

Di cosa parliamo quando parliamo fondamentale?

ISABELLA CHIARI

(Università La Sapienza, Roma)

ALESSANDRO OLTRAMARI

(CNR – Lab Applied Ontology, Trento)

GUIDO VETERE

(IBM Center di Studi Avanzati di Roma)





Il progetto Senso Comune

- Senso Comune è una base di conoscenza della lingua italiana aperta al contributo dei parlanti attraverso il Web
- Senso Comune attualmente si compone di un nucleo costituito dalla fascia del vocabolario fondamentale ([De Mauro 1980](#)) di 2.071 lemmi e circa 13.192 accezioni che coprono più del 90% delle occorrenze in testi parlati e scritti.
- Il progetto Senso Comune, sotto la supervisione di Tullio De Mauro, intende integrare la risorsa lessicografica con altre risorse linguistiche in un progetto collaborativo che si svolge attraverso piattaforma web (<http://www.sensocomune.it>).

Benvenuto *Isabella Chiari*



Cerca

 **ala** 

FO

s. f. sing.

organo mobile del volo, specialmente di uccelli e insetti, e, in diverse religioni e mitologie, di esseri ritenuti capaci di volare, generalmente coperto di penne o membranoso



aprire, muovere, battere le ali apertura d'ali un piccione ferito a un'a. le ali degli angeli come posso arrivare li? Non ho mica le ali Aggiungi esempio d'uso




Aggiungi una relazione lessicale



Aggiungi una sotto accezione



Oggetto naturale 

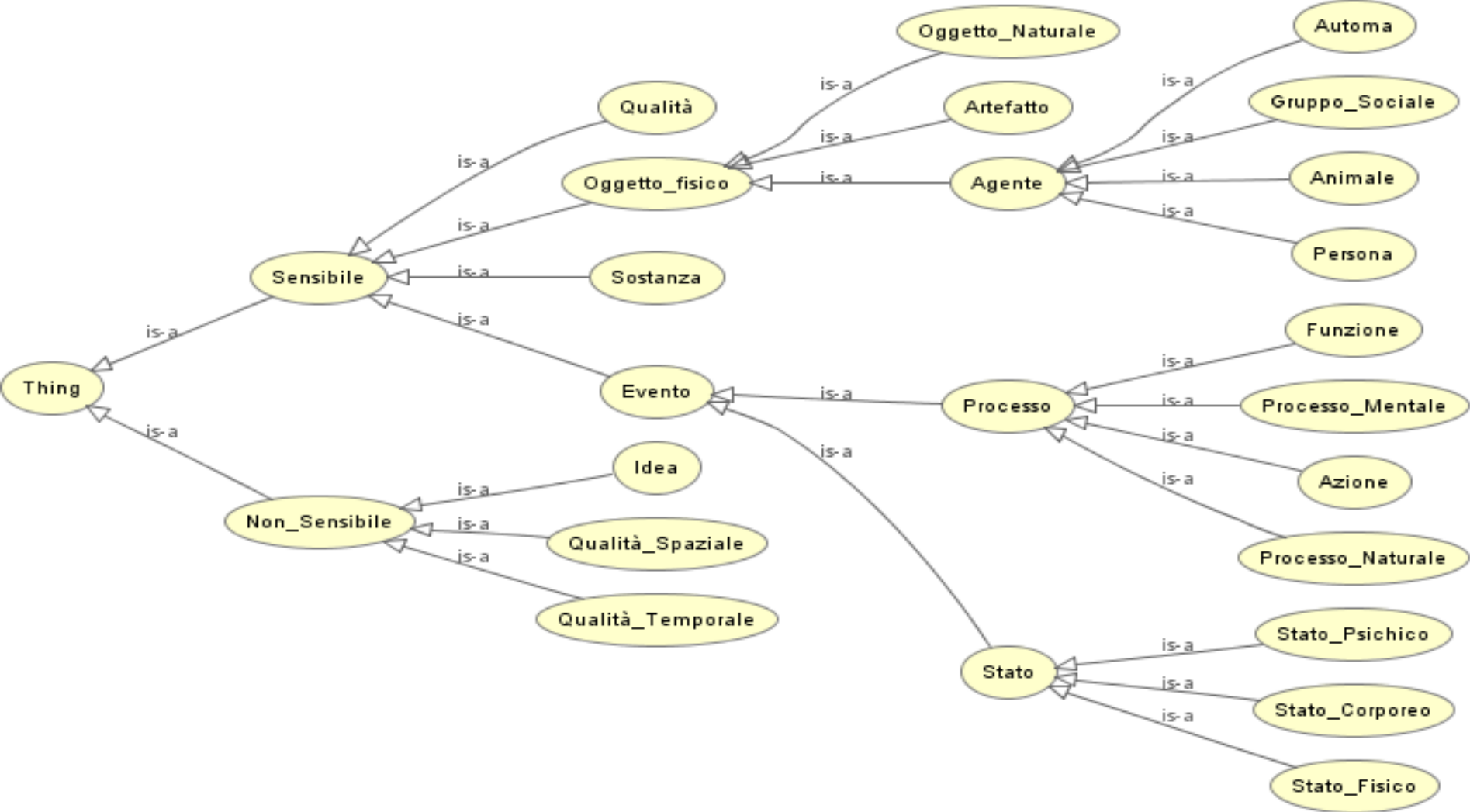


Apri discussione

Accesso al database di Senso Comune

L' ontologia di SC

- In Senso Comune, ciascuna accezione può essere associata ad una categoria proveniente da un' ontologia “fondazionale”.
- L' ontologia di SC è essenzialmente basata su una versione semplificata dell' ontologia OWL-lite DOLCEspray, una versione semplificata dell' ontologia DOLCE (Masolo et al, 2002)



Struttura gerarchica della classificazione

Usata nell' ontologia e in TMEO

Sperimentazione sui sostantivi del vocabolario fondamentale

- **Obiettivo:** osservare la possibilità di associare a ciascuna accezione dei lemmi FO una categoria ontologica generale e verificare quali sono i problemi concreti che si incontrano in tale operazione

La sperimentazione ha previsto tre fasi:

- **I FASE**

- Classificazione primaria, in assenza di istruzioni per preservare una categorizzazione di senso comune, da parte di 12 soggetti (studenti di laurea specialistica)
- Discussione e rimodulazione dell' ontologia

- **II FASE**

- Revisione della classificazione, da parte dei tre autori e di 4 sperimentatori del primo gruppo
- Attribuzione delle classi di confidenza (accettato, controverso, non accettato)
- Apertura delle discussioni

- **III FASE**

- Ultima revisione della coerenza generale, a campione.

TABELLA 1 - ACCEZIONI CLASSIFICATE

Vocabolario fondamentale	2.071
Lemmi con almeno una accezione accezione Sost FO	1.093 (52,8%)
Accezioni dei Sost FO	4.654

Esempio il lemma PESO

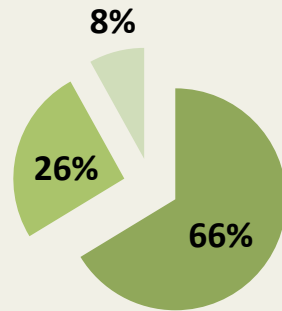
- il lemma sostantivale *PESO* possiede diverse accezioni ciascuna delle quali rimanda a diverse classi ontologiche:
 - come “corpo soggetto alla forza di gravità” (*caricare, portare, sollevare pesi*) è classificato come OGGETTO,
 - “oggetto, specialmente metallico, opportunamente graduato che serve nelle operazioni di pesatura” come ARTEFATTO
 - Come “senso di pesantezza dovuto a cattiva digestione” è classificato come STATO CORPOREO
 - Come “condizione, situazione che reca disagio, fastidio, sofferenza fisica o morale” come STATO PSICHICO
 - “autorità, prestigio, influenza” (*il peso di una posizione sociale, il peso del casato*) come IDEA, ecc.

Il sistema di tutoraggio TMEO 3.0

- Il sistema di tutoraggio, TMEO 3.0, può essere attivato a scelta dell'utente per guidarlo alla classificazione attraverso una serie di domande, sensibili allo storico delle risposte, fino alla classificazione secondo l'ontologia.
 - Ad esempio, sempre per classificare le accezioni di *PESO*, TMEO propone domande quali: *Puoi toccare, vedere, annusare, gustare un "peso" ? Puoi contare diversi "pesi" ? un "peso" è costruito da un agente?, ecc.*

TABELLA 2 - LA POPOLOSITA' DELLE CLASSI			
Classe	Popolosit�	Classe	Popolosit�
IDEA	624	ENTITY (*31)	99
ARTIFACT	585	SOCIAL_GROUP	82
PERSON	497	MENTAL_PROCESS	78
ACTION	457	PROCESS	74
QUALITY	420	FUNCTION	69
NATURAL_OBJECT	211	SENSIBLE	46
PSYCHIC_STATE	184	NATURAL_PROCESS	45
PLACE	177	PHYSIC_STATE	43
SOCIAL_OBJECT	171	CORPORAL_STATE	32
TEMPORAL_QUALITY	170	AGENT	24
EVENT	162	ANIMAL	21
STATE	149	NON_SENSIBLE	1
OBJECT	120	NON_TANGIBLE	1
SUBSTANCE	112		4.654

Certezza di classificazione



Analisi e primi risultati

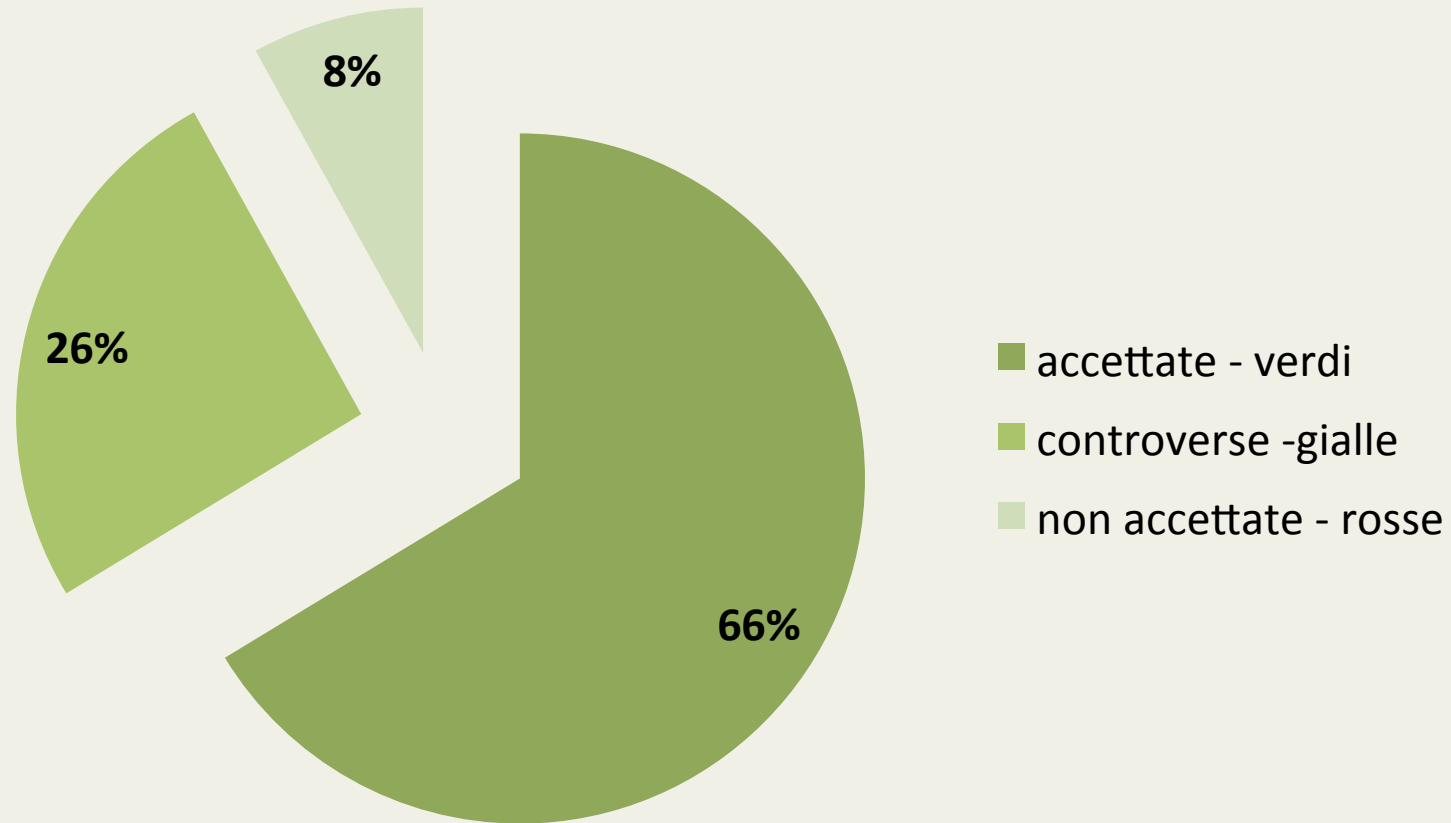
TABELLA 2 - LA POPOLOSITA' DELLE CLASSI ONTOLOGICHE

<i>Classe</i>	<i>Popolosità</i>	<i>Classe</i>	<i>Popolosità</i>
IDEA	624	ENTITY (*31)	99
ARTIFACT	585	SOCIAL_GROUP	82
PERSON	497	MENTAL_PROCESS	78
ACTION	457	PROCESS	74
QUALITY	420	FUNCTION	69
NATURAL_OBJECT	211	SENSIBLE	46
PSYCHIC_STATE	184	NATURAL_PROCESS	45
PLACE	177	PHYSIC_STATE	43
SOCIAL_OBJECT	171	CORPORAL_STATE	32
TEMPORAL_QUALITY	170	AGENT	24
EVENT	162	ANIMAL	21
STATE	149	NON_SENSIBLE	1
OBJECT	120	NON_TANGIBLE	1
SUBSTANCE	112		4.654

SENSIBLE	46	SUBSTANCE	112				
		QUALITY	420				
		OBJECT	120	ARTIFACT	585		
		PLACE	177	NATURAL_OBJECT	211		
				AGENT	24	SOCIAL_GROUP	82
						PERSON	497
						ANIMAL	21
		EVENT	162	PROCESS	74	FUNCTION	69
						MENTAL_PROCESS	78
						NATURAL_PROCESS	45
						ACTION	457
				STATE	149	PSYCHIC_STATE	184
NON_SENSIBLE	1	IDEA	624			PHYSIC_STATE	43
		TEMPORAL_QUALITY	170			CORPORAL_STATE	32
		SOCIAL_OBJECT	171				

POPOLOSITA' DELLE CLASSI NELLA GERARCHIA DI SC

Certezza di classificazione



Livello di certezza nell' attribuzione delle classi dopo le revisioni

Le classi più ‘accettate’

CLASSE	accettate	%	controverse	%	non accettate	%	totali
PERSON	443	89,0%	52	10,4%	3	0,6%	498
ANIMAL	18	85,7%	3	14,3%	0	0,0%	21
NATURAL_OBJECT	177	83,9%	22	10,4%	12	5,7%	211
ARTIFACT	478	81,7%	86	14,7%	21	3,6%	585
PHYSIC_STATE	35	81,4%	6	14,0%	2	4,7%	43
ACTION	350	76,6%	98	21,4%	9	2,0%	457
TEMPORAL_QUALITY	130	76,5%	33	19,4%	7	4,1%	170
AGENT	18	75,0%	6	25,0%	0	0,0%	24
SUBSTANCE	83	74,1%	25	22,3%	4	3,6%	112
NATURAL_PROCESS	33	73,3%	12	26,7%	0	0,0%	45
CORPORAL_STATE	23	71,9%	8	25,0%	1	3,1%	32

Le classi più controverse

CLASSE	accettate	%	controverse	%	non accettate	%	totali
SOCIAL_OBJECT	61	35,7%	96	56,1%	14	8,2%	171
ENTITY	16	51,6%	15	48,4%	0	0,0%	31
PLACE	94	53,1%	79	44,6%	4	2,3%	177
PROCESS	37	50,0%	31	41,9%	6	8,1%	74
FUNCTION	26	37,7%	28	40,6%	15	21,7%	69
EVENT	90	55,6%	65	40,1%	7	4,3%	162
STATE	68	45,3%	59	39,3%	23	15,3%	150
PSYCHIC_STATE	113	61,4%	64	34,8%	7	3,8%	184
MENTAL_PROCESS	49	62,8%	27	34,6%	2	2,6%	78
QUALITY	241	57,2%	140	33,3%	40	9,5%	421
SOCIAL_GROUP	53	64,6%	26	31,7%	3	3,7%	82

Le classi per cui l'ipotesi è già un azzardo

CLASSE	accettate	%	controverse	%	non accettate	%	totali
SENSIBLE	29	63,0%	5	10,9%	12	26,1%	46
FUNCTION	26	37,7%	28	40,6%	15	21,7%	69
OBJECT	60	50,0%	36	30,0%	24	20,0%	120
STATE	68	45,3%	59	39,3%	23	15,3%	150
IDEA	362	58,0%	173	27,7%	89	14,3%	624
QUALITY	241	57,2%	140	33,3%	40	9,5%	421
SOCIAL_OBJECT	61	35,7%	96	56,1%	14	8,2%	171

SENSIBLE	46	SUBSTANCE	112						
		QUALITY	420						
		OBJECT	120	ARTIFACT	585				
		PLACE	177	NATURAL_OBJECT	211				
				AGENT	24	SOCIAL_GROUP	82		
						PERSON	497		
						ANIMAL	21		
				EVENT	162	PROCESS	74	FUNCTION	69
								MENTAL_PROCESS	78
								NATURAL_PROCESS	45
						ACTION	457		
NON_SENSIBLE	1			STATE	149	PSYCHIC_STATE	184		
		IDEA	624			PHYSIC_STATE	43		
		TEMPORAL_QUALITY	170			CORPORAL_STATE	32		
		SOCIAL_OBJECT	171						

CLASSI NELLA GERARCHIA DI SC PER 'CERTEZZA' DI CLASSIFICAZIONE DELLE ACCEZIONI

Numero di classi ontologiche diverse	Lemmi	nr. medio di accezioni
1	165	3,0
2	275	3,3
3	175	5,0
4	119	7,5
5	57	9,1
6	31	10,6
7	19	13,8
8	5	14,2
9	1	19,0

Esempi di lemmi monocategoria sono:

balcone, calza, coltello, ingegnere, ministro, nipote, principessa, rabbia

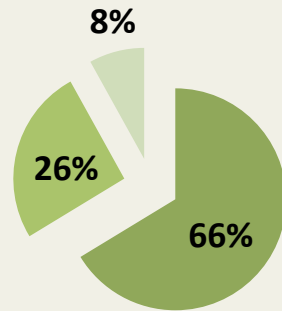
Le classi ontologiche più popolate da lemmi monocategoria sono, in ordine decrescente:

PERSONA (55), ARTEFATTO (21), IDEA (14), AZIONE (13)

La varietà di classe ontologica delle accezioni FO

TABELLA 2 - LA POPOLOSITA' DELLE CLASSI			
Classe	Popolosit�	Classe	Popolosit�
IDEA	624	ENTITY (*31)	99
ARTIFACT	585	SOCIAL_GROUP	82
PERSON	497	MENTAL_PROCESS	78
ACTION	457	PROCESS	74
QUALITY	420	FUNCTION	69
NATURAL_OBJECT	211	SENSIBLE	46
PSYCHIC_STATE	184	NATURAL_PROCESS	45
PLACE	177	PHYSIC_STATE	43
SOCIAL_OBJECT	171	CORPORAL_STATE	32
TEMPORAL_QUALITY	170	AGENT	24
EVENT	162	ANIMAL	21
STATE	149	NON_SENSIBLE	1
OBJECT	120	NON_TANGIBLE	1
SUBSTANCE	112		4.654

Certezza di classificazione



Discussione e problemi

Principali problemi di classificazione

- Riguardano soprattutto gli astratti
 - Ad esempio prodotti scientifici, unità di misura, funzione
- Un problema ancora non affrontato è quello degli oggetti semiotici che in molte accezioni sono vaghi in quanto all' associabilità con una specifica classe
 - *messaggio, discorso, scritto, biglietto, ecc*
- Usi pragmatico-discorsivi (spesso introdotti da *in locuzione pragmatica*):
 - *guarda caso! Dio mio!*

Problemi ricorrenti nella classificazione

- luogo/artefatto/funzione (*bagno, albergo*)
- ma anche luogo/oggetto naturale (*bosco come “terreno”*)
- processo/azione/risultato
- sostanza (naturale e artificiale) e artefatto

Prospettive

- Affinare l' ontologia in modo da permettere classificazioni agevoli delle classi più problematiche
 - Permettere la marcatura con più di una classe ontologica, laddove necessario, o guidare in una categorizzazione sovraordinata
 - affinare l' ontologia per le classi troppo astratte
 - In particolare QUANTITA' e QUALITA' SPAZIALE
- Riflettere sull' opportunità di dividere alcune accezioni in sottoaccezioni che tengano conto delle diverse classi ontologiche
- Riflettere sulle condizioni di miglioramento degli strumenti tecnici e degli strumenti di supporto e/o eventuale addestramento degli utenti nei compiti di classificazione.

Prospettive

- Confrontare l' applicabilità dell' ontologia formale in senso stretto con le etichette applicate mediante classificazione di senso comune.
 - Separare i criteri di 'confidenza' e 'concordanza' dell' annotazione delle classi.
- Conciliare gli scopi lessicografici tradizionali con l' integrazione in applicazioni computazionali, tenendo conto della complessità e indeterminatezza dei confini tra i sensi
- Proiettare la classificazione in maniera controllata sui **testi**, passando dal livello sistemico a quello dell' uso effettivo.

Grazie!

- Un ringraziamento particolare ai nostri **sperimentatori**:
 - Nicola Amabile, Valentina Arena, Valentina Cristini, Marina D' Auria, Flavio De Giusti, Valentina Di Marco, Lorena Mascara, Angela Napoleone, Federico Riccardi, Marco Scarino, Tiziana Taboga, Edoardo Vanni
- Ringraziamo anche per lo **sviluppo e interfaccia**: Piero Cangialosi, Andrea Mencancini