Check for
updates

Linguistic sleuthing for innovators

Francesca Greco[1] · Ken Riopelle[2] · Francesca Grippa[3] ·
Andrea Fronzetti Colladon[4] · Julia Gluesing[2]

Abstract

For centuries "innovation" has been a topic of book authors and academic researchers as documented by Ngram and Google Scholar search results. In contrast, "innovators" have had substantially less attention in both the popular domain and the academic domain. The purpose of this paper is to introduce a text analysis research methodology to linguistically identify "innovators" and "non-innovators" using Hebert F. Crovitz's 42 relational words. Specifically, we demonstrate how to combine the use of two complementary text analysis software programs: Linguistic Inquiry and Word Count and WORDij to simply count the percent of use of these relational words and determine the statistical difference in use between "innovators" and "non-innovators." We call this the "Crovitz Innovator Identification Method" in honor of Herbert F. Crovitz, who envisioned the possibility of using a small group of 42 words to signal "innovation" language. The Crovitz Innovator Identification Method is inexpensive, fast, scalable, and ready to be applied by others using this example as their guide. Nevertheless, this method does not confirm the viability of any innovation being created, used or implemented; it simply detects how a person's language signals innovative thinking. We invite other scholars to join us in this linguistic sleuthing for innovators.

Keywords Text mining methods · Innovation · Innovators · Computational linguistics · Language use · Natural language processing (NLP)

1 Introduction

For centuries, "innovation" has been a topic of book authors and academic researchers as documented by Google Books Ngram and Google Scholar search results. In contrast, "innovators" have had substantially less attention. Figure 1 is the Google Books Ngram search results for the terms: "innovation" and "innovators" from 1800 to 2008, which is the most recent date for which data are available.[1] A visual inspection of the graph makes it obvious that authors

---

[1] Google Ngram Book search last accessed on April 5, 2020.
https://books.google.com/ngrams/graph?content=innovation%2Cinnovators%2C&year_start=1800&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cinnovation%3B%2Cc0%3B.t1%3B%2Cinnovators%3B%2Cc0#t1%3B%2Cinnovation%3B%2Cc0%3B.t1%3B%2Cinnovators%3B%2Cc0

✉ Francesca Greco
francesca.greco@uniroma1.it

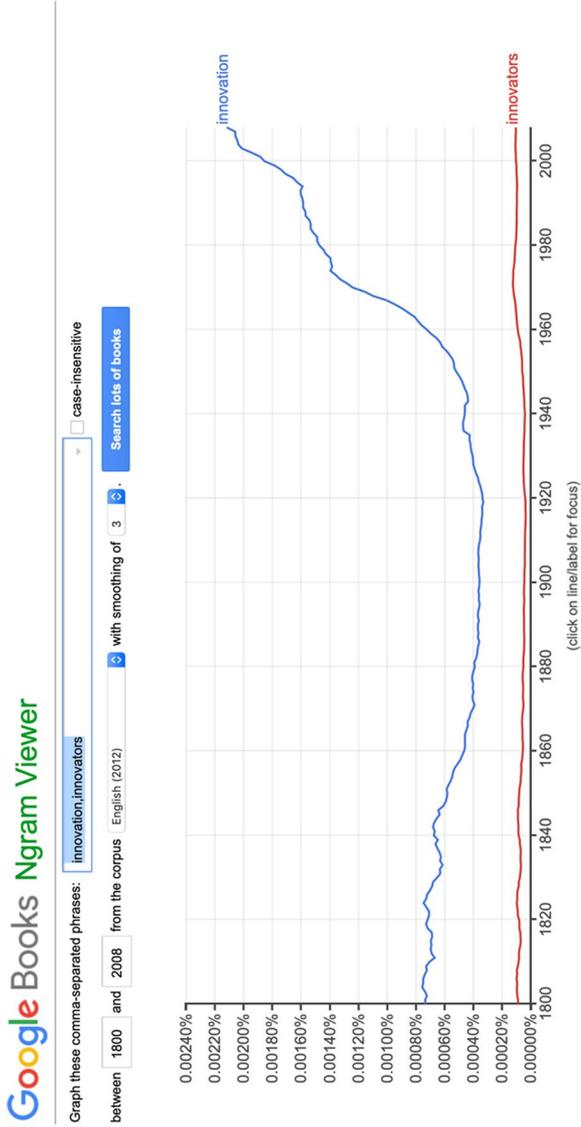Extended author information available on the last page of the article

Fig. 1 Google Books Ngram Viewer Search Results for: "Innovation" and "Innovators" from 1800 to 2008

have used the term "innovation" far more than they use the word "innovators," which has remained relatively flat during that same 200+year period. In fact, book authors' use of the word "innovation" is at an all-time high in 2008.

Likewise, a Google Scholar search on the words "innovation" and "innovators" provides another reference point about the overwhelming disparity that exists in the number of academic articles. Specifically, Google Scholar found 4,210,000 articles for a search on "innovation", compared to a meager 361,000 articles on "innovators", which represents a gap of 3,849,000 articles. Clearly, authors in the popular domain and in the academic domain have had a disproportionate focus on the process of "innovation" as compared to the "innovators", the individuals responsible for initiating the creative process. This gap in the popular and academic literature suggests that the object of discovery has been perceived as more important than the discoverers themselves. The complexity of the "innovation process" requires a more intensive discussion among practitioners and academics compared to the traits of innovators. The success of the innovation process does not depend exclusively on the characteristics of innovators, though individual traits play a big role in determining the success of a creative endeavor. We argue that scholars produced more articles on the "innovation process" because of the more complex interplay of factors in the entrepreneurial ecosystem that could influence the innovation outcome.

There seems to be a shared, well-established agreement among innovation scholars of the main behavioral traits of innovators. This agreement might have led to a decreased interest in understanding new ways to classify creative individuals, beyond the well-known classifications of innovators, early adopters, lighthouse customers, early majority and so forth (Rogers et al. 2019). The diffusion of innovation theory proposed by Rogers (1962) as well as the exploration of various traits including the so-called DNA of innovators (Amabile 1996; Dyer et al. 2011; Fürst and Grin 2018) represent important contributions to understand what makes some people more creative and innovative than others. The focus has then moved to the study of the collective efforts of teams in the innovation process, from early studies on communities of practice (Wenger 1998) to collaborative innovation networks (Gloor 2006).

Moreover, while talking about innovation could imply a *subject* evaluation of the product, a focus on the innovator calls for a specific focus on people and their collaborative networks.

For this reason, this paper makes a step to close that gap and focus on the language of "innovators" in the innovation process using the Crovitz 42 Relational Words in conjunction with two text analysis programs: Linguistic Inquiry and Word Count (LIWC) and WORDij. We call this method the Crovitz Innovator Identification Method.

Identifying and supporting innovative individuals within organizations can help managers provide the necessary autonomy and discretion required for innovation to emerge (Dyer et al. 2011). The recognition of individuals who have creative and innovative mindsets is often associated with improved motivation and increased performance (Gagné and Deci 2005). How can organizations identify the most creative and innovative individuals? The majority of the studies conducted thus far, which we highlight in the following section, focus primarily on qualitative observations, surveys and identification of personality traits. In this study, our goal is to demonstrate that methodologies based on computational linguistics can help identify innovators based on the language they use.

2 Literature review

Observing the specific behaviors and personality traits of innovators, and understanding how these traits differ from those of non-creative individuals, has been the focus of many empirical studies over the past few decades (Amabile 1988; Dyer et al. 2011; Fürst and Grin 2018; Kandemir and Kaufman 2019; Keller and Holland 1978). Traits such as imagination, interests in aesthetics, openness and intellect (Fürst and Grin 2018; Woo et al. 2017), along with personal initiative and social competence (Keller and Holland 1983; Keller 2017) have been associated with innovative behaviors and creative outcomes.

Going back to the diffusion of innovations theory proposed by Everett Rogers (Rogers 2003; Rogers et al. 2019), innovators can be identified through very distinct characteristics. They are risk-takers, venturesome, interested in experimenting with ideas and developing new ones, all traits that set them apart from early adopters, early majority, late majority, and laggards. A key limitation of Rogers' theory is that it does not take into consideration the social support resources available to them to adopt new behaviors and innovations. In addition to individual factors such as persistence, curiosity, energy, and intellectual honesty, Amabile (1996) identified specific relational traits that differentiate innovators from others, including team working, listening, building trust and positive relationships with others, both formally and informally, and building political capital. Scholars interested in the diffusion of innovation within organizations found that innovators were eager to communicate with others, and were less apprehensive about a variety of communication situations compared to their colleagues (Dyer et al. 2011; Ray et al. 1997). Innovators invest time and energy to cultivate new connections outside of their social networks, finding ideas through a network of diverse individuals that will expose them to different perspectives. Innovators seem to possess good interpersonal skills, such as being able to develop numerous contacts with others and act as boundary spanners (Fleming and Waguespack 2007).

Most of the studies have used traditional methods to offer a comparative analysis of behaviors and traits of innovators. Only a few scholars have been trying to study the uniqueness of their language use. For example, a recent longitudinal study has focused on the email exchange of R&D employees and managers to assess their online communication behaviors. Gloor and colleagues were able to differentiate online behavior of different types of innovators by measuring indicators such as employees' network positions, messages sent versus received, and their response times. The two distinct categories found to be significantly different based on individual communication behaviors were: innovators who are prolific with scientific publications or patents, and innovators who are mainly motivated by political and institutional recognition (Gloor et al. 2020). Innovators concerned with internal recognition were more central in the email networks, exchanging more emails with a higher number of contacts, acting as information brokers.

Pennebaker (2011) used computational linguistics to identify mathematically similar patterns in the language used by songwriters. By examining the Beatles' song lyrics, Pennebaker (2011) demonstrated that two people working together could produce works that are very different than if they were writing independently. He analyzed the lyrics Paul McCartney and John Lennon wrote. McCartney and Lennon collaborated on 15 of the 160 songs they wrote. The songs they wrote together were more positive, and the words they used were also different. They used more "I" words and fewer "we" words as well as shorter words than either of them used on their own. This example demonstrates how much lyrics can tell us about the personalities of creative individuals, painting a different picture from what their behaviors or their media construction would want us to believe.

Table 1 Crovitz's 42 Relational Words (From Crovitz 1970, p. 100)

about	At	for	of	round	to
across	Because	From	off	still	under
after	Before	if	on	so	up
against	Between	in	opposite	then	when
among	But	near	or	though	where
and	By	not	out	through	while
as	Down	now	over	till	with

A less explored area of investigation appears to be the application of big data analytics and the use of semantic and sentiment analysis to calculate language-related indicators to measure how the words people choose could influence innovation, creativity and social interactions (Fronzetti Colladon et al. 2020). Pennebaker et al. (2014) offered an interesting contribution in this direction, analyzing over fifty thousand essays from twenty-five thousand students and tracked college grades over four years. By using Linguistic Inquiry and Word Count (LIWC), they found that higher grades were associated with greater article and preposition use, while lower grades were associated with greater use of auxiliary verbs, pronouns, adverbs, conjunctions, and negations. Crovitz (1967) published a two-page paper titled, "The Form of Logical Solutions" in *The American Journal of Psychology*, where he referred to Polya's (1957) principles that described the aim of heuristics as the search for methods and rules that will help with both discovery and invention. Crovitz continued his argument by emphasizing that creativity and the solution to creative problems often occur when two things are brought together in a new relationship to one another. He posited that there might be a set of words that could help foster new thoughts about relations, especially thoughts that could lead to innovation.

Crovitz's answer was 42 relational words that he compiled from Ogden's 1934 book, *The System of Basic English*.[2] Table 1 reports those 42 relational words.

Weick (1979) suggests that users construct a word-wheel and put two problem concepts on discs, with the 42 relational words between them, and spin it to discover new solutions by juxtaposing the components of the problem into new relationships with one another.

Tishman and Perkins (1997) explore various ways that our own thought is "talked about" in a discourse, and describe the language of thinking as embracing the variety of descriptions we might have for our own and others' thinking and mental states. For example, the language of thinking could be used by individuals—innovators or not—when they talk about the thinking processes involved in developing a new product, examining literature, making a decision, or creating a piece of art.

As Ireland et al. (2011) demonstrated in their study on how to measure language style matching, the words we use are often a reflection of the relationship we have with the person in front of us, and the words this person is using. Corroborating the results of the emerging mirror neuron research (Iacoboni 2009), Ireland and colleagues illustrate the strong mirroring effect that occurs when two individuals communicate with each other: the value of their language style matching (LSM) will be higher if there is harmony and reciprocal interest. As stated by Greco and Polli (2020a), this style matching is not just the result of imitation but is an indication of the way relationships shape people's mental functioning and their communication style. For this reason, it is possible to profile people

---

[2] *Basic* is an acronym for "British American Scientific International Commercial".

analyzing word choices and their association (Greco and Polli 2020b), identifying their mental functioning.

As creativity is a specific way of thinking, we might expect that innovators within organizations might display common characteristics in terms of the words they use, mirroring each other especially if they work in the same R&D department and have the opportunity to interact with each other or email each other on a regular basis.

The central question that we aim to explore in this paper is the following: can the Crovitz 42 Relational Words discriminate those who are innovators from those who are not? Our search of the literature did not find any empirical studies that offered any quantitative support for Crovitz's heuristic perspective and choice of words. Our selection of Crovitz's (1970) heuristic was reinforced when he re-examined twelve problems that had previously been solved and illustrated how they could be solved alternatively using these 42 Relational Words. Weick (1979) again positioned the Crovitz 42 Relational Words as a heuristic tool for managers. A primary purpose of this study is to provide a quantitative validation of this heuristic method to support problem-solving and innovation as posited by Crovitz and Weick. Based on the literature presented in this section and the aim of our study, the hypothesis we set out to test using LIWC and WORDij is:

H1 The Crovitz relational words will discriminate between employees classified as innovators and non-innovators based on their forum text postings.

3 Case study

The case study was conducted over the course of 18 months at a European Multinational Company (hereafter EMO), whose name will remain anonymous for confidentiality purposes. The study began with EMO's HR using internal criteria to identify high potential employees as "innovators". The assessment involved various EMO stakeholders including the internal communication group, HR managers, and the Senior Leadership team, who identified and compiled a list of the most innovative individuals. The HR team finalized the assessment using the following criteria: current and past role within the organization; current job performance; involvement in innovative projects; hiring profiles and managerial perception of employees' engagement. The inclusion criteria used by the HR and leadership team all had to be satisfied in order for an individual to be classified as an innovator. The team of judges ranked the employees based on these criteria and assessed their creativity and innovation capabilities based on the innovative projects they had worked on and based on performance over the previous few years. The judges worked on these assessments independently. Subsequently, the judges met to examine discordant judgments and come to a final agreement about their classification of employees as innovators and non-innovators and then reported these evaluations to the researchers.

Based on this employee assessment, we used a private internal online communication forum to collect 16,626 posts in the Italian language from 3754 employees resulting in a large size corpus (token = 2,110,758) with 94,054 type (hapax = 51,998). Table 2 profiles the case study population. Of the 3,754 employees, 173 (5%) were classified as "innovators," posting 38% of the messages, and 3,581 were classified as "non-innovators," posting 62% of the messages.

Table 2 Case study population profile

Employee classification	Count	Classification Percent	Forum Posts	Percent Forum Posts
Innovators	173	5%	6,280	38%
Non-innovators	3,581	95%	10,346	62%
Total	3,754	100%	16,626	100%

One of the most important purposes of the online forum was to support knowledge sharing. A subset of employees regularly used the platform to share information about their work, seek and provide work-related advice, share knowledge, and help other teams in the company. Employees contributed to discussions for the diffusion of innovation and generation of new ideas. In the following section, we describe the software that we used to analyze the forum's text postings.

3.1 LIWC and WORDij software overview

To identify "innovators" who used the Crovitz 42 relational words in the forum postings we followed a method of analysis using the two software programs LIWC and WORDij. First, we provide a brief introduction of each of the software programs, followed by the steps we used to analyze the data in the context of our case study.

3.2 LIWC software

LIWC was originally designed by Pennebaker (2011) to understand how some patients recover from traumatic experiences by writing about those experiences and the emotions associated with them at the time they occurred and then afterwards. LIWC consists of a dictionary of words which assesses the percent that they occur in a given text for one or more categories. For example, the LIWC 2015 default English dictionary includes over 100 categories including parts of speech such as prepositions and conjunctions. Users also can create a customized LIWC dictionary with words and categories of their choosing. In this case, we created a LIWC custom dictionary with just the Crovitz 42 words translated into 37 Italian words with some LIWC categories having numerous entries.

In addition to the custom categories that we have defined, LIWC will automatically add a set of 16 default categories in the results. They are: Word Count (WC), Words Per Sentence (WPS), Six letter words or more (Sixltr), Dictionary (DIC), and 12 punctuation categories; All Punctuation (AllPunc), Period, Comma, Colon, Semi Colon (SemiC), Question Mark (QMark), Exclamation Mark (Exclam), Dash, Quote, Apostrophe (Apostro), Parentheses (Parenth), and Other punctuation (OtherP).

In sum, LIWC performs the task of counting the number of times a word appears in one or more categories and calculates a proportionate percent. However, LIWC does not perform any statistical tests on the results. Here is where WORDij serves as a complementary software to statistically determine if a word is used proportionally in a different way between two text corpora. The next section describes how WORDij can help conduct this type of comparative analysis.

3.3 WORDij software

WORDij[3] is text analysis software designed to determine if the relative frequency and counts of a word or word pair in two text files are statistically similar or different from one another. Two statistical tests are performed: A Z-test for relative proportions and a Chi-Square test on the counts. WORDij has two additional features that are important for this case: the ability to recode words and the ability to use an include file. WORDij enables users to create a recode file that contains a set of words that can be standardized as common names, abbreviations or compound words. For example, "United States" can be recoded to "U.S." or to "United_States", or the Italian word "tra" can be recoded to "fra" (as both have the same use and meaning). The recode function enabled us to translate the Crovitz 42 relational English words into Italian, where often there was more than one Italian word for the English word, such as translating the English word "among" into the Italian "tra" and "fra". WORDij is fast, scalable and is free for academic use.

4 The Crovitz innovator identification method

Table 3 shows the Italian vocabulary we created, starting from the English Crovitz 42 Relational Words. It is important to notice that some words had multiple translations, which were all considered thanks to the recoding function of WORDij.

5 Results

LIWC 2015 was run on the innovator and non-innovator text files using the custom Crovitz LIWC Italian Dictionary of 37 Categories. See Table 4 and Table 5. WORDij's Z-Utility Word module was run comparing the relative frequencies and counts between the innovator and non-innovator files. This analysis produced Table 6.

5.1 LIWC results

The LIWC results are presented in two tables. Table 4 shows the LIWC Five Standard Default Measures and the differences between innovators and non-innovators as well as a T-test of their significance. Table 5 shows the LIWC Twelve Standard Punctuation results and the differences in the use of punctuation between innovators and non-innovators as well as a T-test of their significance. We calculated the t-statistics according to the results of preliminary Levene's tests, indicating whether equal variances could, or could not, be assumed.

Table 4 results indicate three significant differences highlighted in green: the 173 innovators write substantially more than the 3,581 non-innovators by 1,481,707 words or 356%. They also write much longer sentences (42.5 vs 19.48 WPS) and use longer words, which indicate more complex language to describe concepts. Innovators use about 17.69% more six letter (Sixltr) words than the 3581 non-innovators. In addition, innovators have about a 1% less dictionary match rate as compared to non-innovators, 17.04 vs 18.01 respectively, which is close to being significant with a score of 0.055. This gap might be attributable to

---

[3] WORDij software is complimentary for academic use.

Table 3 Italian Translation of the Crovitz 42 Relational English Words and WORDij Recode and Include Lists

11	Italian Translation	WORDij Recoded To	WORDij Italian List n=37
About	circa, grossomodo, incirca	circa	circa
Across	attraverso		attraverso
After	dopo		dopo
Against	contro		contro
Among	tra, fra	tra_fra	tra_fra
And	e		e
As	come		come
At	a, allo, alla, agli, alle	a	a
Because	poiché, poichè, perché, perchè, siccome	perché	perché
Before	prima		prima
Between	*See: Among*		
But	ma, però	ma	ma
By	entro		entro
Down	giù, basso	giù	giù
For	per		per
From	da		da
If	se		se
In	in, nel, nella, nelle	in	in
Near	vicino		vicino
Not	non		non
Now	ora		ora
Of	di		di
Off	spento	via	via
On	su, sul, sullo, sulla, sulle, sugli, sopra	su	su
Opposite	opposto, contrario	opposto	opposto
Or	or		or
Out	fuori		fuori
Over	*See: ON*		
Round	dietro, intorno,	dietro	dietro
Still	ancora		ancora
So	così		così
Then	allora, poi	poi	poi
Though	sebbene, nonostante, benché, benchè, tuttavia, seppure, eppure	sebbene	sebbene
Through	*See: Across,*		
Till	fintanto, fino, finché, finchè	fino	fino
To	*See: AT, See: FOR*		
Under	sotto		sotto
Up	*See: ON*		
When	quando		quando
Where	dove		dove
While	mentre		mentre
With	con		con

Table 4 LIWC five standard default measures

Category	Innovators n=173	Non-innovators n=3,581	(Innovators – Non Innovators)	Percent Difference (Innovators – Non Innovators) /Non Innovators %	T-test Significance (2-tailed)
Segment	1	1			
WC	1898034	416327	1481707	355.90	0.000
WPS	42.50	19.48	23.02	118.17	0.000
Sixltr	34.59	29.39	5.20	17.69	0.000
Dic	17.04	18.01	−0.97	−5.39	0.055

Table 5 LIWC standard punctuation measures

Category	Innovators	Non innovators	(Innovators – non innovators)	Percent difference (Innovators - non - innovators) /non innovators	T-test significance (2-tailed)
Colon	0.02	0.52	−0.50	−96.15%	0.000
Question mark	0.13	0.94	−0.81	−86.17%	0.000
Exclam	0.09	1.41	−1.32	−93.62%	0.000
Dash	1.48	0.57	0.91	159.65%	0.243
Quote	0.67	0.18	0.49	272.22%	0.846
Apostrophe	2.620	2.460	0.160	0.065	0.000
Parenthesis	1.330	1.100	0.230	0.209	0.018
Other punctuation	3.170	2.260	0.910	0.403	0.000

Table 6 WORDij Z-word results for the Crovitz 37 Italian/English Words

Crovitz italian	Crovitz english	Innovators Frequency	Non innovators frequency	Innovators proportion	Non innovators proportion	Z-score	Chi-square	T-test significance (p value)
non	not	7829	5896	0.024769	0.080567	−73.75***	272.24***	0.00
se	if	2083	1584	0.006590	0.021645	−37.99***	67.90***	0.00
ma	but	4220	2437	0.013351	0.033301	−37.51***	477.56***	0.00
quando	when	664	550	0.002101	0.007516	−23.67***	10.71**	0.00
come	as	5508	1986	0.017426	0.027138	−17.23***	1655.26***	0.00
perchè	because	1329	678	0.004205	0.009265	−17.22***	211.16***	0.740
ancora	still	1438	692	0.004549	0.009456	−16.21***	261.28***	0.00
poi	then	1053	532	0.003331	0.007270	−15.08***	171.26***	0.00
ora	now	656	381	0.002075	0.005206	−14.81***	72.93***	0.00
dopo	after	802	403	0.002537	0.005507	−13.03***	132.12***	0.00
fuori	out	212	158	0.000671	0.002159	−11.77***	7.88**	0.00
così	so	1139	367	0.003604	0.005015	−5.54***	395.74***	N/A
attraverso	across/ through	0	6	0.000000	0.000082	−5.09***	NA	0.00
dove	where	1028	328	0.003252	0.004482	−5.09***	361.36***	0.01
prima	before	1542	461	0.004879	0.006299	−4.84***	583.41***	0.801
sebbene	though	336	120	0.001063	0.001640	−4.11***	102.32***	0.481
a	at_to	34980	8399	0.110669	0.114770	−3.18**	16287.83***	0.00
sotto	under	301	97	0.000952	0.001325	−2.85**	104.56***	0.67
per	for_to	27444	6551	0.086826	0.089518	−2.32*	12840.64***	0.01
dietro	round	164	51	0.000519	0.000697	−1.85*	59.39***	0.633
vicino	near_by	89	29	0.000282	0.000396	−1.61	30.51***	0.834
opposto	opposite	91	23	0.000288	0.000314	−0.38	40.56***	0.680
via	off	1056	237	0.003341	0.003239	0.43	518.76***	N/A
fino	till	1002	214	0.003170	0.002924	1.07	510.64***	0.582
contro	against	298	55	0.000943	0.000752	1.55	167.28***	0.829
giù	down	162	22	0.000513	0.000301	2.38*	106.52***	0.499
con	with	15326	3317	0.048488	0.045326	3.61***	7735.67***	0.035
da	from	14784	3101	0.046773	0.042374	5.12***	7631.67***	0.451
or	or	174	7	0.000550	0.000096	5.14***	154.08***	0.074
mentre	while	749	93	0.002370	0.001271	5.77***	511.09***	0.712

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

the fact that innovators use new words that are associated with novel products and ideas, terms that are not yet mapped in common dictionaries. Note: the category "Segment" value of 1 indicates an entire file was processed.

The twelve LIWC Default Punctuation Results also provide insights into how the innovators differ from the non-innovators. First, four of the twelve LIWC punctuation categories are not applicable due to data processing procedures. Specifically, the comma and semicolon were used as a "csv" data format separator in exporting the data from the online forum; a period was added at the end of every post to accommodate the WORDij slide procedure, which made the All Punctuation category not applicable. Nevertheless, six punctuation marks stand out that significantly differentiate the two groups: non-innovators use significantly more often the colon, question mark, and exclamation point than innovators (highlighted in red), while innovators use an apostrophe, parentheses, and other punctuation more often than non-innovators (highlighted in green). While the Dash and the Quote mark had a large numeric difference in favor of the innovators' usage they were not found to be significant.

5.2 WORDij results

The WORDij results are presented in Table 6 and are sorted by Z-Score from low to high.

Table 6 presents three statistical tests: two from WORDij—the Z-score of two population proportions and the Chi-Square for goodness of fit based on counts, which are calculated at the file level. Appended for comparison are the LIWC T-test of means based upon an individual's posts. The Crovitz words that indicate one or more of the three significant test differences are highlighted in red where non-innovators indicate a higher use of certain words, and those words that are used more by innovators are highlighted in green. The rows highlighted in gray indicate no significant difference. Again, we calculated the t-statistics according to the results of preliminary Levene's tests, indicating whether equal variances could be assumed.

Overall, 32 of 37 (86%) of the Crovitz Relational Words have a significant Z-score and a Chi-Square score indicating there exists a clear unambiguous difference in the use of particular words between the 173 innovators and 3581 non-innovators. The only exception is for the word "across_through" where the count for the innovators indicated in Column D has a count of zero "0" and thus no Chi-Square can be calculated.

The 20 rows shaded in red indicate where the non-innovators use the Crovitz words significantly more often than innovators (in proportion), with eleven words having a negative Z-score of greater than 10. They are listed in order of magnitude from highest to lowest difference: "not, if, but, when, as, because, still, then, now, after, [and] out."

The five rows in italic indicate a mixed result. There are no significant Z-Scores or T-test Scores for these five words: "near_by, opposite, off, till, [and] against." However, the Chi- Square values are significant for the same words.

The 12 rows shaded in green indicate where the innovators use the Crovitz words significantly more often than non-innovators (in proportion), with four words having a positive Z-Score greater than 10. They are in order of magnitude: "of, in, and, [and] among_between."

To extend the analysis we also evaluated the significance of mean differences, by using the T-tests reported in the last column of Table 6.

Seventy percent of Crovitz words used by non-innovators more than the innovators have significant T-test values. They are: "not, if, but, when, as, still, then, now, after, out, across, where, at, [and] for." The remaining 30% of the Crovitz words that do not have significant T-test values are: "because, so, before, though, under, [and] round."

The T-test results are consistent with the Z-Score results when indicating that there is no difference between the two groups in four of the five Crovitz words, which have been highlighted in gray. They are: "near, opposite, till [and] against." The T-test could not be calculated for the word "off" because of the zero numerator as indicated by N/A in the table.

In seven of the twelve or 58% of Crovitz words used by innovators more often than non-innovators we find a correspondingly significant T-test value. They are: "of, in, and, among, about, on [and] with." The remaining five of twelve or 42% of Crovitz words in the T-tests are not in agreement with the Z-Scores and Chi-Squares. They are: "by, while, or, from, [and] down."

There is considerable overlap in the results of the T-tests and the Z-Scores and Chi-Square Scores. For Z-Scores and Chi-Squares we considered the text written by innovators and non-innovators as a whole, whereas T-tests were used to compare group means. Based on our results, we accept the hypothesis that the Crovitz 42 Relational Words discriminate employees classified as "innovators" and "non-innovators" based on their forum text postings.

6 Discussion

The results of our analyses demonstrate that there is indeed a difference between the text corpora associated with innovators and the ones linked to non-innovators in the company. Innovators use more "of, in, and" (*di, in, e*) while non-innovators use more "not, if, but" (*non, se, ma*). In line with the literature, the two groups of employees have different patterns in the written language they use (Pennebaker 2011). Innovators seem to use specific prepositions and conjunctions (Pennebaker et al. 2014), and this difference in the lexical profile is the result of both a specific way of thinking and of a specific relational context (Greco and Polli 2020a, b). The language production is the result of both individual characteristics and the context, which can influence the use of words and the communication style. In the context of an intranet forum, whose main goal is to facilitate knowledge sharing, a combination of formal and informal language is at the basis of communication.

It is important to notice that the linguistic style is not only related to individual characteristics, but also to the context, which can strongly influence the quality of speech (Boje et al. 2004). Furthermore, personal characteristics and context can interact with each other, thus influencing linguistic choices (Greco and Polli 2020a, b). In this study the context is represented by an Intranet forum, which offers an informal setting for a collaborative exchange of ideas, rather than a formalized *medium* such as company's memos, institutional emails or written reports. In order to shed light on this difference and to speculate about it, we used the Italian dictionary definitions of Crovitz's 42 Relational Words (Treccani 2014). According to the dictionary definitions, the words mostly used by innovators highlight a specific way to relate concepts while communicating and thinking (Polya 1957; Weick 1979). Innovators associate concepts defining their space characteristics and their belonging. In fact, according to the Italian dictionary, the use of the preposition *of* connects two concepts establishing the belonging, or the ownership, of one concept to the other. The preposition *in* is used when we have to establish the relationship between container and content, in which we define the place where a component is located or in which it tends to be, and the use of the conjunction *and* associates the components connecting them. The prepositions *of* and *in* help innovators describe concepts in relationship to each other, elaborating their thinking.

While innovators tend to create new connections and to specify the characteristics of the concepts involved in the communication, non-innovators seem to focus more on distinguishing and disambiguating the concepts. In fact, the most used Crovitz Relational Word is *not*, a negative particle that negates and excludes, which is the opposite of the conjunction *and* characterizing the language of innovators.

These results are aligned with empirical studies showing how innovators display behaviors of inclusivity, open communication and relationship-building (Amabile 1996; Rogers et al. 2019). Innovators have been depicted as boundary-spanners and information-brokers, as individuals who establish positive relationships with others (Fleming and Waguespack 2007) and create bridging ties within and outside the organizational boundaries (Gloor et al. 2020).

Innovators' ability to connect individuals to each other might be a reflection of an inner predisposition manifested through a use of inclusive language, selecting prepositions that connect concepts to each other (use of *and, in*), rather than disconnecting them. This preference for interconnectedness rather than exclusion can also be explained through the lens of Rogers' theory of diffusion of innovation (Rogers 2003). Innovators, and possibly early adopters, are people who are intrigued by the idea of trying something new, exploring connections between old and new products and ideas. In order to experiment and develop new ideas, innovators communicate using prepositions that *unite* concepts, as if every piece of knowledge they receive or think of could help them find new solutions to a problem. This choice for inclusivity is also a reflection of innovators' traits of openness, imagination, exploration, and reflection (Fürst and Grin 2018; Woo et al. 2017). Likewise, non-innovators mostly use the adversative and restrictive conjunction *but* that expresses an explicit opposition, exception or correction to the previous concept, mostly expressed negatively. The relationship between the two ideas is possible but it modifies and restricts the field of interaction. Its use entails strengthening the second concept at the expense of the first one. For example, the sentence "Andy is a good engineer, but he is slightly bureaucratic" focuses on the second concept: "Andy is slightly bureaucratic", and it doesn't have the same meaning as, "Andy is slightly bureaucratic but he is a good engineer", in which the focus is on the fact that Andy is a good engineer. In line with this hypothesis, non-innovators also use more the conjunction *if* that has a conditional, hypothetical value, in which the relationship between the two concepts is possible only under a specific condition. Following this hypothesis, non-innovators seem to focus more on differences (*not*) while innovators tend to associate (*and*). Even though both of them specify, innovators include concepts (*of, in*) while non-innovators are adverse to them or pose conditions (*but, if*) that tend to create doubt. The idea of inclusive language is consistent with the LIWC results as well, which indicate that innovators use more words in general, writing longer posts, and more complex words (six letter words or longer) to elaborate the concepts that they are discussing.

Our results also indicate that some punctuation marks can help differentiate between innovators and non-innovators. While non-innovators more often use colons, question marks and exclamation points, innovators employ an apostrophe, parentheses, dash, quote and other punctuation. On the other hand, non-innovators seem to be adding emotional content by expressing excitability and emphasizing a statement of fact. Written comments lack the ability to transfer emotions and emphasis, and non-innovators' use of exclamation points could help reduce ambiguity and serious misunderstanding in transmitting strong emotions (Choi et al. 2011).

7 Conclusions, Limitations and Future Research

The method we present in this study allows for an immediate identification of innovators and non-innovators through looking at their lexical choices. We believe that our approach to linguistic sleuthing for innovators using the Crovitz Innovator Identification Method shows promise for researchers and organizations. This method could improve self-awareness and self-reflection by suggesting which words convey inclusivity and interconnectedness, and which ones transfer negativity and exclusion. The specificity in the lexical profile is the result of a particular mental functioning (Greco and Polli 2020a, b; Laricchiuta et al. 2018) that is associated with personal characteristics and traits (Rogers 2003; Fürst and Grin 2018; Woo et al. 2017), as well as with personal initiative and social competence (Keller and Holland 1983; Keller 2017; Greco and Polli 2020a, b).

The specific context in which the innovators and non-innovators expressed their opinions and shared knowledge (an Intranet forum) is certainly going to influence which words, prepositions, and linguistic tone they used. It would be interesting in future studies to compare language used by innovators and non-innovators in multiple contexts and in other media, such as emails, and written reports. Moreover, as a third of the Crovitz's 42 Relational Words were of no use in distinguishing innovators from their colleagues, maybe not all of Crovitz's relational words are necessary to detect innovators and a reduced list could be defined in further studies.

Although there are other statistical procedures evaluating the characterization of the language (e.g., Misuraca et al. 2020), the Crovitz Innovator Identification Method focuses not only on the language itself but also on its contextual nature. Exploring differences in the use of words and punctuation marks could reveal more about emerging use in computer-mediated communication of language related to innovation. For example, future research could focus on the comparative analysis of text corpora in emails, instant messaging and forum posts, and understand whether innovators and non-innovators use prepositions, punctuation marks and exclamation points consistently in different written workplace communications. We might discover, for example, that question marks are less used in emails and more used via informal communication channels such as forums and instant messaging, which lead usually to a less formal use of language. Further research could be conducted to pinpoint differences between types of innovators and creative individuals in various disciplines. For example, Zijlmans et al. (2015) examined the titles of more than 900,000 papers in the medical journal and in other disciplines and found that clinicians used more question marks than non-medical researchers, which might suggest that clinicians have a question-driven approach to research while scientists engaged in basic research show a hypothesis-driven approach. In addition, the use of other statistical techniques, such as matched samples, could provide further validation of our results – while considering individual characteristics that we could not analyze in this study due to privacy reasons (such as age, gender, tenure, education, etc..).

Another opportunity for future research is the application of this method to advance the language analysis of personality traits. While self-report questionnaires have been the gold standard for measuring personality traits, methods like the Crovitz Innovator Identification Method represent viable alternatives that avoid biases and survey limitations. As suggested by Boyd and Pennebaker (2017) in their review of language-based personality studies, language contains a lot of information about important psychological constructs. In words there are deeply embedded attentional and social processes that are critical to our understanding of personality.

By focusing on relational words and punctuation marks, we were able to demonstrate how valuable the Crovitz Innovator Identification Method could be. This scalable and inexpensive method could be used as a complementary method to traditional survey-based methods to understand linguistic traits that differentiate creative individuals from others.

The main limitations of this study are associated with the use of a single case study to explore the potential use of the Crovitz Innovator Identification Method, particularly the issues with researcher subjectivity and external validity. We suggest replicating this study in multiple organizational contexts where individuals speak a language different from Italian to see whether our results are confirmed or whether national cultures play a role in the use of relational words. Could it be the case that more collectivistic societies use more relational words than individualistic societies? How do national differences influence the way innovators communicate? For example, collectivist societies tend to be more relationally oriented than individualist societies, which suggests they could be more innovative. Yet, it is individualist societies that are thought to produce more innovations and probably more innovators, while collectivist societies are good at copying the innovations produced by individualist societies. It would be valuable to determine if the findings about innovators at the organizational level would generalize to the societal level. We also suggest that other studies could investigate how the Crovitz 42 Relational Words vary over time in the diffusion of an innovation. For example, do some words enter the conversation early and remain, while others enter later or drop out of the discussion as the adoption of an innovation gains momentum?

As suggested by Pennebaker and Graybeal (2001), "linguistic sleuthing" through the examination of language use may help us uncover cognitive mechanisms and can tell us a lot about human nature, perhaps even more than some of the traditional psychological measures. Methods such as the Crovitz Innovator Identification Method have the potential to offer additional insights into social behaviors and cognitive models, despite their inability to account for syntax, context and linguistic idiosyncrasies.

References

Amabile, T.: A model of creativity and innovation in organizations. Res. Organ. Behav. **10**, 123–167 (1988)

Amabile, T.: Creativity and innovation in organizations. Harv. Busi. Sch. Backgr. **396–239**, 1–15 (1996)

Boje, D.M., Oswick, C., Ford, J.D.: Language and organization: the doing of discourse. Acad. Manag. Rev. **29**(4), 571–577 (2004)

Boyd, R.L., Pennebaker, J.W.: Language-based personality: a new approach to personality in a digital world. Current opinion in behavioral sciences **18**, 63–68 (2017)

Choi, J.N., Sung, S.Y., Lee, K., Cho, D.S.: Balancing cognition and emotion: Innovation implementation as a function of cognitive appraisal and emotional reactions toward innovation. J. Org. Behavior **32**(1), 107–124 (2011)

Crovitz, H.F.: Galton's walk: Methods for the analysis of thinking, intelligence, and creativity, vol. 100. Harper & Row, New York (1970)

Crovitz, H.F.: The form of logical solutions. Am. J. Psychol. **80**(3), 461 (1967)

Dyer, J.H., Gregersen, H.B., Christensen, C.M.: The innovator's DNA: mastering the five skills of disruptive innovators. Harv. Bus. Rev. **87**(12), 304 (2011)

Fleming, L., Waguespack, D.M.: Brokerage, boundary spanning, and leadership in open innovation communities. Organ. Sci. **18**(2), 165–180 (2007). https://doi.org/10.1287/orsc.1060.0242

Fronzetti Colladon, A., Saint-Charles, J., Mongeau, P.: From words to connections: word use similarity as an honest signal conducive to employees' digital communication. J. Inf. Sci. (2020). https://doi.org/10.1177/0165551520929931

Fürst, G., Grin, F.: A comprehensive method for the measurement of everyday creativity. Think. Skills Creat. (2018). https://doi.org/10.1016/j.tsc.2018.03.007

Gagné, M., Deci, E.L.: Self-determination theory and work motivation. J. Organ. Behav. **26**, 331–362 (2005). https://doi.org/10.1002/job.322

Gloor, P.A.: Swarm creativity. Competitive advantage through Collaborative Innovation Networks. Oxford University Press, NY (2006)

Gloor, P.A., Fronzetti Colladon, A., Grippa, F.: The digital footprint of innovators: using email to detect the most creative people in your organization. J. Bus. Res. **114**, 254–264 (2020). https://doi.org/10.1016/j.jbusres.2020.04.025

Greco, F., Polli, A.: Security perception and people well-being. Soc. Indic. Res. (2020a). https://doi.org/10.1007/s11205-020-02341-8

Greco, F., Polli, A.: Emotional text mining: customer profiling in brand management. Int. J. Inf. Manage. **51**, 101934 (2020b). https://doi.org/10.1016/j.ijinfomgt.2019.04.007

Iacoboni, M.: Mirroring people: the science of empathy and how we connect with others. Picador, New York, NY (2009)

Ireland, M.E., Slatcher, R.B., Eastwick, P.W., Scissors, L.E., Finkel, E.J., Pennebaker, J.W.: Language style matching predicts relationship initiation and stability. Psychol. Sci. (2011). https://doi.org/10.1177/0956797610392928

Kandemir, M.A., Kaufman, J.C.: The Kaufman domains of creativity scale: Turkish Validation and relationship to academic major. J. Creat. Behav. (2019). https://doi.org/10.1002/jocb.428

Keller, R.T., Holland, W.E.: Individual characteristics of innovativeness and communication in research and development organizations. J. Appl. Psychol. **63**(6), 759–762 (1978). https://doi.org/10.1037/0021-9010.63.6.759

Keller, R.T., Holland, W.E.: Communicators and innovators in research and development organizations. Acad. Manag. J. **26**(4), 742–749 (1983). https://doi.org/10.2307/255920

Keller, R.T.: A longitudinal study of the individual characteristics of effective R&D project team leaders. R and D Manag. (2017). https://doi.org/10.1111/radm.12272

Laricchiuta, D., Greco, F., Piras, F., Cordella, B., Cutuli, D., Picerni, E., Assogna, F., Lai, C., Spalletta, G. & Petrosini, L.: "The grief that doesn't speak": Text mining and brain structure. In: Iezzi, D.F., Celardo, L., Misuraca, M. (eds.) JADT' 18: Proceedings of the 14th International Conference on Statistical Analysis of Textual Data, pp. 419–427. Roma: Universitalia. (2018) ISBN 978-88-3293-137-2

Misuraca, M., Scepi, G., Spano, M.: A network-based concept extraction for managing customer requests in a social media care context. Int. J. Inf. Manage. **51**, 101956 (2020). https://doi.org/10.1016/j.ijinfomgt.2019.05.012

Ogden, C.K.: The system of basic english. Harcourt Brace, New York (1934)

Pennebaker, J.W.: The secret life of pronouns. New Sci. **211**(2828), 42–45 (2011). https://doi.org/10.1016/S0262-4079(11)62167-2

Pennebaker, J.W., Chung, C.K., Frazee, J., Lavergne, G.M., Beaver, D.I.: When small words foretell academic success: The case of college admissions essays. PLoS ONE (2014). https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J.W., Graybeal, A.: Patterns of natural language use: disclosure, personality, and social integration. Curr. Direct. Psychol. Sci. **10**(3), 90–93 (2001). https://doi.org/10.1111/1467-8721.00123

Polya, G.: How to solve it: a new aspect of mathematical method. Doubleday, Garden City (1957)

Ray, R.G., Ugbah, S.D., Brammer, C., DeWine, S.: Communication behaviors, innovation, and the maverick leader. J. Leadersh. Organ. Stud. **3**(3), 20–30 (1997). https://doi.org/10.1177/107179199700300304

Rogers, E.M.: Diffusion of innovations, 1st edn. Free Press of Glencoe, New York (1962)

Rogers, E.: The diffusion of innovations (fifth edition). The Free Press, New York (2003). https://doi.org/10.1007/s10460-007-9072-2

Rogers, E.M., Singhal, A., Quinlan, M.M.: Diffusion of innovations. In: Stacks, D.W., Salwen, M.B., Eichhorn, K.C. (eds.) An integrated approach to communication theory and research. Routledge, Abingdon (2019). https://doi.org/10.4324/9780203710753-35

Tishman, S., Perkins, D.: The language of thinking. Phi Delta Kappan **78**(5), 368 (1997)

Treccani, G.: In Treccani, L'Enciclopedia Italiana Online. The Online Italian Encyclopedia (2014). www.treccani.it

Weick K. E.: Social Psychology of Organizing. In: The Social Psychology of Organizing (1979). ISBN: 978-0075548089

Wenger, E.: Communities of practice: Learning as a social system. Systems thinker **9**(5), 2–3 (1998)

Woo, S.E., Keith, M.G., Su, R., Saef, R., Parrigon, S.: The curious dynamic between openness and interests in creativity. In: Feist, G.J., Reiter-Palmon, R., Kaufman, J.C. (eds.) The Cambridge Handbook of Creativity and Personality Research. Cambridge University Press, Cambridge (2017). https://doi.org/10.1017/9781316228036.004

Zijlmans, M., Otte, W.M., van't Klooster, M.A., van Diessen, E., Leijten, F.S., Sander, J.W.: Do clinicians use more question marks? JRSM Open (2015). https://doi.org/10.1177/2054270415579027

Affiliations

Francesca Greco[1] · **Ken Riopelle[2]** · **Francesca Grippa[3]** ·
Andrea Fronzetti Colladon[4] · **Julia Gluesing[2]**

Ken Riopelle
kenriopelle@me.com

Francesca Grippa
f.grippa@northeastern.edu

Andrea Fronzetti Colladon
andrea.fronzetticolladon@unipg.it

Julia Gluesing
j.gluesing@wayne.edu

[1] Department of Communication and Social Research, La Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

[2] Department of Industrial and Systems Engineering, College of Engineering, Wayne State University, Detroit, MI 48201, USA

[3] College of Professional Studies, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

[4] Department of Engineering, University of Perugia, Via G. Duranti 93, 06125 Perugia, Italy