

dipartimento
economia
Marco Biagi



DEMB Working Paper Series

N. 88

**Documentare e comunicare l'attività di trasferimento tecnologico.
Analisi testuale della comunicazione dei poli di innovazione**
Pasquale Pavone*, Valentina Fiordelmondo**, Margherita Russo***

June 2016

* Progetto Poli.in and University of Padova

email: pasquale.pavone@unipd.it

** Progetto Poli.in

*** University of Modena and Reggio Emilia

CAPP (Centro Analisi Politiche Pubbliche)

Address: Viale Berengario 51, 41121 Modena, Italy

email: margherita.russo@unimore.it

ISSN: 2281-440X online



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Dipartimento di Economia Marco Biagi

Università degli studi di Modena e Reggio Emilia

Via Berengario 51 | 41121 Modena

tel. 059 2056711 | fax. 059 2056937

info.economia@unimore.it | www.economia.unimore.it

Documentare e comunicare l'attività di trasferimento tecnologico.

Analisi testuale della comunicazione dei poli di innovazione

Pasquale Pavone^{*}, Valentina Fiordelmondo[^] e Margherita Russo[§],

^{*} contatto: pasquale.pavone@unipd.it, Progetto Poli.in e Università di Padova

[§] Dipartimento di Economia Marco Biagi e CAPP - Università di Modena e Reggio Emilia, margherita.russo@unimore.it

[^] Progetto Poli.in

Abstract

There is an increasing attention on the needs to support SMEs in enhancing their innovation opportunities and capabilities. Through a policy measure to foster the regional innovation system, 12 innovation poles were active in Tuscany in the period 2011-2014 to provide to their members (affiliation is needed) a range of knowledge-intensive services such as knowledge and technology mapping, R&D partnership formation, technical assistance in R&D projects, technology transfer. Each pole was created as a consortium of organizations operating as public or private research centres and service centres (universities, innovation centres or technology transfer centres and firms). In this paper we adopt a statistical analysis of textual content produced by the innovation poles to identify distinctive or common elements in the various texts they produced in three years of activity and to draw some assessment of their communication on their activities.

Documents under analysis are of different types: designed as written texts (on Smart Specialization Strategy and monitoring the activities of the poles), transcripts of spoken language (the recordings of interviews); web communication. Through an automatic analysis we propose a systematic comparison of all these documents that would not be possible through direct reading of texts: on the whole it is over 56,000 graphic forms, for a total of over two million occurrences. To compare both the intra diversity across the same type of document and across the different types of documents, first we analyse each of the four body separately, in order to identify the specific content and the four languages used by the poles of innovation: "report", texts structured in the format of the monitoring; "design", the documents on smart specialization strategy; "reflection and analysis", in the transcription of interviews; and "communication", that characterizes the web sites. For this analysis, each document is associated with one or more categories (such as, for example, pole' band category, date of the document) that allow us to group or isolate relevant content in different contexts.

In this work we first introduce the set of processing of texts aimed at the selection of graphic forms on which we focus our analysis. Then, we present for each corpus the description of the analysed documents, the results of calculations performed for the treatment of the text and the analysis of the main components that explain the variability of language within each corpus. These analyses (represented by the factorial of two main components) interpret the selection of graphic forms being analysed with respect to categorical variables, defined for each document in each of the corpora. The analysis concludes with some ideas for the modelling of regional system of innovation clusters in Tuscany.

Key words: linguistic analysis; web communication; innovation poles; regional innovation policies

JEL codes: R10 General Regional Economics; O25 Industrial Policy; Y8 Related Disciplines; C88 Other Computer Software

Premessa

Vi è una crescente attenzione alla necessità di sostenere le piccole e medie imprese perché migliorino le loro opportunità e capacità di innovazione. Attraverso una misura politica per favorire il sistema di innovazione regionale, la Toscana ha finanziato 12 poli di innovazione che sono stati attivi nel periodo 2011-2014 per fornire alle imprese aderenti una serie di servizi ad alta intensità di conoscenza, quali la mappatura di tecnologie, la ricerca e la formazione, l'assistenza tecnica in progetti di R & S, il trasferimento di tecnologia. Ogni polo è stato costituito come consorzio di organizzazioni che operano come centri pubblici o privati di ricerca o centri di servizio (università, centri di innovazione o centri di trasferimento tecnologico e imprese). In questo lavoro si adotta un'analisi statistica di contenuto testuale prodotta dai poli di innovazione per identificare gli elementi distintivi o comuni nei vari testi che hanno prodotto in tre anni di attività e trarre qualche valutazione della comunicazione sulle loro attività.

Dell'ampia base documentale testuale che abbiamo raccolto per analizzare e modellizzare il sistema dei poli di innovazione fanno parte quattro tipi di documenti. Vi sono innanzitutto i documenti di monitoraggio, strutturati su una griglia tematica, prodotti dai poli di innovazione ogni sei mesi per rispondere ai requisiti per l'ottenimento del finanziamento della regione Toscana¹. Vi sono poi documenti prodotti dai poli con le proposte progettuali elaborate nel giugno 2013 per contribuire alla Smart Specialization Strategy della regione Toscana. Un'altra serie di documenti testuali è costituita dalla trascrizione dalle interviste che abbiamo realizzato nel corso del progetto, utilizzando una traccia di domande per condurre il dialogo con i nostri interlocutori sui temi della ricerca Poli.in. Un'ultima serie di testi è tratta dai siti web dei poli.

Da tutti questi documenti, letti man mano che venivano acquisiti, abbiamo tratto le informazioni utilizzate per completare i dossier sui singoli poli, o per qualificare alcuni dati, ad esempio quelli presentati nelle tabelle di performance o nell'analisi delle reti di relazioni².

Avvalendoci di strumenti di analisi automatica dei testi possiamo produrre una comparazione sistematica di tutti questi documenti, che non sarebbe possibile effettuare attraverso la lettura diretta dei testi: nel complesso si tratta di oltre 56 mila forme grafiche, per un totale di oltre due milioni di occorrenze. L'analisi statistica dei contenuti testuali nel contesto, ci consente di individuare elementi distintivi o comuni nei vari testi. È possibile anche evidenziare a quali singoli poli di innovazione o fascia di poli sono associati, traendo spunti per la modellizzazione del sistema regionale di innovazione.

Sia i documenti per la Smart Specialization Strategy che quelli dei monitoraggi sono concepiti come testi scritti (anche se questi ultimi sono strutturati su un format definito dalla Regione Toscana), mentre nel caso delle interviste si tratta di trascrizioni delle registrazioni, e quindi di linguaggio parlato. La comunicazione web usa altri canoni e registri specifici³. Mettere a confronto i contenuti testuali dei diversi tipi di documenti presenta quindi problemi che derivano dalle diversità interne alla base documentale testuale che stiamo prendendo in esame. Abbiamo quindi analizzato ciascuno dei quattro corpus separatamente, in modo da individuare i contenuti specifici e i quattro linguaggi utilizzati dai poli di innovazione: quello rendicontativo/amministrativo, strutturato dalla Regione attra-

¹ Questo requisito era previsto dal Decreto 6377/2010 con cui la regione Toscana ha dato avvio alla costituzione dei poli.

² La relazione completa prodotta dal gruppo di ricerca è consultabile on line http://www.poliinnovazione.unimore.it/wp-content/uploads/2015/10/Poli.in_Relazione-finale_2015.pdf

³ L'estrazione dei contenuti testuali dai siti web è stata realizzata con le tecniche di text mining descritte in Matteo Di Cristofaro *et al.* (2015), Promozione del sistema dei poli di innovazione nello spazio web: analisi dei contenuti e delle reti di relazioni virtuali, DEMB-Working Paper series n. 36 http://merlino.unimo.it/campusone/web_def/wpdemb/0074.pdf

verso i format dei monitoraggi; quello progettuale, dei documenti sulla smart specialization strategy; quello della riflessione e analisi, delle interviste; e infine quello della comunicazione, che caratterizza i siti web. Per tale analisi ogni documento è associato a una o più categorie (quali, ad esempio, polo, fascia di appartenenza, data del documento) che consentono di raggruppare o isolare contenuti rilevanti nei diversi contesti.

In questo lavoro introduciamo innanzitutto la serie di elaborazioni necessarie per il trattamento dei testi finalizzato alla selezione delle forme grafiche su cui concentriamo la nostra analisi. Presentiamo poi, per ciascun corpus, la descrizione dei documenti analizzati, i risultati delle elaborazioni effettuate per il trattamento del testo e l'analisi delle componenti principali che spiegano la variabilità del linguaggio all'interno di ciascun corpus. Attraverso tale analisi (rappresentata con i piani fattoriali delle due componenti principali) interpretiamo la selezione di forme grafiche oggetto di analisi rispetto alle variabili categoriali, definite per ciascun documento in ognuno dei corpora. L'analisi si conclude con alcuni spunti per la modellizzazione del sistema regionale dei poli di innovazione in Toscana.

1 Trattamento del testo

La Tabella 1 elenca il numero di documenti oggetto di analisi per corpus e polo a cui si riferiscono⁴.

Tabella 1 Documenti oggetto di analisi per corpus e polo

	date di riferimento del monitoraggio							Monitoraggi _totale documenti	Smart_Spe cialization _Strategy	Interviste	Siti web	Totale
	2011.	2012.	2012.	2013.	2013.	2014.	relazio ne finale					
	12	06	12	06	12	06						
1 OPTOSCANA	1	1	1	1	1	1	1	7	1	1	1	17
2 INNOPAPER	1	1	1	1	1	1	1	7	1	1	1	17
3 OTIR 2020	1	1	1	1	1	1	1	7	1	1	1	17
4 VITA	1	1	1	1	nu	1	1	6	nu	1	1	14
5 PIETRE	nd	nd	1	1	1		nd	3	1	1		8
6 PENTA	1	1	1	1	1	1	1	7	1	1	1	17
7 POLIS	1	1	1	1	1	1	1	7	1	1		16
8 NANOXM	1	1	1	1			1	5	1	1	1	13
9 CENTO	1	1	1	1	1	1	1	7	1	1	1	17
10 PIERRE	1	1	1	1	1	1	nd	6	1	1	1	15
11 POLO12	1	1	1	1	1	1	nd	6	1	1	1	15
12 POLITER	1	1	1	1	1	1	1	7	7	1	1	23
Distretto ferroviario									1	1	1	3
CCIAA Lu										1		1
Associazioni										3		3
Amministrazione regionale										1		1
Combinazioni su più poli										4		4
totale	11	11	12	12	10	10	9	75	18	22	11	201

nd: non disponibile; nu: file immagine non utilizzabile

Ciascuno dei 208 documenti che costituiscono i quattro corpora oggetto di analisi (Monitoraggio, Smart Specialization Strategy, Intervista, Sito web) è stato classificato in base alle categorie elencate nella Tabella 2. Per non appesantire l'analisi della matrice forme per subtesti, abbiamo deciso di condurre l'elaborazione finale solo su due variabili categoriali: il polo a cui si riferisce il documento e la fascia in cui è classificato il polo, ometten-

⁴ Tre file erano immagini di documenti testuali e li abbiamo scartati perché non siamo riusciti a farci inviare dai poli i file in formato testo. La non usabilità di uno di questi file ha comportato l'esclusione del polo Vita dall'analisi dei documenti di SSS. Gli altri due file erano relativi a relazioni di monitoraggio (una del polo Pietre e una del polo Vita) e non crediamo che la perdita di informazioni sia rilevante.

do in questa versione finale il dettaglio delle elaborazioni sulle altre categorie nell'analisi⁵.

Tabella 2 Categorie per la classificazione dei documenti acquisiti per l'analisi dei testi

Variabile	Modalità	Descrizione
<i>Esaminabile</i>	2	si/no
<i>Nome polo</i>	12	12 poli di innovazione
<i>polo_aggregazione settoriale</i>	6	Aerospazio, Automazione, ICTGestionale, Media, OSOD, Telecomunicazioni
<i>Fascia in cui è classificato il polo</i>	3	1 ^a , 2 ^a , 3 ^a fascia come da decreto di costituzione dei poli
<i>Monitoraggio_tipo di documento</i>	2	Relazione semestrale o Relazione finale
<i>monit_data</i>	6	semestre del monitoraggio
<i>appartenenza_a_distretto</i>	2	si/no
<i>distretto</i>	5	Beni_Culturali, Ditecfer, Energie_Rinnovabili, FORTIS, Scienze_Vita

La documentazione testuale oggetto di analisi è descritta nella Tabella 3 che riporta la consistenza dei quattro corpora in termini di numero di documenti, forme grafiche e occorrenze in ciascun corpus. Considerato come un unico corpus, la base testuale in esame contiene 56.551 forme grafiche: rispetto alla somma dei singoli corpora la differenza è spiegata dalla presenza delle stesse forme grafiche in più corpora.

Tabella 3 Descrizione del corpus di testi: numero di documenti, forme grafiche e occorrenze, per tipo di corpus

Corpus	Documenti	Forme Grafiche	Occorrenze
Monitoraggio	75	26.833	562.765
Smart Specialization Strategy	17	15.232	188.085
Intervista	22	14.728	222.018
Siti web	10	27.939	1.095.719
Corpus Completo	124	56.551	2.248.833

Per ciascun corpus preso in esame, le principali operazioni di trattamento del testo hanno riguardato: l'individuazione di multiwords, l'analisi del linguaggio peculiare, l'annotazione grammaticale e l'annotazione tematica. Attraverso l'analisi dei segmenti ripetuti nel testo si è proceduto alla ricerca di entità nominali di tipo multiwords. In questo modo sono state lessicalizzate le forme grafiche che costituiscono una parte del dizionario terminologico del corpus (ad esempio: beni culturali, obiettivo triennale, etc...). Successivamente si è proceduto al riconoscimento del linguaggio peculiare del corpus mediante il calcolo dello scarto quadratico medio fra le frequenze del vocabolario in analisi rispetto ad un modello di riferimento, nello specifico si è utilizzato il linguaggio della stampa periodica. Le forme grafiche sovrautilizzate nel corpus in analisi rispetto al modello di riferimento costituiscono il linguaggio peculiare del corpus: possono essere considerate “parole chiave” del testo analizzato in quanto sono presenti nel testo in misura molto maggiore rispetto a quanto ci si sarebbe potuto aspettare. Attraverso la fase di annotazione grammaticale delle forme del vocabolario è stato possibile calcolare le occorrenze dei lemmi dei verbi in ciascun corpus di testi. È stata anche generata un'annotazione delle forme del vocabolario sulla base di alcune tematiche emergenti dalle analisi svolte, in particolare sono state fatte le annotazioni tematiche elencate nella Tabella 4.

⁵ I risultati sulle altre variabili categoriali non aggiungevano informazioni di rilievo per l'analisi testuale. Nella Relazione intermedia sono riportate le analisi preliminari in cui abbiamo utilizzato anche altre variabili categoriali.

Tabella 4 Lista dei termini utilizzati per l'annotazione tematica

Categorie

Tecnologia
Innovazione
Ricerca
Progetti
Energia
Ambiente
Cultura
Finanziamenti
Università
Sostenibilità
Internazionale
Sociale
Sperimentazione
TOPONIMI INTERNAZIONALI
CONDIZIONALE

L'annotazione delle tematiche è avvenuta ricercando esclusivamente il morfema lessicale, senza prefisso o suffisso, attraverso delle query lessicali, tralasciando per questioni di tempo la possibilità di annotare semanticamente significanti diversi. Esempi di annotazione tematica:

Query: *sostenib* ; risultato: sostenibilità, sostenibile, edilizia sostenibile, ecosostenibile, auto sostenibilità...

Query: *ambient*; risultato: ambientale, impatto ambientale, ambiente, tutela dell'ambiente...

In due casi l'annotazione non è riferita alla ricerca dei morfemi, ma, rispettivamente, alle forme verbali coniugate al condizionale e ai toponimi internazionali: nel primo caso, per avere un indicatore della propositività del linguaggio e, nel secondo, per rilevare l'ambito spaziale citato dai poli. Nel caso dell'individuazione delle forme condizionali dei verbi, dopo una prima operazione di tagging grammaticale si è proceduto alla selezione e conteggio delle forme verbali annotate come condizionali nel campo imprinting, come ad esempio: *potrebbe, dovrebbero, sarebbe, consentirebbe, permetterebbe, consentirebbero*.

Per quel che riguarda l'individuazione dei toponimi, nella fase di normalizzazione sono state utilizzate le risorse di Taltac per annotare toponimi internazionali.

Completato il trattamento di ciascun corpus, per mettere a confronto il linguaggio usato nei diversi documenti, si è generata una matrice "forme per subtesti" da sottoporre ad analisi fattoriale delle corrispondenze semplici, rappresentata nel piano fattoriale. Le "parole" in analisi all'interno della matrice sono state ottenute selezionando esclusivamente le forme pretrattate, ovvero: le multiword, il linguaggio peculiare, i verbi peculiari e le forme con annotazioni tematiche, i verbi condizionali e i toponimi internazionali.

2 I sei monitoraggi semestrali e le relazioni finali

Descrizione del corpus

Il primo corpus di testi analizzato è costituito dai 75 documenti relativi ai sei monitoraggi e alle relazioni finali, prodotti dai poli (vedi Tabella 1). Il corpus in analisi è risultato essere formato da 26.833 forme grafiche per un totale di 562.765 occorrenze (Tabella 3). Le variabili associate ai documenti di questo corpus sono descritte nella Tabella 7.

Tabella 5 Monitoraggi: elenco variabili categoriali con numero di modalità e descrizione

Variabile	Modalità	Descrizione
Nome polo	12	12 poli di innovazione
Fascia di appartenenza del polo	3	1 ^a , 2 ^a , 3 ^a fascia come da decreto di costituzione dei poli
Data	6	2011.12, 2012.01, 2012.06, 2012.12, 2013.06, 2013.12, 2014.06

Trattamento del testo

Attraverso la ricerca di entità nominali di tipo multiwords sono state lessicalizzate 991 forme grafiche che costituiscono una parte del dizionario terminologico del corpus (ad esempio: *beni culturali, obiettivo triennale, attività di scouting, patrimonio culturale, capitale umano, energie rinnovabili, etc...*). Le forme grafiche che maggiormente caratterizzano il linguaggio peculiare dei monitoraggi sono riportate nella Tabella 6.

Tabella 6 Monitoraggi: forme grafiche con maggiore scarto dal modello di riferimento

Forma grafica	Occorrenz	LiPe	Forma grafica	Occorrenz	LiPe
contrattualizzati	328	550,66	partenariato	162	153,00
aggregate	374	499,97	rendicontazione	85	142,63
cluster	228	398,38	innovazione	1008	133,74
semestri	247	247,36	fotonica	65	131,09
workshop	357	222,66	imprese	3145	127,96
filiera	244	203,68	triennale	603	123,87
disseminazione	176	189,70	consuntivo	367	106,71
semestre	1796	172,20	contattate	138	105,89
progettuali	259	169,83	supporto	739	105,12
contrattualizzazione	108	160,86	aziende	2481	104,99
bandi	388	154,99	implementazione	73	104,06
attività	4652	153,84	aderenti	701	103,58

Dall'elenco dei termini del linguaggio peculiare di questo corpus di testi si può evincere come i termini *contrattualizzati* e *contrattualizzazione*, *aggregate*, *cluster*, *semestri*, *workshop*, *filiera*, *disseminazione*, siano quelli che caratterizzano il linguaggio della rendicontazione. Da osservare anche i termini *innovazione*, *imprese* e *aziende* che si caratterizzano per un altissimo numero di occorrenze all'interno dei documenti analizzati.

Nella Tabella 7 sono riportati i lemmi dei verbi più utilizzati nei monitoraggi, mentre la Tabella 8 riporta i principali lemmi dei verbi evidenziati in base alla peculiarità. Da un'analisi delle occorrenze emerge una predominanza dei verbi *erogare*, *implementare*, *supportare*, *contattare*, ma anche *incrementare*, *aderire*, *promuovere*, *favorire* e *sviluppare*. Emerge quindi una propositività da parte dei poli, che si evince anche dall'analisi degli scarti quadratici medi dei lemmi dei verbi che descrivono le azioni su cui i poli sono stati chiamati ad agire e caratterizzano il loro ruolo di sostegno ai processi di innovazione.

Tabella 7 Monitoraggi: lemmi dei verbi con maggiore frequenza di occorrenze

Lemma	Occorrenze	Lemma	Occorrenze	Lemma	Occorrenze
erogare	490	finanziare	149	generare	72
effettuare	293	coinvolgere	140	stipulare	69
favorire	255	collaborare	131	approfondire	66
individuare	254	raccogliere	129	aggregare	65
presentare	252	operare	124	concretizzare	62
svolgere	243	organizzare	113	incentrare	60
contattare	230	incrementare	110	attuare	58
aderire	202	valutare	104	valorizzare	55
partecipare	181	investire	103	diffondere	55
promuovere	177	segnalare	98	concernere	54
sviluppare	170	supportare	97	denominare	48
realizzare	169	facilitare	88	includere	48
fornire	164	stimolare	83	rimandare	45
riportare	162	completare	76	attivare	40
utilizzare	158	evidenziare	75	ampliare	40
valere	150			consolidare	40

Tabella 8 Monitoraggi: lemmi dei verbi con maggiore scarto dal modello di riferimento

Lemma	Scarto	Lemma	Scarto
erogare	74,13	sviluppare	14,78
implementare	53,04	conteggiare	13,75
supportare	41,21	facilitare	13,63
contattare	33,24	incentrare	13,41
incrementare	19,18	monitorare	13,41
aderire	19,07	computare	13,37
promuovere	17,92	concernere	12,54
favorire	17,90	incubare	12,22
individuare	16,36	concretizzare	11,92
veicolare	16,21	riportare	11,85
effettuare	15,75	stimolare	11,62
interpolare	15,37	evidenziare	10,85
aggregare	15,16	divulgare	10,49

Le annotazioni tematiche di questo corpus sono elencate nella Tabella 9. L'analisi delle tematiche evidenzia una predominanza dei temi riferiti a *tecnologia*, *innovazione*, *ricerca*, *progetti*. Molto numerose sono anche le parole che rientrano nelle tematiche dell'*energia* e dell'*ambiente*. Minore è invece la frequenza dei termini relativi alla *cultura*, alla *sperimentazione*.

Le annotazioni tematiche sui toponimi internazionali ci consentono di approfondire quale sia l'ambito spaziale a cui si fa riferimento nel corpus di testi oggetto di analisi, che presenta ben 82 unità lessicali. Per quel che riguarda l'individuazione dei toponimi internazionali, la Tabella 10 riporta la lista delle aree geografiche e dei paesi citati nei documenti di monitoraggio, con occorrenze maggiori di 2. I toponimi contenuti nelle relazioni di monitoraggio riguardano i capoluoghi e paesi europei (es. Bruxelles, Catalogna, Germania, Marsiglia, etc) e località extra-europee (es. USA, Miami, Tokio, Silicon Valley, Delhi, etc)⁶. La Mappa 1 illustra lo spazio assai ampio in cui si proiettano le attività documentate dai poli nei documenti di monitoraggio: vi è una concentrazione di riferimenti in Europa, ma anche USA, India e Cina sono aree citate in maniera relativamente consistente.

Tabella 9 Monitoraggi: classificazione tematica dei termini, numero di unità lessicali e occorrenze totali

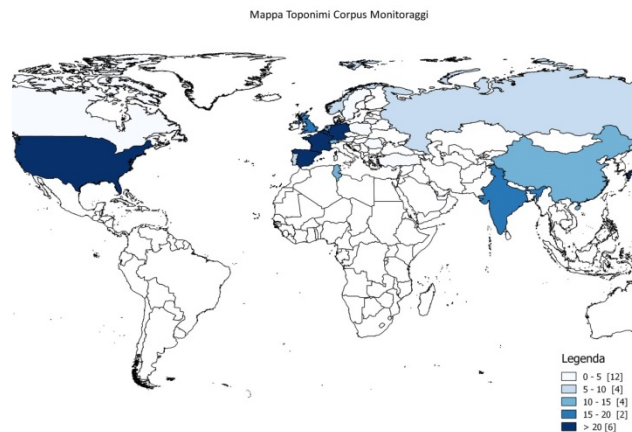
Categorie	Numero di Occorrenze	
	unità lessicali	totali
Tecnologia	128	4.455
Innovazione	121	3.605
Ricerca	83	3.092
Progetti	63	3.009
Energia	70	1.608
Ambiente	29	1.112
Cultura	32	978
Finanziamenti	48	919
Università	57	824
Sostenibilità	32	757
Internazionale	13	555
Sociale	24	420
Sperimentazione	8	306
TOPONIMI INTERNAZIONALI	82	332
CONDIZIONALE	74	616

⁶ Un approfondimento dell'analisi potrebbe categorizzare i contesti in cui i toponimi sono stati utilizzati. Tale elaborazione richiederebbe un intervento manuale che si potrà realizzare se richiesto dalla Regione Toscana.

Tabella 10 Monitoraggi: lista dei toponimi internazionali

Forma grafica	Occ.	Forma grafica	Occ.
Bruxelles	48	Germany	6
Germania	31	Amsterdam	6
USA	24	Barcelona	6
Spagna	20	Russia	6
Miami	17	Berlino	6
Belgio	16	Dusseldorf	6
Catalogna	15	Israele	6
Galles	14	Hannover	5
Tunisia	12	Virginia	5
Cina	11	Canada	5
Svizzera	11	Spain	5
Giappone	10	Turchia	4
France	10	California	4
Monaco	9	INDIA	4
Westfalia	9	Sweden	4
Kyoto	9	Netherlands	4
Tokyo	8	Dortmund	3
Marsiglia	8	Delhi	3
India	8	Gran Bretagna	3
Norvegia	8	Barcellona	3
Silicon Valley	7	Marseille	3
Portogallo	7		

Mappa 1 Monitoraggi: toponimi internazionali presenti nel corpus

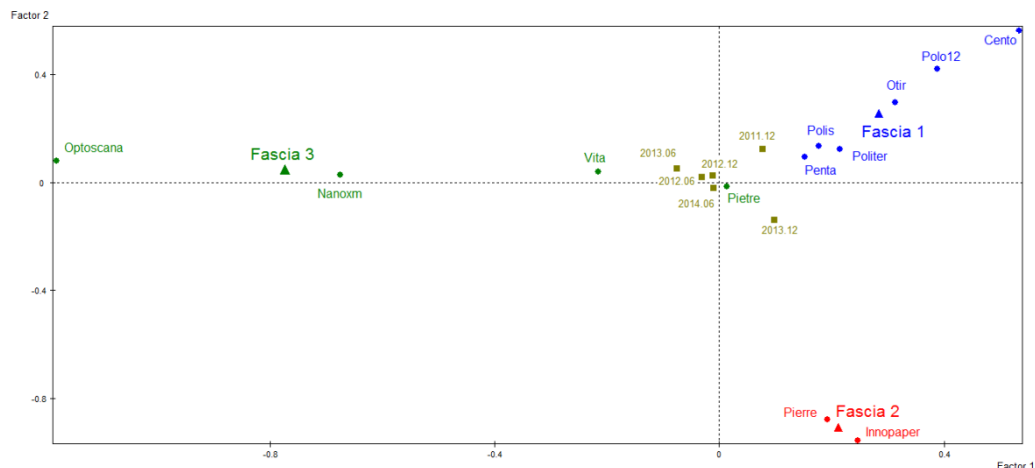


Analisi fattoriale

Nell'analisi fattoriale delle corrispondenze semplici, che rappresentiamo nel piano fattoriale forme per subtesti, abbiamo considerato esclusivamente 2.557 forme, dalle 26.833 dell'intero vocabolario, per un totale di 187.887 occorrenze, rispetto alle 562.765 del corpus oggetto di analisi. Le categorie usate nella matrice per selezionare i subtesti sono: nome del polo, fascia in cui è classificato il polo, data del documento.

La Figura 1 presenta innanzitutto la distribuzione sul piano fattoriale della struttura categoriale della matrice forme per subtesti.

Figura 1 Monitoraggi: distribuzione sul piano fattoriale della struttura categoriale della matrice



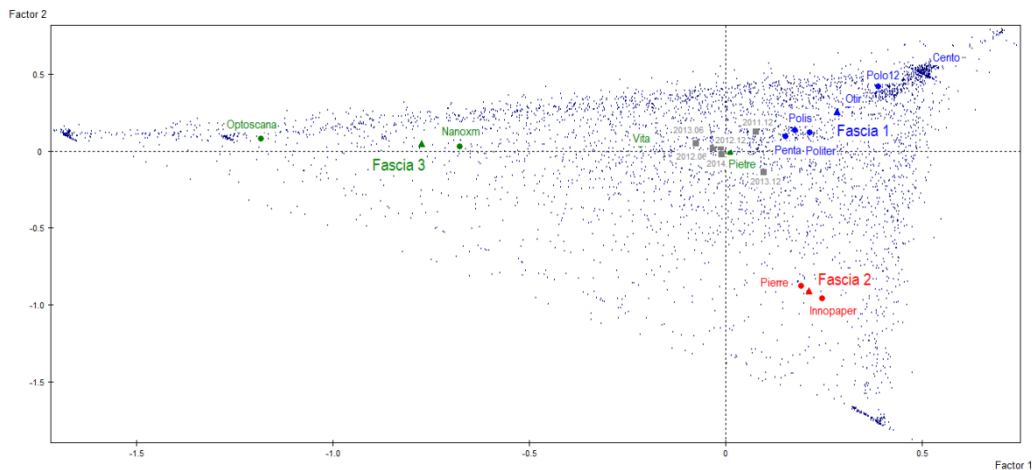
L'analisi del linguaggio in base alle categorie in cui abbiamo classificato i testi evidenzia, in primo luogo, che il linguaggio usato nelle relazioni semestrali si colloca nella zona centrale del piano fattoriale, con una differenza maggiore delle relazioni di fine anno nel 2011 e nel 2013.

Vi è una sostanziale differenza di linguaggio tra i documenti di monitoraggio dei poli di fascia 3 e di fascia 1 e 2 (fattore 1, sull'asse orizzontale). Significativa è poi la polarizza-

zione tra i poli di fascia 1 e 2 (sull'asse verticale). Sebbene le relazioni di monitoraggio avessero una struttura comune, i poli hanno linguaggi molto diversi tra loro, e questo anche all'interno della singola fascia, come nel caso della fascia 3 in cui troviamo che il linguaggio di Optoscana e il più distante, mentre quello di poli della stessa fascia, quali Vita e Pietre, diversissimi in termini di struttura, condotta e performance, è mediamente quello dell'intero corpus.

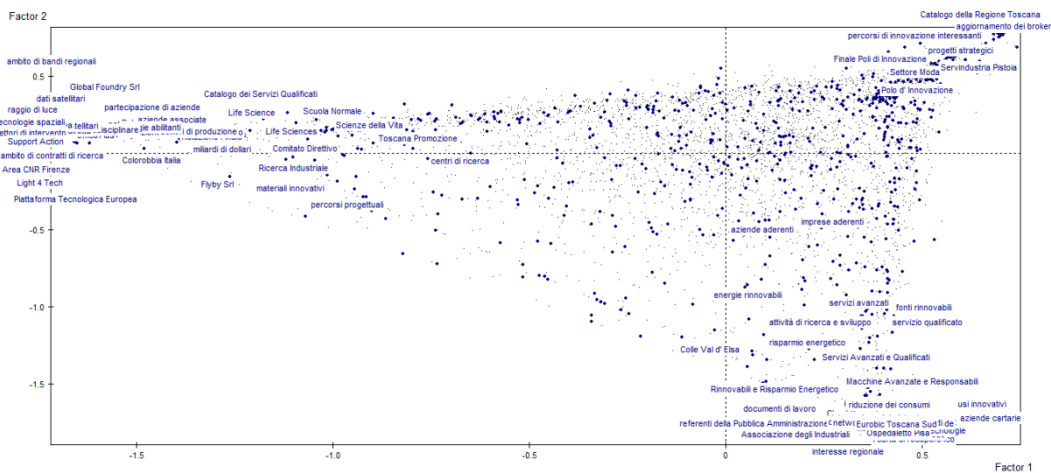
La Figura 2 riporta il piano fattoriale con la distribuzione del lessico oggetto di analisi. Ogni punto rappresenta una delle 2.557 forme grafiche selezionate. La posizione delle modalità vista in sovrapposizione con la nuvola dei punti indica la corrispondenza fra i lessici e le variabili categoriali (sub-testi).

Figura 2 Monitoraggi: distribuzione del lessico in analisi e della struttura categoriale.



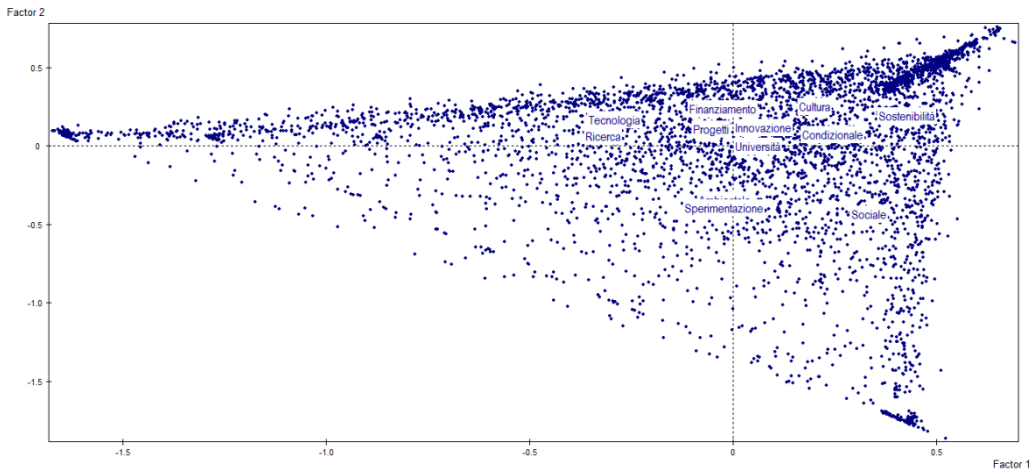
Le successive sei figure evidenziano, sullo sfondo della distribuzione del lessico in analisi, rispettivamente, la terminologia creata attraverso l'analisi delle multiwords (Figura 3), i morfemi tematici (Figura 4), una selezione dei principali lemmi dei verbi peculiari (Figura 5) e la distribuzione dei toponimi internazionali (Figura 6).

Figura 3 Monitoraggi: selezione di multiwords



Dalla distribuzione delle multiwords sul piano fattoriale è possibile iniziare a dare una prima interpretazione degli assi fattoriali. Nei quadranti a sinistra sono evidenziate le terminologie caratterizzanti i Poli della Fascia 3, in particolare Optoscana e Nanoxm. Nello specifico i termini utilizzati richiamano dei concetti di alta tecnologia, in contrapposizione ai lessici dei quadranti a destra del piano, riferiti a settori classici, con una forte attenzione alla sostenibilità.

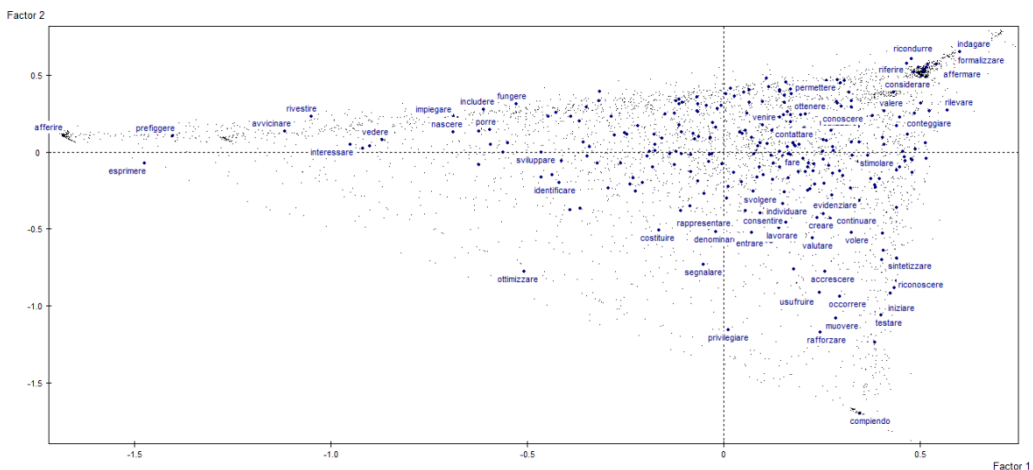
Figura 4 Monitoraggi: selezione dei morfemi tematici



I termini *tecnologia* e *ricerca* sono maggiormente usati nei poli di fascia 3, mentre *sostenibilità* e *cultura* in quelli di fascia 1. *Sperimentazione* e *tematiche sociali* sono i temi che caratterizzano maggiormente i poli di fascia 2. *Università* e *progetti di innovazione* sono morfemi tematici mediamente presenti in tutti i documenti. *Finanziamento* è peculiare dei poli di più grandi dimensioni (fascia 1). Tra i morfemi tematici *innovazione* è comune a tutti i documenti. *Cultura* è il tema distintivo di Polis, *sostenibilità* e *sociale* caratterizza maggiormente il linguaggio di Penta e Politer: per Pierre è invece *energia* il tema peculiare. *Sperimentazione*, *ricerca* e *tecnologia* sono maggiormente presenti nei poli Nanoxm e Optoscana.

Se associamo le forme condizionali dei verbi alla propositività, questa è mediamente presente nel linguaggio dei poli e maggiormente nei poli di fascia 1 e 2.

Figura 5 Monitoraggi: verbi a maggior contributo



Lo spazio dei toponimi, Figura 6, presenta un denso campo internazionale che riguarda tutti i poli. In particolare: Optoscana spazia in lungo e largo in regioni e città europee, vari stati degli Stati Uniti, e Tokio; i poli di fascia 1 citano particolarmente Corea, Nord e sud America, Russia, Svezia e Olanda (che per Optoscana sono Sweden e Neherland).

Tabella 12 SSS: forme grafiche con maggiore scarto dal modello di riferimento

Forma grafica	Occorrenze totali	Lipe	Forma grafica	Occorrenze totali	Lipe
Smart	314	907,13	robotica	91	206,93
Cluster	102	570,65	Fotonica	31	200,21
filiera	211	564,26	Max	64	191,24
policy	183	543,93	polis	91	188,31
cluster	95	531,48	ambiti	201	185,49
filiere	89	460,96	ottimizzazione	97	174,28
implementazi	88	401,93	score	82	172,66
Toscana	459	328,22	Robotica	75	170,50
target	274	256,46	sottosistema	29	169,40
validazione	44	241,13	nano	117	159,58
smart	76	219,42	eccellenze	42	159,52
tecnologie	636	209,80	sistemi	629	158,94

Per quanto riguarda i verbi, dalla Tabella 13 e dalla Tabella 14 si può vedere come, nonostante quelli con maggiore frequenza siano i verbi più comuni (*essere, avere, potere, dovere, favorire, fare, etc.*), quelli più specifici del linguaggio del corpus e quindi quelli con uno scarto maggiore siano molto più propositivi: *specificare, supportare, sviluppare, favorire, innovare, implementare, interconnettere, incentivare, massimizzare, promuovere etc.* È evidente come il linguaggio, facendo comunque riferimento a delle strategie Smart, utilizzi delle forme verbali che rimandano allo sviluppo di azioni.

Tabella 13 SSS: lemmi dei verbi con maggiore frequenza di occorrenze:

Lemma	Occorrenze totali	Lemma	Occorrenze totali
essere	2071	operare	87
avere	522	rendere	86
potere	519	riguardare	84
dovere	201	venire	84
favorire	178	rappresenta	81
sviluppare	165	promuovere	74
fare	140	fornire	73
specificare	120	ridurre	73
utilizzare	115	vedere	71
individuare	113	trovare	66
raggiungere	89	costituire	65
creare	88	portare	65

Tabella 14 SSS: lemmi dei verbi con maggiore scarto dal modello di riferimento

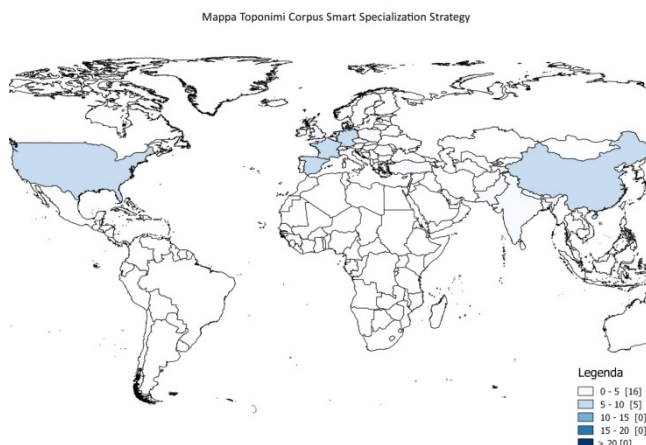
Lemma	Occorrenze totali	Lemma	Occorrenze totali
specificare	120	ottimizzare	15
supportare	42	integrare	40
sviluppare	165	elencare	42
favorire	178	operare	87
innovare	24	monitorare	13
implementare	6	focalizzare	12
interconnettere	11	impattare	7
incentivare	37	valorizzare	37
facilitare	50	abilitare	6
massimizzare	13	utilizzare	115
promuovere	74	generare	44
individuare	113	identificare	59

Per quel che riguarda l'individuazione dei toponimi internazionali, la Tabella 15 riporta la lista delle aree geografiche e dei paesi citati nei documenti di Smart Specialization, con le relative occorrenze. La numerosità dei toponimi contenuti in questi documenti è molto minore rispetto a quella dei documenti di monitoraggio: questa evidenza è ragionevole se si pensa al fatto che in questo caso il contenuto dei documenti sono strategie da attuare soprattutto all'interno del territorio regionale. Nella Mappa 2 sono stati messi in evidenza i paesi citati nei documenti.

Tabella 15 SSS: lista dei toponimi internazionali

Forma grafica	Occ.	Forma grafica	Occ.
Cina	10	Grecia	1
Francia	7	Medio Oriente	1
Germania	6	Paesi Bassi	1
Spagna	6	Slovenia	1
USA	6	America Latina	1
Asia	5	Nord Europa	1
Polonia	4	Turchia	1
Regno Unito	3	Pakistan	1
Austria	3	Danimarca	1
Giappone	3	Kyoto	1
Basilea	2	Taiwan	1
Svizzera	2	Nord Africa	1
India	2	Baviera	1
Africa	2	Illinois	1
Bruxelles	2	Ungheria	1
Portogallo	1		

Mappa 2 SSS: toponimi internazionali presenti nel corpus

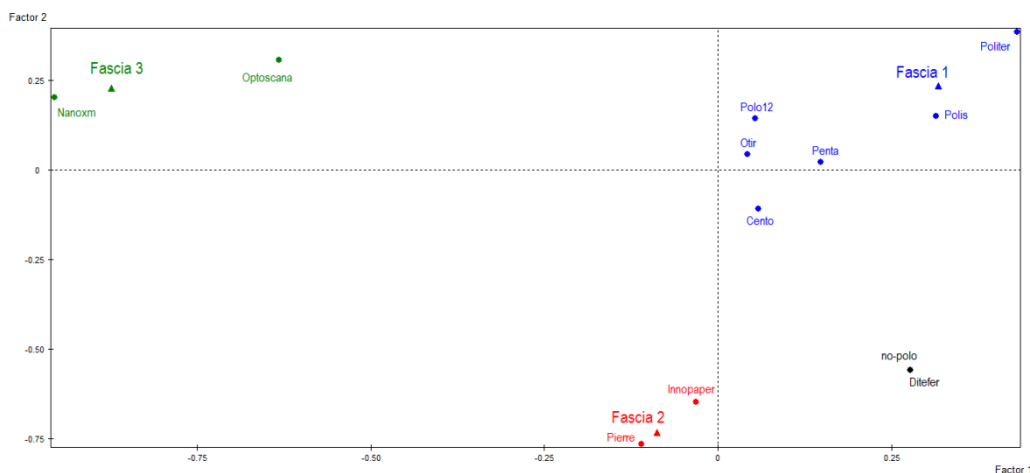


Analisi fattoriale

Nell'analisi fattoriale sono state selezionate 1.520 diverse forme grafiche (dalle 15.232 dell'intero vocabolario), per un totale di 62.708 occorrenze (rispetto alle 188.085 dell'intero corpus sulla Smart Specialization Strategy). Le categorie usate nella matrice per selezionare i subtesti sono: nome del polo, polo-fascia, data documento.

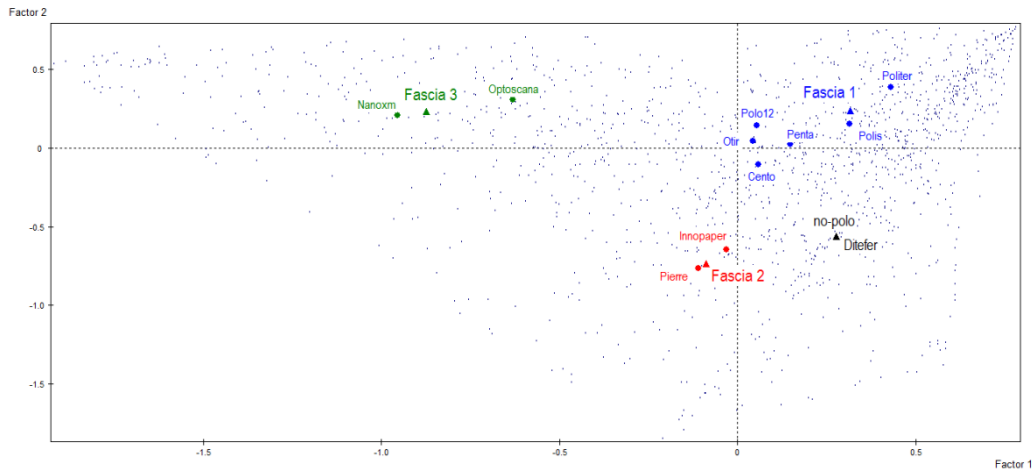
La Figura 7 presenta innanzitutto la distribuzione sul piano fattoriale della struttura categoriale della matrice forme per subtesti relativa al corpus dei documenti delle Smart Specialization Strategies. Rispetto al precedente corpus di testi, in quello ora in esame viene ancor più accentuata la polarizzazione dei linguaggi dei poli delle diverse fasce. Da segnalare come il linguaggio del distretto ferroviario sia distante da tutti i linguaggi dei poli di innovazione.

Figura 7 SSS: distribuzione sul piano fattoriale della struttura categoriale della matrice



La Figura 8 riporta il piano fattoriale con struttura categoriale della matrice del nuovo corpus di testi oggetto di analisi. La forma della nuvola dei punti appare analoga a quella del corpus dei monitoraggi.

Figura 8 SSS: distribuzione del lessico in analisi



Le successive quattro figure evidenziano, sullo sfondo della distribuzione del lessico in analisi, rispettivamente la terminologia creata attraverso l'analisi delle multiwords (Figura 9), i morfemi tematici (Figura 10), una selezione dei principali lemmi dei verbi peculiari (Figura 11), i toponimi internazionali (Figura 12).

Figura 9 SSS: distribuzione multiwords

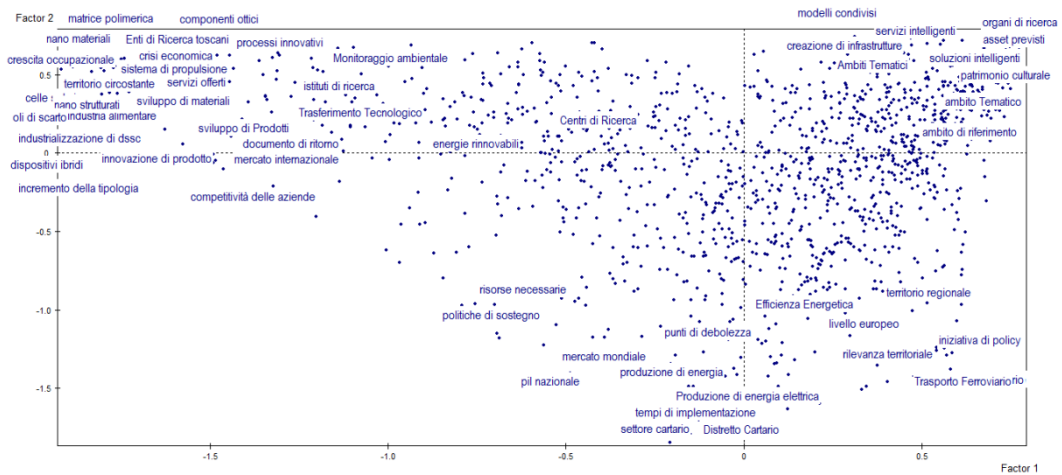


Figura 10 SSS: selezione dei morfemi tematici

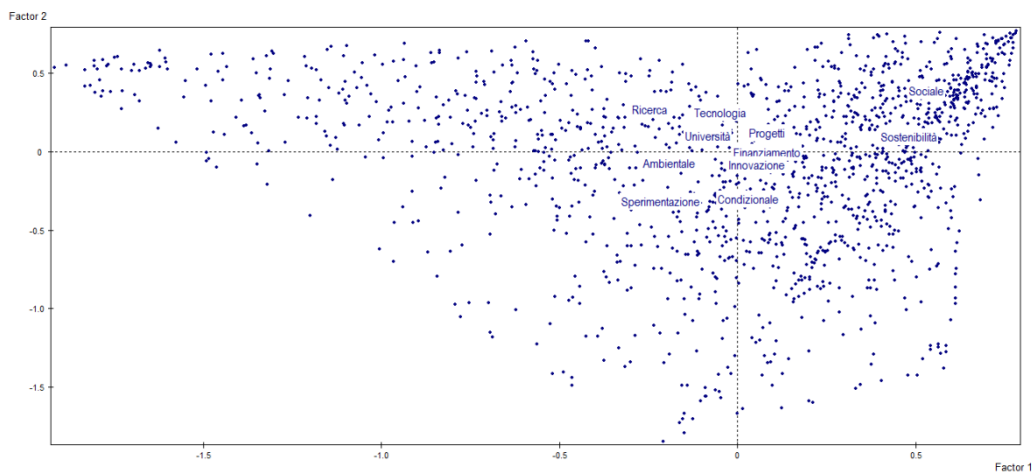
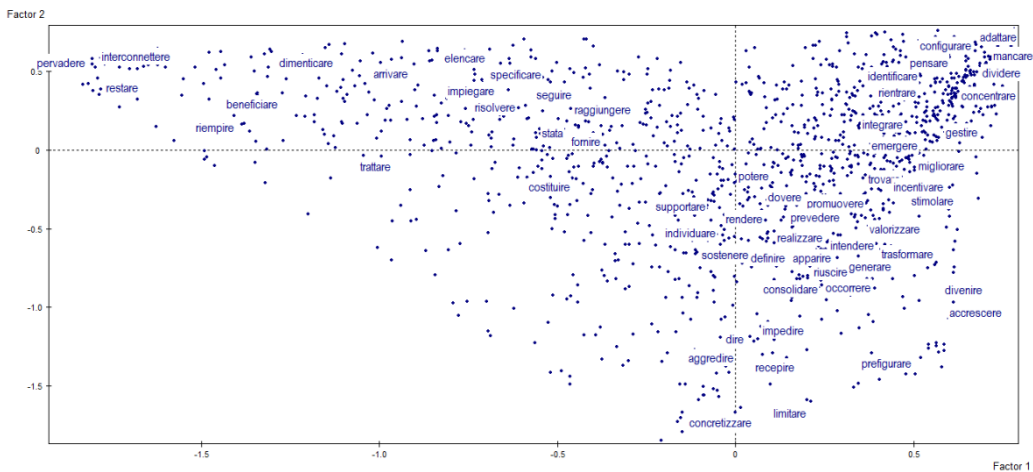
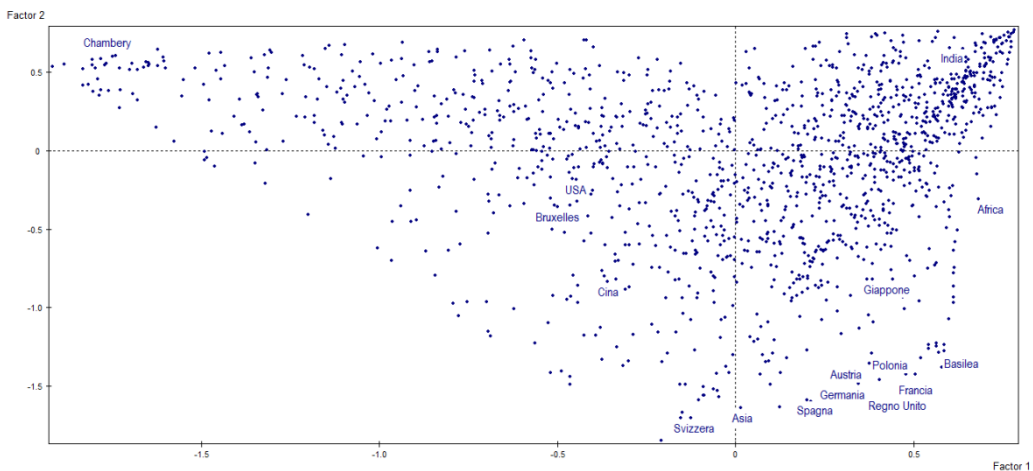


Figura 11 SSS: distribuzione del 20 percentile della distribuzione dei verbi



La distribuzione dei toponimi internazionali, Figura 12, mostra una prevalenza dei toponimi nel terzo quadrante, in corrispondenza soprattutto del distretto ferroviario e dei poli di fascia 2.

Figura 12 SSS: distribuzione dei toponimi internazionali



4 Le interviste

Descrizione del corpus

Il corpus delle interviste si compone di 22 documenti per un totale di 14.728 forme grafiche e di 222.018 occorrenze. Le variabili associate ai documenti di questo corpus sono descritte nella Tabella 16.

Tabella 16 Interviste: elenco variabili categoriali con numero di modalità e descrizione

Variabile	Modalità	Descrizione
polo_nome	22	11 poli, 5 altro, 4 multi-polo (Cento-Otir-Penta-Pierre, Cento-Penta, Polis-Politer, Politer-Polo12-Vita)
polo_distretto	4	altro (associazioni di categoria, CCAA, Regione), , multi-polo, polo
polo_fascia_ufficia	3	1°; 2°; 3° fascia come da decreto di costituzione dei poli
tipo_soggetto	4	Associazione Categoria, Gestore, Capofila, Regione

Trattamento del testo

Attraverso l'analisi dei segmenti ripetuti nel testo si è proceduto alla ricerca di entità nominali di tipo **multiwords** e sono state lessicalizzate 205 forme grafiche che costituiscono una parte del dizionario terminologico del corpus (ad esempio: *trasferimento tecnologico*,

punto di vista, centri di ricerca, distretti tecnologici, scienze della vita, distretto tecnologico, etc...). La Tabella 17 elenca una selezione delle multiwords con maggiore frequenza di occorrenze.

Tabella 17 Interviste: Lista delle multiwords

Forma Grafica	Occorrenze	Forma Grafica	Occorrenze
trasferimento tecnologico	181	Sant' Anna	33
punto di vista	167	Università di Firenze	32
Regione Toscana	141	Università di Pisa	30
servizi qualificati	106	scienze della vita	27
poli di innovazione	78	distretto tecnologico	25
centri di ricerca	77	attività di ricerca	25
associazioni di categoria	70	valore aggiunto	25
polo di innovazione	61	milioni di euro	24
ricerca e sviluppo	51	catalogo dei servizi	24
Camera di Commercio	50	livello europeo	23
soggetto gestore	43	innovazione tecnologica	23
distretti tecnologici	38	comitato tecnico	22
livello nazionale	36	livello regionale	22
centri di competenza	36	centro di ricerca	21
soggetti gestori	35	spin off	19

Anche se si possono chiaramente individuare anche delle forme grafiche tipiche del linguaggio parlato (come, ad esempio, *diciamo, quindi*), dall'elenco dei delle forme grafiche con maggiore scarto (Tabella 18) si evince che il linguaggio peculiare delle interviste è più simile al linguaggio delle Smart Specialization Strategies che al linguaggio che caratterizza i monitoraggi. Nel linguaggio peculiare delle interviste vi sono parole molto rappresentative delle tematiche principali che hanno dovuto affrontare i poli di innovazione come la *rendicontazione*, il *trasferimento tecnologico*, il *partenariato*, i *bandi*, i *progetti*, gli *incubatori* e i *servizi qualificati*. Sono emerse però anche tematiche specifiche di settore: tra le forme grafiche caratteristiche del corpus vi sono anche *cluster*, *distretto/distretti*, *filiere/filiere*, *cartario*, *fotonica*, *nautica*.

Tabella 18 Interviste: forme grafiche con maggiore scarto dal modello di riferimento

Forma Grafica	Occorrenze	Scarto	Forma Grafica	Occorrenze	Scarto
cluster	69	370,7975	filiere	22	96,74867
poli	343	218,0823	bando	234	96,64637
innovazione	343	181,644	imprese	832	95,9018
distretto	218	160,3854	qualificati	27	92,53214
polo	490	157,7874	soggetti	340	91,11742
diciamo	653	153,0233	abbiamo	1347	82,46133
rendicontazione	32	146,0932	partenariato	30	77,00001
incubatore	35	125,6814	trasferimento	62	67,48579
tecnologico	48	122,1941	nautica	55	66,69401
quindi	1747	119,1383	incubatori	12	62,42033
bandi	100	117,2743	criticità	27	61,4782
distretti	94	110,4955	attività	617	61,42807
filiere	46	104,3762	start	25	60,86527
cartario	26	99,20364	progetti	306	60,58835
fotonica	18	98,74965	servizi	393	59,8684

Per quel che riguarda i lemmi dei verbi presenti nel linguaggio peculiare delle interviste, Tabella , si richiamano chiaramente molto delle attività in capo ai poli di innovazione: *relazionare*, *finanziare*, *mappare*, *interagire*, *sviluppare*, *lavorare*, *strutturare*, *innovare*, *etc.*

Tabella 19 Interviste: lemmi dei verbi con maggiore scarto dal modello di riferimento

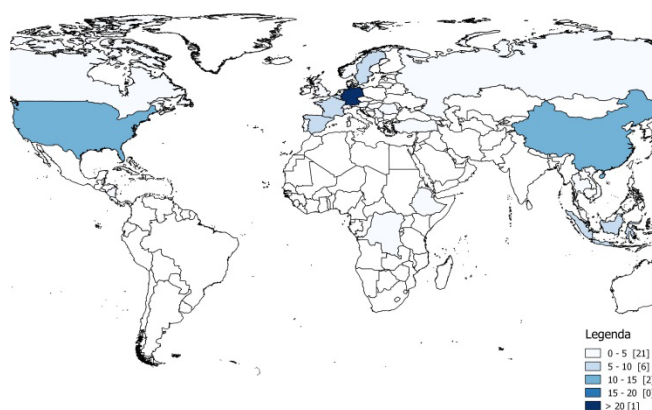
Lemma	Scarto	Occorrenze totali	Lemma	Scarto	Occorrenze totali
fare	75,01	2094	strutturare	19,35	12
essere	62,80	10640	lavorare	17,28	241
dire	39,54	1287	andare	15,96	641
finanziare	35,55	181	riuscire	14,91	151
incubare	34,99	10	innovare	13,69	13
relazionare	34,65	17	creare	13,42	150
implementare	30,11	8	sapere	12,43	383
mappare	30,00	5	contattare	12,23	35
interagire	24,91	26	funzionare	12,11	99
sviluppare	22,66	98	potere	11,61	1016
supportare	20,46	19	gestire	11,58	80
avere	20,44	4679	erogare	11,44	32

Infine, anche in questo caso si è proceduto all'individuazione dei toponimi internazionali citati nei testi. Nella è riportata Tabella 19 la lista delle aree geografiche e dei paesi citati nelle interviste, con le relative occorrenze: si fa principalmente riferimento a paesi e città dell'Unione Europea, anche se fanno eccezione gli USA e la Cina.

Tabella 19 Interviste: toponimi internazionali presenti nel corpus

Forma Grafica	Occorrenze	Forma Grafica	Occorrenze
Germania	15	Vietnam	1
Cina	13	Barcellona	1
Francia	10	Berlino	1
Spagna	8	Laos	1
Brasile	8	Madrid	1
Indonesia	7	Etiopia	1
Giappone	6	Manhattan	1
Svezia	5	Mosca	1
Monaco	4	Armenia	1
Portogallo	4	Oslo	1
Boston	3	Olanda	1
Bruxelles	3	Romania	1
Sud America	3	Turchia	1
Belgio	3	Palestina	1
Hannover	2	Stoccolma	1
Inghilterra	2	Africa	1
Amburgo	2	Nicaragua	1
Croazia	2	Montenegro	1
Tokio	2	Balcani	1
Congo	2	Grecia	1
America	2	Stoccarda	1
Seattle	2	Austria	1
Svizzera	2	Lugano	1
Urss	2	Israele	1
Canada	2	Toronto	1

Mappa 3 Interviste: toponimi internazionali presenti nel corpus

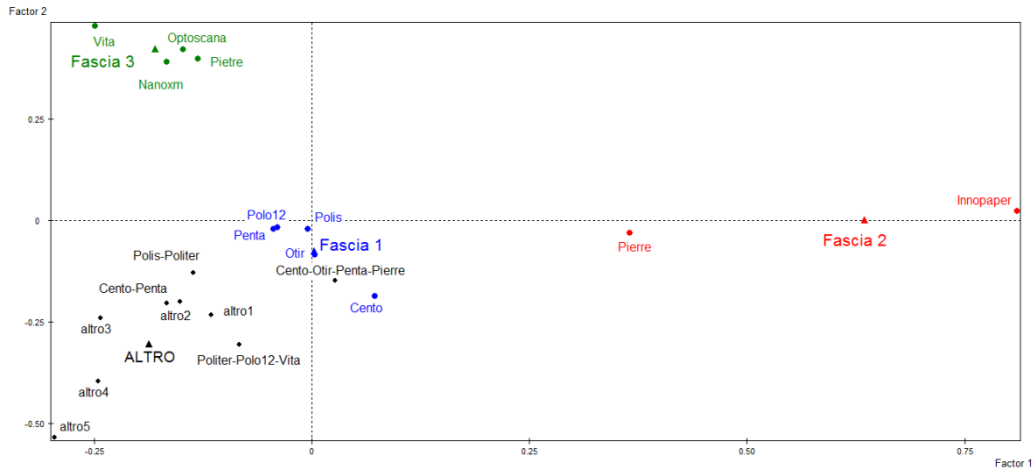


Analisi fattoriale

Dal trattamento preliminare dei testi delle interviste sono state selezionate 1.776 forme grafiche per un totale di 41.230 occorrenze. Le categorie usate nella matrice per selezionare i subtesti sono: nome del polo, polo-fascia.

Nell'analisi fattoriale delle interviste la matrice forme per subtesti (Figura 13) mostra che anche in questo corpus è la fascia di appartenenza a spiegare la gran parte della variabilità dei contenuti testuali: i poli di terza fascia, in particolare, hanno un linguaggio molto simile tra di loro, che si differenzia in modo sostanziale da quello dei soggetti capofila e gestori degli altri poli di innovazione e degli altri stakeholders intervistati.

Figura 13 Distribuzione sul piano fattoriale della struttura categoriale della matrice del corpus interviste



La Figura 14 rappresenta la distribuzione del lessico che caratterizza il corpus in analisi sottostante al piano fattoriale precedente, mentre nella Figura 15 i due piani sono stati sovrapposti evidenziando la collocazione dei poli nel piano del lessico di riferimento. È possibile quindi vedere come i linguaggi di tutte le persone intervistate si collocano nell'area centrale del piano, mostrando un certo appiattimento dei linguaggi con cui i rappresentanti dei poli di innovazione si esprimono. Unica eccezione è Innopaper, e in parte anche Pierre, il cui linguaggio si differenzia molto da quello degli altri poli: evidenza che viene spiegata dalla particolarità dell'intervista, molto più lunga delle altre e in cui sono stati invitati a partecipare anche diversi rappresentanti delle imprese.

Figura 14 Interviste: Distribuzione del lessico in analisi

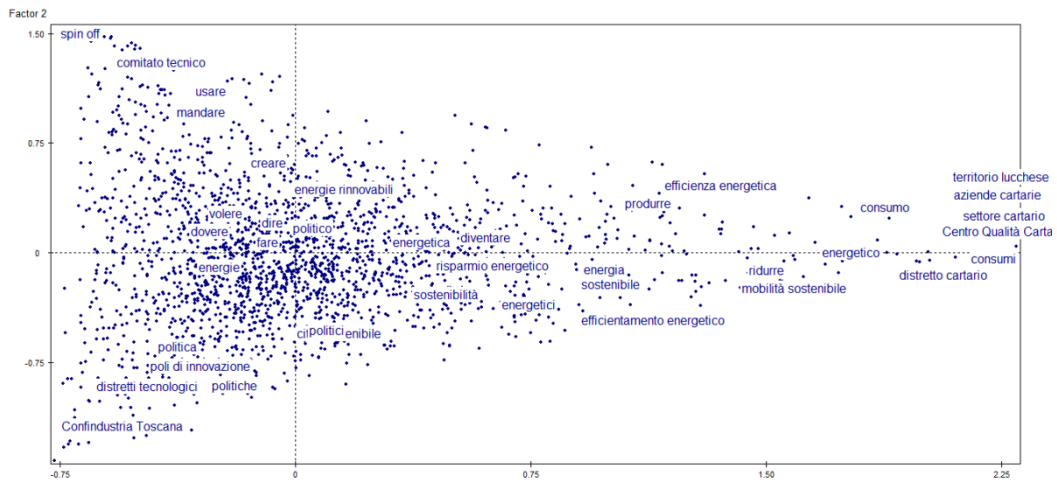
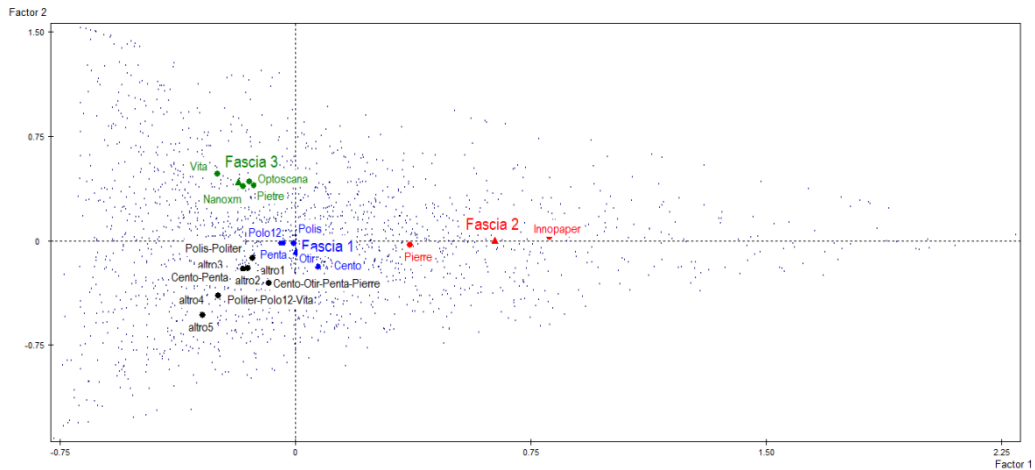


Figura 15 Interviste: Strutture delle variabili e lessico in analisi



Le successive tre figure evidenziano, sullo sfondo della distribuzione del lessico in analisi, rispettivamente, la terminologia creata attraverso l'analisi delle multiwords (Figura 16) i morfemi tematici (Figura 17), una selezione dei principali lemmi dei verbi peculiari (Figura 18).

Figura 16 Interviste: distribuzione multiwords

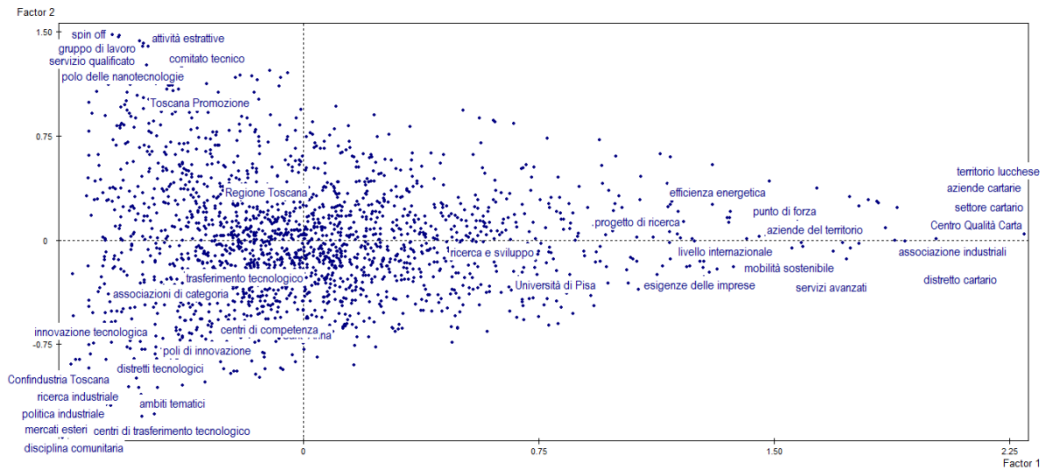


Figura 17 Interviste: selezione dei morfemi tematici

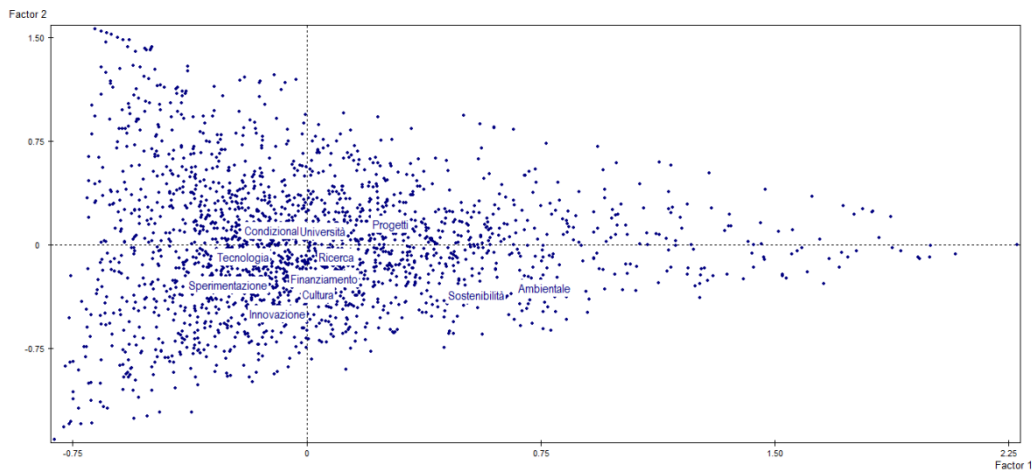
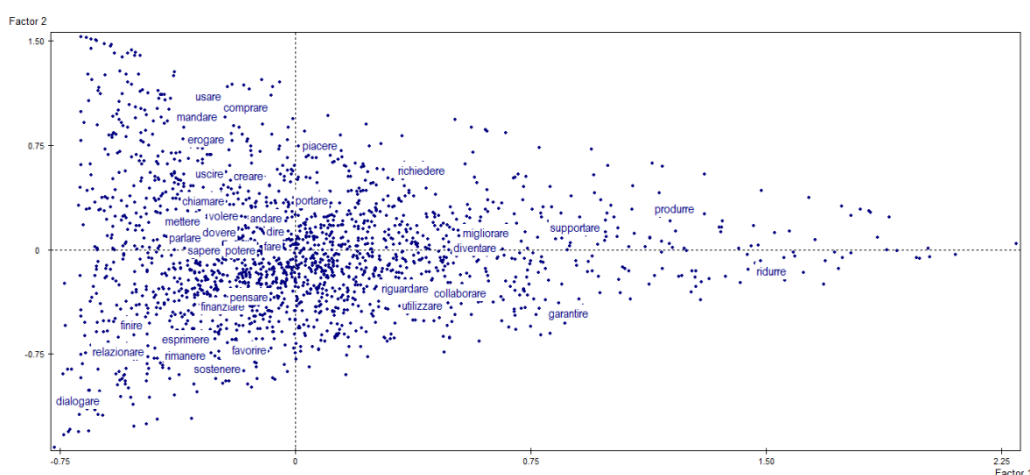


Figura 18 Interviste: distribuzione del 20 percentile della distribuzione dei verbi



5 I contenuti dei siti web

Descrizione del corpus

È stato analizzato il contenuto dei siti web di 10 poli. Abbiamo escluso il polo Pietre perché non attivo nel periodo di analisi dei siti web (maggio 2015). Il polo Polis invece ha una ricchissima comunicazione web, strutturata in decine di migliaia di pagine ciascuna con pochissimo testo⁸, che non si presta all'analisi automatica realizzata con Taltac⁹.

Il corpus analizzato è composto da un totale di 27.939 forme grafiche corrispondenti a 1.095.719. Nella Tabella 20 sono elencate le forme grafiche e il totale delle occorrenze per ogni sito web e quelle complessive. Da tenere in considerazione, nelle successive analisi del testo della differenza di consistenza dei diversi corpus di ogni sito web: il polo Vita o il polo Penta, ad esempio, hanno un peso molto maggiore rispetto al polo Otir2020 o al polo Innopaper¹⁰.

Tabella 20 Siti web: Forme grafiche e occorrenze suddivise per sito web e totali

Descrizione	Numero forme	Occorrenze
Polo - Optoscana	3.117	25.681
Polo - Innopaper	2.527	9.641
Polo - Otir2020	1.389	9.642
Polo - Vita	9.689	659.276
Polo - Penta	2.990	92.977
Polo - Nanoxm	6.522	45.237
Polo - Cento	6.070	63.203
Polo - Pierre	3.316	16.732
Polo - Polo12	3.286	82.325
Polo - Politer	11.057	91.005
Vocabolario	27.939	1.095.719

Trattamento del testo

Per quanto riguarda l'individuazione delle multiwords, sono state lessicalizzate 260 forme grafiche che costituiscono una parte del dizionario terminologico del corpus (ad

⁸ Nel complesso sono state rilevate 1.380 forme grafiche per un totale oltre un milione e mezzo di occorrenze (si veda l'appendice metodologica in Di Cristofaro et al. 2015 (op.cit.)).

⁹ Il contenuto delle pagine web di Polis è preso in esame in Di Cristofaro et al. 2015 (op.cit.) in un'analisi mirata su specifici termini.

¹⁰ Si tenga presente che, nel caso del polo Vita, si è preso in considerazione il sito del nuovo "distretto tecnologico", in quanto il polo di innovazione è stato considerato come un progetto del più ampio distretto tecnologico.

esempio: *ricerca e sviluppo, linee guida, Unione Europea, progetti di ricerca, programmazione europea, centri di ricerca, etc...*). La Tabella 21 elenca una selezione delle multiwords con maggiore frequenza di occorrenze.

Tabella 21 Siti web: Lista delle multiwords

Forma grafica	Occorrenze	Forma grafica	Occorrenze
ricerca e sviluppo	1235	Unione Europea	212
Produzione e sviluppo di processo	585	progetti di ricerca	198
Azienda Biomedicale	507	scuola superiore	185
esecuzione delle prove	474	scienze della vita	169
piccole e medie imprese	389	programmazione europea	168
coltura cellulare	382	imprese innovative	162
linee guida	342	territorio regionale	159
Polo dell' innovazione della nautica	329	leggi Biomedicale	152
conferenza stampa	238	pubblici e privati	150
Corsi di specializzazione	237	on line	149
caratterizzazione specialistica	236	Analisi genetiche	146
cantieri toscani	231	finanza di rischio	129
piano operativo	231	centri di ricerca	120

Le forme grafiche con maggiore scarto sono riportate nella Tabella 22.

Tabella 22 Siti web: forme grafiche con maggiore scarto dal modello di riferimento

Forma grafica	Occorrenze	Lipe	Forma grafica	Occorrenze	Lipe
contact	1376	1401,74	Biomedicale	842	515,00
Events	1748	1304,36	http	2501	504,57
main	1760	1133,88	Cluster	499	486,63
About	2204	953,29	Contact	471	479,75
Services	2012	933,62	Toscana	3600	449,01
Sciences	935	931,52	March	378	438,00
being	1005	907,46	This	1008	437,67
mission	1444	869,18	it	4259	426,71
Event	1112	774,95	Read	401	418,05
Other	1082	731,06	workshop	1183	412,02
small	1466	598,27	Expo	563	384,19
us	2654	588,97	analysis	389	371,67
Dom	934	571,29	Magona	294	351,14
Prec	1042	543,10	contatta	899	338,32
need	1005	533,42	affiliato	1436	336,85

Quello che salta subito all'occhio è la predominanza dei termini inglesi, evidenza prevedibile data la particolare tipologia di corpus analizzato. I siti web, per loro natura sono internazionali e i loro contenuti dovrebbero essere quanto più comprensibili ad un pubblico differenziato dal punto di vista linguistico e culturale.

Dall'elenco dei termini della Tabella 22, si possono inoltre individuare sia i termini propri della comunicazione web come ad esempio *contact, events, contatta*, sia dei termini che fanno maggiormente riferimento alle tematiche dei poli: *services, sciences, biomedicale, cluster, Toscana, etc.*

Come per le precedenti analisi, oltre alle forme grafiche complessive, si sono analizzate anche le sole forme verbali: la Tabella 23 mostra le forme che presentano scarti maggiori dalla media. Anche all'interno dei siti web, il linguaggio rispecchia molto le attività in capo ai poli di innovazione: *contattare, innovare, supportare, abilitare, promuovere, finanziare, sviluppare, etc.* Nel caso dei siti web però, si nota anche uno stimolo che i poli vogliono dare a chi visita il sito web all'essere pro-attivi nella partecipazione alla vita del polo: *contattare, cliccare, abilitare, compilare, partecipare, confrontare, scaricare, etc*

Tabella 23 Siti web: lemmi dei verbi con maggiore scarto dal modello di riferimento

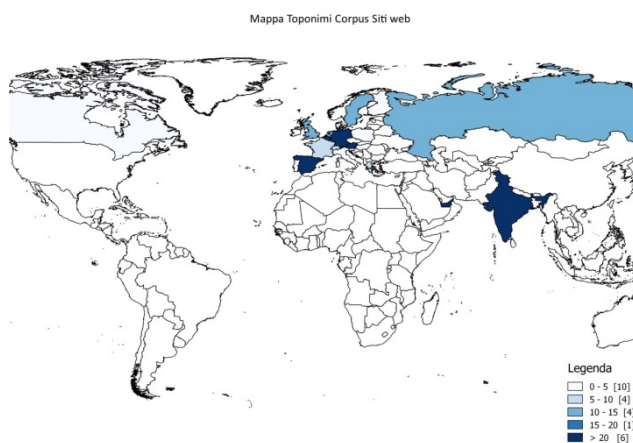
Lemma	Scarto	Occorrenze totali	Lemma	Scarto	Occorrenze totali
contattare	215,54	2718	promuovere	21,55	394
innovare	156,60	695	finanziare	17,80	460
visualizzare	92,71	348	sviluppare	17,38	369
clickare	86,76	268	partecipare	16,17	674
piaggiare	75,14	138	gloriare	16,07	28
planare	60,19	142	vistare	15,91	35
supportare	58,59	258	approfondire	15,47	189
pentirsi	56,18	535	porre	15,36	580
abilitare	51,83	112	presentare	15,25	723
binare	33,16	45	crescere	14,90	580
compilare	24,60	195	confrontare	14,73	290
allocare	24,12	35	scaricare	14,12	196

Infine, anche in questo caso si è proceduto all'individuazione dei toponimi internazionali citati nei testi. Nella Tabella 24 è riportata la lista delle aree geografiche e dei paesi citati nei siti web, con le relative occorrenze. Come evidenziato anche nella Tabella 24 e nella Mappa 4, nei siti web dei poli di innovazione si fa principalmente riferimento a paesi e città dell'Unione Europea, ed in particolare Germania, Spagna e Austria. Fanno eccezione però la Cina e la Russia, che vengono nominate più di 10 volte la prima e bene 83 volte la seconda.

Tabella 24 Siti web: lista dei toponimi internazionali

Forma grafica	Occorrenze	Forma grafica	Occorrenze
Cina	83	Netherlands	7
Vienna	75	Grecia	6
Miami	64	Israele	6
China	48	Tunisia	5
USA	42	Berlino	5
Germania	32	Sweden	4
Stoccarda	31	Canada	4
Dubai	30	Olanda	3
INDIA	28	Belgio	3
New York	24	Spain	3
Barcellona	24	Nord America	3
Chicago	24	Gran Bretagna	3
Spagna	23	America	3
California	22	BRASILE	3
Bruxelles	22	Palestina	2
Illinois	18	Londra	2
Silicon Valley	18	Portogallo	2
Monaco	14	Croazia	2
Barcellona	13	Mosca	2
Catalogna	11	Belgium	2
Hannover	10	Boston	2
Asia	10	India	2
Brasile	10	Taiwan	1
Tokyo	9	Polonia	1
Giappone	9	Usa	1
Svezia	9	Finlandia	1
Cambridge	9	Austria	1
Westfalia	9	Amsterdam	1
MOSCA	9	Ungheria	1
Svizzera	8	Los Angeles	1
France	7	Russia	1
Germany	7		

Mappa 4 Siti web: toponimi internazionali presenti nel corpus

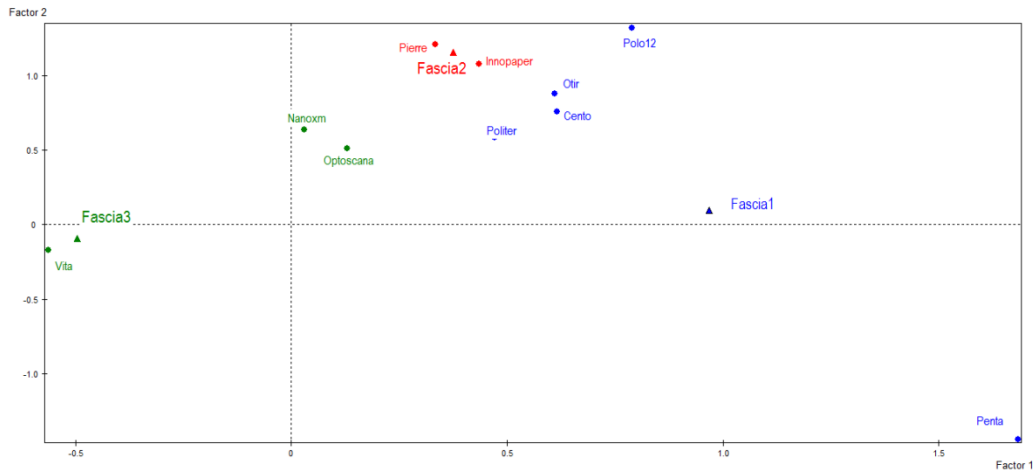


Analisi fattoriale

L'analisi fattoriale dei siti web (Figura 19) evidenzia una certa differenza di linguaggio rispetto ai tre corpora già analizzati. La diversa distribuzione del lessico dipende in primo luogo dalle caratteristiche strutturali del corpus in esame, in quanto non si tratta di un documento discorsivo, più o meno strutturato, ma di un'estrazione di testi composti per dei siti web, con tutte le limitazioni che ne derivano; in seconda battuta anche dalla diversa tipologia di comunicazione e scopo che hanno i quattro corpus (vedi par. 4.0), dai diversi destinatari e di conseguenza dal diverso standard linguistico utilizzato.

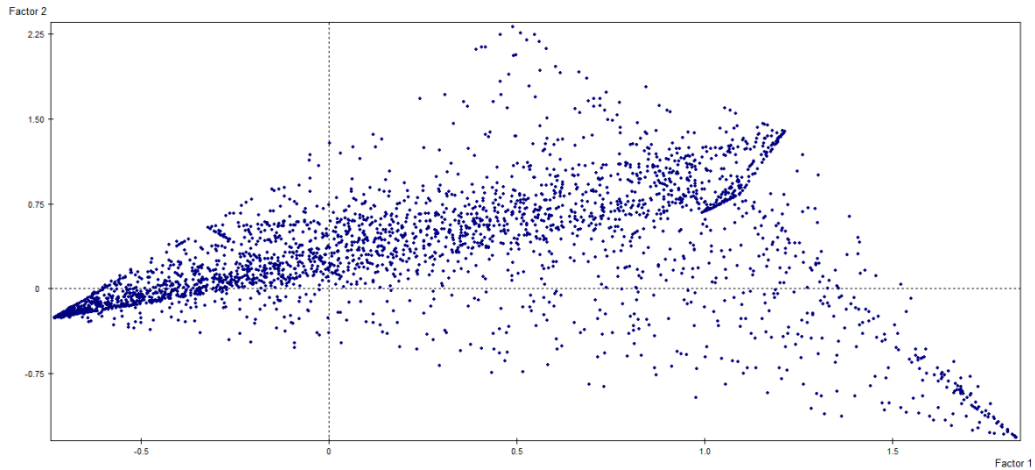
Dalla Figura 19 è evidente come ci sia ancora la presenza di una certa polarizzazione tra i poli di fascia diversa ma un sostanziale avvicinamento di tutti i poli ad un linguaggio medio. Fa eccezione il polo Penta, che si colloca molto distante da tutti gli altri poli di innovazione e, seppur in maniera più attenuata, il polo Vita.

Figura 19 Siti web: distribuzione sul piano fattoriale della struttura categoriale della matrice



La Figura 20 rappresenta la distribuzione del lessico che caratterizza il corpus in analisi sottostante al piano fattoriale precedente: viene ancor più messa in luce la polarizzazione dei due outsider e la distribuzione del linguaggio degli altri nell'ascissa del piano.

Figura 20 Siti web: distribuzione del lessico



Le successive tre figure evidenziano, sullo sfondo della distribuzione del lessico, la terminologia creata attraverso l'analisi delle multiwords (Figura 21), una selezione dei principali lemmi dei verbi peculiari (Figura 22), e la distribuzione dei toponimi internazionali utilizzati dai poli nei loro siti (Figura 23).

Figura 21 Siti web: distribuzione multiwords

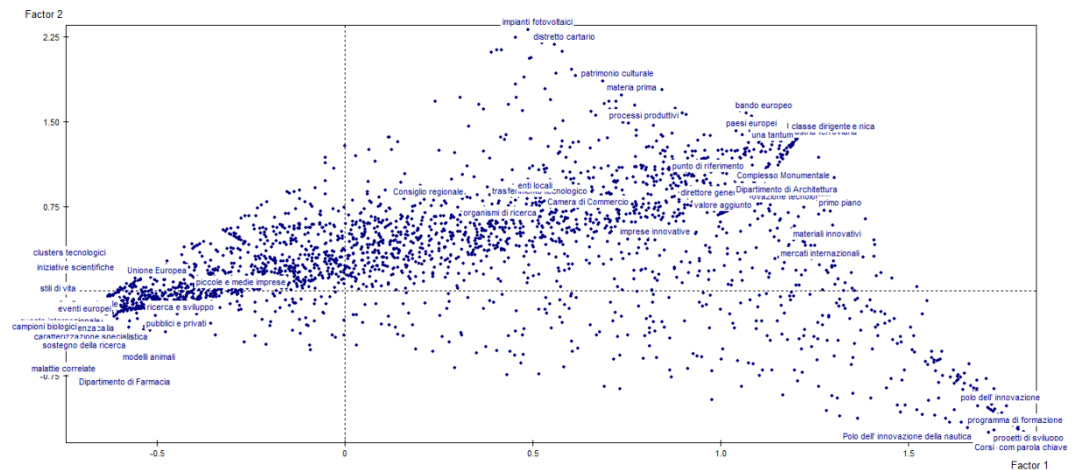
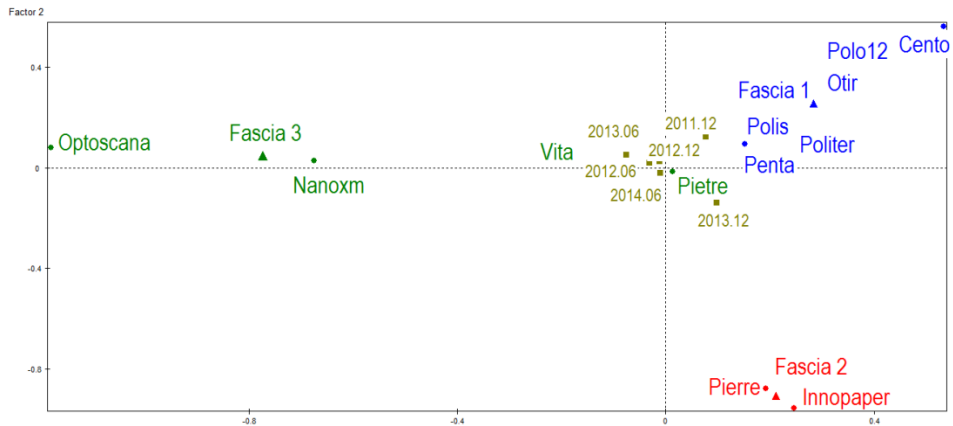
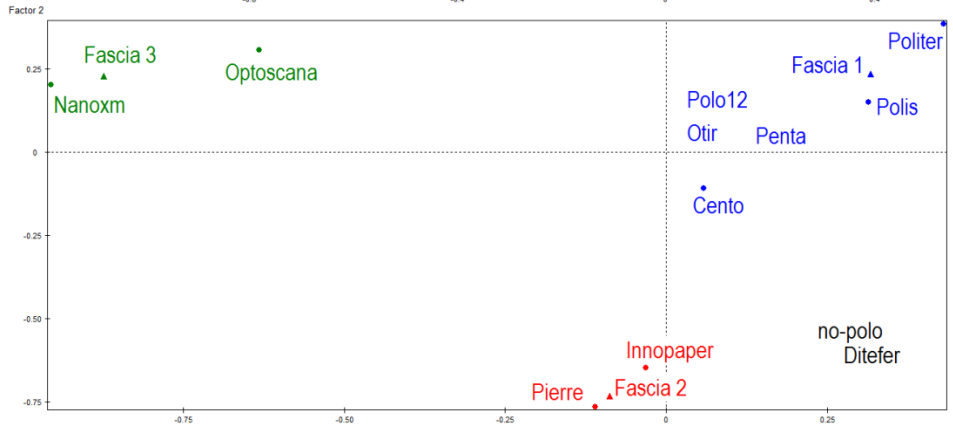


Figura 24 Differenze nei linguaggi dei poli: confronto tra i quattro corpora

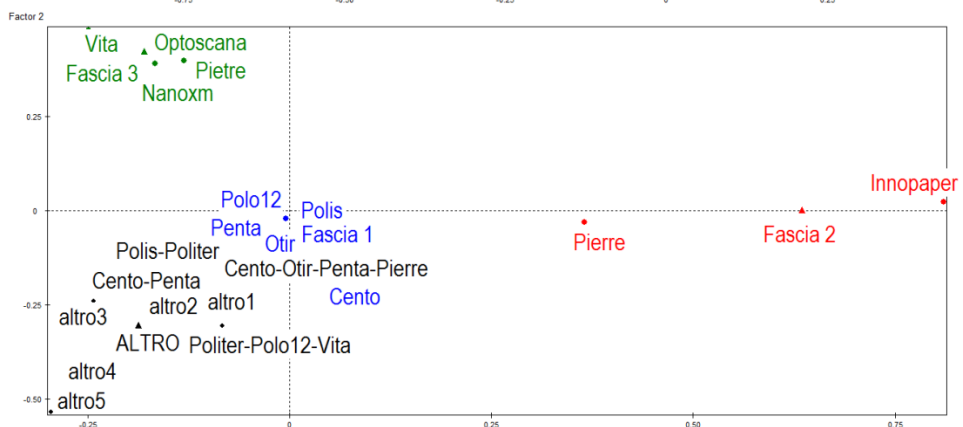
Monitoraggi



Smart Specialization Strategy



Interviste



Siti web

