

**Strategie, pubbliche e private, in
azione per ri-costruire meglio.
Analisi dei testi di quattro interviste**

Pasquale Pavone, Margherita Russo

CAPPaper n. 156
luglio 2017



Università di Modena e Reggio
Emilia Facoltà di Economia
Marco Biagi



Università di Bologna
Dipartimento di Scienze
Economiche

CAPP - Centro di Analisi delle Politiche Pubbliche
Dipartimento di Economia Politica - Università di Modena e Reggio Emilia
Ufficio 54 - Ala Ovest
Viale Berengario, 51 41100 Modena - ITALY
phone: +39 059 2056854 fax: +39 059 2056947
email capp@unimo.it

DEMB Working Paper Series

Strategie, pubbliche e private, in azione per ri-costruire meglio.
Analisi dei testi di quattro interviste

Pasquale Pavone* e Margherita Russo♦

*Centro di Analisi delle Politiche Pubbliche, Università di Modena e Reggio Emilia, pasquale.pavone@unipd.it

♦Dipartimento di Economia Marco Biagi e Centro di Analisi delle Politiche Pubbliche, Università di Modena e Reggio Emilia, margherita.russo@unimore.it

ISSN on line: **2281-440X**

Abstract

In this paper we present the methodology and results of the automatic analysis of the transcripts of the interviews carried out under the project "ideas and projects for better reconstruction" (Esposito et al., 2017). Four strategic areas are considered: the Territorial Cohesion Agency, the Casa Italia Plan, the Civil Protection, and that of a community model of action in the community. The goal is twofold. On the one hand, it is intended to offer a systematic reading of the specificities and elements common to the various aspects considered in the interviews. On the other hand, we want to explore some methodological ideas in the automatic analysis of the texts on the following issues: integration of analytical tools, constraints resulting from the size of the corpus, interpretative potential of graphical representation of results, integration between automatic information retrieval, auto-coding of a corpus and expert reading.

Abstract

In questo lavoro presentiamo la metodologia e i risultati dell'analisi automatica delle trascrizioni delle interviste realizzate nell'ambito dell'istant book "BBB: idee e progetti per ricostruire meglio" (Esposito et al. a cura di, 2017). Sono presi in esame quattro ambiti strategici: dell'Agenzia della Coesione Territoriale, del Piano Casa Italia, della Protezione Civile, e quello di un'impresa modello di azione nella comunità. L'obiettivo del lavoro è duplice. Da un lato si intende offrire una lettura sistematica delle specificità e degli elementi comuni ai diversi ambiti presi in esame. Dall'altro lato, si vogliono approfondire alcuni spunti metodologici nell'analisi automatica dei testi in merito a: integrazione di strumenti di analisi, limiti che derivano dalla ridotta dimensione del corpus, potenzialità interpretative della rappresentazione grafica dei risultati, integrazione tra analisi automatica dei testi, annotazione automatica di un corpus di testi e lettura esperta.

Keywords: automatic information retrieval, content analysis,

JEL code: C81 Methodology, for Collecting, Estimating, and Organizing Microeconomic Data; Data Access; C88 Other Computer Software; R58 Regional Development Planning and Policy

Ringraziamenti

Questo working paper è stato realizzato nell'ambito del progetto di ricerca Energie Sisma Emilia dell'Università di Modena e Reggio Emilia (www.energie.unimore.it). Il progetto è stato cofinanziato da Fondazione Cassa di Risparmio di Modena, Azienda Unità Sanitaria Locale di Modena and Università di Modena e Reggio Emilia.

Gli autori desiderano ringraziare Maria Ludovica Agrò, Giovanni Azzone, Fabrizio Curcio, Enrico Loccioni per aver autorizzato, nell'analisi sviluppata in questo lavoro, l'uso delle trascrizioni delle loro interviste realizzate nell'ambito del progetto *Building Back Better: idee per ri-costruire meglio*. Un ringraziamento particolare a Fulvio Esposito, Massimo Sargolini, Laura Sartori e Vania Virgili che con Margherita Russo hanno curato il progetto di cui le interviste fanno parte.

Introduzione

Nell'ambito dell'istant book *BBB: idee e progetti per ricostruire meglio* sono stati presi in esame quattro ambiti strategici: quello in capo all'Agenzia della Coesione Territoriale, che coordina le azioni della Strategia nazionale aree interne, quello del Piano Casa Italia, promosso dal Governo per realizzare soluzioni che nei prossimi 30 anni completino la messa in sicurezza degli edifici, quello che nell'azione sistematica della fase di emergenza vede l'azione della Protezione Civile, e quello di un'impresa. In particolare, per approfondire gli specifici ambiti strategici, sono state realizzate quattro interviste ai responsabili delle strategie pubbliche messe in atto per ri-costruire meglio e ad una impresa marchigiana, Loccioni, modello di impresa attiva nella comunità e quindi soggetto promotore di soluzioni che migliorino la capacità di risposta di individui e della comunità.

Agli intervistati sono state poste quattro domande a cui hanno risposto in maniera libera selezionando gli spunti più specifici della loro strategia: lo specifico ambito di azione spiega le differenze nei temi e nel linguaggio utilizzato dagli intervistati.

In questo documento presentiamo la metodologia e i risultati dell'analisi automatica delle trascrizioni delle loro interviste, con l'obiettivo di offrire, da un lato, una lettura sistematica delle specificità e degli elementi comuni ai diversi ambiti presi in esame, dall'altro, di approfondire alcuni spunti metodologici nell'analisi automatica dei testi: l'integrazione di strumenti di analisi, i limiti che derivano dalla dimensione del corpus, le potenzialità interpretative della rappresentazione dei risultati, l'integrazione tra analisi automatica dei testi e annotazione automatica di un corpus di testi.

Il working paper è organizzato in quattro sezioni. Nella sezione 1 si presenta la metodologia adottata nell'analisi; nella sezione 2 verranno richiamate le domande attorno a cui sono state realizzate le interviste aperte e si descrivono le caratteristiche del corpus oggetto di analisi; la sezione 3 discute i risultati dell'analisi automatica che riguarda sia i testi completi delle trascrizioni che i contenuti dei testi dei singoli intervistati e offre alcune considerazioni sui risultati ottenuti dalla analisi automatica e di quelli elaborati nella lettura esperta. La sezione 4 conclude l'analisi proponendo un intreccio del punto di vista degli intervistati su obiettivi, condizioni e azioni necessarie per ricostruire meglio.

1. Metodologia

Sviluppata nell'ambito della linguistica computazionale, l'analisi automatica del contenuto testuale nello specifico contesto è una tecnica ormai molto usata nell'analisi di grandi corpus di testi per i quali non sarebbe possibile procedere nella classificazione e interpretazione dei contenuti attraverso la lettura integrale del corpus (Bolasco, 2013). Nel caso preso in esame in questo lavoro, il corpus è piccolo e relativamente poco complesso. Inoltre, i curatori del volume, che peraltro hanno partecipato alle interviste, hanno poi anche letto e discusso tra di loro i contenuti delle trascrizioni, su cui hanno elaborato delle sintesi pubblicate nel volume in corso di stampa (Esposito *et al.* a cura di). Le ragioni che hanno spinto all'uso dell'analisi automatica sono duplici.

Da un lato, per quanto i testi non siano di lunghezza eccessiva, si intendeva affiancare i curatori del volume con spunti specifici di lettura che li ponessero alla stessa distanza dagli specifici testi: non tutti e cinque avevano partecipato alle interviste e non

tutti gli intervistati avevano seguito la traccia delle domande che venivano poste nel corso dell'intervista. L'eterogeneità del corpus era quindi maggiore di quella attesa. I curatori del volume avevano progettato di realizzare un dialogo strutturato tra i quattro intervistati, idealmente presenti nello stesso contesto tematico delle domande formulate dai curatori del progetto, ma sostanzialmente lontani per ambiti e competenze di intervento, e fra l'altro non contemporaneamente presenti nel corso dell'intervista. L'analisi automatica consente di esplorare in maniera sistematica la specificità di temi che caratterizzano ognuna delle interviste.

Una seconda motivazione è di carattere più strettamente didattico¹. Il piccolo corpus si presta ad una lettura completa e può far cogliere agli studenti, che si avvicinano alle tecniche di analisi automatica, lo scarto tra la lettura diretta e l'interpretazione di risultati dell'analisi automatica. Si potrà apprezzare come l'analisi sistematica del testo – con il trattamento lessicale e l'elaborazione del linguaggio specifico - evidenzia chiaramente le caratteristiche delle tematiche espresse nei discorsi degli intervistati. Si potranno valutare i vantaggi delle rappresentazioni grafiche dei risultati: consentono di cogliere simultaneamente molte dimensioni che, ad esempio, richiederebbero una tabella con molte serie di dati. Si potrà apprezzare, per converso, come sia più facile cogliere, nella lettura diretta delle trascrizioni, sfumature lessicali, toni e colori del parlato che non emergono integralmente nell'analisi automatica proposta in questo lavoro, ma che potrebbero emergere in analisi di tipo linguistico-computazionale (che esulano dall'obiettivo di questo lavoro).

L'analisi automatica dei testi è stata realizzata su due corpus: quello dei testi integrali delle trascrizioni e quelli da cui sono stati esclusi i frammenti di testo che si riferiscono agli interventi dei curatori del volume presenti all'intervista. Per l'esplorazione del corpus e la rappresentazione dei risultati, l'analisi è stata realizzata integrando una serie di strumenti software: Taltac2, Iramuteq, Gephi e Atlas.ti. Le fasi di analisi automatica dei testi sono essenzialmente tre: analisi lessicale, estrazioni di informazioni dal testo, estrazione dei frammenti di testo per l'analisi comparata.

Analisi lessicale

La prima fase è quella dell'analisi lessicale, stata realizzata utilizzando il software Taltac2 (Bolasco, 2010). Questa analisi prevede diverse specifiche elaborazioni: l'individuazione delle unità di analisi lessicale (“parole”), intese sia come parole singole che multiword; l'attribuzione delle meta-informazioni di tipo grammaticale alle “parole” costituenti il Vocabolario; la selezione delle parole di contenuto (sostantivi, verbi, aggettivi) la selezione su base probabilistica dei termini significativi del linguaggio specifico dei singoli intervistati e degli intervistatori.

Estrazione delle informazioni dal testo: linguaggio specifico, tematiche, similitudini

Nella seconda fase, è stata elaborata la visualizzazione del linguaggio specifico, rappresentato sul piano fattoriale, che elabora graficamente le combinazioni di una ma-

¹ Altri riferimenti per un uso didattico sono quelli proposti da Balbi e Misuraca (2005) e da Bolasco (2013). Le applicazioni dei metodi di analisi dei testi sono molte, ad esempio quelle pubblicate dalla rivista on line *Lexicometrica*, (<http://lexicometrica.univ-paris3.fr/jadt>) che raccoglie gli atti del convegno europeo di studiosi di analisi dei testi. Si segnalano tre applicazioni esemplificative delle potenzialità, rispettivamente, nella ricerca in campo economico e sociale: Gambarotto e Cortelazzo, a cura di (2013); nella classificazione dei contenuti di documenti, Pavone *et al.* (2016), che riguarda l'analisi dei contenuti e delle citazioni normative delle ordinanze della struttura commissariale per la gestione dell'emergenza post sisma in Emilia.

trice *forme caratteristiche* × *sub-testi* (nel nostro corpus i sub-testi sono quelli dei singoli intervistati e del gruppo degli intervistatori). La posizione delle parole sul piano fattoriale è funzione dell'associazione delle loro occorrenze nei sub-testi, pertanto esprime la similarità o la diversità dei diversi sub-testi: due parole sono vicine perché presenti negli stessi sub-testi; e al centro del piano fattoriale si trovano i termini più comuni fra i diversi linguaggi caratteristici dei vari sub-testi. La dimensione delle parole rappresentate nel piano fattoriale è proporzionale al corrispondente valore del χ^2 . Questa rappresentazione è stata elaborata graficamente con Iramuteq (Ratinaud, 2009)².

È stata anche realizzata un'analisi che mira ad individuare quali argomenti caratterizzino le varie parti del corpus di testi (in questo caso i singoli intervistatori). Con Iramuteq, è stato possibile individuare gli argomenti del Corpus delle quattro interviste, attraverso l'applicazione del metodo ALCESTE³. Tale metodo consiste in una procedura di classificazione gerarchica discendente di record individuali (frasi o enunciati) caratterizzati da singole parole⁴. Le frasi da classificare (Unità di Contesto Elementare) sono individuabili tramite un processo di frammentazione del corpus ricavato attraverso la punteggiatura e la lunghezza dei segmenti di testo considerati (nel nostro caso, in assenza di punteggiatura, le UCE hanno al più una lunghezza di 40 parole). Nella rappresentazione grafica di questa elaborazione delle tematiche (ottenuta con il metodo ALCESTE e realizzata con Iramuteq), viene visualizzato il dendrogramma dei cluster tematici, con i dizionari che caratterizzano ogni cluster. Per ogni cluster: l'altezza dei riquadri è proporzionale al peso percentuale del cluster (in termini di occorrenze) nell'intero corpus; la dimensione delle parole è proporzionale al valore del Test-Value nel cluster.

Sempre attraverso il software Iramuteq è stato possibile effettuare un'Analisi di Similitudine (ADS) delle parole (Flament, 1962; Flament, 1981; Marchand & Ratinaud, 2012). Lo scopo della ADS è quello di studiare la vicinanza e la relazione tra gli elementi di un insieme. Nel caso testuale, quindi, consente di evidenziare le relazioni fra le parole in base alle co-occorrenze delle stesse all'interno degli enunciati. Tale tecnica utilizza la teoria dei grafi per elaborare le reti di relazioni tra le parole di contenuto. La rappresentazione grafica delle reti di co-occorrenze è stata successivamente elaborata con il software Gephi⁵ (Bastian et al., 2009). Nella matrice di co-occorrenze sono presenti le parole di contenuto con più di tre occorrenze. Data la piccola dimensione del testo, il grafo rappresenta tutte le co-occorrenze. L'algoritmo di visualizzazione del grafo è il Fruchterman-Reingold. I colori delle diverse parti del grafo evidenziano le comunità di parole, generate attraverso l'algoritmo della modularità, che identifica le sotto reti di parole all'interno del corpus e delinea quindi la struttura della rete di co-occorrenze.

² Attraverso SPAD sarebbe possibile elaborare l'output del linguaggio specifico elaborato con Taltac2, ma non è possibile ottenere la stessa qualità informativa che si ottiene con Iramuteq, in termini di colori diversi associati ai sub-testi e di dimensione delle parole proporzionale al numero di occorrenze. Nel complesso la lista dei termini che caratterizzano il linguaggio specifico elaborato con i due software non è diversa e quindi si è scelto di elaborare il linguaggio specifico con Iramuteq per la più efficace rappresentazione dei risultati sul piano fattoriale, che ne facilita l'interpretazione.

³ ALCESTE è l'acronimo di 'Analyse des Lexèmes Cooccurrentes dans les Enoncés Simples d'un Texte', un metodo di analisi delle co-occorrenze testuali introdotto da Max Reinert nel 1985.

⁴ "La procedura gerarchica discendente considera come risultato solo le classi di unità di contesto (UC) che per due volte risultano associate insieme nella clusterizzazione", Bolasco (2013, p. 202).

⁵ Gephi è un software per la visualizzazione ed esplorazione di grafi e reti, <https://gephi.org/>.

Estrazione dei frammenti per l'analisi delle tematiche specifiche e di quelle comuni

Nella terza fase, l'analisi degli ambiti tematici specifici e di quelli comuni alle quattro interviste è stata realizzata attraverso l'estrazione automatica, con Atlas.ti, dei frammenti di testo delle interviste, applicando la modalità di *autocoding* e selezionando come unità di testo la frase. L'obiettivo è produrre il minimo testo che faciliti la lettura comparata dei contenuti che, nelle quattro interviste, si riferiscono ai termini individuati come comuni nella elaborazione fatta con Taltac2. Dalla lista completa dei frammenti estratti con Atlas.ti, sono stati selezionati manualmente - con una nostra decisione informata dalla lettura dei testi estratti - quelli con contenuti significativi per l'analisi comparata. Di questa analisi non riportiamo i risultati in questo lavoro perché non è ancora stata validata dagli intervistati la trascrizione dell'intervista e quindi non è possibile riprodurre i frammenti di citazioni.

2. Le domande sulle strategie e il corpus di testi

Il dialogo con i quattro interlocutori intervistati si è concentrato su quattro principali domande che hanno sullo sfondo il *Sendai Framework for Disaster Risk Reduction 2015-2030* delle Nazioni Unite⁶. Secondo quanto affermato nell'Accordo di Sendai, la gravità di un disastro è strettamente correlata alle scelte che facciamo per le nostre vite e per l'ambiente. Scelte che riguardano come produciamo cibo, come e dove costruiamo le nostre case, che tipo di politiche sociali adottiamo, come funziona il nostro sistema economico. Ogni azione è in grado di renderci più vulnerabili o più resilienti ai disastri. L'Accordo di Sendai descrive sette obiettivi⁷, da raggiungere entro 15 anni, che richiedono un impegno di tutti i paesi con un'azione mirata in cinque settori prioritari: comprensione del rischio di catastrofi; rafforzamento della governance e della gestione del rischio di catastrofi; investimento nella riduzione del rischio di catastrofi e nella resilienza; miglioramento della preparazione alle catastrofi per una risposta efficace e per 'ricostruire meglio' nella fase di recupero, ripristino e ricostruzione. Ad oggi, hanno aderito al *Sendai Framework* solo 18 paesi, i cui governi si sono impegnati a realizzare azioni per raggiungere gli obiettivi di riduzione del rischio a livello nazionale o regionale; solo sei sono paesi europei, e tra questi non ancora c'è formalmente l'Italia, anche se al *Sendai Framework* si ispirano le azioni strategiche messe in atto a livello nazionale.

L'obiettivo delle interviste era raccogliere elementi per sviluppare una riflessione comune per collocare all'interno di quel *framework* le strategie e le iniziative dell'organizzazione che ciascuno intervistato rappresenta. Le domande poste agli intervistati sono state raggruppate in quattro macro temi:

(a) In che modo le strategie e le iniziative dell'organizzazione che lei rappresenta si relazionano con le priorità indicate nel *Sendai Framework*?

⁶ <http://www.unisdr.org/we/coordinate/sendai-framework>

⁷ I sette obiettivi sono: riduzione sostanziale della mortalità mondiale legata alle catastrofi; riduzione sostanziale del numero di persone colpite; riduzione delle perdite economiche in rapporto al PIL mondiale; riduzione sostanziale dei danni causati alle infrastrutture e dell'interruzione dei servizi di base, in particolare alle strutture sanitarie ed educative, in caso di catastrofe; aumento del numero di paesi dotati di strategie nazionali e locali per ridurre i rischi di catastrofe entro il 2020; rafforzamento della cooperazione internazionale sul tema della riduzione del rischio di disastri; accesso accresciuto ai sistemi di allarme precoce multi-rischi e alle informazioni e valutazioni sui rischi di catastrofi.

- (b) Nell'ambito della strategia e delle iniziative dell'organizzazione che lei rappresenta, in che modo si raggiunge un punto di equilibrio fra le necessità della gestione dell'emergenza e le politiche di medio e lungo periodo per la prevenzione e lo sviluppo di territori e comunità resilienti?
- (c) In che modo il coinvolgimento delle comunità locali può favorire l'azione strategica e le iniziative dell'organizzazione che lei rappresenta nella direzione auspicata dal *Sendai Framework*?
- (d) La tecnologia e la scienza dei dati mettono a disposizione informazioni sempre più precise che, opportunamente analizzate, permettono l'efficace modellizzazione di scenari a breve, medio e lungo termine. In che misura la trasformazione digitale ed i big data sono utilizzabili in maniera efficace per le iniziative di prevenzione, preparazione, emergenza, ricostruzione e sviluppo? Come si possono rendere questi 'strumenti' accessibili alle autorità locali, alle comunità, agli imprenditori e ai singoli cittadini?

Le persone intervistate sono quattro: Maria Ludovica Agrò, Direttore Generale dell'Agenzia per la Coesione Territoriale, che coordina le azioni della Strategia Nazionale Aree Interne, Giovanni Azzone, coordinatore del Piano Casa Italia, Fabrizio Curcio, Capo del Dipartimento della Protezione Civile, e Enrico Loccioni, fondatore e presidente di Loccioni Group. Alle interviste hanno preso parte i curatori del volume. Le interviste, realizzate tra maggio e giugno del 2017, presso le sedi delle persone intervistate, sono state registrate e trascritte e i testi rivisti dagli intervistati⁸.

Le caratteristiche del corpus sono riassunte nella Tabella 1 che riporta le diverse forme (parole) del vocabolario dell'intero corpus delle quattro interviste e delle singole interviste, nelle due modalità prese in esame che riguardano, rispettivamente, l'intero testo e quello che esclude i frammenti di testo relativi agli interventi dei curatori che partecipavano all'intervista. La lunghezza dei sub-testi nel corpus (le singole interviste) varia significativamente, ma se escludiamo i testi degli intervistatori le differenze si riducono. Nell'intervista al professor Giovanni Azzone, i curatori del volume stavano avviando il loro progetto di analisi delle strategie e hanno discusso con Azzone varie tematiche della ricerca in corso. Nell'intervista a Enrico Loccioni, in chiusura del progetto di elaborazione dell'*instant book*, i curatori erano affiancati da altri partecipanti all'incontro che hanno contribuito con le loro osservazioni ad arricchire il confronto.

Tabella 1 Riepilogo delle occorrenze

Descrizione	corpus complessivo intervistati e intervistatori			corpus solo intervistati			% forme (g)=d/a	% occorrenze (h)=e/b
	Numero forme	Occorrenze	(c)=b/a	Numero forme	Occorrenze	(f)=e/b		
	(a)	(b)		(d)	(e)			
Corpus 4 Interviste	5.735	38.369	6,7	4.586	26.675	5,8	0,8	0,7
Agenzia Coesione territoriale_Agrò	1.767	7.137	4,0	1.592	5.996	3,8	0,9	0,8
Piano casa Italia_Azzone	2.225	10.426	4,7	1.542	6.161	4,0	0,7	0,6
Protezione Civile_Curcio	2.177	10.064	4,6	1.860	8.243	4,4	0,9	0,8
Loccioni Group_Loccioni	2.584	10.742	4,2	1.766	6.275	3,6	0,7	0,6

Nel complesso si tratta di un corpus piccolo, la cui analisi si presta facilmente ad una lettura e rielaborazione dei contenuti, che è quella presentata nel volume nell'*instant book*. L'analisi automatica del corpus che presentiamo in questo lavoro ha lo

⁸ Le trascrizioni sono state fatte da Chiara Florini (interviste ad Azzone, Agrò e Curcio), in stage presso il CAPP-DEMB, Stefano Tripi con Lucia Cambo' e Giulia Mennuti (intervista a Loccioni), del corso di specializzazione EmTASK, in stage nel progetto Energie Sisma Emilia-DEMB.

scopo di offrire una lettura sistematica di tutti i contenuti e connessioni tra i temi, oltre a quanto è stato ritenuto specifico oggetto del progetto editoriale.

3. I risultati dell'analisi automatica dei testi: i contenuti specifici

Il corpus integrale

Attraverso l'analisi delle specificità è possibile evidenziare le parole caratteristiche di ciascuna intervista, ovvero le parole sovrautilizzate in un'intervista (sub-testo) rispetto alle altre. È possibile osservare i dizionari specifici all'interno di una tabella oppure su un piano fattoriale calcolato su una matrice lessicale *forme caratteristiche* × *sub-testi*.

Nella Figura 1 abbiamo affiancato, a sinistra, il piano fattoriale relativo all'intero corpus di testi (Figura 1a), mentre a destra l'analisi fattoriale riguarda solo i testi degli intervistati (Figura 1b).

Nella Figura 1a osserviamo una polarizzazione principale (fattore 1, sull'asse orizzontale) tra il linguaggio delle agenzie pubbliche e quello dell'imprenditore: ai tre ambiti che caratterizzano le strategie della Protezione Civile, del Piano Casa Italia e dell'Agenzia della Coesione Territoriale, a sinistra nel piano fattoriale, si contrappone l'ambito dell'azione dell'imprenditore Loccioni (più a destra nel piano fattoriale, in colore viola). Un secondo fattore vede invece la polarizzazione tra il lessico della programmazione (Agenzia della coesione territoriale, in alto, in rosso) e quello dell'azione (Protezione Civile, in basso, in verde).

Prima di descrivere in dettaglio i contenuti che caratterizzano i quattro ambiti di analisi delle strategie, mette conto osservare in che misura il lessico utilizzato dagli intervistatori intervenga nel delineare la posizione relativa del lessico degli intervistati. Nel confronto tra la Figura 1a e Figura 1b, appare chiaro come i contenuti proposti dagli intervistatori (in blu nella figura a sinistra) costituiscano quasi il baricentro del linguaggio, agendo da mediatori nella conversazione e popolando il vocabolario di termini presenti nelle quattro interviste, anche se maggiormente l'ambito delle strategie pubbliche, e in particolare quelle della programmazione e del Piano Casa Italia, ma contribuendo anche al linguaggio condiviso con l'impresa.

L'elaborazione delle tematiche che emergono dall'analisi dei frammenti di testo per l'intero corpus di testi evidenzia cinque cluster (individuati dall'algoritmo *ALce-ste*). Confermando l'interpretazione degli specifici linguaggi, emersa già nel piano fattoriale, il dendrogramma nella Figura 2 presenta i termini specifici dei principali argomenti trattati dagli intervistati, indicando anche il peso percentuale che hanno i singoli cluster in termini di occorrenze e indicando la dimensione delle parole in proporzione al Test-Value nel cluster. In particolare, possiamo delineare i contributi dei quattro intervistati. Dei cinque cluster, il dendrogramma evidenzia una separazione netta del cluster 1, che caratterizza il lessico dell'imprenditore, dagli altri quattro cluster, che caratterizzano l'azione delle agenzie pubbliche, rispettivamente, la Protezione Civile (cl-2), il Piano Casa Italia (cl-3 e cl-4) e l'Agenzia della Coesione Territoriale (cl-5). La rete delle co-occorrenze di parole Figura 2, a destra, ci offre una traccia di lettura generale dell'intero corpus. Nel grafo sono considerate solo le parole di contenuto (sostantivi, aggettivi e verbi) che nel testo hanno più di quattro occorrenze.

Figura 1 Piano fattoriale calcolato su una matrice lessicale forme caratteristiche × sub-testi

Corpus complessivo con i sub-testi dei quattro intervistati e del gruppo di intervistatori:

Legenda: blue: intervistatori; rosso: Agenzia Coesione Territoriale; grigio: Piano Casa Italia; verde: Protezione Civile; viola: Loccioni Group

Corpus con solo i sub-testi dei quattro intervistati

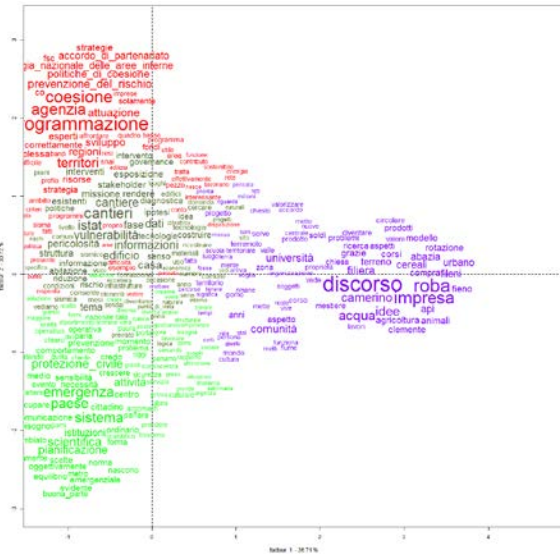
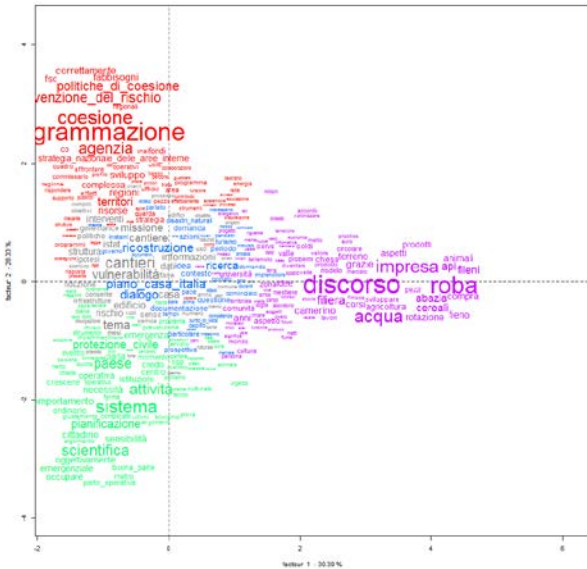
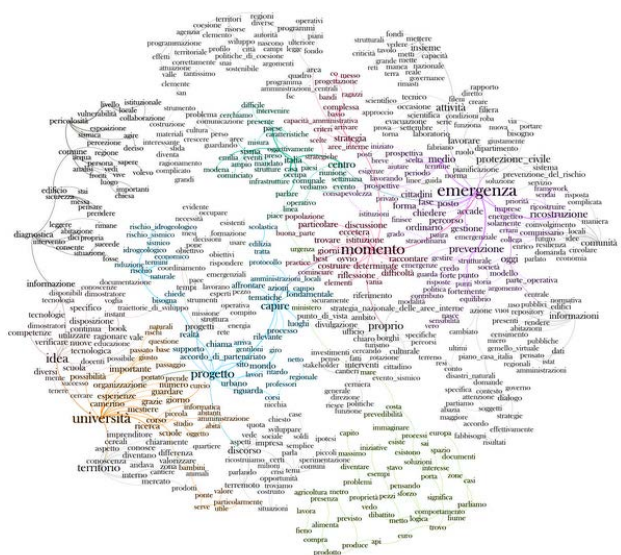
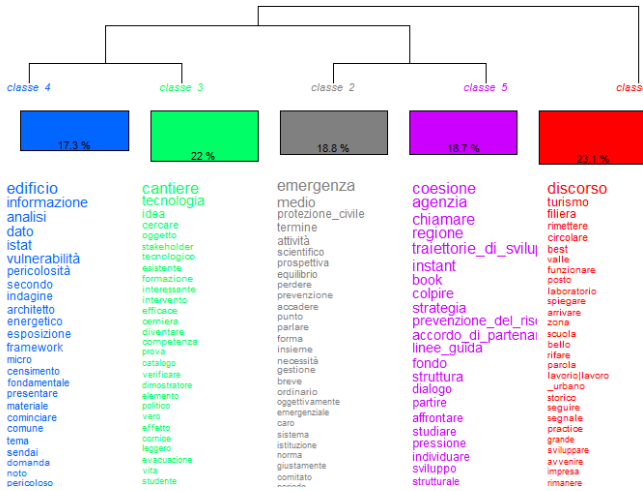


Figura 2 Intero corpus di testi

Dizionari tematici e dendrogramma dei cluster di frammenti di testo ottenuti con il metodo ALCESTE. Per ogni cluster: l'altezza dei riquadri è proporzionale al peso percentuale del cluster (in termini di occorrenze) nell'intero corpus; la dimensione delle parole è proporzionale al valore del Test-Value nel cluster

Analisi delle similitudini delle forme di contenuto. Nel grafo sono considerate solo le parole di contenuto che nel testo hanno più di quattro occorrenze. I colori evidenziano *communities* di parole (modularity)



"Emergenza" e "università" sono poli di una ragionamento che si snoda attraverso le azioni e i processi che sono necessari per "capire" e delineare "progetti" nel tempo e nei luoghi (richiamati dal termine "momento" nelle molte declinazioni temporali e di contesto). Nel piccolo corpus di testi in esame, lo sviluppo tematico è molto scandito dagli specifici ambiti di intervento dei quattro intervistati e risulta quindi più proficuo rivolgere l'attenzione direttamente all'analisi delle singole interviste.

Le strategie

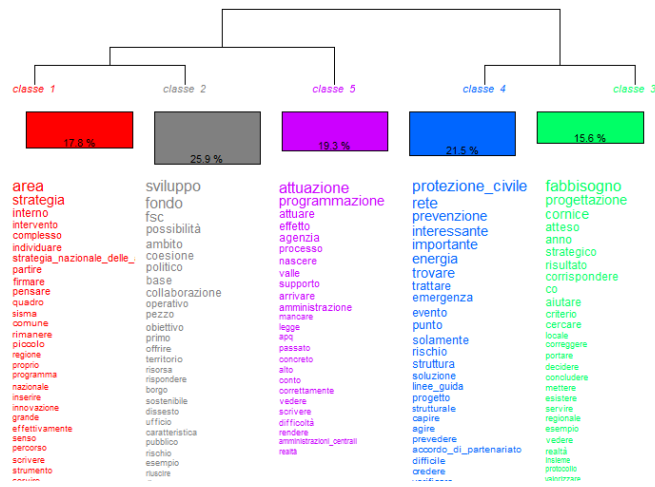
Per meglio focalizzare l'analisi sulle tematiche specifiche dei quattro ambiti strategici presi in esame, nell'analisi che segue faremo riferimento ai contenuti tratti solo da quanto espresso dagli intervistati, escludendo quindi dall'analisi testuale tutti gli interventi degli intervistatori. Per ciascuna intervista commenteremo prima gli argomenti trattati, come sono evidenziati dai cluster tematici elaborati con il metodo ALCESTE, e poi il discorso nel suo complesso, visualizzato dalla rete di co-occorrenze.

Agenzia di Coesione Territoriale. L'azione dell'agenzia si colloca nella specifica fase di attuazione della programmazione, in cui l'agenzia dà supporto alle azioni delle amministrazioni ("concreto" è un termine che caratterizza l'orizzonte dell'azione, come evidenziato nel dendrogramma della Figura 3, cl-5). Le specifiche tematiche dell'Agenzia sono quelle del sostegno allo sviluppo del territorio, con risorse del Fondo Sociale di Coesione (fsc) (cl-2); l'ambito di azione è offrire un quadro di riferimento per interventi complessi, come quelli delle aree interne, che si muovono tra molti livelli (da quello comunale, a quello regionale). Possiamo leggere nel cl-3 il modo in cui lavora l'Agenzia: analisi del fabbisogno, progettazione, cornice di attuazione, aiuto, ma anche decisioni, condivisioni). L'intreccio con altre strategie lo possiamo leggere nel cl-4, che evidenzia in particolare l'ambito strategico della protezione civile, con particolare riferimento alla prevenzione e all'emergenza, in cui l'azione delinea linee-guida, accordi di partenariato, soluzioni che riguardano interventi strutturali. Il discorso ruota attorno a due nuclei principali: "coesione-sviluppo-attuazione" e "strategia-quadro-progetti-programma". È attorno al primo nucleo che troviamo il riferimento alla collaborazione attivata con il Commissario alla ricostruzione e il nesso sulle azioni di prevenzione che sostengono l'operare della Protezione Civile. Dal secondo nucleo si dipana la descrizione dell'azione degli esperti che accompagnano l'azione dell'agenzia nei territori, l'operare nell'ambito dei Fondi Strutturali di Coesione, le specifiche azioni che delincono una cornice entro cui inscrivere fabbisogni e risultati attesi del piano di azione.

Piano Casa Italia. La specifica azione del Piano Casa Italia, che possiamo leggere dal dendrogramma della Figura 5, è l'intervento sulla tecnologia di costruzione, antisismiche (cl-1). L'università sarà un soggetto di questa azione. Workshop, dimostratore, soluzione, idea, prevedere e progettare sono parole chiave di questa azione che partirà da specificità locali. L'ambito di azione è quello del rischio (naturale, ma anche specificamente idrogeologico) che caratterizza l'intervento sul patrimonio edilizio (cl-2). In tale ambito si situano stakeholder per intraprendere l'azione che ha un riferimento nel territorio, e si caratterizza per una serie di modalità (*guardare, pensare, toccare, ragionare, ma anche il cercare di capire, di creare, di descrivere, di fare*). Uno specifico ambito riguarda le informazioni per la diagnostica (cl-3) e i dati (cl-4). Il primo tema si focalizza sulle informazioni che riguardano l'edificio, e in questo tema troviamo l'attività con l'agire che diventa operativo (volere, mettere, consentire, cominciare, trovare, decidere, vedere, evitare), fa riferimento a normative e risorse, ed evidenzia la questione della vulnerabilità. Le informazioni si collegano ad uno tema specifico che riguarda i *dati* e qui il principale soggetto è *Istat*. Parole specifiche che caratterizzano questo tema sono *repository, utilizzare, micro dati, esistente, amministrazione*: a fotografare il quadro di riferimento per l'analisi dei luoghi. L'azione specifica e gli ambiti in cui viene collocata appaiono chiari nella rete di co-occorrenze (Figura 5).

Figura 3 Agenzia coesione territoriale_Agrò

Dizionari tematici e dendrogramma dei cluster di frammenti di testo ottenuti con il metodo ALCESTE. Per ogni cluster: l'altezza dei riquadri è proporzionale al peso percentuale del cluster (in termini di occorrenze) nell'intero corpus; la dimensione delle parole è proporzionale al valore del Test-Value nel cluster



Analisi delle similitudini delle forme di contenuto. Nel grafo sono considerate solo le parole di contenuto che nel testo hanno più di quattro occorrenze. I colori evidenziano *communities* di parole (modularity)

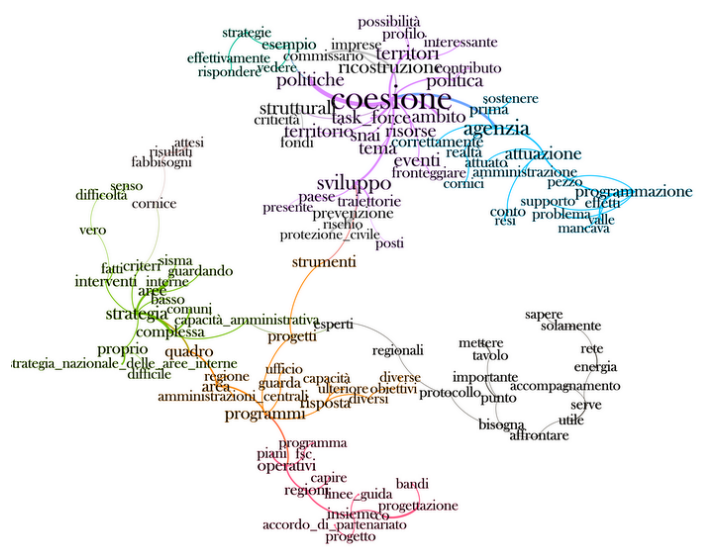
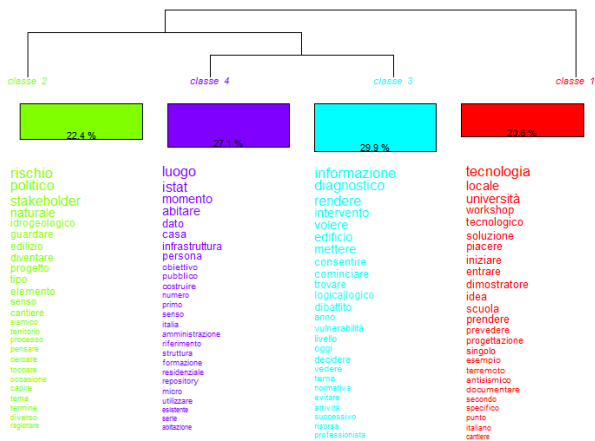
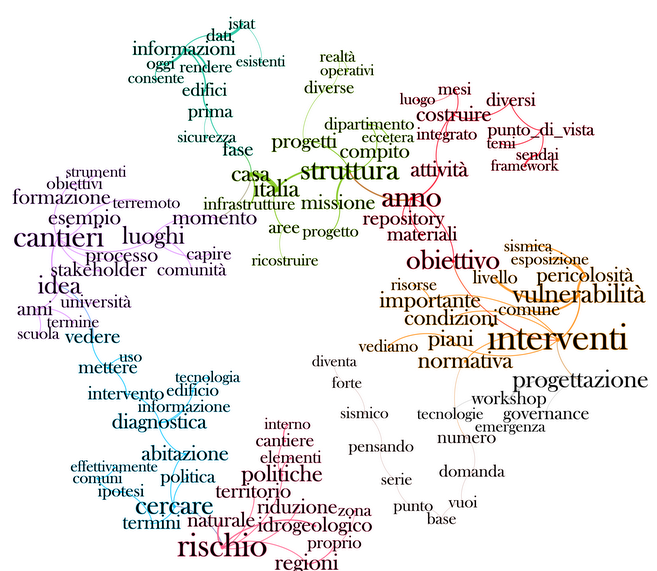


Figura 4 Piano Casa Italia_Azzone

Dizionari tematici e dendrogramma dei cluster di frammenti di testo ottenuti con il metodo ALCESTE. Per ogni cluster: l'altezza dei riquadri è proporzionale al peso percentuale del cluster (in termini di occorrenze) nell'intero corpus; la dimensione delle parole è proporzionale al valore del Test-Value nel cluster



Analisi delle similitudini delle forme di contenuto. Nel grafo sono considerate solo le parole di contenuto che nel testo hanno più di quattro occorrenze. I colori evidenziano *communities* di parole (modularity)



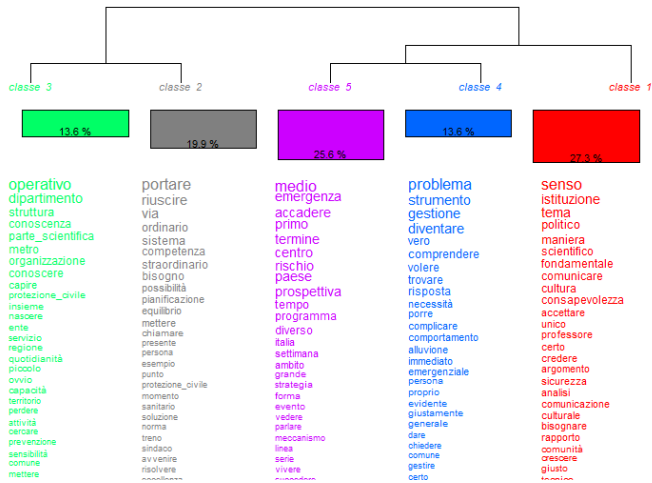
I quattro nuclei principali del discorso possono essere letti a partire dalla descrizione del Piano Casa Italia, nel contesto dei compiti della struttura di missione e con il riferimento al Sendai Framework. La base informativa e gli specifici dati per l'analisi sono una parte del discorso che, come abbiamo visto nell'analisi tematica, sostiene l'ambito degli "interventi" del Piano. Gli interventi hanno un obiettivo e il ragionamento distingue tra le condizioni di esposizione, vulnerabilità e pericolosità. Il tema dei "cantieri" si collega alla formazione (con obiettivi e strumenti), all'università, e a luoghi, che si intrecciano alle comunità nelle loro specificità nel tempo e nello spazio, colte dalle diverse declinazione del termine "momento". La parte del discorso sui i cantieri apre verso il tema del rischio, dei soggetti (le "regioni" che hanno proprie zone di rischio) e delle azioni da intraprendere, tra cui l'informazione accanto alla diagnostica.

Protezione Civile. Le azioni della Protezione Civile si sviluppano lungo due piani. Nel dendrogramma della Figura 6 troviamo, da un lato, l'ambito di azione dell'emergenza (cl-5) - rischio, prospettiva, programma - e le azioni (cl-4) (*trovare risposta, problemi, comprendere, complicare, comportamenti*), insieme con le modalità dell'agire (cl-1) (*senso, tema politico, maniera scientifica, cultura, comunicare/comunicazione, consapevolezza, argomento, analisi, accertare*). Dall'altro lato, troviamo l'azione operativa di struttura (*dipartimento, struttura, organizzazione, servizio regionale, territorio*) (cl-3), e di progettazione (*portare, riuscire, sistema di competenze*) (cl-2). Al centro del discorso, sintetizzato dalla rete di co-occorrenze della Figura 6, troviamo "il Paese", con le sue specifiche emergenze sismiche, i problemi di sicurezza. "Emergenza" (anche in questo caso nel tempo e nello spazio richiamati dal termine "momento") è lo specifico ambito di azione della "protezione civile", di cui riconosciamo gli elementi strutturali evidenziati dall'analisi cluster. A contorno del ragionamento vi è una serie di riferimenti – tracciati dal filo di collegamenti visibile nella rete di co-occorrenze - che si snodano lungo i cambiamenti degli ultimi anni per passare dalla sensibilità culturale al rapporto di fiducia nelle istituzioni, alla crescita culturale e di consapevolezza dei cittadini, ma anche al mondo della politica e della cultura scientifica. E ancora, un altro filo del ragionamento è quello che, sempre a partire dagli ultimi anni passa attraverso le scuole.

Loccioni: un imprenditore nella comunità. Il dendrogramma che caratterizza i contenuti discussi da Loccioni (Figura 7) mostra da un lato un cluster tematico indipendente dagli altri accanto ad una serie di specifici approfondimenti del discorso che caratterizza il cl 6 e si specifica a cascata negli altri cluster. Iniziamo la lettura del dendrogramma proprio dal cl-6, dove *comunità* e *luogo* entrano nel discorso sul paesaggio, l'agricoltura, l'ambiente, e in questo contesto i verbi nei frammenti di testo sottolineano il campo di azione del progettare (*cominciare, capire, spiegare, funzionare, bisognare (!)*) che delinea modelli e best practice. Altro cluster è quello che racconta del rapporto con l'Università (di Camerino) e le esperienze con le scuole. Qui il focus è attorno al formare (*imparare, raccontare, pensare, stimolare, riguardare, scoprire*), ancora *idee, progetti, territorio, reti*. Il cluster 4, che parla di *rimanere* e di *casa* apre alla questione del senso [della ricostruzione che si avvierà], ripreso nel tema del cl-1 e del cl-3. Il cl-1 si struttura attorno all'esempio dell'abazia di Sant'Urbano di Apiro, e della variegata esperienza che collega il lavoro degli antichi monaci al progetto che impegna il gruppo Loccioni nel valorizzare quel territorio; il cl-3 riferisce di un altro esempio di intervento: quello della Valle di San Clemente nella parte centrale delle Marche, di cui Apiro nel 1227 fu capoluogo della giurisdizione. L'esempio entra in dettagli del progetto, che intreccia persone, agricoltura e cultura. Un altro esempio si

Figura 5 Protezione Civile_Curcio

Dizionari tematici e dendrogramma dei cluster di frammenti di testo ottenuti con il metodo ALCESTE. Per ogni cluster: l'altezza dei riquadri è proporzionale al peso percentuale del cluster (in termini di occorrenze) nell'intero corpus; la dimensione delle parole è proporzionale al valore del Test-Value nel cluster



Analisi delle similitudini delle forme di contenuto (soglia 4). Nel grafo sono considerate solo le parole di contenuto che nel testo hanno più di quattro occorrenze. I colori evidenziano *communities* di parole (modularity)

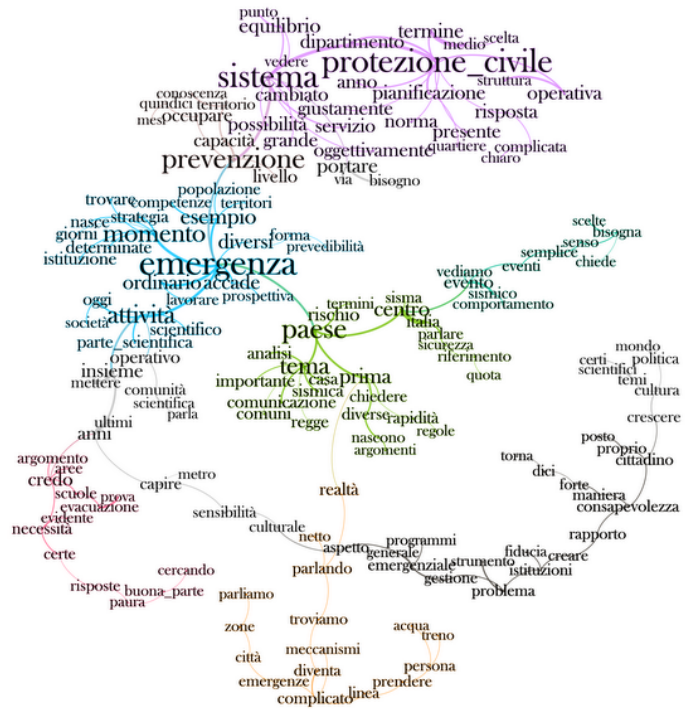
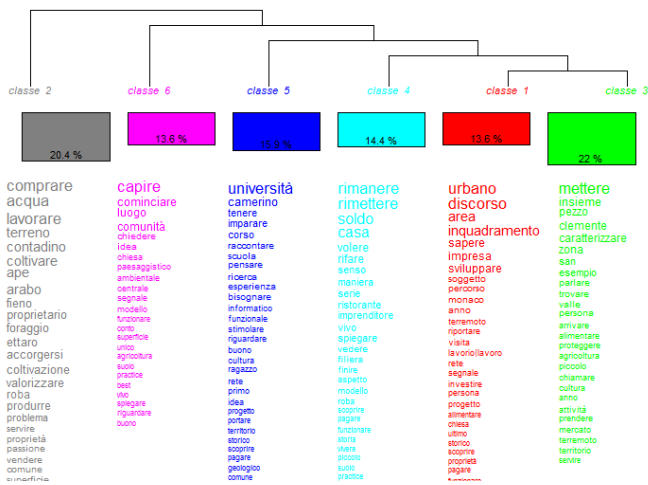


Figura 6 Loccioni

Dizionari tematici e dendrogramma dei cluster di frammenti di testo ottenuti con il metodo ALCESTE. Per ogni cluster: l'altezza dei riquadri è proporzionale al peso percentuale del cluster (in termini di occorrenze) nell'intero corpus; la dimensione delle parole è proporzionale al valore del Test-Value nel cluster



Analisi delle similitudini delle forme di contenuto (soglia 4). Nel grafo sono considerate solo le parole di contenuto che nel testo hanno più di quattro occorrenze. I colori evidenziano *communities* di parole (modularity)

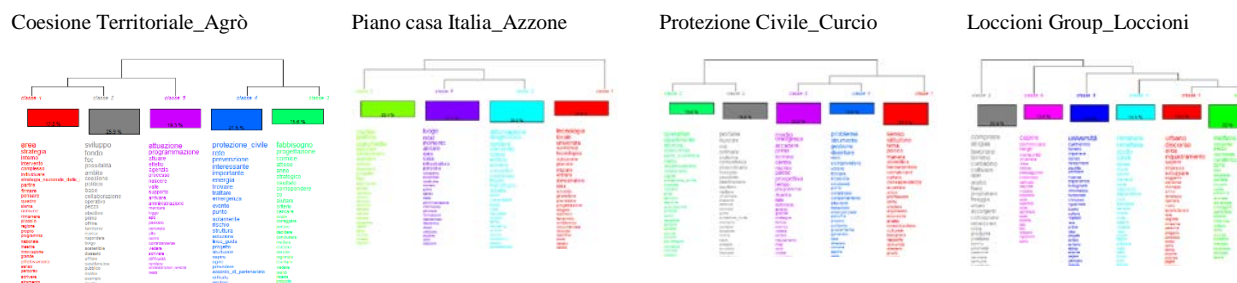


snoda lungo il tema dell'acqua, nella coltivazione del fieno destinato ad acquirenti arabi che lo comprano per i loro cavalli in Medio Oriente. Anche qui, valorizzare il territorio è tema centrale del ragionamento. Dalla rete del co-occorrenze riportata nella Figura 7 si coglie immediatamente che al centro del discorso di Loccioni è il nesso tra impresa, comunità e territorio, che si collega, da un lato, ai progetti (con esempi di interventi sull'acqua) e, dall'altro lato, al ruolo dell'università, della scuola e della cultura, essenziali per lo sviluppo di attività. Al centro dell'intervista non sono le sofisticate produzioni meccaniche realizzate da Loccioni Group, ma quelle che entrano in sintonia con i luoghi - le api, essenziali per le coltivazioni, le attività agricole e zootecnica, strettamente collegate a scuola, cultura, e sviluppo della persona. Il racconto delle attività di snoda lungo esempi che descrivono un percorso di comprensione dall'ideazione al prodotto, ma anche ai valori che questo processo contribuisce a formare. "Roba" è l'espressione che allude a processi di trasformazioni messe in gioco dagli interventi "sull'acqua".

Analisi automatica e lettura esperta

L'analisi automatica dei contenuti delle quattro interviste ci consente di individuare alcune differenze strutturali del discorso: nel caso delle strategie pubbliche, la tematizzazione è per blocchi di ambiti di azioni e specifici interventi, mentre nel caso dell'imprenditore è strutturata su due macro ripartizioni di esempi di azioni, una delle quali si suddivide in dettagli specifici, con un lessico più colloquiale che accompagna la descrizione delle esperienze realizzate con riferimenti a luoghi e persone.

Figura 7 Dendrogramma: cluster semantici dei frammenti di testo



L'articolazione narrativa dei quattro intervistati è abbastanza diversa, e abbiamo presentato come si sviluppa attraverso la lettura del grafo elaborato sull'analisi delle similitudini. Abbiamo osservato che le dimensioni del corpus sono leggermente diverse, ma soprattutto è differente la varietà linguistica che è massima per Loccioni (in media 3,5 occorrenze per parola) e minima per Curcio (in media 4,4 occorrenze per parola), a cui però è associato un corpus oggetto di analisi decisamente più vasto degli altri tre intervistati. Ma è soprattutto diverso il discorso nel suo insieme: la struttura narrativa che possiamo tracciare in termini di co-occorrenze delle parole all'interno dei frammenti (con una soglia davvero minima di almeno quattro occorrenze) mostra dei grafi molto concentrati attorno a insiemi di parole collegate tra loro nel discorso ("*communities*", nel lessico di analisi delle reti). In particolare, dai grafi leggiamo in modo distinto alcune di queste *communities*: per l'Agenzia di coesione territoriale, *coesione*, *programmazione*, *strategia*; per il Piano Casa Italia, *struttura di missione*, *Casa Italia*, *interventi*, *cantieri*, *rischio*; per la Protezione Civile, *emergenza*, *sistema*, *protezione civile*, *paese*; per Loccioni, *impresa* e *territorio*.

Nel complesso, la sintesi delle tematiche e gli specifici dizionari sono quelli che emergono dalla lettura dei testi (si veda Esposito *et al.* a cura di): un risultato a sostegno

dell'analisi automatica dei testi, che diventa indispensabile quando la dimensione del corpus sia oltre la portata dell'analisi esperta.

Dal punto di vista metodologico, l'integrazione tra gli strumenti di Analisi delle Somiglianze e la rappresentazione delle tematiche con gli strumenti di grafica sviluppati da Iramuteq rende più immediata la lettura degli ambiti di analisi, azione e intervento che caratterizzano il linguaggio specifico. Infine, la rappresentazione dei grafi realizzata con Gephi introduce verso una più agile lettura della narrazione sulle tematiche che caratterizzano il corpus e i sub-testi, ma apre anche verso analisi quantitative delle caratteristiche delle strutture delle reti di co-occorrenze delle parole nei sub-testi, come ad esempio attraverso gli indici descrittivi delle reti (in particolare, di densità del grafo e modularità).

5. Il punto di vista degli intervistati su obiettivi, condizioni e azioni necessarie per ricostruire meglio

Il dialogo attorno alle quattro problematiche, creato nella realizzazione delle interviste, evidenzia come ciascuno degli intervistati sia sostanzialmente concentrato sullo specifico ambito di azione.

Tutte le strategie entrano in relazione con le priorità indicate nel *Sendai Framework*. Sebbene con ruoli e modalità di azione diverse, sono elementi comuni a tutte le strategie: la comprensione del rischio di catastrofi, l'investimento nella riduzione del rischio e nella resilienza; il miglioramento della preparazione alle catastrofi per una risposta efficace e per 'ricostruire meglio' nella fase di recupero, ripristino e ricostruzione. Specifico della Protezione Civile è il rafforzamento della governance e della gestione del rischio di catastrofi.

Sul coinvolgimento delle comunità, presente in tutte le strategie, troviamo prospettive differenti. Per l'Agenzia della coesione territoriale, le comunità sono 'oggetto' dell'azione: sono attori rilevanti da accompagnare con strumenti di inquadramento normativo, per rendere possibili ed efficaci il complesso di azioni necessarie. Per le altre strategie sono i 'soggetti' dell'azione: nel Piano Casa Italia sono chiamate a praticare le soluzioni che verranno messe a punto, e in qualche misura, nella fase di sperimentazione, entrano in gioco nella stessa messa a punto delle soluzioni; per la Protezione Civile sono la forza essenziale perché la prevenzione si radichi in modo diffuso nella popolazione, e l'emergenza sia più efficace; per Loccioni devono diventare il motore del processo di salvaguardia dei luoghi, nel quale si identificano nuove pratiche e si consolidano opportunità di sviluppo.

La scienza dei dati (*big data science*), le informazioni disponibili, la trasformazione digitale, i big data aprono scenari nuovi su cui si stanno incamminando tutte le strategie, ma che richiedono l'entrata in gioco di molte istituzioni: le organizzazioni che producono ricerca (Università ed Enti di ricerca), le istituzioni di ricerca che generano, organizzano e rendono i dati accessibili, ma anche i soggetti che hanno i dati amministrativi (dal Registro delle Imprese alla Agenzia delle Entrate) e che dovranno diventare soggetti attivi perché da quei dati si possano generare informazioni. Occorre infatti poter accedere e integrare i dati (che dovranno essere open data), facendo interagire quei dati con il modo della ricerca, ovvero con chi i dati li elabora attraverso teorie e modelli, li analizza, li interpreta. Il tema della conoscenza e della condivisione delle informazioni è presente in tutte le strategie, ma finora non è percepito quale asse prioritario dell'azione di nessuna di quelle prese in esame.

Il delicato equilibrio fra gestione dell'emergenza e politiche di medio e lungo periodo - per la prevenzione e lo sviluppo di territori e comunità resilienti - è giocato nella separazione dei diversi ambiti di azione. L'Agenzia per la Coesione Territoriale delinea il quadro della programmazione - e quindi dell'orizzonte di medio periodo - e ne sostiene l'attuazione. Il Piano Casa Italia mette in campo una prima fase sperimentale per definire soluzioni che potranno cambiare le effettive capacità del paese di prevenire i disastri naturali, aumentare la resilienza degli edifici e contribuire al rafforzamento della capacità delle comunità locali di scegliere il sentiero di sviluppo che vorranno intraprendere e quindi mitigare le conseguenze delle devastazioni. Per la Protezione Civile, la qualità dell'emergenza si gioca sui cambiamenti che si potranno realizzare nel medio-lungo periodo, e che dipendono non solo dall'azione specifica e contingente della Protezione Civile, ma dell'intero sistema di interconnessioni tra consolidamento delle conoscenze e della formazione, cambiamenti culturali (di individui, comunità e istituzioni), e azioni istituzionali.

Riferimento bibliografici

- Balbi S. and Misuraca A. (2005), *Pesi e Metriche nell'Analisi dei Dati Testuali*, Quaderni di Statistica, vol. 7, pp. 55-68.
- Bastian M., Heymann S., Jacomy M., (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM 8*, 361-362.
- Benzecri J.-P. (1973), *L'Analyse des données*. Tome 1 : la taxinomie, Dunod.
- Benzecri J.-P. (1973), *L'Analyse des données*. Tome 2 : l'analyse des correspondances, Dunod.
- Bolasco S. (1999), *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*, Carocci, Roma.
- Bolasco, S. (2010), *TaLTaC 2.10. Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi*. Milano: LED.
- Bolasco S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Carocci, Roma.
- Esposito, F., Russo, M., Sargolini, M., Sartori, L., e Virgili, V., a cura di (in stampa), *Building Back Better: idee per ri-costruire meglio*, Carocci, Pressonline, Roma.
- Flament, C. (1962), *L'analyse de similitude*. *Cahiers du centre de recherche opérationnelle*, 4, 63-97.
- Flament, C. (1981), *L'Analyse de Similitude, une Technique pour les Recherches sur les Représentations Sociales*. *Cahiers de Psychologie Cognitive*, 1, 375- 395.
- Gambarotto, F. e Cortelazzo, M., a cura di (2013), *Parole, Economia, Storia. I discorsi dei Presidenti di Confindustria dal 1945 al 2011*, Marsilio Ed., Venezia.
- Marchand P., Ratinaud P. (2012), *L'analyse de similitude appliquée aux corpus textuels?: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011)*. In : *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT 2012*. Liège. p. 687-699.
- Pavone, P., Russo, M., Righi, S., Righi S. (2016), *Text Mining and network analysis to support improvements in legislative action. The case of the earthquake in Emilia-Romagna*, *Lexicometrica, Actes JADT (2016)*, pp.1-11, ISSN: 1773-0570, disponibile on line <http://lexicometrica.univ-paris3.fr/jadt/jadt2016/01-ACTES/83702/83702.pdf>.
- Ratinaud, P. (2009), *Iramuteq: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*. www.iramuteq.org.
- Reinert, M. (1983), *Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte*. *Les cahiers de l'analyse des données*, VIII (2), 187-198.
- Reinert, M. (1990), *ALCESTE: Une méthodologie d'analyse des données textuelles et une application: «Aurélia» de Gérard de Nerval*. *Bulletin de méthodologie sociologique*, 26, 24-54.