

**A 78 anni, il papà del più moderno dizionario della lingua italiana si butta nella Rete con *Senso Comune*.**

Un nuovo progetto online tutto wiki, interattività e intelligenza artificiale. E rigorosamente in *Creative Commons*, perché le parole con cui comunichiamo (all'incirca 260mila) sono di tutti

— LUCA SOFRI

## CHAT

### Che cos'è *Senso Comune*?

«Allora, la premessa personale è che io avevo fatto dei dizionari».

**Premessa superflua, abbiamo sfogliato il "De Mauro" tutti quanti. Però allora approfitto e chiedo come si fa un dizionario.**

«Il mitico Johnson del Settecento era fatto da una sola persona. Oggi, per garantire la coerenza nella stesura delle parole, si costituiscono delle redazioni e si coordina il loro lavoro».

**E come si fa a essere sicuri di non perdersi delle parole?**

«No: siamo sicuri di perdercele. Si cerca di non perdersi quelle essenziali».

**Ci sono gerarchie di parole?**

«Le parole che hanno una circolazione vera nella lingua si stimano a molti milioni. Molte però hanno un "indice di cercabilità" bassissimo. Altre, che intuitivamente ci sembrano frequenti, compaiono poco: come "abbottonare" e "sbottonare". Sono un migliaio, forse 1500. In tutto, nel nostro Grande Dizionario, ci sono 260mila parole. Il grande Oxford inglese in 12 volumi ne ha circa 330mila ed è stato fatto in quarant'anni».



**Voi quanto ci avete messo?**

«L'informatizzazione ci ha permesso rapidità e calcolabilità dei tempi: nove anni. Tra il 1990 e il 1999. Poi ne abbiamo fatto una seconda edizione e una terza».

**E quella che era online?**

«Il dizionario monovolume di Paravia finito online è poi stato tolto per beghe editoriali con Utet, la società della De Agostini di Novara che aveva pubblicato il Grande Dizionario (sei volumi nella prima edizione e otto nella seconda). La storia è complessa, forse rientrerò nei diritti: adesso però pensiamo di rimetterci le mani e farne una nuova versione online. Con un gruppo di informatici e lessicografi, che si chiama appunto *Senso Comune* ([sensocomune.org](http://sensocomune.org)), stiamo progettando la messa online del vocabolario "di base": le 2mila parole più frequenti che coprono il 90 per cento delle occorrenze, più altre 2500 di grande frequenza».

**Ah, la coda lunga del dizionario è esigua quindi.**

«Ancora di più se pensa che il 5-6 per cento è coperto dalle altre 2500 che ho chiamato di grande frequenza».

**Ma è solo un dizionario online o c'è altro nel progetto?**

«C'è altro eccome. Ci interessa che il materiale messo a disposizione consenta l'interattività, cioè che permetta all'utente di intervenire criticamente e propositivamente sull'impianto delle definizioni, proponendo correzioni, integrazioni, precisazioni».

**Non mi dica che affidiamo agli utenti il dizionario e la definizione dei termini: demagogia linguistica?**

«No, no: non consegniamo questo materiale ai lettori, ma a contributi che in un sistema wiki controlliamo e selezioniamo. Raccogliamo informazioni, ma gli utenti non modificano immediatamente il repertorio. È una specie di consultazione popolare che ha però l'opportunità del controllo».

**E non temete che questo controllo si traduca in giustizialismi grammaticali, nel trollismo da Internet?**

«Non sono così preoccupato, filtreremo: i rompiscatole possono esistere, ma l'accesso non è "liberi tutti". Chiediamo un'autoidentificazione, ma non mi aspetto grandi molestie. È molto difficile sparare sentenze. Se poi si scopre che c'è un colonnello in pensione di Novara che tormenta il dizionario, vedremo...».

**Se è di Novara è sospetto, però...**

«Perché? Ah, no, noi siamo buoni amici della De Agostini».

**A che punto siete?**

«Abbiamo già pronta una redazione idonea delle duemila parole fondamentali, stiamo ancora lavorando su quelle di base. Tecnicamente se ne occupa Guido Vetere a Roma. Può parlare con lui».

**Pronto, sono Luca Sofri, mi ha detto di chiamarla il professor De Mauro.****Mi spiega cosa fa?**

«Sono il direttore del centro di studi avanzati di Ibm, dove facciamo ricerca in collaborazione con l'università soprattutto sull'intelligenza artificiale e le cosiddette tecnologie semantiche».

**Ovvero?**

«Ovvero cose su cui i sistemi devono saper ragionare in rapporto col mondo reale: e questo ha una declinazione linguistica, perché la conoscenza linguistica è centrale in queste operazioni. L'informatica di oggi è molto interessata allo schema classico dell'ontologia, ma ne ha una concezione un po' ingenua».

**E lei come è finito a occuparsi di queste cose?**

«Mi sono laureato con De Mauro con una tesi di linguistica computazionale, e da lì sono entrato in Ibm».

**Mi parli del dizionario online di *Senso Comune*.**

«Ci sono altri dizionari online, non siamo a zero: ma ne manca uno della ricchezza e della preparazione di quello che la Paravia aveva messo online senza però averne i diritti, e che quindi è stato rimosso».

**Lo chiedo anche a lei: è solo un dizionario consultabile online?**

«No. L'idea è di non avere solo un

vocabolario, ma una base di conoscenza: un lavoro di associazione a qualcosa di ontologico. Costruiamo una rappresentazione della conoscenza che sta dietro il linguaggio».

**E dopo cosa ci fate?**

«Potenzialmente un sacco di cose. Potremo usarla come base per i motori di ricerca delle informazioni o per i sistemi di classificazione automatica di documenti di testo. Il tutto grazie alla partecipazione degli utenti, come le avrà detto il professor De Mauro».

**Me lo ha detto, e mi ha spiegato che non è come Wikipedia.**

«Io sono un grande ammiratore di Wikipedia e penso che funzioni bene così com'è. Ma nel nostro caso il sistema è più "controllato". Noi pensiamo che la conoscenza dietro il lessico sia molto strutturata e tecnicamente più specifica, e che quindi ci voglia la conoscenza di elementi lessicografici che un utente qualsiasi non padroneggia. Gli utenti però possono testimoniare il loro uso linguistico».

**Chi paga tutto questo?**

«Un parziale sostegno della fondazione Ibm, 10mila euro l'anno, e poi il lavoro volontario mio e dei miei studenti in Ibm. Ma vogliamo collaborare con le università e i privati: le conoscenze linguistiche appartengono a tutti».

**Ci sono prospettive di business?**

«No, non nel senso di un progetto remunerativo in sé. Ma in prospettiva potranno nascere opportunità sfruttabili anche commercialmente. Il contenuto avrà comunque una licenza *Creative Commons*».

**E quando partite sul serio?**

«Siamo online in alpha per una comunità ristretta di utenti, ma non immaginiamo un lancio vero e proprio. È una politica di piccoli passi, non grandi annunci a cui non segue niente, vedi Wolfram Alpha».

**Io però ho cercato "ontologia" e *Senso Comune* mi ha detto «il lemma cercato non esiste».**

«Perché ci sono ancora solo le parole fondamentali».

**E io come faccio a sapere cosa vuol dire "ontologia"?**

«Aspetta, o cerca su Wikipedia».

**Il progetto è già online in alpha con duemila parole che coprono il 90% degli usi**