Caratteristiche testuali

- Discontinuità
False partenze, interruzioni, cambi di progetto
- Uso della deissi
Manifestazione della multimodalità all'interno del testo
- Ridondanza
Eccesso, sia di lingua sia di contenuto, rispetto a quanto necessario per la comunicazione
- Ripetizione
- Prosodia come strumento di coesione testuale
- Marcatori di discorso
Cioè e simili

Caratteristiche lessicali e sintattiche

- Uso di lessico e strutture polisemiche
- Bassa densità lessicale e frequenza delle parti del discorso
- Uso di sintassi additiva
- Frasi senza verbo

Il caso particolare dell'italiano parlato

- Qui il discorso in sostanza si ferma
- Lo possiamo portare avanti noi, ricordando che, se in linea di massima è possibile scrivere tutto ciò che si dice in italiano (nei limiti del sistema di scrittura!), in pratica alcuni tratti si manifestano solo nel parlato e alcuni solo nello scritto
- Due concetti (credo) già familiari: standard e neostandard

«Neostandard» (Berruto) o «italiano dell'uso comune»

- In sostanza, è *ancora* la lingua delle grammatiche
- Si differenzia da questa, però, per una sessantina di tratti (pochissimi! Anche se alcuni sono molto visibili)
- Alcuni di questi tratti sono entrati ormai stabilmente anche nell'uso scritto formale (per es., «lui» soggetto)
- Tavoni ha proposto di usare queste etichette:
 - «Neostandard» (NS): normale nella scrittura giornalistica
 - «Substandard» (SS): presente nel parlato ma assente nella scrittura giornalistica

Tratti neostandard (pronomi)

- SS. *Te* in funzione di soggetto (*Questo lo dici te*).
- NS. *Lui, lei, loro* in funzione di soggetti (*Lui ha detto proprio questo, Non è detto che lei sia d'accordo, Loro non sono venuti*).
- NS-SS. *Lui, lei, loro* riferiti a cose (*L'automobile anche lei fa quello che può con queste salite*).
- SS. *Gli* per 'a lei' (*Ho visto Maria e gli ho detto di venire anche lei*).
- NS. *Gli* per 'a loro', masch. e femm. (*Ho visto quelli della III B e gli ho detto di venire anche loro*).
- NS. Declino di *loro* come allocutivo di cortesia (*Loro si rendono conto / Voi vi rendete conto*).
- NS. Dimostrativi usati come pronomi personali, spesso con una sfumatura negativa (*Ora questo mi si presenta, e cosa gli dico? Quelli non ne vogliono sapere*).
- NS-SS. Forme rafforzate *questo qui, quello lì* (*Se continua a tenere quell'atteggiamento lì, fa poca strada*).
- *uno di quei* + sostantivo
- *questa cosa di...*
- *un qualcosa*
- *lì in* + indicazione di luogo (deittico pleonastico)
- SS. Forme *'sto, 'sta*, ecc., per *questo, questa* (*E così c'è 'sto problema del rimborso*).
- NS. *Niente* usato come aggettivo (*Niente scherzi, mi raccomando!*).

Scritto e parlato

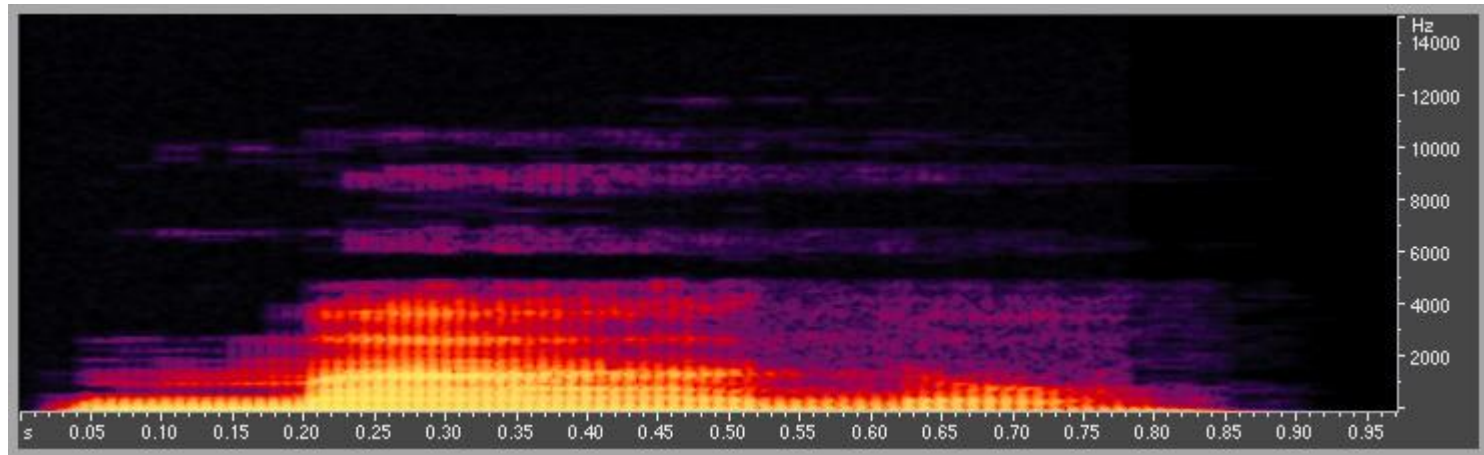
- Nella definizione di Tavoni, sono «neostandard» proprio i tratti che (quantitativamente) differenziano il parlato dallo scritto
- Una buona parte della differenziazione si trova comunque a livello lessicale... con un numero imprecisabile di parole (centinaia? migliaia?) usate spesso nel parlato ma rare nello scritto
- Attenzione! stiamo parlando solo della situazione dell'italiano – lasciando da parte i dialetti

Naturalmente, non è un vincolo strutturale

- La differenza è solo una differenza nell'uso
- Per puro caso (più o meno) oggi nello scritto si conserva la distinzione tra «gli» e «le»... ma potrebbe essere cancellata, e potrebbe essere conservata al suo posto la distinzione tra «gli» e «a loro»
- Anche in un testo scritto può essere usato «te» per «tu», eccetera
- Il rapporto con il parlato è evidente (in particolare per le esigenze di semplificazione) ma non ci sono vincoli strutturali

La scrittura non riproduce tutti i tratti del parlato

- Il parlato può essere riprodotto graficamente («scritto») da uno spettrogramma con opportuni metadati



- Tuttavia, nessun sistema di scrittura storico (o utilizzabile facilmente da esseri umani) ha questa fedeltà di riproduzione
- Le trascrizioni appena viste usano segni speciali per riprodurre le pause, ecc., ma anche così il risultato è parziale
- I sistemi storici eliminano alcuni tratti... e introducono nuove possibilità

Che cosa si perde

La classificazione di Halliday (rielaborata e contestualizzata per la comunicazione elettronica in italiano da Pistolesi 2004) identifica come i principali elementi persi :

- i tratti indicali (i tratti cioè che identificano il singolo parlante: estensione di tono e altezza della voce, cadenza e così via)
- l'intonazione
- i segni paralinguistici che accompagnano il discorso (gesti, espressioni della faccia...)

Nella scrittura tradizionale, e nella scrittura elettronica, sono state elaborate varie strategie di sostituzione

Tratti indicali

- La voce umana non è mai «media»: contiene tratti indicali ed è sempre la voce di qualcuno (uomo o donna, adulto, bambino o anziano, gradevole o sgradevole...)
- Gli esseri umani usano abitualmente le caratteristiche della voce per individuare un parlante, anche non visibile
- Nella scrittura ordinaria italiana, la voce è trasformata in un insieme di elementi standardizzati... Il che rappresenta un problema, o un'opportunità comica, per chi vuole rappresentare con la scrittura scene di interazione umana reale
- Da una fan fiction:

< < Scusatemi... > >

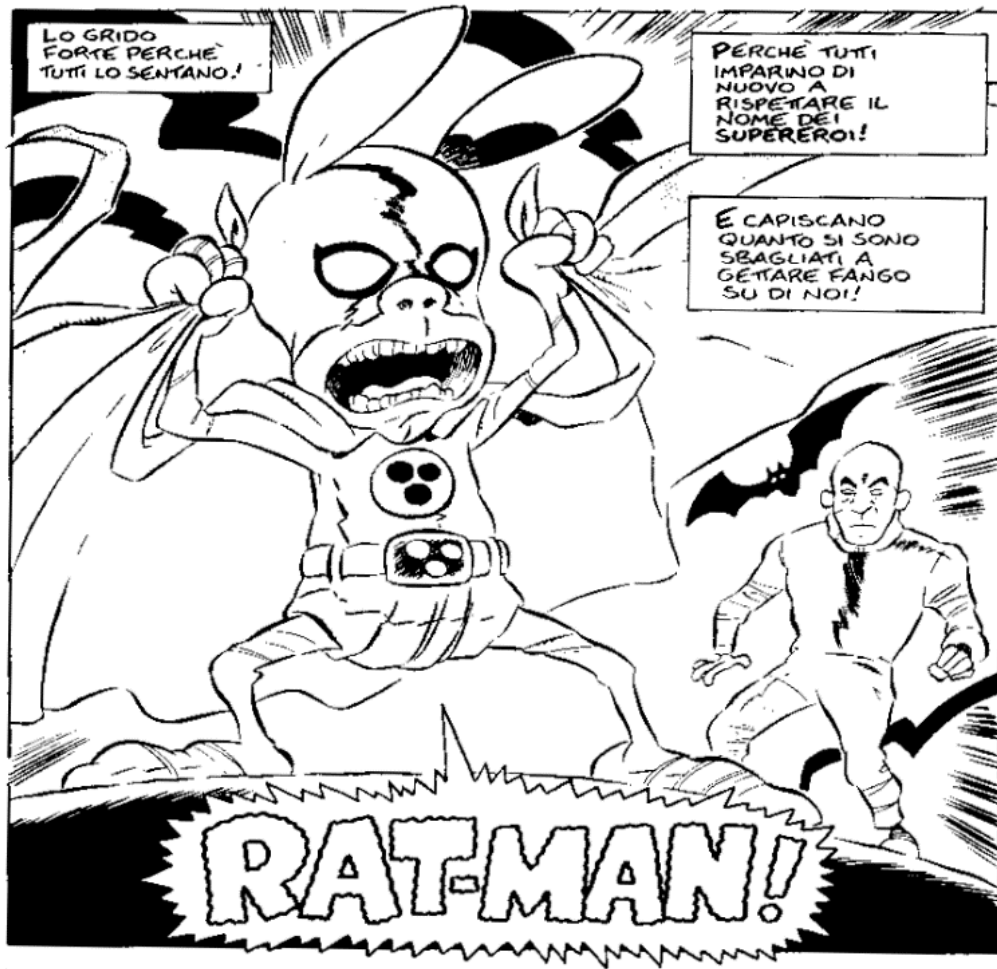
Sento una voce lontana...

< < Harry... > >

Uhm? Quella voce...

< < Non pensare male... > >

Hermione?

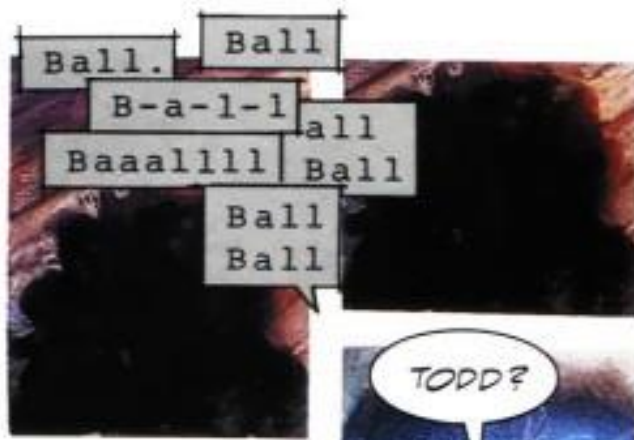


Leo Ortolani, *Un uomo in calzamaglia*, in *Rat-Man*, 19, 2000, p. 32.

In aggiunta al gioco basato sull'impossibilità di identificare i parlanti dalla voce: le variazioni di intensità della voce sono indicate con aumento delle dimensioni fisiche dei caratteri



E ORA (SUL SERIO!) LA CONCLUSIONE!
SUL PROSSIMO NUMERO:
IL SUPEREROE!



Bill Sienkiewicz, *Stray Toasters*: variazioni nel tipo di carattere e nella forma e nel colore di sfondo del fumetto (*balloon*).

Intonazione

L'ortografia italiana tradizionale include solo due elementi per indicare le variazioni di tono:

- Il punto interrogativo «?»»
- Il punto esclamativo «!»»

Un ruolo minore hanno le virgolette per indicare la presa di distanza da una parola o da un'espressione (cui può corrispondere o meno una variazione di tono nel parlato)

Nella scrittura espressiva, dai fumetti alle chat, questi segni di interpunzione sono integrati da

- imitazione delle spezzature del parlato
- puntini di sospensione
- punteggiatura creativa («!!!», «!?!», ecc.)

Segni paralinguistici

- In un testo scritto non ci sono le espressioni del volto, i gesti di accompagnamento e così via
- In un romanzo, o in un copione teatrale, si possono fornire indicazioni su tutti questi elementi
- In un fumetto, i disegni possono mostrare posizioni del corpo, espressioni, movimenti...
- Nella comunicazione scritta informale le emoticon permettono spesso di integrare rapidamente il testo scritto



Soluzioni parziali

- Tutte le soluzioni sono parziali
- Usarle per trasmettere e interpretare messaggi richiede (sorpresa!) un po' di pratica
- La comunicazione elettronica ha coinvolto una buona parte della popolazione in problemi di questo tipo
- Esempio tipico: i *flame* della posta elettronica delle origini