



# UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

Relazione

**COMode: creazione e analisi del Corpus di  
*L'Officiel de la Mode***

**Candidato:** Isabella Prisco

**Relatore:** Alessandro Lenci

**Correlatore:** Felice Dell'Orletta

Anno Accademico 2015-2016

# Indice

## Introduzione

### 1. Moda e linguaggio

#### 1.1 “ *Il sistema della Moda* ” di Roland Barthes

- 1.1.1 Priorità del linguaggio e della scrittura
- 1.1.2 Langue e parole
- 1.1.3 La prova di commutazione
- 1.1.4 Struttura del significante
- 1.1.5 Struttura del significato
- 1.1.6 Struttura del segno
- 1.1.7 Accenno al sistema retorico
- 1.1.8 Il sistema duplice della Moda

### 2. *COMode*: il Corpus di *L’Officiel de la Mode*

- 2.1. Breve storia del giornale
- 2.2 Come si costruisce un corpus linguistico
  - 2.2.1 Perché l’uso del file di testo come documento di partenza
- 2.3 *COMode* : costruzione e caratteristiche del Corpus

### 3. Talismane

- 3.1 Che cosa è Talismane
- 3.2 Analisi di *COMode* attraverso Talismane
  - 3.2.1 Titoli e costruzione nominale

### 4. Analisi della terminologia

- 4.1 EXTra: estrazione e trattazione della terminologia
  - 4.1.1 Analisi di *COMode* attraverso EXTra: i nomi comuni
- 4.2 Uno sguardo agli stilisti protagonisti in *L’Officiel de la Mode*
- 4.3 La contaminazione della lingua in Francia
  - 4.3.1 Il *franglais* in *COMode*

## Conclusioni

## Bibliografia

## Appendici

# Introduzione

La Moda è un concetto multiforme e in continuo cambiamento che trae la propria essenza dall'intersezione tra cose e parole. La sua natura ontologica e quella linguistica coesistono, poiché essa, per affermarsi, non può prescindere né dalla realtà né dal discorso. Per capire come si racconta la Moda e come essa viene comunicata al destinatario, è necessario analizzare il canale principale tramite cui si diffonde: il giornale. Infatti, se si escludono immagini, fotografie, video, e quindi pubblicità e sfilate, il mezzo di comunicazione ufficiale è quello scritto: lontano da ogni libera interpretazione o lettura soggettiva, l'inchiostro su carta permette di decifrare l'indumento nella sua totalità, tramite la descrizione.

L'oggetto della tesi è dunque il linguaggio giornalistico settoriale di Moda: è stato per questo costruito *COMode*, un corpus in lingua francese composto da articoli selezionati dall'archivio di *L'Officiel de la Mode, 1995-2013*.

L'analisi, attraverso gli strumenti della linguistica computazionale, mira all'identificazione di alcune costruzioni significative e alcuni tratti peculiari della terminologia.

Se infatti il modo di esprimere la Moda rappresenta un linguaggio settoriale, quali sono le sue principali caratteristiche? Come la creazione di un Corpus, mirato alla sua rappresentatività, può essere il punto di partenza dell'elaborazione dei dati? In che modo gli strumenti del *Natural Language Processing* (NLP) possono essere adoperati per l'esplorazione di questo linguaggio tecnico?

Per arrivare a risposte, più o meno significative, la tesi si articola in 4 parti principali: nella prima, si fa riferimento allo studio sistematico e approfondito che il linguista francese Roland Barthes condusse riguardo al linguaggio di Moda nel suo "*Il Système de la Mode*", individuando i diversi stadi tramite cui significante e significato conducono al segno di Moda.

Nella seconda parte, vengono descritti i parametri e le caratteristiche da definire durante la costruzione di un corpus; si fa successivamente riferimento alla genesi del corpus di *COMode*.

Nella terza sezione, viene illustrato il funzionamento dell'analizzatore sintattico Talismane e la sua applicazione al Corpus di *L'Officiel de la Mode*; in seguito al processo di tokenizzazione e di individuazione delle dipendenze, emerge un tratto peculiare della scrittura giornalistica: la costruzione nominale dei titoli.

Nella quarta e ultima parte, i soggetti della trattazione sono: l'elaborazione terminologica dei nomi comuni avvenuta tramite l'analizzatore EXTra, un breve approfondimento sulle occorrenze dei nomi propri e uno studio sulla contaminazione linguistica del francese da parte soprattutto dell'inglese.

Seguono le conclusioni, la bibliografia, la sitografia e le appendici.

# 1. Moda e linguaggio

Attribuire un inizio cronologico alla Moda risulta difficile e quasi impossibile in quanto si potrebbe affermare che essa sia nata agli albori della specie umana, quando vestirsi di pellami e tessuti era una necessità legata alla sopravvivenza. Già allora, sicuramente, vanità e stimoli estetici inducevano a migliorare sempre più la propria esteriorità, tuttavia è solo in un secondo momento che la Moda ha assunto in maniera sempre più crescente valenza sociale e componente estetica, diventando più che un bisogno primordiale, un fatto culturale. Proprio come la cultura, la Moda non è un patrimonio fisso e immutabile ma si trasforma col tempo, rispecchiando quelle che sono le attitudini della società e la sensibilità di ciascun individuo. Proprio per il suo carattere variabile e interdisciplinare, ha bisogno di uno strumento altrettanto mutevole e multiforme, in grado di assecondare di pari passo il ritmo del suo cambiamento e della sua evoluzione: il linguaggio.

Quando si fa riferimento a tale associazione, risulta naturale il richiamo a colui che viene considerato il precursore degli studi sulla moda in ambito semiotico, Roland Barthes.

Per la prima volta in maniera sistematica, il linguista e semiologo francese affronta il discorso intorno a Moda-Linguaggio nel suo libro intitolato “*Il Système de la Mode*”, pubblicato nel 1967. Egli afferma infatti che la Moda esiste soltanto in virtù del discorso che si fa su di essa.

## 1.1 “*Il sistema della Moda*” di Roland Barthes

Nella stesura del libro “*Il sistema della Moda*”, Roland Barthes ha come obiettivo quello di condurre un’analisi strutturale dell’indumento femminile descritto dalle riviste di Moda. In particolare l’oggetto del suo studio è un corpus sincronico, giugno 1958 - giugno 1959, composto da articoli presi dai periodici “*Elle*”, “*Jardin de mots*”, da alcune pubblicazioni di “*Vogue*”, “*Echo de la Mode*” e da altre pagine settimanali di quotidiani dedicati alla Moda, escludendo il settore bellezza, le pubblicità e le indicazioni di fabbricazione dei vestiti. La selezione del materiale ha come scopo la differenziazione dei segni del vestire, non la loro ripetizione.

### 1.1.1 Priorità del linguaggio e della scrittura: l’indumento-scritto

La sua nuova esplorazione, iniziata nel 1957 e terminata nel 1963, è stata condotta seguendo il metodo utilizzato dalla scienza dei segni, quella che Ferdinand De Saussure ha chiamato semiologia.

Il lavoro di Barthes è stato una vera e propria prova di apprendistato; davanti al primo bivio affrontato nella scelta del materiale da indagare, ha dovuto prediligere la Moda scritta, quella dell’indumento-scritto (giornali o riviste), alla Moda reale, quella dell’indumento-immagine (indossato o fotografato), concentrandosi appunto sull’aspetto descrittivo, di struttura verbale, piuttosto che quello visivo, di struttura plastica. (v. fig.1).

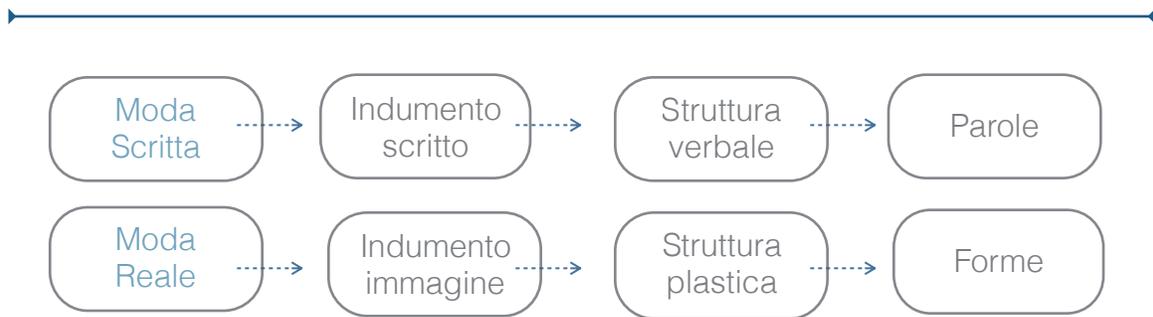


Figura 1. Distinzione Moda scritta / Moda reale

Data la distinzione, si potrebbe pensare che indumento-scritto e indumento-immagine, trovino corrispondenza in un terzo termine, l'indumento-reale: ciò non accade perché vi è tra tutti differenza fondamentale di materiali e di rapporti. Per analizzare l'indumento reale in termini sistematici bisogna risalire infatti alla fabbricazione e quindi alla sua struttura tecnologica. Le sue unità non possono che essere le tracce della produzione, una cucitura o un taglio. Per uno stesso oggetto corrispondono quindi tre strutture diverse e autonome, ognuna dotata di un codice proprio.

Barthes sceglie dunque l'abito scritto perché vede nel linguaggio, rispetto all'immagine, funzionalità migliori. La parola è infatti in grado di:

- fissare i livelli di percezione a un certo livello di intelligibilità, scegliendo dove focalizzare l'attenzione al posto dell'occhio;
- fornire informazioni che la fotografia dà imperfettamente o non dà affatto (funzione di conoscenza), cogliendo dettagli e impieghi che solo il carattere astratto del linguaggio riesce a fare;
- evidenziare alcuni aspetti nascosti che l'immagine non sottolinea (funzione di enfasi) e svelare l'indumento secondo un certo ordine di importanza o rilievo.

### 1.1.2 Langue e parole, indumento e toilette.

Il punto di partenza negli studi del linguista Ferdinand de Saussure, è la distinzione tra *langue* e *parole*. Il primo termine della dicotomia rappresenta l'aspetto sociale del linguaggio, in quanto si tratta di un sistema comune a tutti, composto da significati e significanti condivisi. Il secondo costituisce invece l'aspetto individuale del linguaggio, fa riferimento alla singola esecuzione ed è infatti unico e irripetibile. Se da una parte la *langue* è una istituzione, la *parole* non è altro che la parte momentanea di questa istituzione, che il parlante attualizza per comunicare.

Barthes applica lo stesso ragionamento al concetto di vestito, per meglio spiegare il rapporto tra indumento-scritto e indumento-immagine. L'oggetto della Moda scritta corrisponde alla *langue* per la sua purezza strutturale, in quanto mantiene la sua integrità

al di là delle manifestazioni singole. Invece la *parole* coincide con la Moda reale per la sua individualità, dato che la sua attuazione è esclusiva e isolata. Barthes chiama anche “indumento” la forma istituzionale e formale dell’abito, e “toilette” la forma soggettiva e circoscritta ad un momento.

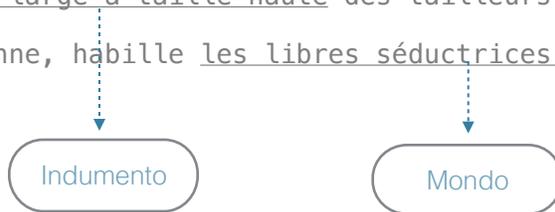
Tuttavia Barthes afferma che «(...) questa opposizione non è valida che a al livello del sistema vestimentario; poiché al livello del sistema linguistico è evidente che la descrizione è a sua volta sorretta da una ‘*parole*’ particolare (quella di *questo* giornale di Moda, in *questa* pagina); è, se si vuole, un indumento astratto affidato a una parola concreta; (...) (Barthes 1967, pp.21)». La natura dell’indumento scritto è duale poiché al livello dell’indumento è istituzione o *langue* ma al livello del linguaggio è anche atto o *parole*.

### 1.1.3 La prova di commutazione

Barthes, nel suo studio sistematico, espone come modello operativo, la prova di commutazione: per cogliere le più piccole unità strutturali del sistema di significazione dell’indumento-scritto; egli fa variare artificialmente un termine e osserva se questa variazione introduce un mutamento nella lettura o nell’uso della struttura data. Osserva che, anche se il giornale non parla mai in modo esplicito e diretto di “indumento” e di “mondo”, sono queste le prime due classi commutative individuate nel testo di moda. La classe “indumento”, comprende i tratti vestimentari (forme, colori, materie) mentre la classe “mondo”, è costituita dai tratti circostanziali e caratteriali (situazioni, occupazioni, stati, umori). Nel giornale di Moda, indumento e mondo entrano in relazione secondo rapporti diversi, (finalità, circostanza, transitività, casualità), rimanendo due entità separate: esse non sono identiche ma equivalenti.

Esempio commutazione indumento/mondo:

Le pantalon large à taille haute des tailleurs Saint Laurent, inspirés de la garçonne, habille les libres séductrices d'aujourd'hui.



Non sempre però il tratto vestimentario è correlato a situazioni relative al mondo e si potrebbe pensare che la prova di commutazione non sia verificata.

Tuttavia entra in gioco un’altra classe commutativa che nella maggior parte dei casi è sottintesa: il concetto “alla Moda” (e “fuori Moda”). Tale espressione è implicita e non viene enunciata ma rappresenta il secondo termine della prova di commutazione. Anche in questo caso il rapporto è di equivalenza e non di identità.

Esempio commutazione indumento/moda:

Une capeline encre de chine, un tailleur noir cafard, un perfecto corbeau. ( \_\_\_\_\_ ).



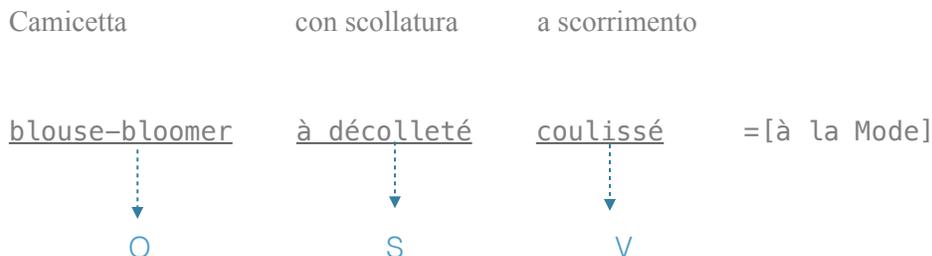
La classe “indumento” è costituita da elementi materiali e di numero finito mentre le classi “mondo” e “moda” sono composte da unità astratte e illimitate. Il visibile e l’invisibile trovano sintesi nell’enunciato che serve a noi lettori per cogliere il rapporto di equivalenza tra il “mondo” o la “moda” e l’ “indumento”.

Secondo Barthes quindi «(...) ogni enunciato del giornale costituisce un sistema di significazioni, composto di un significante, dai termini discreti, materiali, numerabili e visibili: l’indumento; e di un significato, immateriale, che è secondo i casi il mondo (insieme A) o la moda (insieme B); conformemente alla nomenclatura saussuriana, chiameremo segno<sup>1</sup> la correlazione dei due termini, del significante vestimentario e del significato mondano o di moda. (Barthes 1967, pp.28)».

#### 1.1.4 Struttura del significante

Barthes, confrontando fra loro degli enunciati del significante vestimentario, si accorge che alcuni elementi risultano in comune; questo vuol dire che tali elementi sono mobili e che partecipano a sensi diversi a seconda della combinazione fra l’oggetto e la sua qualificazione. L’enunciato del significante può così essere scomposto in unità più piccole: un oggetto (O), su cui ricade la significazione, un supporto (S), che serve a sostegno della significazione e un terzo elemento, detto variante (V)<sup>2</sup>.

Esempio matrice OSV:



<sup>1</sup> Segno, nel senso saussuriano, è l’unione del significante e del significato, e non solo del significante come viene creduto comunemente.

<sup>2</sup> Si farà riferimento all’unione dei tre elementi, detta matrice OSV.

Tuttavia questo rappresenta l'esempio più semplice, poiché oggetto, supporto e variante si trovano uno accanto all'altro e dunque non c'è ambiguità di senso. Non sempre però la collocazione delle parole è così ordinata nello spazio e, soprattutto nel caso del giornale di Moda, succede che oggetto e supporto non siano adiacenti e che il senso venga così irradiato a distanza, come capita ad esempio per le matrici concatenate fra loro.

Il supporto che, come l'oggetto, è un corpo materiale (indumento, inserto o accessorio), è il primo termine a ricevere l'alternativa del senso da trasmettere all'oggetto in questione; la variante o vestema è invece immateriale, serve a modificare le caratteristiche del supporto ed è il punto della matrice attraverso cui la significazione si estende su tutto l'enunciato.

Tra gli elementi della matrice esiste un rapporto di solidarietà<sup>3</sup>, o di doppia implicazione: l'oggetto obbliga il supporto e il supporto obbliga la variante. Nessuno dei tre può essere trovato isolato perché l'enunciato perderebbe di significato.

### 1.1.5 Struttura del significato

Come risultato della prova di commutazione<sup>4</sup>, Barthes nell'analisi del significato rimanda a quei due tipi di enunciato il cui significante faceva riferimento o al mondo (insieme A) o al significato implicito "alla moda" (insieme B).

Secondo il semiologo francese il secondo caso sfugge alla regola di strutturazione che prevede la scomposizione in unità minime, trattandosi sempre dello stesso significato (si parla di isologia).

Per strutturare i significati dell'insieme A, unico indagabile, Barthes li ritaglia in unità irriducibili, chiamate unità semantiche. Queste sono suddivisibili a loro volta in unità usuali e unità originali: le prime sono quelle che si prestano alla ripetizione nei vari enunciati e che quindi risultano mobili, le altre sono uniche e isolate (hapax).

Esempio unità semantiche usuali/originali:

pour l'été	----->	Unità usuale
cocktails à Capri	----->	Unità originale

Le unità usuali sono di natura lessicale e servono a determinare le funzioni reali dell'indumento portato, svelano dunque le finalità pratiche dell'abito più comuni; le unità originali sono invece di natura fraseologica, proprio perché suggeriscono contingenze complesse, quasi utopiche, dell'abito.

Barthes, concentrandosi sulle unità usuali, osserva come queste siano spesso concatenate e distingue tre tipi di combinazione: AUT, ET e VEL.

---

<sup>3</sup> Barthes fa riferimento alla metafora della porta chiusa a chiave: v. Barthes 1967, pp.20

<sup>4</sup> Cfr 2.1.3

La relazione AUT è disgiuntiva esclusiva, pone intorno a due termini, un'alternativa di due significati, una doppia variazione concomitante. Es: On adoptera la robe le matin ou le soir et sous tous les aspects: robe tunique ou robe-manteau.

La relazione ET è cumulativa, stabilisce un rapporto di complementarità fra un certo numero di significati, amalgamando sia quelli affini che quelli apparentemente contraddittori. Attraverso la combinazione, ET consente di trarre da una riserva generale di funzioni usuali, una contingenza particolare, sfruttando al massimo le potenzialità del linguaggio. Es: Pratique et chic, cette inspiration militaire.

La relazione VEL è sia disgiuntiva, perché le unità che lega non possono essere attualizzate nello stesso tempo, e sia inclusiva, perché esse appartengono ad una classe unica. Es: Drapé, ouvert, sexy et plein de fantaisie, le maillot de bain se porte le jour ou le soir.

La relazione AUT può essere neutralizzata se i due termini alternativi diventano retti da un unico significante (*dominance*); in questo modo le unità opposte o si fondono, relazione ET, o si livellano, relazione VEL. Il meccanismo della neutralizzazione porta progressivamente le unità semantiche della Moda ad annullare i tratti distintivi e a perdere il senso particolare in favore di uno sempre più generale. La piramide<sup>5</sup> è percorribile anche a ritroso, la Moda consente di convertire ET e VEL in AUT, tornando alle opposizioni pertinenti, sdoppiando il significante. Secondo Barthes questo percorso a due sensi rende illusorio ogni lessico della Moda proprio perché il passaggio dal sintagma al sistema dei significati risulta vano. Il linguista aggiunge: «È come se la Moda preparasse al livello dei suoi enunciati dei segni forti, numerosi, decisi e durevoli, ma, affidandoli a una memoria versatile, li dimenticasse immediatamente (Barthes 1967, pp. 212) ». Qual è allora il significato finale di ogni enunciato? Lo troviamo non al livello del codice vestimentario, dice in ultimo Barthes, ma a quello retorico: il senso globale è la Moda in sé.

### 1.1.6 Struttura del segno

Il segno è l'unione del significante e del significato; tuttavia il segno vestimentario, anche se ridotto a unità, può comprendere più frammenti di significanti (combinazioni di matrici e elementi della matrice stessa) e più frammenti di significati (combinazioni di unità semantiche): risulta essere un sintagma completo, composto da una sintassi di elementi. Barthes, correggendo Saussure, afferma che il segno linguistico non è arbitrario<sup>6</sup> e che non ci sia margine di ambiguità nell'equivalenza tra significante e significato; questo, però, non accade nel sistema della Moda, in quanto il segno viene elaborato dagli "addetti ai lavori" del *fashion system*, non coinvolgendo la totalità dei parlanti. Nasce ogni anno dalle intenzioni di pochi, non dal consenso collettivo e

---

<sup>5</sup> Barthes per spiegare il sistema di neutralizzazione, lo paragona ad una piramide composta da una scala di neutralizzazioni progressive, l'opposizione pertinente si fa sempre più rara. v.Barthes 1967 pp.208

<sup>6</sup> Saussure afferma che il segno è arbitrario e immotivato, il significato non ha nessun legame naturale con il significante.

asseconda la natura della Moda stessa: non evolve, ma cambia. Nel caso in cui il significato sia mondano, il segno di Moda è sempre motivato<sup>7</sup>. Nei tre regimi individuati, la motivazione del rapporto di significazione o è una funzione utilitaria o si limita a imitare un modello culturale ed estetico. Per quanto riguarda invece gli enunciati con significato implicito “alla moda”, il segno è immotivato: sarebbe infatti una tautologia affermare che “il vestito alla moda è la Moda”.

### 1.1.7 Accenno al sistema retorico

Procedendo nell’analisi, Barthes affronta il sistema retorico di Moda, indagando l’aspetto connotativo del codice vestimentario. In questo caso il significante è la scrittura di Moda che, attraverso semplici espedienti stilistici e terminologici, è in grado di suggerire immagini al di là del manifesto senso letterale e denotativo. Il significato vestimentario è invece l’ideologia di Moda, esso assume caratteri generali, non è né esplicito né implicito, ma latente. Per questo il significato retorico apre il paradosso della significazione connotata poiché il lettore la riceve pur non avendola letta. Si tratta quindi di una ricezione soggettiva e individuale, il significato retorico risulta confuso proprio perché dipende dal lettore che ha manipolato il messaggio, dal suo stato d’animo, dalla sua storia, dalla sua cultura. In questo modo, il significato retorico si apre al mondo, toccando il limite stesso dell’intero sistema.

### 1.1.8 Il sistema duplice della Moda

Al termine della sua indagine Barthes si trova di fronte ad una fatale contraddizione: sul piano della denotazione la Moda è un sistema di segni la cui giustificazione è puramente semantica, mentre sul piano della connotazione comunica col mondo, diventando piuttosto ideologia. Fatto riferimento all’insieme A e all’insieme B<sup>8</sup>, troviamo corrispondenza con questi due livelli: quando il significato è mondano, esso diventa connotante e partecipa indirettamente ad un sistema aperto di segni (sistema naturalistico), quando invece la Moda è il significato implicito, esso diventa denotante e partecipa direttamente ad un sistema chiuso di segni (sistema logico).

È questa ambiguità simmetrica che caratterizza il sistema della Moda ed è proprio per la divergenza dei suoi insieme A e B che il giornale di Moda oscilla tra la stampa a largo pubblico popolare, che predilige la Moda naturalizzata e quindi contaminata dal mondo, e la stampa elitaria, che preferisce la Moda pura. Secondo Barthes «(...) la Moda deve proiettare il modello aristocratico, fonte del suo prestigio: è la Moda pura; ma essa deve nello stesso tempo rappresentare, in maniera euforica, il mondo dei suoi consumatori, trasformando le funzioni intramondane in segni (lavoro, sport, vacanze, stagioni, cerimonie): è la Moda naturalizzata, dai significati nominati (Barthes 1967, v. pp. 292-293)».

---

<sup>7</sup> Il segno è motivato quando il significante sta in rapporto razionale o naturale con il suo significato.

<sup>8</sup> Cfr, 2.2.5

## 2. COMode: il Corpus di *L'Officiel de la Mode*

Il materiale sottoposto all'analisi computazionale è stato estrapolato da alcuni articoli tratti dalla rivista parigina "*L'Officiel de la Mode*", nei numeri usciti tra il 1995 e il 2013. La scelta dell'analisi è ricaduta su scritti in lingua francese, perché trovare raccolte sistematiche e ordinate di materiale in italiano non è stato possibile; infatti gli archivi dei giornali di Moda, che ripropongono in versione digitale i contenuti del cartaceo, o sono consultabili esclusivamente online, a volte a pagamento, o sono accessibili solamente ai professionisti del mestiere (questo succede ad esempio per il *Vogue Archive*<sup>9</sup>). Le riproduzioni digitali di *L'Officiel* sono custodite e ordinate per anno e numero di uscita nel *Jalou Archive*<sup>10</sup>, un archivio creato da *Jalou Media Group* nell'intento di rendere accessibile i numeri di diverse testate giornalistiche di moda francesi (in più, *l'Officiel Hommes*, *Jalouse*, *L'Optimum*, *la Revue des montres...*), per incentivare la curiosità del lettore a guardarsi indietro e per conservare parte del patrimonio culturale nazionale francese.

### 2.1. Breve storia di *L'Officiel*

*L'Officiel de la couture et de la mode de Paris*, oggi abbreviato in *L'Officiel* nasce a Parigi nel lontano maggio 1921 dall'iniziativa di Max Brunhes, con l'obiettivo di avvicinare le lettrici di età compresa tra i 25 e i 49 anni al mondo della Moda. È infatti negli anni '20 che il desiderio di tendenza cessa di appartenere ad un'élite ristretta di persone abbienti, coinvolgendo un pubblico sempre più ampio; aprono i primi grandi magazzini e vengono creati abiti vendibili a prezzi inferiori. Parigi, dove trionfano Coco Chanel, la rivoluzionaria, e Paul Poiret, l'avanguardista, diventa meta per tutti coloro che vogliono rimanere aggiornati sulle mode del tempo. *L'Officiel* ufficializza attraverso i suoi numeri le ultime tendenze, i capi cult, le icone di stile e negli anni '30, presenta al mondo intero le collezioni dell'*haute couture*. Queste creazioni dalla sartorialità minuziosa e dalla preziosità smisurata possono essere ammirate da tutti, grazie soprattutto all'introduzione nel 1933 delle fotografie a colori. Durante la Seconda Guerra Mondiale, la rivista si schiera apertamente, pubblicando in copertina le foto della bandiera francese; passati quegli anni violenti e difficili, si può dire che nel 1950 *L'Officiel* sia il giornale di moda più influente di tutta Francia. Nel corso degli anni vi collaborano infatti gli scrittori Colette, Jean Cocteau, l'illustratore René Gruau e il fotografo Patrick Demarchelier, che accrescono il valore artistico della rivista. Nel 1980, *L'Officiel* è il primo giornale francese ad essere venduto in Cina; ad oggi è pubblicato in 20 Paesi esteri e ha dato vita a nuove edizioni, come i trimestrali *L'Officiel Hommes* e *L'Officiel Art*. Il sito ufficiale e i canali social (Instagram e Facebook) sono quotidianamente aggiornati.

---

<sup>9</sup> Consultabile con restrizione : <http://archivio.vogue.it/>

<sup>10</sup> Consultabile gratuitamente : <http://patrimoine.editionsjalou.com/lofficiel-de-la-mode-sommairepatrimoine-13.html>

## 2.2 Come si costruisce un corpus linguistico

Prima di vedere come è stato creato il Corpus di *L'Officiel*, vengono considerate tutte le tappe fondamentali da percorrere per arrivare alla costruzione di un corpus.

Innanzitutto è necessario stabilire quale sia il fine ultimo di tale creazione e quindi l'uso destinato al corpus, in modo che la ricerca dei testi di partenza sia pertinente e azzeccata: i principali usi dei corpora sono di tipo applicativo, se l'obiettivo è quello di progettare strumenti, e di tipo analitico, quando si vuole analizzare la distribuzione delle costruzioni e dei fenomeni all'interno di un linguaggio.

Bisogna poi soffermarsi sul grado di generalità, che riflette la volontà di rappresentare un linguaggio in parte o nella sua interezza: si parla di corpus specialistico quando è orientato alla descrizione di una particolare varietà del linguaggio (sublanguage) o a un ristretto dominio applicativo, mentre un corpus si definisce generale o di riferimento quando mira a rappresentare le caratteristiche del linguaggio nella sua completezza.

Per quanto riguarda il materiale di indagine, è necessario inquadrare la modalità, ovvero la natura del linguaggio considerato nei testi selezionati: scritto, parlato (trascrizioni), misto (scritto e parlato), audio (*speech database*) e multimediale (audio-video).

Bisogna poi stabilire la/e lingua/e rappresentata/e: si hanno corpus monolingue e bi/plurilingue (L1-L2-ecc) in cui si distinguono corpus parallelo, quando lo stesso testo è rappresentato (in traduzione) in più di una lingua, corpus allineato se ciascuna frase (parola) della lingua L1 è esplicitamente collegata col suo traduttore nella lingua L2, corpus comparabile se vi sono testi in più lingue (non in traduzione) appartenenti alle stesse tipologie.

Nella creazione del corpus è necessario inoltre fissare un'asse di tempo (cronologia) e scegliere tra: corpus sincronico se si vuole rappresentare un particolare stadio del linguaggio, descrivendo una stessa finestra temporale, o corpus diacronico se si desidera delineare il mutamento linguistico lungo un arco di tempo dilatato.

Una volta decisi tali aspetti, bisogna tenere in considerazione altri tre parametri: dimensione, bilanciamento e rappresentatività.

La dimensione dipende in gran parte dalle interrogazioni che si vogliono porre al corpus e al grado di dettaglio e precisione dei dati recuperati. Essa può rimanere invariata nel tempo, è il caso dei corpora chiusi, ma può anche incrementare grazie all'introduzione di nuovi testi selezionati secondo gli stessi criteri usati per determinare la collezione iniziale, si parla in questo caso di corpora aperti (o di monitoraggio).

Il bilanciamento è la condizione essenziale che garantisce la rappresentatività del corpus; un buon inizio può essere prefissare il numero complessivo di parole dei testi da selezionare. In ogni caso un bilanciamento corretto richiede una quantità consistente di materiale scelto per le diverse tipologie individuate nella popolazione.

Un corpus è rappresentativo se tiene traccia dell'intero ambito di variabilità dei tratti e delle proprietà di una lingua o di una sua varietà. Ovviamente, non è possibile raccogliere un intero linguaggio per testare la rappresentatività di un corpus ma possiamo sfruttare il concetto di 'saturazione': un corpus è saturo se, dividendolo in sezioni uguali, troviamo lo stesso numero di parole in ciascuna partizione.

### 2.2.1 Perché l'uso del file di testo come documento di partenza

La maggior parte dei software che analizzano i corpora non sono in grado di leggere le tipologie complesse di formattazione che caratterizzano i pacchetti di elaborazione di testi più comuni, come Microsoft Word e PDF. Mentre la tecnologia è in continua evoluzione, attualmente i documenti che compongono un corpus, sono infatti salvati in formato .txt. È importante ricordare che qualsiasi documento che venga preparato per l'analisi computazionale è solo una rappresentazione dell'originale; per questo motivo la conversione in *plain text* può causare la perdita di informazioni. Per superare questo impoverimento del dato testuale, si può ricorrere all'annotazione, quindi alla codifica ad alto livello che usa i linguaggi di *mark-up* come XML. Il testo viene quindi annotato attraverso delle etichette o *tag* di marcatura per conservare le informazioni strutturali del documento iniziale. L'annotazione ha tre vantaggi principali: portabilità, interscambiabilità e massimo grado di espressività.

### 2.3 COMode : costruzione e caratteristiche del Corpus

Per selezionare il materiale di partenza per la creazione di *COMode*, è stato necessario inquadrare la finalità: analizzare il linguaggio giornalistico di moda, focalizzando l'attenzione sulla variazione del lessico e sulla costruzione delle dipendenze tipiche del giornale del settore. L'uso del Corpus è dunque analitico.

Proprio perché l'argomento di indagine si restringe al linguaggio giornalistico di moda, il Corpus è specialistico; il materiale di partenza è costituito da documenti scritti (corpus scritto), in particolare da articoli redatti unicamente in francese (corpus monolingue). I testi presi in considerazione appartengono ad un arco temporale piuttosto esteso: dal primo numero uscito<sup>11</sup> nel 1995 all'ultimo numero uscito nel 2013, in modo da osservare il cambiamento del linguaggio nel corso dei 19 anni (corpus diacronico).

Il Corpus è formato da 190192 tokens, una dimensione adeguata ai fini stabiliti e per il tempo a disposizione; ad ogni modo non è esclusa la possibilità in futuro di ampliarlo con l'aggiunta di articoli nuovi (corpus aperto). (v. tab.1)

Per verificare il bilanciamento e la rappresentatività è necessario fare riferimento ai parametri di selezione applicati alla scelta del materiale. All'interno di *L'Officiel* per quanto riguarda gli argomenti si trovano più settori trattati: moda, bellezza (*make-up* e *hair-style*), cinema, arte, viaggi, arredamento e oroscopo. Anche rispetto alla lunghezza dei testi si trovano diverse dimensioni: editoriali<sup>12</sup>, articoli standard (in media 550 battute), rubriche a tema, interviste, didascalie, pubblicità, e brevi presentazioni degli ultimi *must have* di stagione.

---

<sup>11</sup> Si fa riferimento alle uscite di *L'Officiel* che l'archivio Jalou ha reso disponibili sul sito; per motivi di difettosa digitalizzazione OCR, alcuni numeri non sono stati presi in considerazione.

<sup>12</sup> Si fa riferimento al testo in apertura di ogni numero scritto dal Direttore della testata e alle introduzioni monotematiche dei servizi fotografici sulle tendenze del mese.

I testi selezionati appartengono alla categoria ‘moda’, mentre per quanto riguarda le tipologie, hanno avuto precedenza gli editoriali, gli articoli standard, le presentazioni su sfilate e tendenze, le interviste a stilisti e le storie sulla loro carriera. Tuttavia è inverosimile immaginare che la struttura grafica dei contenuti del giornale sia rimasta invariata nel tempo, per questo la selezione del materiale ha dovuto considerare i cambiamenti dell’impaginazione e degli spazi dedicati ai testi: nei primi numeri a prevalere è la componente scritta ma gradualmente si assiste ad un rovesciamento, arrivando alla predominanza dell’immagine e dei titoli.

L’Officiel 1995, n° 793 - Tendenze



L’Officiel 2012, n° 970 - Tendenze



L’Officiel 1996, n° 805 - Editoriale Moda



L’Officiel 2010, n° 946 - Editoriale Moda



Quindi la questione del bilanciamento è delicata: per superare la perdita del numero di battute dei contenuti, è necessario selezionare gradualmente un numero maggiore di articoli.

La fase successiva alla scelta del materiale, è quella della preparazione del testo; per prima cosa è necessario ripulire i documenti digitalizzati da tutti gli elementi che possono alterare le successive elaborazioni computazionali (*text cleaning*). Dunque sono stati eliminati le didascalie delle figure, le pubblicità, i crediti degli indumenti e i links. Successivamente sono stati corretti gli errori di conversione dovuti alla digitalizzazione OCR<sup>13</sup>, come lo scambio della “t” con la lettera “l” e sono state eliminate le convenienze grafiche che compromettevano il significato delle parole, come il trattino “-” usato per andare a capo. Un altro esempio frequente, di questo tipo, è rappresentato dalla preposizione francese “À” con cui iniziano molti periodi che, per conversione errata, è diventato “A”, verbo avere della terza persona singolare. In questo caso con la sostituzione è stato evitato ogni tipo di equivoco possibile nell’assegnazione dei POS tag. Un altro passaggio è stato recuperare i nomi propri e le parole ad inizio frase con la lettera maiuscola (normalizzazione). Una volta eliminato ogni tipo di “rumore”, il Corpus è stato sottoposto alla tokenizzazione, prima vera fase dall’analisi computazionale.

Caratteristiche del corpus di L'Officiel	
Uso/finalità	Analitico
Generalità	Corpus specialistico - linguaggio giornalistico settore moda
Modalità	Corpus scritto
Lingua	Corpus monolingue
Cronologia	Corpus diacronico - 1995/2013
Dimensione	190192 tokens
Annotazione Strutturale	no

Tabella 1. Caratteristiche del Corpus

<sup>13</sup> *Optical Character Recognition*, sistema di riconoscimento ottico dei caratteri che converte l'immagine in testo.

### 3. Talismane

Talismane è stato utilizzato come analizzatore sintattico di *COMode*, corpus creato *ad hoc* per la realizzazione della tesi. La scelta è ricaduta su di esso perché permette l'analisi computazionale in lingua francese.

#### 3.1 Che cosa è Talismane

Talismane<sup>14</sup> è un analizzatore sintattico sviluppato da Assaf Urieli all'interno del laboratorio CLLE-ERSS dell'Università di Tolosa in Francia. Viene distribuito attraverso la licenza *AGPL open-source license*<sup>15</sup> per favorire l'accesso libero e gratuito.

È scritto interamente in Java, per questo risulta facilmente integrabile in diverse applicazioni e funziona sui principali sistemi operativi tra cui Linux, MacOS e Windows.

Talismane è composto da quattro stadi principali che trasformano un testo, in formato semplice, in una serie di alberi di dipendenza sintattica, attraverso un'analisi a cascata.

Essi sono: suddivisione in frasi, tokenizzazione, pos-tagging e parsing. (v.fig.2)

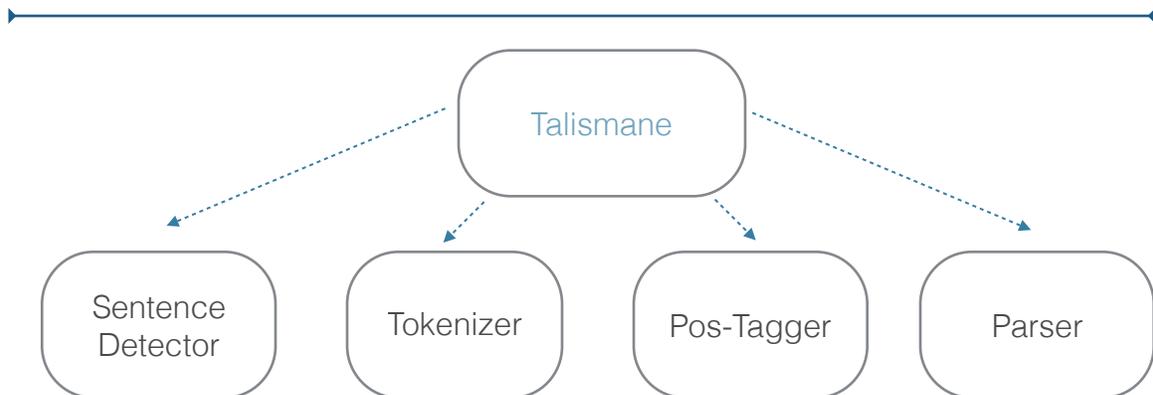


Figura 2. Talismane processing chain

L'obiettivo di ciascuna fase è definito come un problema di classificazione e viene risolto statisticamente attraverso un training corpus annotato che nel caso specifico del francese è il *French TreeBank*<sup>16</sup>.

Talismane, per poter classificare correttamente un dato linguistico, si serve di *features* e di *rules*: le prime sono attributi che forniscono automaticamente un'informazione utile

<sup>14</sup> Assaf Urieli. Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talismane toolkit. PhD thesis, Université de Toulouse II le Mirail, 2013.

<sup>15</sup> GNU Affero General Public License, licenza di software libero: <https://www.gnu.org/home.html>

<sup>16</sup> French TreeBank, corpus annotato sintatticamente e morfologicamente: <http://www.lif.cnrs.fr/Gens/Abeille/French-Treebank-fr.php>

---

alla classificazione del dato linguistico mentre le seconde vengono applicate quando Talismane analizza i nuovi dati, consentendo al sistema di bypassare il modello statistico e automaticamente assegnare una certa classificazione ad un determinato fenomeno.

Per quanto riguarda il lessico, la principale risorsa di riferimento per corpora in francese è il *Leff*<sup>17</sup>, che troviamo di default all'interno del *French language pack*; ogni volta che Talismane inizia la sua analisi, carica in memoria l'intero *Leff* per individuare la classificazione di un determinato dato linguistico.

Il *tagset*<sup>18</sup> è basato sulla lista di Marie Candito e di Benoît Crabbé, con piccole modifiche apportate e l'output di default del parsing è basato sul *CoNll Format*<sup>19</sup>.

### 3.2 Analisi di *COMode* attraverso Talismane

Talismane, in quanto rende accessibile il pacchetto il lingua francese, è stato utilizzato per eseguire la prima parte dell'analisi linguistica di *COMode*.

Dalla suddivisione in unità minime, il corpus risulta essere composto da 190192 tokens. Attraverso l'annotazione morfo-sintattica, è stata assegnata a ciascuno di esso, l'informazione relativa alla categoria grammaticale (POS-tag) di appartenenza; successivamente è stata effettuata la lemmatizzazione, con cui ogni parola del testo è stata ricondotta al relativo esponente lessicale o lemma.

Il Corpus è stato oggetto anche dell'annotazione sintattica, attraverso cui sono state individuate le dipendenze e quindi le relazioni binarie grammaticali (soggetto, oggetto, modificatore ecc..) che interessano le parole del testo.

#### 3.2.1 Titoli e costruzione nominale

Se la funzione del POS tagging e della lemmatizzazione rimandano ad un'analisi più approfondita<sup>20</sup>, l'assegnazione delle dipendenze evidenzia immediatamente un tratto peculiare del linguaggio giornalistico: la costruzione nominale.

Si assiste infatti ad una frequente “de-verbalizzazione” nei titoli degli articoli e in maniera più ricorrente per quelli degli editoriali. Infatti nella costruzione delle dipendenze la *root*, ovvero la radice da cui si diramano le relazioni sintattiche, viene assegnata al sostantivo principale se risulta assente il verbo. (v.fig.3)

---

<sup>17</sup> Leff, Lexique des Formes Fléchies du Français: <http://alpage.inria.fr/~sagot/leff.html>

<sup>18</sup> Tagset di riferimento, vedi *Appendici*

<sup>19</sup> CoNLL, The SIGNLL Conference on Computational Natural Language Learning: <http://www.conll.org/>, vedi *Appendici*

<sup>20</sup> cfr. 4.1.1

1	Trois	trois	2	det	
2	carrières	carrière	0	—	La dipendenza 0, o <i>root</i> , è assegnata, nel caso in cui non vi sia forma verbale, al sostantivo principale dell'enunciato.
3	singulières	singulier	2	mod	
4	,	,	3	ponct	
5	trois	trois	6	det	
6	personnalités	personnalité	2	mod	
7	opposées	opposer	6	mod	
8	,	,	7	ponct	
9	trois	trois	10	det	
10	façons	façon	2	mod	
11	d'	de	10	dep	
12	envisager	envisager	11	prep	
13	une	une	15	det	
14	même	même	15	mod	
15	passion	passion	12	obj	
16	:	:	15	ponct	
17	la	la	18	det	
18	mode	mode	2	mod	
19	.	.	18	ponct	

Figura 3. Esempio di analisi delle dipendenze di un enunciato nominale

Per la sua natura breve e incisiva, il messaggio del titolo è veicolato da enunciati che prediligono il nome e l'aggettivo, a discapito della forma verbale. Affinché sia garantita la comunicazione, il significato complessivo non risulta compromesso, anzi, l'attenzione è focalizzata su uno o più termini maggiormente rilevanti.

Gli esempi più rappresentativi di titoli sono i seguenti:

Costruzione nominale

“ Pantalon la liberté à deux jambes! ”

“ Pour les femmes, des coordonnés souples, pantalons, chemisiers, jupes, pulls et petits talons.”

“ Elizabeth Street au coeur de Belgravia, une nouvelle enclave du luxe britannique. ”

“ Le Beach Fashion Festival de Rio: le carnaval de la mode à la plage.”

“ Phénomène nouveau dans les collections de prêt-à-porter : le grand retour de la robe et l'ébauche d'une nouvelle féminité joueuse et triomphante. Les nouvelles femmes fatales. ”

“ Tendances: avant-première printemps-été 2006. Coup d'oeil panoramique sur les meilleurs modèles et les grandes tendances de l'été prochain. ”

## 4. Analisi della terminologia

L'analisi attraverso Talismane ha avuto come obiettivo principale l'individuazione dei POS tag all'interno del Corpus ed è proprio dal formato CoNLL che parte la trattazione terminologica per mezzo di EXTra.

### 4.1 EXTra: estrazione e trattazione della terminologia

EXTra è uno strumento in grado di estrarre combinazioni di parole: i “*multiword terms*”; così facendo, identifica espressioni più articolate tenendo conto sia della struttura linguistica generale sia della loro complessità interna. L'unità minima di EXTra non è rappresentata dalla singola parola, perché la concatenazione di termini risulta meno ambigua dal punto di vista del significato e più rappresentativa rispetto al contenuto del testo preso in considerazione.

Per riconoscere ed estrarre particolari combinazioni di parole, è necessario definire una serie di patterns composta da liste di sequenze di POS tag<sup>21</sup> annidate; EXTra usa questa informazione per controllare l'ordine in cui verranno calcolate le forze di associazione, procedendo dai componenti più innestati verso quelli più esterni.

Pattern: `[[ 'nc', 'pd', 'nc'], 'p', 'nc']`  
Combinazione di termini: `“retour du chemisier en crêpe”`

`[[nome('NC'), prep. articolata('PD'), nome('NC'), prep('P') nome('NC')]]`

`[[retour('NC'), du('PD'), chemisier('NC')], en('P') crêpe('NC')]]`

Per prima cosa EXTra calcola la forza di associazione  $\sigma$  della combinazione di termini (*terms*)  $\langle w_1, w_2 \rangle$ , secondo la misura scelta tra Pointwise Mutual Information (PMI), Local Mutual Information (LMI) o Log Likelihood; se il grado di associazione supera una soglia precedentemente stabilita, allora la combinazione di termini viene aggiunta all'insieme dei termini  $T = \{t_1, \dots, t_n\}$ . L'algoritmo che calcola il grado di “*termhood*” funziona ricorsivamente sulla struttura dei patterns e misura la forza di associazione  $\sigma$  di ogni n-parola all'interno della concatenazione dei termini, combinata con la forza di associazione dei suoi sotto-elementi, secondo la seguente formula:

$$\sigma(\langle w_1, w_2 \rangle) = S(w_1) * S(w_2)$$

dove:  $S(w_i) = 1$  se  $w_i$  è una parola

$$S(w_i) = \log_2(\sigma(w_i)) / k \text{ se } w_i \in T$$

Il parametro  $k$  controlla il contributo del sotto-elemento sull'intera combinazione di termini; il suo valore è inversamente proporzionale alla precedenza assegnata alle combinazioni più lunghe.

<sup>21</sup> Lo standard riconosciuto è il CoNLL format.

Nell'esempio riportato non si tiene conto della preposizione semplice e quella articolata perché sono presenti nella lista dei POS da ignorare. La configurazione di EXTra prevede un'altro filtro applicabile arbitrariamente all'estrazione: la lista delle *stop words*, nella quale si indicano i lemma delle parole per cui l'indicizzazione si deve arrestare.

---

`[[retour('NC'), du('PD'), chemisier('NC')], en('P') crêpe('NC')]`

---

- nel **primo passaggio** EXTra calcola la forza di associazione tra i suoi sotto-elementi:

`[retour-NC, du-PD, chemisier-NC]`

- nel **secondo passaggio** EXTra, appartenendo l'associazione dei suoi sotto-elementi all'insieme T, calcola la forza di associazione tra:

`[retour_du_chemisier, en, crêpe]`

- il calcolo della "*termhood*":

$$\sigma(\text{retour\_du\_chemisier, en\_crêpe}) = S(\text{retour\_du\_chemisier}) * S(\text{en\_crêpe})$$

---

L'identificazione e l'estrazione della terminologia risulta interessante per diverse aree applicative come la costruzione di dizionari e thesauri, l'indicizzazione di testi, la traduzione automatica e l'apprendimento.

#### 4.1.1 Analisi di *COMode* attraverso EXTra: i nomi comuni

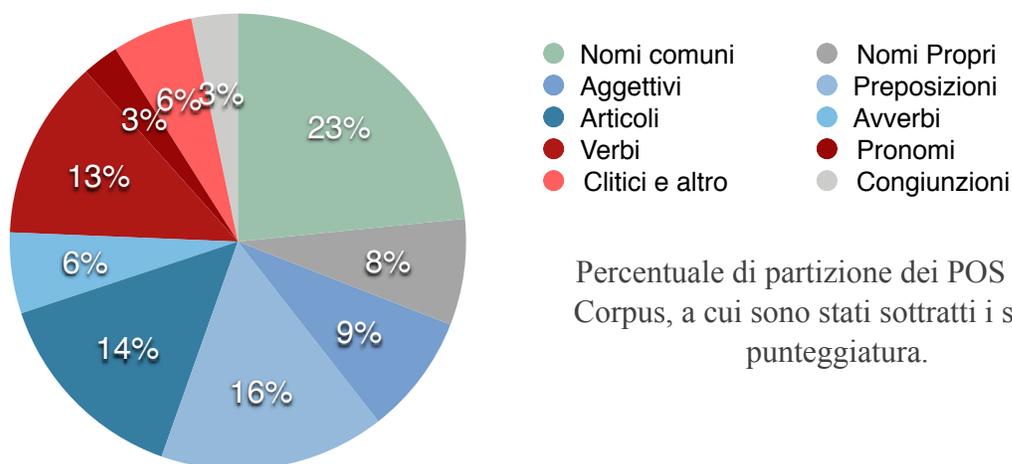
EXTra risulta essere facilmente adattabile all'elaborazione di documenti scritti in lingue diverse: seppur sia stato adoperato per la prima volta su un corpus in italiano, il PAWaC<sup>22</sup> (Public Administration Web as Corpus), è risultato idoneo alla trattazione di *COMode*. Partendo dal parsing del Corpus, ottenuto attraverso l'analisi di Talismane, è bastato definire i patterns utilizzando i POS tag del francese.

Il focus dell'estrazione terminologica è ricaduto sui nomi comuni, contrassegnati dal POS tag "NC", e sulle loro combinazioni. Infatti, essendo il linguaggio settoriale di moda altamente iconico, è proprio attraverso i nomi e in parte anche dagli aggettivi, che viene veicolata l'informazione. Il lettore deve percepire ogni minimo dettaglio dell'indumento dalla descrizione precisa del modello, del tessuto e della fantasia. Il linguaggio assume così un ruolo fondamentale nel raccontare oggetti concreti, suggerendone un'immagine corrispondente il più possibile alla realtà.

---

<sup>22</sup> v. Lenci, Passaro 2015. pp.4-7

---



Percentuale di partizione dei POS tag nel Corpus, a cui sono stati sottratti i segni di punteggiatura.

La configurazione di EXTra per l'estrazione dei termini da *COMode* prevede, oltre ai patterns<sup>23</sup> dichiarati, una lista di *stop words*: in questo caso il sistema fa riferimento al lemma del token, quindi per includere ogni forma verbale ausiliare, sono stati specificati l'infinito "être" e "avoir". I POS tag invece da ignorare sono quelli che individuano la preposizione semplice (P) e la preposizione articolata (PD). Il valore soglia della forza di associazione è stato impostato a 0 e la frequenza minima, perché le combinazioni di termini siano estratte, è di 5 ma, lavorando sulla relazione d'ordine stretto, nella configurazione è uguale a 4. Il parametro k è impostato a 1, valore per cui il peso dei termini più lunghi è massimale; la misura di associazione scelta è la Positive Local Mutual Information<sup>24</sup> che dimostra avere una maggiore propensione per le espressioni frequenti. (v. tab.2)

Configurazione di EXTra per <i>COMode</i>	
Stop words	être - avoir
Pos ignored	P - PD
Association measure threshold	0
Minimum n-gram frequency	4
K-division	1
Chosen association measure	PLMI

Tabella 2. Configurazione di EXTra per *COMode*

<sup>23</sup> Patterns per estrazione EXTra, vedi Appendici.

<sup>24</sup>  $LMI = O * \log_2(O/E)$ , dove O è la frequenza osservata della combinazione ed E la frequenza attesa se i suoi sotto-elementi fossero considerati indipendenti l'uno dall'altro.

I patterns dell'estrazione prevedono nomi comuni, aggettivi / participi passati, e verbi. Dai risultati ottenuti, emergono:

- terminologie tipiche del linguaggio di moda;

		PLMI
mise-NC en-P scène-NC	messa in scena	65.4839 <sup>25</sup>
goût-NC de-P jour-NC	gusto del giorno	54.7124
mode-NC masculin-ADJ	moda maschile	49.4839
grand-ADJ magasin-NC	grande magazzino	47.8083
campagne-NC publicitaire-ADJ	campagna pubblicitaria	44.3178

- espressioni che indicano alcune professioni del sistema;

		PLMI
directeur-NC artistique-ADJ	direttore artistico	78.6356
top-NC models-ADJ	top model	78.6356
créateur-NC de-P mode-NC	creatore di moda	54.2703
grand-ADJ couturier-NC	stilista	41.8322

- costruzioni utilizzate per definire le tempistiche delle collezioni;

		PLMI
premier-ADJ collection-NC	prima collezione	70.2048
collection-NC croisière-ADJ	collezione crociera	49.9567
nouveau-ADJ ligne-NC	nuova linea	45.0472
premier-ADJ défilé-NC	prima sfilata	37.4216

- concatenazioni di termini per definire capi d'abbigliamento.

		PLMI
robe-NC de-P soir-NC	abito da sera	74.7249
maillot-NC de-P bain-NC	costume da bagno	65.4839
veste-NC court-ADJ	giacca corta	44.3178
petit-ADJ talon-NC	tacchi bassi	29.6878

Sono da approfondire le combinazioni di termini che hanno valore di forza di associazione più alto e numero di occorrenze piuttosto elevato: *haut-ADJ couture-NC* (alta moda), con PLMI pari a 361.6412 e frequenza uguale a 78 e *petit-NC robe-NC noir-ADJ* (tubino), con PLMI pari a 361.3612 e frequenza uguale a 10. Perché proprio queste due espressioni hanno un così elevato grado di associazione? La risposta può essere celata dietro la loro storia e dietro il contesto in cui sono state inventate.

Il termine *haute couture* nasce infatti nel XVII secolo alla corte del re Luigi XIV ed oggi indica le creazioni d'alta sartorialità delle case di moda di lusso più importanti al mondo. In Francia questo appellativo è giuridicamente protetto da un decreto emanato nel 1945, per cui si possono avvalere di tale *status* solo le aziende che, grazie al lavoro di un

<sup>25</sup> I valori relativi alla forza di associazione PLMI sono qui riportati fino a 4 cifre decimali.

numero definito di sarti, producono abiti confezionati a mano, garantendo l'unicità del capo, la preziosità dei tessuti e la realizzazione di due sfilate l'anno. Inoltre, nella maggior parte dei casi, nei testi di riviste e comunicati stampa del settore, l'espressione *haute couture* non viene tradotta perché viene riconosciuta facilmente dai lettori proprio per la sua unicità di significato. Anche la seconda concatenazione di termini, *petite robe noir*, vanta un'origine che spiega la coesione dei suoi sotto-elementi: a coniarla fu infatti, nel 1926, Coco Chanel, il cui tubino nero divenne presto capo cult riconosciuto in tutto il mondo come sinonimo di eleganza e raffinatezza. Raramente, quest'espressione trova un corrispondente in traduzione, perché è portatrice di un valore connotativo che va oltre l'indumento in sé: è infatti il simbolo di quella rivoluzione che diffuse una nuova idea di vestiario femminile e di donna emancipata.

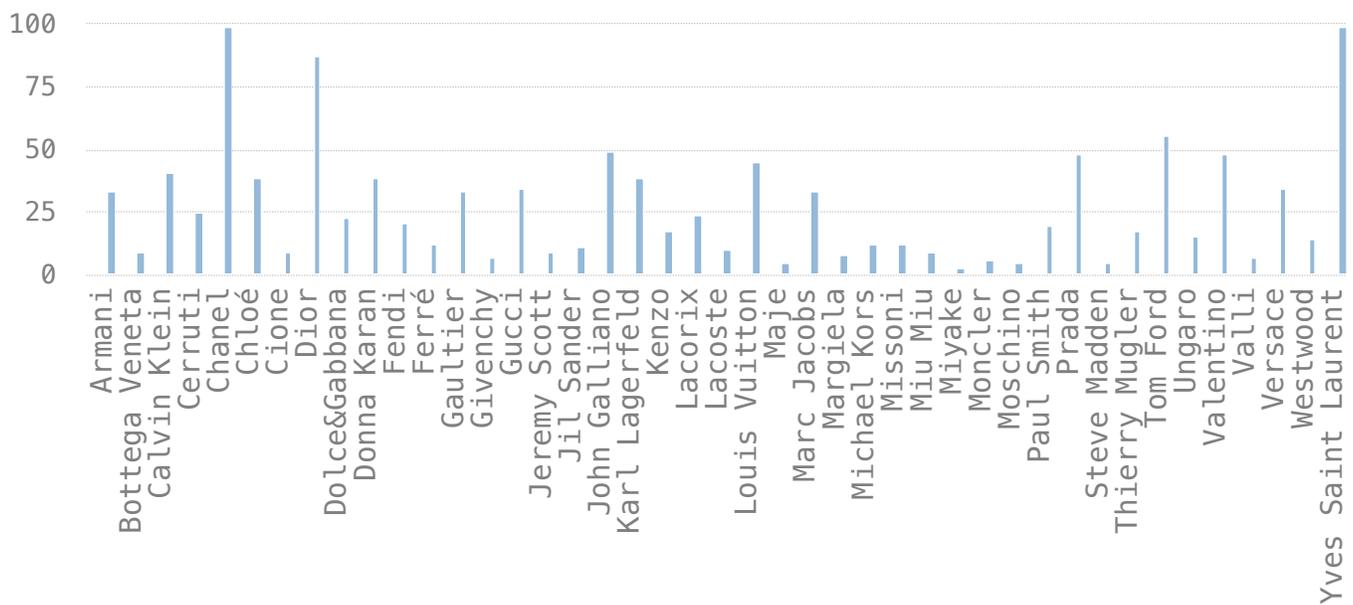
Qui di seguito, le prime 20 combinazioni di termini estratte:

Terms	PLMI	Frequency
haut-adj#couture-nc	361.641235261	78
petit-adj#robe-nc#noir-adj	361.361222252	10
monde-nc#entier-adj	202.65632656	33
premier-adj#fois-nc	117.169981832	23
été-nc#prochain-adj	98.7309937093	18
rendre-v#hommage-nc	84.9866963613	11
même-adj#temps-nc	78.9379948648	13
directeur-nc#artistique-adj	78.6356682932	10
top-nc#models-adj	78.6356682932	10
air-nc#de-pd#temps-nc	78.4996648132	12
petit-adj#pièce-nc	77.18837634	13
robe-nc#de-pd#soir-nc	74.7249525158	11
premier-adj#collection-nc	70.2048707632	20
nouveau-adj#génération-nc	67.5708728704	12
jeune-adj#fille-nc	66.6193296815	10
mise-nc#en-p#scène-nc	65.4839593937	8
valeur-nc#sûr-adj	65.4839593937	8
maillot-nc#de-p#bain-nc	65.4839593937	8
clin-nc#de-p#oeil-nc	65.4839593937	8

## 4.2 Uno sguardo agli stilisti protagonisti in *L'Officiel de la Mode*

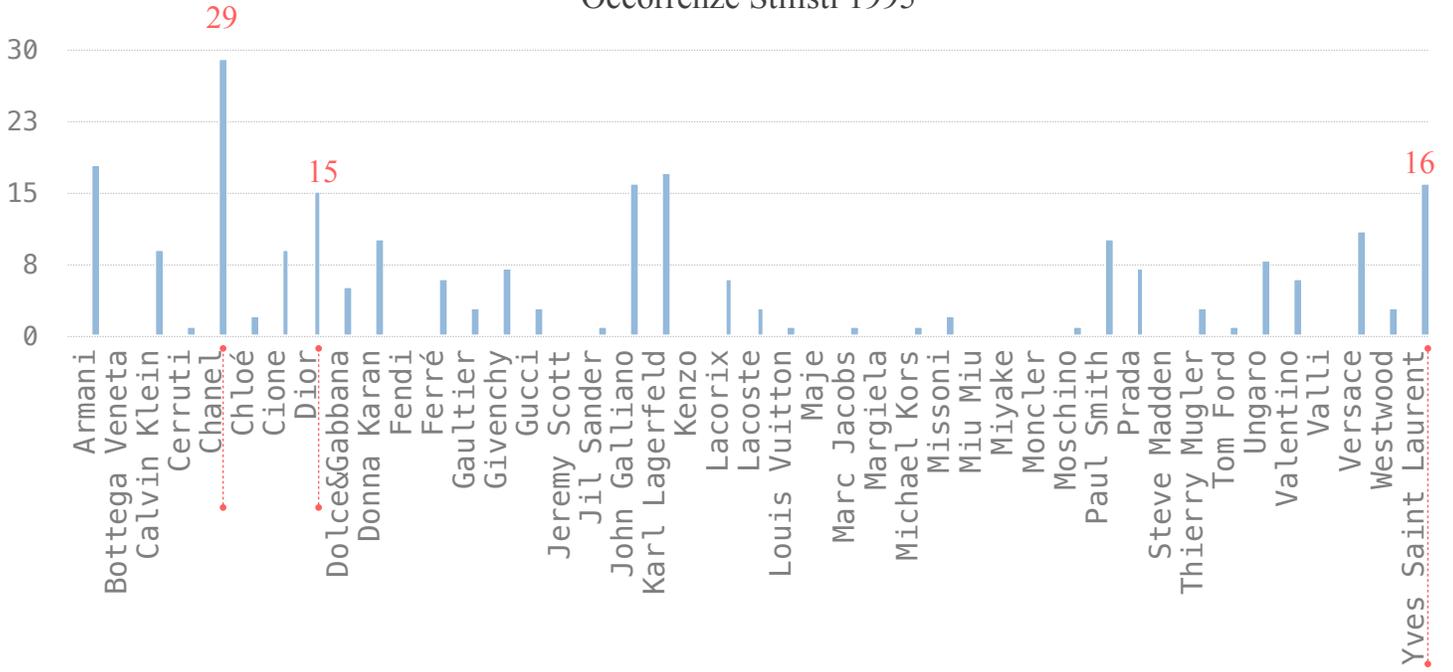
La diacronia di *COMode* permette di osservare una finestra temporale piuttosto ampia, individuando un arco cronologico di 19 anni. Per sfruttare tale caratteristica, il contenuto del Corpus è stato suddiviso in 4 sotto-corpora, rispettando il bilanciamento dei dati e la suddivisione per anno. Dato il settore del giornale, la moda, è stato esplorato il campo dei nomi propri (tag “NPP”): in particolare sono stati estratti gli stilisti che, con le loro collezioni, hanno segnato tendenze e costumi dal 1995 al 2013. Si tratta di una lista dei maggiori esponenti e *fashion influencer* del periodo, francesi e non, le cui storie e creazioni hanno preso vita attraverso le pagine di *L'Officiel*. In generale, il giornale di moda, decidendo di citare o raccontare uno stilista piuttosto di un altro, ci suggerisce implicitamente quali siano le tendenze “alla Moda” a cui ciascuna lettrice deve far fronte per seguire il *trend*; è evidente, dai dati estratti, che se alcuni nomi propri, mantengono alto il loro grado di occorrenza, altri costituiscono nel testo semplici citazioni di poco rilievo. Nei grafici riportati (v. fig. 4), sono evidenziati i tre designer che, sull'intero Corpus, ricorrono più volte: Chanel, con occorrenza 99, Yves Saint Laurent, con occorrenza 99 e Christian Dior, con occorrenza 87. Non a caso, si parla di tre illustri stilisti francesi che hanno dettato le regole dell'alta moda in tutto il mondo, creando capi iconici simbolo dell'*haute couture* francese del XX secolo. Inoltre, se da una parte alcuni nomi mantengono il loro numero di occorrenza costante nello spettro diacronico, di altri, si nota un picco seguito da un calo: è il caso del marchio Donna Karan che, dopo gli anni d'oro alla fine degli anni '90, ha attraversato un periodo difficile dal punto di vista economico e burocratico, sfociato, nel 2001, con la cessione dell'azienda al gruppo francese LVMH<sup>26</sup>.

Occorrenze Stilisti sull'intero COMode

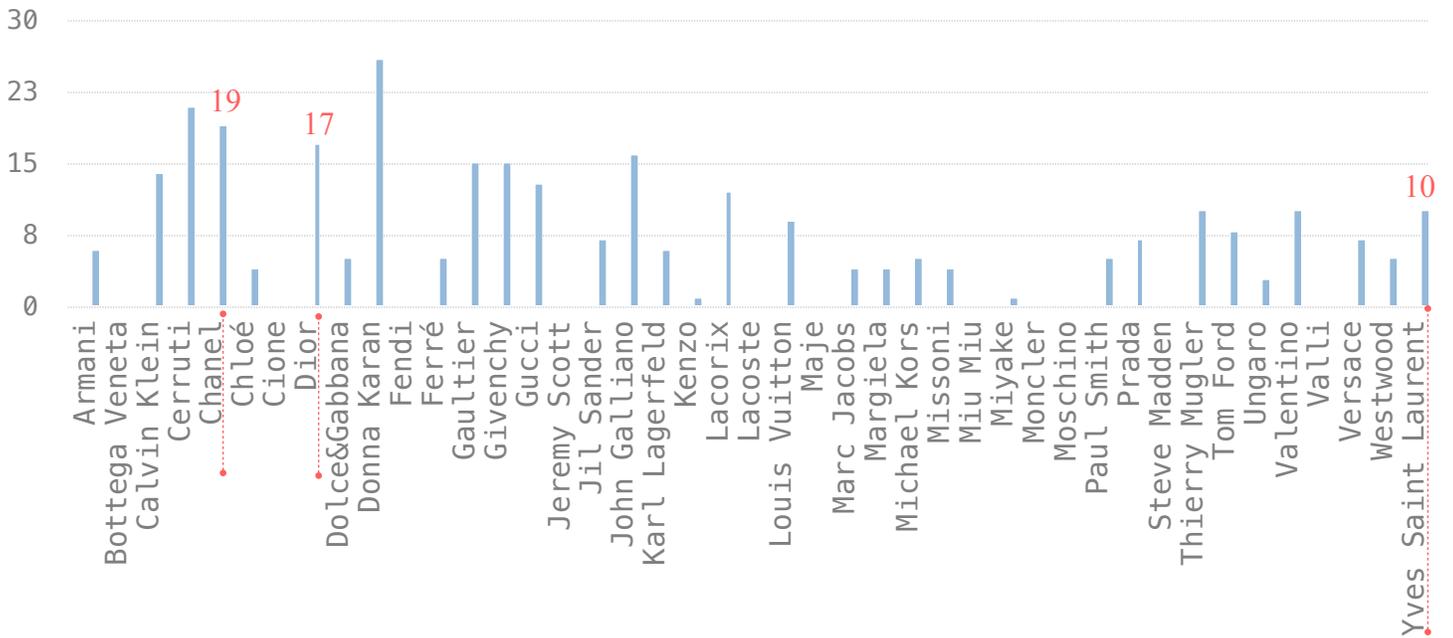


<sup>26</sup> LVMH, Moët Hennessy Louis Vuitton S.A, multinazionale operante nel settore dei beni di lusso.

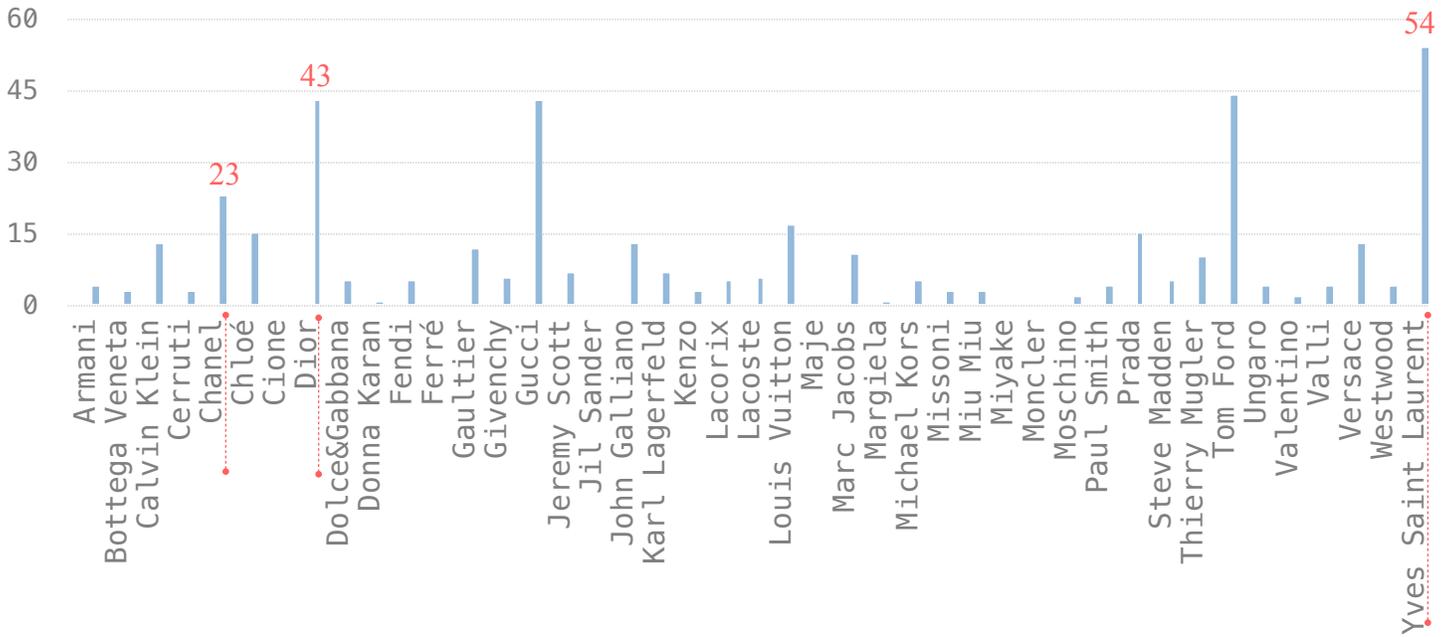
### Occorrenze Stilisti 1995



### Occorrenze Stilisti 1996/1999



### Occorrenze Stilisti 2000/2006



### Occorrenze Stilisti 2007/2013

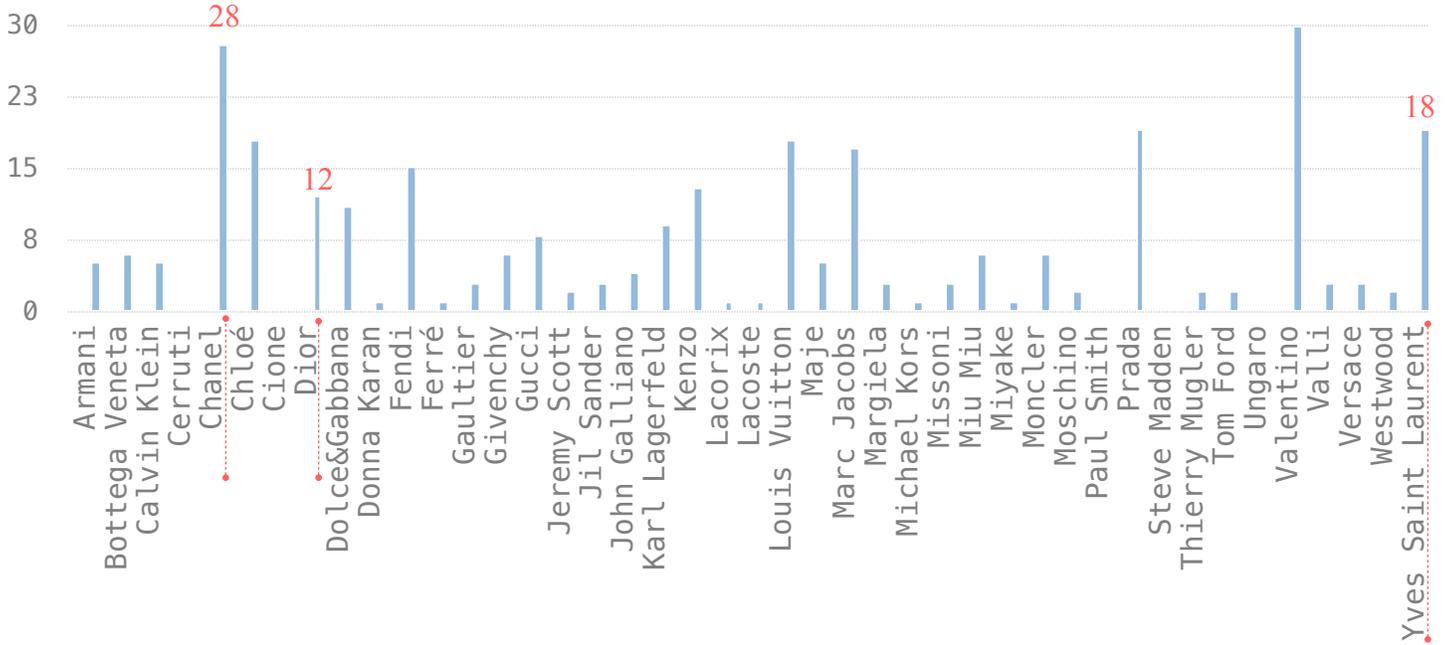


Figura 4. Grafici di rappresentazione delle occorrenze degli stilisti nominati su COMode

### 4.3 La contaminazione della lingua in Francia

La globalizzazione portata dai *mass media* ha compromesso l'integrità delle lingue nazionali ufficiali di tutta Europa, inibendo la diversità e provocando una graduale omologazione dei tratti peculiari del linguaggio.

Il modello di riferimento utilizzato per la comunicazione tra ambienti linguistici diversi, è l'inglese. La diffusione di nuove espressioni, dette neologismi, i prestiti e i calchi linguistici, hanno modificato, sotto la diretta influenza dell'angloamericano, il lessico delle diverse lingue. Tuttavia in Francia è stata applicata una politica piuttosto rigida nei confronti delle contaminazioni linguistiche, che si sforza di assicurare la presenza e l'integrità della lingua nazionale. Per questo motivo sono nate nel corso degli anni, diverse associazioni che mirano alla custodia e al mantenimento del francese puro: già nel 1937 fu fondato *L'Office de la langue française*, per arrivare in ultimo, alla creazione nel 1989 della *Délégation générale à la langue française*, dove sono confluiti diversi organismi preesistenti interessati alla difesa della terminologia.

A tal proposito è stata approvata nel 1994 la legge Toubon<sup>27</sup> che impone l'uso della lingua francese per tutti i contratti e tutti i messaggi destinati al pubblico. Tuttavia la battaglia contro la diffusione del “*franglais*” è, in molti settori, persa in partenza; basta pensare alla terminologia del ramo economico, tecnologico, informatico e quello della comunicazione di massa che in maniera crescente, adotta l'uso dell'inglese.

Per quanto riguarda il settore Moda, il francese risulta essere, insieme all'italiano, la lingua ufficialmente riconosciuta. In principio, l'impatto con gli esotismi è stato meno significativo rispetto ad altri ambiti perché gran parte dei termini legati al mondo della moda sono nati proprio in Francia, in concomitanza con la creazione dei primi giornali di tendenze verso la metà del XVIII secolo.

Tuttavia è la stessa natura mutevole della moda che rende necessario un continuo adeguamento del lessico; spesso gli indumenti, i tessuti, i modelli e lo stile cambiano così velocemente che entrano a far parte della vita di ognuno di noi senza che sia trascorso un tempo sufficiente a farli esistere anche sul piano linguistico. Per questo si diffondono forestierismi, si ricorre ai calchi e si formano neologismi.

Ad esempio: *dessinateur textile*, affianca il più diffuso “*styliste*” ma rimanda al verbo inglese “*design*”.

L'esterofilia linguistica può essere anche dovuta ad una scelta stilistica: alcune neoformazioni permettono di descrivere al meglio l'indumento su carta poiché, essendo espressioni altamente iconiche, garantiscono immediatezza e incisività.

---

<sup>27</sup> Per il testo integrale della legge: <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000005616341&dateTexte=vig>

#### 4.3.1 Il *franglais* in *L'Officiel de la Mode*

Per osservare l'influenza della terminologia inglese nelle pagine di *L'Officiel*, il Corpus è stato suddiviso ed elaborato distintamente anno per anno. L'obiettivo è quello di monitorare, diacronicamente, l'adozione di forestierismi, prevedendo un incremento nel corso degli ultimi anni. (v. fig. 5)

Nell'analisi, sono considerati anche quei termini di origine inglese che già sono stati riconosciuti e adottati dai parlanti; compaiono infatti nel dizionario<sup>28</sup> francese.

*Es: Club:* in questo caso ha precedenza sul termine francese "équipe".  
"Marseille, son port, ses parties de pétanque, son club de football, son Bernard Tapie et son savon."

Dai risultati ottenuti, si osserva che l'andamento non segnala picchi evidenti e che le percentuali, calcolate in base al numero di tokens per ciascun anno, siano basse. Infatti il numero di termini inglesi utilizzati è mediamente inferiore all'1% del numero totale di parole di ciascun sotto-corpora: gli articoli di *L'Officiel* sembrano contenere la diffusione del *franglais*, preferendo agli esotismi, espressioni fedeli al francese. Se da una parte, tale scelta possa trovarsi in linea con la politica della difesa della lingua incontaminata, dall'altra, potrebbe essere interpretata come semplice scelta stilistica. L'uso sporadico di parole non appartenenti alla lingua principale permette che il testo risulti comprensibile ad un pubblico di lettrici più ampio. Inoltre, facilita la lettura rendendola più scorrevole: se si assiste ad una totale "invasione" degli esotismi, essa può infatti risultare pesante e, a volte, grottesca. Al contrario, se l'uso del forestierismo è calibrato e contenuto, l'introduzione di un termine che, in maniera evidente, non appartiene al contesto linguistico di base, rappresenta una nota vivacizzante sull'intero testo.

*Es: Glamour e Cool:* espressioni vivaci per intendere "à la mode".  
"Il y a du glamour dans l'air."  
"Nouvelle tendance 'cool', des pastels froids melon pâle, citron doux."

L'ingresso di esotismi arricchisce prevalentemente il vocabolario del settore moda, in cui vengono coniate particolari espressioni di largo uso.

*Es: "Top models et robes de fête, on veut toujours en savoir plus."  
"En détails ou total look, le tweed sera partout l'été prochain."  
"Et les fashion victims commencent à bouder logos et labels."  
"Vestes, jupes, pantalons, font évoluer l'approche du sportswear."*

Si utilizzano anglicismi anche per descrivere i colori degli abiti e dei tessuti: troviamo il black, il white, il dark e il gold.

*Es: "En cuir 'gold' tressé façon crochet, surmonté d'une poche en cuir lisse et d'un fermoir de métal."*

---

<sup>28</sup> Dizionario di lingua francese: <http://www.larousse.fr/dictionnaires/francais>

Altri prestiti derivano dal linguaggio del settore economico: si incontrano spesso le parole *marketing* e *business*, vengono usate le espressioni *leadership* e *lobbying*.

In altre occasioni, ricorrere al prestito straniero risulta essere l'unica strada percorribile: si tratta dei casi in cui non vi sia un esatto corrispondente nella lingua originale. Per la maggior parte delle volte, tale situazione, si verifica per i tecnicismi, il cui alto grado di specificità, rende il forestierismo riconoscibile interamente dalla comunità parlante.

**Es:** Patchwork: tecnica di cucitura che prevede l'assemblaggio di più tessuti (patch = pezze + work = lavoro).

"Un **patchwork** survolté sur lequel on reconnaît le mini LV en alligator corail."

**Es:** Stretch: detto di un tessuto particolarmente elasticizzato.

"Une veste en coton **stretch** blanc associée à une jupe fendue en mousseline rosé poudré."

L'espressione "New Look"<sup>29</sup> rappresenta un caso emblematico: si tratta infatti dello stile rivoluzionario lanciato da Christian Dior nella sua collezione presentata nel 1947, all'indomani della fine della Guerra. Capo iconico del New Look, è il tailleur bar, acclamato e desiderato dalle donne di tutto il mondo. Conosciuto così, per la prima volta dalle lettrici, il nome di questo stile non è mai stato cambiato o tradotto in francese, nonostante il suo creatore fosse proprio uno dei principali couturier di Francia.

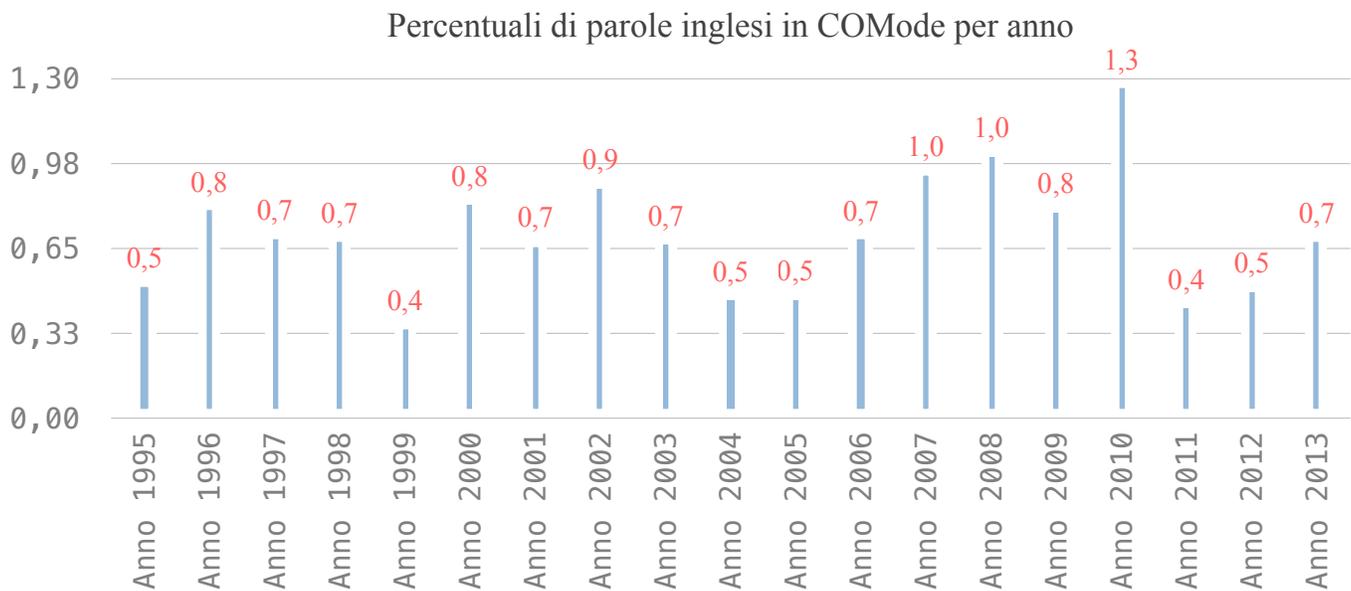


Figura 5. Grafico di rappresentazione delle percentuali di parole inglesi in COMode

<sup>29</sup> Maison Dior, la rivoluzione del New Look: [http://www.dior.com/couture/it\\_it/la-maison-dior/le-storie-dior/la-rivoluzione-del-new-look](http://www.dior.com/couture/it_it/la-maison-dior/le-storie-dior/la-rivoluzione-del-new-look).

## Conclusioni

A conclusione di questa prima analisi svolta sul corpus *COMode* è possibile delineare i risultati più rilevanti della ricerca.

La costruzione del corpus *COMode* è risultata funzionale ad un'analisi mirata, per la tipologia del materiale, ma contenuta, per la quantità di dati considerati, del linguaggio giornalistico di Moda in francese.

Essendo quest'ultimo, settoriale, sono emerse come ci si poteva aspettare, alcune caratteristiche peculiari: dall'analisi dei titoli degli articoli, si è manifestata la predilezione della costruzione nominale, che, per immediatezza e impatto, risulta essere, per gli incipit, la forma più adatta alla comunicazione giornalistica.

L'assenza della forma verbale, fa prevalere il numero di sostantivi e di aggettivi, in modo che la funzione iconica e connotante del linguaggio sia espressa al massimo grado. A conferma di questo, abbiamo l'indice di variazione linguistica, calcolato come il rapporto tra il numero di sostantivi e di verbi, che è pari a 2.416.

In base ai patterns definiti, l'elaborazione della terminologia ha estratto, come combinazioni di termini ad alta forza di associazione, espressioni inerenti al sistema Moda, individuando capi, oggetti e stili-culto che, associandosi stabilmente ad un'epoca o ad un particolare contesto, entrano a pieno titolo nel patrimonio culturale e linguistico riconosciuto dalla collettività.

L'analisi dei nomi propri e in particolare quella degli stilisti, sfrutta la diacronia del Corpus, mostrando l'andamento delle loro occorrenze. In questo caso si è trattato di un'indagine più approssimativa che potrebbe, se si aumentasse la quantità di dati analizzati dagli strumenti del *Natural Language Processing* (NLP), condurre a risultati interessanti in grado di fornire informazioni sui dettami della moda.

Infatti, se la presenza più o meno assidua del nome di uno stilista su un giornale, ci suggerisse il suo grado di successo o di influenza del momento, allora le applicazioni della linguistica computazionale potrebbero essere utilizzate per le ricerche di tendenze del settore (*coolhunting*).

Per quanto riguarda il fenomeno della contaminazione linguistica, gli articoli che compongono *COMode*, sembrano seguire la linea di principio della politica francese contraria all'assorbimento dei forestierismi. Le percentuali degli anglicismi riconosciuti, anno per anno, sono piuttosto basse e si può notare, dai dati estratti, che i prestiti appartengano principalmente al mondo della Moda (abiti, tessuti, modi di dire).

Non a caso, è proprio il vocabolario di questo settore che, per la natura mutevole e multiforme della materia trattata, deve essere continuamente aggiornato; tuttavia, abbiamo osservato come, in alcuni casi, terminologie nate in determinati contesti si carichino di una particolare forza rappresentativa, in grado di garantirne l'immutabilità della forma espressiva. Il linguaggio, sembra essere l'unico mezzo in grado di assecondare ed esaltare i cambiamenti che stili e tendenze subiscono continuamente; la Moda, per potersi affermare ed esistere, ha infatti bisogno della dimensione temporale che solo la rappresentazione scritta le può conferire.

## Bibliografia

- Barthes, Roland. Sistema della Moda. La Moda nei giornali femminili: un'analisi strutturale. Traduzione a cura di Lidia Lonzi, Einaudi Editori, 1970.
- Blignaut, Hélène e Ciuni, Luisa. La comunicazione della moda. Significati e metodologie. Franco Angeli Editore, 2009.
- Corbucci, Gloria. La lingua della moda. Studi di glottodidattica. Università per stranieri di Perugia, 2008.
- De Saussure, Ferdinand. Corso di linguistica generale. A cura di Tullio de Mauro, Editori Laterza, 2005, pp.28-43, pp.85-88.
- Evans, David. Corpus building and investigation for the Humanities. University of Nottingham, 2007, Unit 2.
- J. Nivre, J. Hall, S. Kubler, R. McDonald, J. Nilsson, S. Riedel, D. Yuret. The CoNLL Shared Task on Dependency Parsing. In Proc. of CoNLL, 2007.
- Lenci, Alessandro, Montemagni, Simonetta e Pirrelli, Vito. Testo e computer. Elementi di linguistica computazionale. Carocci Editore, 2005.
- Lenci, Alessandro e Passaro, Lucia. Extracting terms with EXTra. University of Pisa, CoLing Lab, 2015.
- Principato, Aurelio. Breve storia della lingua francese. Dal cinquecento ai giorni nostri. Carrocci Editore, 2000.

## Sitografia

- Les Éditions Jalou, Paris 1921. Archives: <http://patrimoine.editionsjalou.com/lofficiel-de-la-mode-sommairepatrimoine-13.html> (Consultato il 15-12-2015).
- Talismane User's Manual. Version 2.3.1b: [http://redac.univ-tlse2.fr/applications/talismane/talismane\\_doc.html](http://redac.univ-tlse2.fr/applications/talismane/talismane_doc.html) (Consultato il 10-01-2016).
- Dizionario di lingua francese: <http://www.larousse.fr/dictionnaires/francais> (Consultato il 29-01-2015)

# Appendici

## Tagset di riferimento

Tag	Part of speech
ADJ	Adjective
ADV	Adverb
ADVWH	Interrogative adverb
CC	Coordinating conjunction
CLO	Clitic (object)
CLR	Clitic (reflexive)
CLS	Clitic (subject)
CS	Subordinating conjuncton
DET	Determinent
DETH	Interrogative determinent
ET	Foreign word
I	Interjection
NC	Common noun
NPP	Proper noun
P	Preposition
P+D	Preposition and determinant combined
P+PRO	Preposition and pronoun combined
PONCT	Punctuation
PRO	Pronoun
PROREL	Relative pronoun
PROWH	Interrogative pronoun
V	Indicative verb
VIMP	Imperative verb
VINF	Infinitive verb
VPP	Past participle
VPR	Present participle
VS	Subjunctive verb

## CoNLL Format Patterns per estrazione con EXTra

1	Chez	chez	P	P	
2	Vuitton	_	NPP	NPP	_
3	,	,	PONCT	PONCT	
4	la	la	DET	DET	n=slg=fl
5	nouvelle	nouveau	ADJ	ADJ	n=slg=fl
6	toile	toile	NC	NC	n=slg=fl
7	monogram	_	ADV	ADV	_
8	se	se	CLR	CLR	n=plp=3l
9	décline	décliner	V	V	n=slt=PSlp=13l
10	en	en	P	P	
11	kaki	kaki	NC	NC	n=slg=ml
12	.	.	PONCT	PONCT	

In ordine orizzontale sono date le seguenti informazioni:

- numero di token (inizia da 1 in concomitanza dell'inizio di una frase)
- forma originale della parola ( o “ \_ ” se si tratta di un token vuoto)
- lemma corrispondente trovato nel lessico ( o “ \_ ” se sconosciuto)
- part of speech tag
- categoria grammaticale corrispondente trovata nel lessico
- addizionali informazioni morfo-sintattiche trovate nel lessico tra cui:
  - g=m|f: genere = maschile or femminile
  - n=p|s: numero = plurale or singolare
  - p=1|2|3|12|23|123: persona = 1st, 2nd, 3rd
  - poss=p|s: numero del possesso = plurale or singolare
  - t=pst|past|imp|fut|cond: tempo verbale = presente, passato, imperfetto, futuro o condizionale.

## Patterns per estrazione con EXTra

['v','nc']	['nc','adj'],'p','npp']
['v','npp']	['nc','p','nc'],'p','npp']
['v',['npp'],'npp']]	['nc','p','nc'],'p','nc']
['nc','adj']	['nc','p',['nc','adj']]
['nc','vpp']	['adj','nc']
['nc',['adj'],'vpp']]	['adj',['nc','adj']]
['nc',['vpp'],'adj']]	['adj',['nc'],'npp']]
['nc','npp']	['adj','p','nc']
['nc','p','nc']	['adj','pd','nc']
['nc','pd','nc']	['vpp','nc']
['nc','p','npp']	['det','nc'],'p',['det','nc']]
['nc','adj'],'p','nc']	