

PLAGIO Nel testo con cui nel 2008, già deputata, ottenne il dottorato all'Imt di Lucca

La Madia e la tesi fabbricata col copia-incolla

Il lavoro della futura ministra Pd contiene ben 4 mila parole rubate da ricerche altrui (mai neppure citate)

▪ Interi blocchi riportati da altri studi o da documenti Ue e Fmi, come se fossero scritti da lei, nella ricerca su "Effetti della flessibilità nel

mercato del lavoro". La replica: "Non spetta a me giudicare la qualità del mio lavoro, è tutto in bibliografia"

◦ MARGOTTINI A PAG. 2 - 3

L'INCHIESTA Interi blocchi con il "copia e incolla"

Madia, dottorato con 4 mila parole copiate nella tesi

Nella ricerca finale del 2008 all'Imt la futura ministra prese ampi brani da testi altrui senza indicare la citazione

Le risorse utilizzate Tra i tanti paper saccheggiate anche documenti del Fmi e della Commissione

Karl zu Guttenberg La Merkel ha fatto dimettere il titolare della Difesa per un caso simile

» **LAURA MARGOTTINI**

La tesi di dottorato di Marianna Madia, ministro per la Semplificazione e Pubblica amministrazione nei governi Gentiloni e Renzi, non pare essere tutta frutto della sua creatività. In 35 di 94 pagine della tesi (al netto di bibliografia, figure e tabelle) – titolo: "Essays on the Effects of Flexibility on Labour Market Outcome" – ci sono passaggi pressoché iden-

tici a quelli presenti in altre pubblicazioni. La fonte di quei passaggi non risulta citata là dove il ministro li riporta nella sua tesi. Col risultato che spesso non è possibile distinguere le parole originali della Madia da quelle di altri autori. Da un'indagine del *Fatto*, risulta essere circa 4 mila le parole senza chiara attribuzione nei tre capitoli della tesi.

A FINE 2008, LA MADIA (già parlamentare Pd) ha conse-

guito il titolo di dottorato alla Scuola Imt di Alti Studi di Lucca. Fabio Pammolli, allora rettore dello stesso Imt, e Giorgio



Rodano, già professore di ordinario di Economia all'Università Sapienza, erano i relatori della tesi. Che dovrebbero essere garanti della sua originalità e della conformità alle regole che l'accademia si dà per preservare l'integrità della ricerca.

Nell'analisi, il *Fatto* ha escluso dal conteggio tutte le frasi che hanno riconosciuto di uso comune nell'ambito delle scienze economiche e anche i passaggi che appaiono identici in altre pubblicazioni, ma attribuite tra parentesi dalla Madia alla fonte originale, nel punto in cui sono riportate. Tali passaggi sono stati esclusi dal conteggio anche quando ripresi parola per parola, ma senza virgolette (le regole accademiche impongono di virgolettare se le frasi sono riprese letteralmente). La tesi della Madia è sul sito dell'Imt.

Nel lavoro del ministro passaggi anche di centinaia di parole risultano identici ad altri già apparsi in pubblicazioni scientifiche *peer reviewed* (cioè certificate dal controllo della comunità scientifica), o in articoli che nel 2008 erano ancora *in progress*, in rapporti della Commissione europea, del Fondo monetario internazionale e di centri di ricerca (come l'Istituto Iza per l'Economia del Lavoro di Bonn, in Germania, o il National Bureau of Economic Research di Cambridge (Nber), Massachusetts, negli Usa). "Anche articoli *in progress*, *working paper*, o i rapporti di istituzioni vanno assolutamente citati," spiega Gerhard Dannemann, direttore del Centro di Studi britannici a Berlino, membro del VroniPlag, il gruppo di accademici che ha analizzato le tesi di dottorato di decine di politici e professori tedeschi.

LA PRATICA DI RIPRENDERE interi passaggi senza citare la fonte all'interno del proprio testo è giudicata molto severamente nel mondo accademico. Anche il codice etico che Imt si è dato, con Pammolli rettore, definisce come plagio accademico "la

presentazione delle parole o idee di altri come proprie". E questo "può assumere varie forme" come "appropriarsi deliberatamente del lavoro di altri o non citare correttamente le fonti all'interno del proprio lavoro accademico." Dal 2011 Imt ha messo a disposizione dei docenti un software anti-plagio, in grado di smascherare le parti copiate nelle tesi degli studenti.

In tre sottocapitoli della tesi del ministro, la quantità di passaggi che risultano originariamente presenti in articoli di altri autori non citati dove appaiono nella tesi, è rispettivamente del 40%, del 56% e del 79%. E in sette pagine su 95 si va dal 56% all'89% di testo identico a quello di altri autori, senza virgolette né attribuzione della fonte. In alcuni casi, si cita in parentesi il lavoro di un autore, ma si riprendono intere parti da un altro lavoro del medesimo autore, che però non è citato dove i passaggi sono riportati. Alcune pagine appaiono come collage di più articoli di diversi autori, senza fonte né virgolette, inframmezzate da frasi scritte dall'autrice della tesi.

Oppure, in una serie di frasi riprese *verbatim* (senza fonti né virgolette) vengono cambiate solo alcune parole: "question" nella fonte originale diventa "issue" nella tesi, "step" diventa "stage", "those" diventa "these". "In generale, segnali di questo tipo possono indicare l'intenzionalità da parte dell'autore di non citare correttamente", spiega Dannemann al *Fatto*. E una tecnica nota come "shake and paste", mescola e incolla.

Un altro *modus operandi* riscontrato è quello chiamato *pawn sacrifice*, "l'arte del concedere poco, per nascondere molto". Si cita la fonte all'inizio di un passaggio, ma nel testo che segue — anche per centinaia di parole nel caso della tesi della Madia — non si specifica che si tratta di un testo tratto dalla medesima pubblicazione, né si utilizzano virgolette. E questo rende impossi-

bile riconoscere le frasi scritte dal ministro da quelle di altri autori. Le pubblicazioni da cui sono ripresi i passaggi senza attribuzione tra parentesi e senza virgolette sono elencati nella bibliografia della tesi. Ma non sono citate nel punto esatto in cui vi si attinge.

LA MANCANZA di correttezza nel citare le fonti può spingere le università a revocare il titolo di dottorato o le riviste scientifiche a revocare la pubblicazione di un articolo qualora si riscontri che la mancata attribuzione sia deliberata. In Italia, la legge 475 punisce con pene fino a 3 anni di reclusione chi "in esami o concorsi, prescritti o richiesti da autorità o pubbliche amministrazioni per il conferimento di lauree o di ogni altro grado (...) presenta, come propri, dissertazioni, studi, pubblicazioni, progetti tecnici e, in genere, lavori che siano opera di altri".

Alcune delle criticità riscontrate dal *Fatto* nel caso Madia sono analoghe a quelle trovate nella tesi di dottorato di Karl-Theodor zu Guttenberg, ex ministro della Difesa tedesco che si è dimesso e ha rinunciato al dottorato dopo le accuse di plagio nel 2011. A questo primo scandalo, rivelato dalla *Süddeutsche Zeitung*, ne seguirono altri, riguardanti ministri del governo Merkel e una decina di politici. Fu aperta un'inchiesta dall'Università di Bayreuth, dove Zu Guttenberg aveva conseguito il dottorato, e dalla magistratura, che finì con un patteggiamento e un'ammenda.

Tre settimane fa in Francia *L'Express* ha scovato una decina di passaggi copiati da autori celebri, ma non citati neanche in bibliografia, nell'ultimo libro di Etienne Klein, fisico, filosofo e divulgatore francese. Due mesi prima Klein era diventato presidente dell'Istituto di Studi Scientifici e Tecnologici francese che vuole rafforzare la fiducia dei cittadini nell'impresa scientifica. Sul caso Klein, il ministero della Ricerca francese ha aperto un'inchiesta.

IL METODO

I software anti-furbetti



A caccia di "impronte"
I due software usati dal "Fatto" per l'analisi

LA METODOLOGIA usata dal Fatto per analizzare la tesi del ministro Madia si basa, in prima battuta, su due software "anti plagio", tecnologie in uso nelle università europee e americane dal 2006. Utilizzano algoritmi per creare una impronta digitale di ogni porzione del testo in esame e poi la comparano con le impronte dei passaggi di migliaia di documenti online. Il Fatto ne ha utilizzati due, *PlagScan* e *iThenticate*, incrociandone successivamente i dati per maggiore accuratezza. *iThenticate* è, insieme a *Turnitin*, il più usato nelle università nel mondo, che lo mettono a disposizione non solo dei

professori, ma anche degli studenti, per evitare guai. I due software hanno prodotto ognuno un proprio rapporto in merito alle potenziali similitudini riscontrate con altri testi, evidenziando tutte le possibili frasi identiche o simili in migliaia di documenti. A ogni fonte i software assegnano un colore diverso in base alla percentuale di similitudine che rilevano. A quel punto, per ogni singolo passaggio abbiamo controllato a mano tutte le pubblicazioni segnalate dai software per ogni passaggio. Da esse è stato scartato ciò che è risultato come mero "rumore di fondo", cioè frasi di uso comune o incoerenti rispetto al significato del testo della tesi. La scrematura ha considerato solo le pubblicazioni in cui risultavano identiche o quasi ampie porzioni di testo. Andando a ritroso grazie alle bibliografie delle pubblicazioni e della stessa tesi, si è potuto evidenziare da quale articolo sono state riprese le frasi contestate.

Alcune delle parti plagiate

ities between reforms in product and labour markets. Amable and Gatti (2006) show that engaging in a process of product market deregulation yields an implicit labour market reform, leading to a more intense turnover on the labour market. This mechanism is exacerbated by increased competition on the product market but is dampened by redundancy payments. Similarly, in Amable and Gatti (2004), they develop a dynamic efficiency wage framework where deregulation in product and labour markets boosts labour turnover and reduces job security; this mechanism pushes the incentive compatible real wage schedule upwards and may generate aggregate employment losses.



Cap. 1 pag. 15

incentives to protect jobs are reduced. Amable and Gatti (2006) show that engaging in a process of product market deregulation yields an implicit labour market reform leading to a more intense turnover on the labour market. This mechanism is exacerbated by increased competition on the product market but is dampened by redundancy payments.

In some cases yield perverse effects on employment. Amable and Gatti (2004) develop a dynamic efficiency wage framework where deregulation in product and labour markets boosts labour turnover and reduces job security; this mechanism pushes the incentive compatible real wage schedule upwards and may generate aggregate employment losses.

Amable, Demmou, Gatti (2007)
"Employment, Performance and Institutions:
New Answers to an Old Question", IZA Discussion Paper No.2731

Another important concern is the fact that the inclusion of country fixed effects precludes the inclusion of time-invariant or slowly-changing variables as independent variables. Distinguishing between their influence and the influence of omitted country-specific variables might be difficult. If fixed effects are not included in the model, the time-invariant variables will carry the weight of all country specific factors. To overcome this problem, Plümper and Troeger (2007) propose a procedure for analyzing the effect of time-invariant variables in a model including fixed effects. Their procedure has three stages: (i) estimate a fixed-effect model; (ii) regress the unit effects on the time-invariant variables; (iii) re-estimate the first stage including the error term of the second stage (XTFEVD procedure). Their Monte Carlo experiments suggest that the fixed effect vector decomposition (XTFEVD)



Cap. 1 pag. 21

The second problem concerns the fact that the inclusion of country fixed effects precludes the inclusion of time-invariant or slowly-changing variables as independent variables. Several of the variables we consider in our estimations are either invariant (at least for a non negligible part of the period considered) or change slowly. Distinguishing between their influence on unemployment, inactivity or joblessness and the influence of omitted country-specific variables will thus prove difficult. If one does not include fixed effects in the model, the time-invariant variables will carry the weight of all the country specific factors determining employment and unemployment. Plümper and Troeger (2004) propose a procedure for analyzing the effect of time-invariant variables in a model including fixed effects. Their procedure takes three stages: (i) estimate a fixed-effect model (ii) regress the unit effects on the time-invariant variables (iii) re-estimate the first stage including the error term of the second stage (XTFEVD procedure).

Sul sito
Sul fattoquotidiano.it tutte le 50 slide

Amable, Demmou, Gatti (2007)

ness (Madsen et al. (2006)). We will measure firms' innovative propensity by means of new processes and new products introduced into the market whereas firm innovation intensity can be measured by the share of innovative sales in total sales. **Changes in other types of activity like... (Explanatory Innovation Survey - CIS). It is not possible to distinguish between innovative sales corresponding to products new for the firm (new products) to the market, which can be considered as innovations of products already produced by other competitors, and those corresponding to products only new for the market, which can be regarded as true innovations.**



Cap. 2 pagg. 11-12

Desiring by g_{it} the binary variable indicating if firm i is an innovative firm - that is, a dummy variable indicating whether the firm has introduced at least one product or process time-series - we can write:

$$g_{it} = \begin{cases} 1 & \text{if } g_{it}^* > 0 \\ 0 & \text{if } g_{it}^* \leq 0 \end{cases} \quad (2.3)$$

where $g_{it}^* = \beta_0 + \beta_1 x_{it} + \beta_2 z_{it}$ if $\beta_1 > 0$ and $\beta_2 > 0$ represents the innovation to innovate. z_{it} is a vector of explanatory variables. In a vector of parameter to be estimated, and β_0 is a random error term, which includes the effect of left-out omitted variables. As explanatory variables x_{it} , in addition to the amount

Mohmen, Mairesse e Dagenais (2006)
"Innovativity: A comparison across seven European countries"
Economics of Innovation and New Technology

The second equation of the Tobit (type 2) model is specified in terms of a second latent variable y_{it}^* , which is equal to the actual share of innovative sales y_{it} , if the firm is innovative (i.e. $y_{it} > 0$). Since the share of innovative sales is bounded by 1 and 1, it is preferable to perform a logit transformation of the data and express the second equation in terms of the latent logit share variable $y_{it}^* = \ln(y_{it}/(1 - y_{it}))$ which varies from $-\infty$ to $+\infty$. Thus we can write our second equation as

$$y_{it}^* = \begin{cases} -\infty & \text{if } y_{it}^* > 0 \\ \text{undefined} & \text{if } y_{it}^* \leq 0 \end{cases} \quad (2.2)$$

or equivalently

$$y_{it} = \begin{cases} e^{y_{it}^*}/(1 + e^{y_{it}^*}) & \text{if } y_{it}^* > 0 \\ 0 & \text{if } y_{it}^* \leq 0 \end{cases}$$

where $y_{it}^* = \beta_0 + \beta_1 x_{it} + \beta_2 z_{it}$. β_0 is a vector of explanatory variables, β_1 is a vector of parameters to be estimated and $\beta_2 > 0$ is an error term called omitted variables. Since we have data on sales growth for



Cap. 2 pag. 13

Mohmen, Mairesse e Dagenais (2006)
"Innovativity: A comparison across seven European countries"
Economics of Innovation and New Technology

from the overall equation. The second equation of our generalized Tobit model is then specified in terms of a second latent variable y_{it}^* , which is equal to the actual share of innovative sales y_{it} , if the firm is innovative (i.e. $y_{it} > 0$). Since the share of innovative sales is bounded by 0 and 1 (i.e. $0 \leq y_{it} \leq 1$), it is actually preferable to specify the second equation in terms of the latent logit share variable $y_{it}^* = \ln(y_{it}/(1 - y_{it}))$ which can vary from $-\infty$ to $+\infty$. Because the resulting variable is close to normally distributed, the

The second equation of the generalized Tobit model is then the following:

$$y_{it}^* = \beta_0 + \beta_1 x_{it} + \beta_2 z_{it}$$

where β_0 is a vector of explanatory variables, β_1 is a vector of parameters to be estimated and $\beta_2 > 0$ is a stochastic error term reflecting omitted variables and other sources of heterogeneity. Now that because of the logit share transformation of equation (2), y_{it} values from $-\infty$ to $+\infty$. We can therefore write:

$$y_{it} = \begin{cases} e^{y_{it}^*}/(1 + e^{y_{it}^*}) & \text{if } y_{it}^* > 0 \\ 0 & \text{if } y_{it}^* \leq 0 \end{cases} \quad (2.2)$$

or equivalently

$$y_{it} = \begin{cases} e^{y_{it}^*}/(1 + e^{y_{it}^*}) & \text{if } y_{it}^* > 0 \\ 0 & \text{if } y_{it}^* \leq 0 \end{cases} \quad (2.2)$$