ASAP / As Soon As Possible. Abbreviations in word formation: from form to meaning

Mauro Le Donne
Università per Stranieri di Perugia

Abstract
In morphology, specifically in word formation, it is often distinguished between grammatical and extra-grammatical word formation processes. The former, more salient and productive in natural languages, comprise derivation and compounding beside others; the latter, instead, include ephemeral and irregular formations. In this corpus-based study, it is proposed to analyse and outline structural, semantic and pragmatic features of three extra-grammatical processes of Italian, in which phonological reduction takes place, i.e., initialization, clipping and blending. In particular, a more drastic reduction involves initialisms, which often create an either in-group or out-group effect that depends strictly on the reader/hearer's expertise. Clippings undergo a minor degree of phonological reduction and represent often diaphasic and more informal variants of their full forms. Blends are lexemes generated by the conflation of two (or more) words, in which at least one is reduced. Sometimes the hybrid structure of the output mirrors a blended component of the referent expressed.
Keywords: word formation, extra-grammatical, initialisms, clippings, blends

In morfologia, più precisamente nella formazione delle parole, spesso viene effettuata una distinzione tra processi di formazione di parola grammaticali ed extra-grammaticali. I primi, più salienti e produttivi nelle lingue naturali, comprendono, fra gli altri, la derivazione e la composizione; i secondi, invece, includono formazioni effimere e irregolari. In questo studio *corpus-based* ci si propone di analizzare e delineare le caratteristiche strutturali, semantiche e pragmatiche di tre processi di formazione extra-grammaticali dell'italiano in cui la coniazione avviene mediante riduzione fonologica: l'inizializzazione, l'accorciamento e il *blending*. Nello specifico, una riduzione più drastica coinvolge la formazione degli inizialismi, che possono generare effetti di *in-group* o *out-group*, strettamente dipendenti dalla competenza del lettore/ascoltatore. Gli accorciamenti sono investiti da un minor grado di riduzione fonologica e rappresentano spesso delle varianti diafasiche delle forme estese di partenza. Infine, i *blend* sono lessemi generati dalla fusione di due (o più) parole. Alle volte, la struttura "mescolata" del lessema risultante riflette una componente ibrida nel referente espresso.
Keywords: formazione delle parole, extra-grammaticale, ritagli, inizialismi, parole macedonia

Word formation is a wide field of investigation in morphology which is not exhaustively covered by two fundamental instances such as composition and derivation. In particular, some studies group together word formation processes (from hereon WFPs) on account of a common property: phonological reduction (cfr. López Rúa 2002; Fandrych 2004, Thornton 2005, Mattiello 2012). In this study, three instances of word formation will be analysed: initialization, clipping and blending[1]. Specifically, it is proposed to analyse the features of such WFPs in Italian and to look at the distribution

---

1. Initialization, clipping and blending are by no means the only WFPs that involve phonological reduction. For instance, another WFP that involves reduction is back-formation, which will not be discussed in this paper. Examples of back-formation comprise deverbal nouns involving deletion of a suffix, e.g., It. n. *arrivo* 'arrival' < v. *arrivare* 'to arrive'.

of initialisms, clippings and blends inside different textual genres, assuming a corpus-based perspective. After this brief introduction, followed by a summary on theoretical and terminological issues (1), methodological choices are clarified (2) and features of each WFP by reduction highlighted (3). Subsequently, results of a small, corpus-based analysis of lexemes collected from the 2009 Supplement of the *Grande Dizionario della Lingua Italiana* (from hereon named GDLI 2009) are presented and discussed (4). Finally, some preliminary conclusions based on data are drawn (5).

I. Theoretical and terminological issues

Under the Structuralist and Generativist frameworks, WFPs by reduction are considered «non-grammatical» or «unusual coinages», therefore derogatorily labelled as «oddities» or simply «minor word formation processes» (Marchand 1969, 2f; Aronoff 1976, pp. 20-21; Scalise 1984, p. 98, note 1)[2]. Indeed, it has been repeatedly pointed out that blends, clippings and initialisms are non-analysable lexemes, at least in terms of full linguistic signs. For instance, Aronoff (1976) states clearly that «[t]he main characteristic of this type of word-formation is the fact that the meaning of a word formed by such a process can never be derived regularly» and that this irregularity makes them simply unproductive (1976, p. 21)[3]. In *English word-formation*, Bauer (1983) calls such phenomena «unpredictable», but he observes as well that «[a]s far as English is concerned, these formations are so common […] that it is misleading to consider them out of the ordinary» (1983, p. 232). In the Eighties, some attempts to accommodate WFPs by reduction into a more flexible morphological framework have been made. Under the Natural Morphology framework those instances fall into the domain of extra-grammatical morphology (cfr. Dressler *et alii* 1987; Dressler & Merlini Barbaresi 1994; Dressler 2000, 2005). Here, the notion "extra-grammatical" is employed to group those WFPs that violate different, universal preferences, i.e., "more natural" tendencies of world languages to express the association between *signans*

---

2. A detailed discussion on the debate around WFPs by reduction falls out the purpose of this paper. In this section we will mainly resume the main positions adopted in the literature. For a detailed theoretical discussion cfr. Mattiello (2013, cap. 2).

3. Aronoff's assumption includes blends like *smog* (*smoke* + *fog*), acronyms like *NATO* (< *North Atlantic Treaty Organization*, named *letter words* or *syllable words* in his account) (cfr. 1976, *ibid.*).

and *signatum* (Dressler 2005, p. 268)[4]. Nonetheless, a closer look reveals that each extra-grammatical WFP complies differently on this respect. For instance, taking into account the preference for *morphosemantic transparency*, initialisms, clippings and blends could show gradient degrees of opaqueness, as follows (Dressler 2005, pp. 271-272)[5]:

- initialisms reach a maximum degree of opaqueness in view of a more drastic shortening of their full form, which reduces the output to initial letters, e.g., It. *OPA/opa* (< *Offerta Pubblica di Acquisto* 'public tender offer') (cfr. López Rúa 2002, p. 35);
- clippings often present a truncated disyllabic output and sometimes the head of the input form remains unanalysable, e.g., It. *trilo* (< *trilocale* 'three-room apartment');
- complete blends retain both input forms, but show lowered transparency than compounds due to unexpected conflation of the lexical bases, e.g., It. *webete* 'a person who uses the web recklessly' (< *web* + *ebete* 'idiotic') (see section 3.3).

In the Italian tradition, these phenomena are often neglected or hastily discussed as pertaining to «minor morphology» (Scalise & Bisetto 2008, p. 207). Sometimes, extra-grammatical WFPs are discussed collectively in extensive monographic works on word formation or morphology *in toto* (Grossmann & Rainer 2004, Thornton 2005)[6]. In particular, Thornton (2005) groups initialisms, clippings and blends («*parole macedonia*», see below) classifying them as processes of word formation by reduction («*processi di formazione di parola per riduzione*», 2005, p. 140). In *La formazione delle parole in italiano* (Grossmann & Rainer 2004) clippings and initialisms are presented jointly in a chapter dedicated to instances of «reduction», while blends are discussed under the heading «*parole macedonia*» (lit. 'fruit-salad words') (Thornton 2004a, 2004b). This notion was firstly employed by Migliorini (1949) in a well-known passage that will be reported below (Migliorini 1949, p. 89; Thornton 2004b, p. 569):

In qualche caso una o più parole maciullate sono state messe insieme con una parola intatta: Cogepesca, Fedemetalli, ecc. Così si sono avuti successivamente il Cogefag, il Fabbriguerra, il Miproguerra (né saprei

---

4. The term "natural" refers to «cognitively simple, easily accessible (esp. to children), elementary and therefore universally preferred» (Dressler 2005, p. 167).
5. In this study, the letters which make up the initialism will be capitalized, while the part of each input form retained in the resulting output of blends and clippings will be underlined.
6. Here, initialization, clipping and blending will be often referred to as "extra-grammatical WFPs", notwithstanding, extra-grammatical morphology comprises other morphological phenomena, which are not discussed in this work.

dire quale, fra queste parole macedonia, sia la più orribile. La Sepral (Sezione Provinciale dell'Alimentazione) ci mostra che alle volte non si prende nemmeno una sillaba intera, ma un paio di lettere [...] (Migliorini 1949, p. 78)[7].

Prima facie, there seems to be consistent structural diversity between "fruit-salad words" and blends (cfr. Thornton 2004, p. 571)[8]. However, such terminological discrepancies in the Italian tradition occur often in the literature on extra-grammatical WFPs due to their non-discrete nature which would require a similar, flexible approach (cfr. Merlini Barbaresi 2007, pp. 39-40; Castagneto & Parente 2020, pp. 348-349)[9]. An example of non-discrete approach applied to categories of word formation is given by López Rúa (2002). In this study, the morphological category is conceived as a *continuum*, in which categorial members distribute themselves «according to their varying degrees of representativity» (López Rúa 2002, p. 33). The items that satisfy all categorial criteria or most of them are in close proximity to the core and thus "more representative" members of a given category, whereas items that do not conform to many of those features are closer to periphery and considered "less representative".

II. Methodology

This study proposes a corpus-based approach on Italian extra-grammatical WFPs based upon a scrutiny of the GDLI 2009. Each lexeme found in this dictionary, formed by initialization, clipping or blending has been included into a list[10]. The decision has

---

7. "In some cases, one or more mangled words have been put together with a full word: *Cogepesca*, *Fedemetalli*, etc. Thus, there have been later *Cogefag*, *Fabbriguerra*, *Miproguerra* (nor can I say which, among these fruit-salad words, is the most horrible. *Sepral* (*Sezione Provinciale dell'Alimentazione*) shows us that sometimes not even a whole syllable is taken, but a couple of letters [...]" (mine the translation).

8. Take, for instance, the difference between *Sepral* (see note 7 above), formed by syllables (*se*-), complex onsets (-*pr*-), and phonological segments (-*al*) vs. *rurbanizzazione* 'rurbanization' (< *rurale* 'rural' + *urbanizzazione* 'urbanization'), a blend in which there is overlap of two words around a common phonological string /*ur*/ (see section 3.3).

9. The non-discreteness approach to categories departs originally from the prototype theory developed by Eleanor Rosch in cognitive studies on the perception of form and colour (cfr. Rosch 1973; Ježek 2005, p. 78).

10. Criteria to include lexemes from the dictionary rely essentially on the definition presented below for each WFP (3.1, 3.2, 3.3). It must be noted though that especially in the case of lexemes included as blends, the etymological notes specified in the lemmas' entries in the GDLI 2009 were not always uniform. For instance, given A and B, or AB and

been made to keep in the list also non-native items, in order to understand if and to which extent foreign languages play a role in the spreading of such lexemes in Italian. Each item has been then analysed according to different categorial parameters, most of which are specific for the three morphological categories taken into account. Moreover, a small quantitative analysis has been carried out, using the Perugia Corpus (from hereon named PEC; cfr. Spina 2014)[11]. This has been done on the basis of the occurrences found in the PEC for each item, along with the lexeme distribution in different texts (i.e., the number of texts in which it occurred) and in diverse textual genres. The PEC corpus has been chosen primarily in view of this latter criterion, and it is structured as follows:

- • <u>written textual genres</u>, namely, texts from literature, essays, newspapers, academic texts, documents related to school, administration and the web;

- • <u>oral textual genres</u>, namely, tv dialogues, film scripts and spoken texts (conversations, songs, speeches, conferences and lessons).

It must be noted though, that despite being very well-balanced, the PEC corpus is small in size, if compared with other corpora of Italian, therefore a restricted usage of many items found in the GDLI 2009, for instance, obsolete or old-fashioned terms, resulted often in no occurrence at all[12].

III. Extra-grammatical WFPs

In the following subsections, the features of three extra-grammatical process of word formation, that is, initialization (3.1), clipping (3.2) and blending (3.3), will be briefly sketched. All the examples are drawn out from the scrutiny of the GDLI 2009 (see section 2),

---

CD as input words, one could find divergent entries, such as: *incrocio di A con B* ("cross between A and B"; *da A con sovrapposizione di B* ("from A with overlap of B"); *deriv. da A, sul modello di B* ("derived from A, on the model of B"); *comp. da A[B] e [CD]* ("composed of A[B] and [CD]"). A similar observation is made in Castagneto & Parente (2020, p. 348-349).

11. The PEC (Perugia Corpus) is a corpus of written and spoken Italian made up between 2011 and 2012, it is composed of 26 million words distributed in 10 sections corresponding to different textual genres. It is not a big corpus (26.487.716 tokens, 299.813 types, 41.401 texts), but it is very well balanced and representative of diverse written and spoken Italian varieties. More information on the PEC can be found in Spina (2014).

12. An example of big corpus for the Italian language is the ItTenTen20, a web corpus from the family of TenTen corpora; it contains 12.4 billion of words (cfr. Jakubíček *et alii* 2013).

when not otherwise specified.

III.I. Initialization

Initialization is the process whereby new lexemes, namely *initialisms*, are coined picking up initial letters from an underlying syntagm, phrase, title, compound, or list (Fandrych 2008, p. 108; Mattiello 2013, pp. 82-83)[13]. Traditionally, initialisms are subdivided into the following two main categories (López Rúa 2004, p. 124):

- *acronyms*, i.e., words composed of initial letters with an orthoepic pronunciation, e.g., Eng. *laser* /ˈleɪzəʳ/ < *Light Amplification by Stimulated Emission of Radiation*; It. *Onlus* /ˈɔnlus/ < *Organizzazione Non Lucrativa di Utilità Sociale* 'NPO of social utility';

- *alphabetisms*, i.e., words formed by initials and spelled out letter-by-letter, e.g., Eng. *BBC* /ˌbiːbiːˈsiː/ < *British Broadcasting Corporation*; It. *OGM* /ˌodˌʤiˈɛmme/ < *Organismo Geneticamente Modificato* 'GMO'.

As shown above, the main difference between alphabetisms and acronyms is in pronunciation, which is phonological for the former and alphabetical for the latter[14]. Moreover, acronyms may be further distinguished into another subcategory:

- *syllabic acronyms*, i.e., acronyms formed by more than one letter, often syllables or syllabic constituents, e.g., Eng. *Nabisco* /nəˈbɪskəʊ/ < *NAtional BIScuit COmpany*; It. *co. pro* /koˈpro/ < *COntratto a PROgetto* 'project contract'.

Phonological and orthographical motivations characterize peripheral instances of initialisms that present the following features (in **a**) to increase readability and memorization:

a.   alternative spellings, as It. *C.T.* vs. *CT* vs. *ct* vs. *cittì* (see note 14);

unexpected readings, as It. *FGCI* /ˌfidˌʤitˈʧì/ (< *Federazione -Giovanile*

---

13. Sometimes, the terms *alphabetism* or *acronym* are employed to refer to the whole category of initialisms (Cannon 1989, pp. 106-107; cfr. Thornton 2004a, p. 557). In this study, we decided to use the term *initialization* to name the process rather than its product (López Rúa 2004, p. 124). Nonetheless, this usage is not common in the literature (but cfr. Renner 2020). In Italian, the term *sigla* usually refers to both acronyms and alphabetisms (Thornton 2004a, p. 557; but cfr. Merlini Barbaresi 2007, p. 40).

14. Indeed, alphabetisms rely on spelling rather than on phonemes of their full form, otherwise, in *cittì* /ʧitˈti/ (< *Commissario Tecnico* 'head coach'), the first phoneme would have been /k/ rather than /ʧ/ (cfr. Thornton 2004a, pp. 559-560).

Comunista Italiana 'Italian young communist federation'), in which /fi/ replaces the original Italian spelling /effe/;

<u>inclusion of non-initial letters</u>, as Eng. *Erasmus* (< *EuRopean community Action Scheme for the Mobility of University Students*);
• <u>constituent inversion</u>, as Eng. *MISHAP* (< *Missiles High-Speed Assembly Program*);

• <u>doubly-motivated acronyms or acrostics</u>, as It. *LUCE* < *L'Unione Cinematografica Educativa* 'the educational film union', in which the acrostic is homophone of It. *luce* 'light' and shares a semantic affinity with it (cfr. Fandrych 2004, p. 21; Merlini Barbaresi 2007, p. 40)[15].

III.II. Clipping

Clippings are lexemes formed by truncation of an existing base, generally maintaining the same denotation of the full form, but conveying a different stylistic and pragmatic connotation[16]. Like initialisms, clippings are not to be considered new words *stricto sensu*, in facts, from a semantic perspective, they are often connoted "doublets" of their base lexemes, attested mostly in diaphasic, non-standard varieties (Montermini 2002, p. 310; Fandrych 2004, p. 31; Thornton 2004a, p. 561; Bauer 2006, p. 498).

Truncation of the base may act somewhat randomly as it could affect different word parts, as highlighted in **b**[17]:

b. right-hand clipping: Eng. *champ* < <u>*champ*ion</u>;

• left-hand clipping: Eng. *phone* < <u>*tele*phone</u>;

• ambi-clipping[18]: Eng. *fridge* < *re*<u>*frig*</u>*erator*;

• central clipping: Eng. *Jo'burg* < <u>*Johannes*burg</u>

In structural terms, despite the unexpected output, a large number of Italian clippings conform to the model of the minimal prosodic word viable in this language and firstly detected by Thornton

---

15. All examples except It. *C.T.* are drawn out from Thornton (2004a, p. 559), Brinton & Traugott (2005, p. 42), Fandrych (2008, p. 109).
16. In the Italian tradition it is generally employed the term *accorciamento* 'shortening', which dates back to Migliorini (1957). Montermini (2002) adopts the term *apocope*, admitting its non-traditional usage (2002, p. 305, note 1). Clipped proper nouns, or *hypocoristics*, usually discussed with clippings, will not be further discussed in this study.
17. Examples are drawn out from Fandrych (2004, p. 45).
18. The term has been borrowed from Bauer (2006, p. 499).

(1996): a disyllabic trochaic foot, i.e., «a foot of two syllables with stress on the first one», ending in vowel (1996, p. 83)[19]. Following the parameter of syllabic retention, Italian clippings could be grouped in three main categories, as shown in the table below (**fig. 1**) (similarly in Thornton 1996, 2004a and Montermini 2002):

monosyllabic	*veg* < *vegetarian*
	nick < *nickname*
disyllabic	*bio* < *biologico* 'biologic'
	rasta < *rastafariano* 'Rastafarian'
trisyllabic	*maria* < *marijuana*
	perquisa < *perquisizione* 'frisking'

(**fig. 1**)

According to the minimal prosodic word constraint viable in Italian, within a non-discrete perspective of word formation, disyllabic clippings should be regarded as the nearest members to the categorial core of clipping (cfr. López Rúa, 2002; Montermini, 2002: 316). Moreover, right-hand truncation appears to be the most common solution in Italian.

> However, as other instances of extra-grammatical morphology, the periphery of clipping contains lexemes that display non-prototypical morphosyntactic and semantic features, as shown in **c** below[20]:
>
> **c.** clipped compounds, i.e., compounds formed by at least one clipped constituent, e.g., It. *netdipendenza* 'web addiction' < *Internet* + *dipendenza* 'addiction'[21];
> transcategorization via truncation, i.e., change of lexical class between the full input form and the clipped output, e.g., It. n., *meteo* < adj. *meteorologico* 'metereological'[22];
> semantic ambiguity, i.e., creation of homophonic pairs with different

---

19. The notion of *minimal prosodic word* was originally developed under the Prosodic Morphology framework (McCarthy & Prince 1986, 1990). Very basically, the prosodic word is an Abstract template scheme which specifies the minimal size that a word must display to be considered as such from the speakers of a language.
20. The last example in the list is drawn out from Brinton & Traugott (2005, p. 40).
21. Many scholars are doubtful whether to account those items as pertaining to clippings or to blends (Bauer 1983, p. 233; Fandrych 2004, p. 32). In her study, Beliaeva (2014) addresses this question and highlights differences between the two categories in English, pointing out, for instance, that clipped compounds originate as "contractions of existing compounds", while blends imply "the formation of new notions in the process of conceptual integration" (Beliaeva 2014, p. 51).
22. Eventually, the output may also develop new lexical functions, as in It. n./adj. *bio* < adj. *biologico* 'biologic'.

semantic denotation, e.g., It. *clima* (< *climatizzatore* 'air conditioner') vs. *clima* 'climate'; It. *bio* (< *biologico*) vs. *bio* (< *biografia* 'biography')[23];
semantic dissociation, i.e., the process whereby clippings acquire a new denotation, e.g., Eng. *fan* < Eng. *fanatic* (cfr. Fandrych 2004, p. 31; 2008, p. 114).

III.III. Blending

Blends are lexemes generated by the conflation of two (or more) lexical bases, in which at least one is reduced (Mattiello 2019, p. 3; Castagneto & Parente 2020, pp. 353-354)[24]. Unlike initialisms and clippings, blends are iconic and eye-catching formations that often convey a new, non-compositional meaning, as in *vigoressia* 'bigorexia' (< *vigore* 'vigour' + *oressia* 'desire, appetite'). The common phonological string */-ore-/*, traditionally named *overlap*, is preserved in the output form (cfr. Algeo 1977, pp. 48-49; Mattiello 2019, p. 3). As already mentioned, morphosemantic transparency depends crucially by different structural blending patterns, indeed, retention of one, both source words, or neither of them is possible (cfr. Ronneberger-Sibold 2006)[25]. A synthetic typology will be presented in the table below (**fig. 2**)[26]:

fragment	*merinozio* 'midnight' < Lat. *měrīdǐēs* 'midday' + *equinozio* 'equinox' *botox* < *botulinum* + *toxine*
semi-complete	*archistar* 'starchitect' < *architetto* 'architect' + *star* *webjay* < *web* + *deejay*
complete	*calciotto* '8-a-field football' < *calcio* 'football' + *otto* 'eight' *vigoressia* < *vigore* + *oressia*

23. In this case, clippings violate the preference for biuniqueness identified under the Natural Morphology framework (Dressler 2005, p. 274).
24*. Source word* is the most frequent term used in literature on blending to refer to blend constituents (cfr. Algeo 1977; Fandrych 2004; Beliaeva 2014; in It. *parole fonte*, cfr. Castagneto & Parente 2020). Here, source words will be referred to as W1 and W2 if called into question singularly. In *motel*, *motor* and *hotel* are the source words of the resulting blend.
25. Lehrer (2007) includes «the number and percentage of letters (or phonemes) of the source word [...]» present in the output among the criteria for a correct decoding of blends (Lehrer 2007, p. 126).
26. Morphosemantic transparency is linked as well to some kind of semantic plausibility between the two referents. For instance, *apericena* (< *aperitivo* + *cena*) could be viewed as a more transparent formation than *metrosessuale* (< *metropolitano* + *sessuale*), in view of the semantic proximity between the two source words. In the following typology, terminological choices are borrowed from Ronneberger-Sibold (2006).

It has been previously argued that semi-complete Italian blends retaining the second source word fully are more prototypical, while other patterns may be seen as more distant to the categorial core of blending (Thornton 2004b, p. 571)[27]. The word parts retained in the resulting word are traditionally called *splinters* (Berman 1961, p. 279; cfr. Fandrych 2004)[28]. More recently, Mattiello (2019) has pointed out how analogy could play an important role on blend formation, as we find blends based upon models of existing lexemes, e.g., *Nollywood* (< *Nigeria* + *Hollywood*), modeled after *Bollywood* (< *Bombay* + *Hollywood*); or upon the spreading of lexical series through reanalysis of splinters, e.g., the It. combining form *fanta-*, etymologically derived from the left-hand splinter of the blend *fantascienza* 'sci-fi' (< *fantasia* 'fantasy' or *fantastico* 'fantastic' + *scienza* 'science'), which has given rise to several new neologisms, for instance, *fantaeconomia* 'fictional economy' or *fantafilm* 'sci-fi film' (Mattiello 2019, pp. 19-22). Semantically, blends are classified following the terminology used in compounding (Castagneto & Parente 2020). Here, the typology found in Scalise & Bisetto (2008) has been employed in reference to blends, distinguishing among: *coordinative blends*, in which the two source words are co-hyponyms, as in *aeroboxe* 'sport combining aerobics with boxing' (< *aerobica* 'aerobics' + *boxe* 'boxing'); *subordinative blends*, in which an underlying complement specifies the semantic relationship between the two source words, for instance, *elisoccorso* 'rescue by means of a helicopter' (< *elicottero* 'helicopter' + *soccorso* 'rescue'); *attributive blends* like *liquiletame* 'liquid manure' (< *liquido* 'liquid' + *letame* 'manure'), in which W1 conveys an attribute of W2 (Scalise & Bisetto 2008, pp. 130-131).

To conclude the brief outline of these three extra-grammatical WFPs, a concise presentation of peripheral items pertaining to blending will be presented in d, as follows:

> d. minor structural patterns, as *multiple blends*, formed by more than two source words, e.g., *acesulfame* 'acesulfame-K' < *acetico* 'acetic' + *sulphur* + *lattame* 'lactame'; and *interpolated blends*, in which W2 is inserted

27. Some authors argue that semi-complete blends which retain the second source words are not prototypical (Thornton 1993, pp. 147-148). In this study, they are considered so as far as Italian is our language of reference, in view of the fact that they are more numerous than lexemes following other structural patterns (cfr. the predominance of this pattern in Bertinetto 2001; Castagneto & Parente 2020).
28. In Italian, Migliorini (1949) has named them *tronconi*, lit. 'stumps' (Migliorini 1949, p. 86; cfr. Thornton 2004b, p. 569, note 2). In the English blend *brunch*, the splinters are *br-* and *-unch*.

into W1, e.g., *itangliano* 'Italian variety mixed up with English words' < *italiano* 'Italian' + *inglese* 'English';

• <u>constituent inversion</u>, as in *ligre* 'liger' (< *leone* 'lion' + *ti<u>gre</u>* 'tiger') vs. *tigone* 'tigon' (*tigre* + *le<u>one</u>*)[29];

• <u>semantic ambiguity</u>, as splinters may coincide with homophonic combining forms or affixes, e.g., *birbocrazia* 'rascal democracy' < *<u>birbo</u>* 'rascal' + *dem<u>ocrazia</u>* 'democracy', *archistar* 'starchitect' <*<u>archi</u>tetto* 'architect' + *<u>star</u>*, *pervaporazione* 'pervaporation' <*<u>per</u>meabilità* 'permeability' + *e<u>vaporazione</u>* 'evaporation';

• <u>intra-word code-switching</u>, as source word may pertain to two different languages, e.g., It. and Eng., in *folktronica* (Eng. *folk* + *elet<u>tronica</u>* 'electronic') and *risto-book* (*<u>risto</u>rante* 'restaurant' + Eng. *<u>book</u>*); It. and Lat., in *resveratrolo* 'resveratrol' (< *resorcin<u>olo</u>* 'resorcinol' + Lat. *<u>veratr</u>um* 'hellebore'), *merinozio* (< Lat. *mĕrīdĭēs + equi<u>nozio</u>*) (McArthur 1998).

IV. Analysis

The GDLI 2009 scrutiny has led to the compilation of a list containing 192 items, among which we find 107 blends (55.7%), 47 clippings (24.4%) and 38 initialisms (19.7%). The non-native lexemes are 90 (46.8%) in total: 45 blends (42%), 26 initialisms (68.4%) and 19 clippings (40.4%). Thus, lexical borrowing, especially from English as source-language, carves out an important part in the spreading of extra-grammatical formations in Italian[30].

The data outlined will be highlighted in the graph below (**fig. 3**):
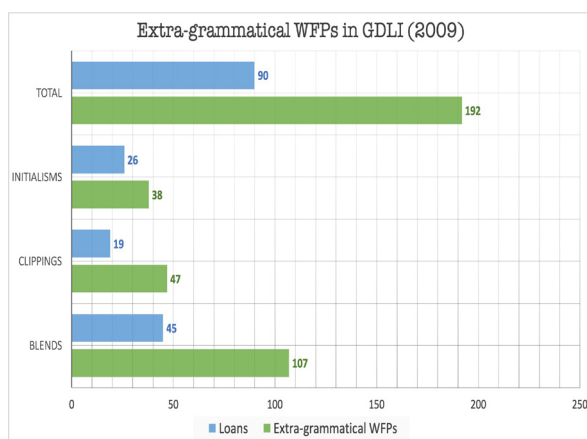
29. The example, *tigone*, is drawn out from Thornton (2004b, p. 571).
30. It must be noted though that lexical borrowing can affect other domains of word formation as well, indeed, the emergent productivity of right-headed compounds in Italian is significant on this respect (cfr. Iacobini 2014).

A more detailed computation of the subclasses identified before (see sections 3.1, 3.2, 3.3) will be shown in the subsequent table:

initialization (38/192)	alphabetisms (20/38): *OGM*	
	acronyms (13/38): *Onlus*	syllabic acronyms (5/38): *co.co.pro.*
clipping (47/192)	monosyllabic (8/47): *veg*	
	disyllabic (36/47): *corto*	
	trisyllabic (3/47): *perquisa*	
blending (107/192)	fragment (31/107): *fantastiliardo*; *botox*	
	semi-complete (66/107): *aromacologia*; *apericena*	
	complete (3/107): *calciotto*	
	multiple (3/107): *acesulfame*	
	interpolated (4/107): *itangliano*	

fig. 4

As it could be noted, alphabetisms (20/38 – 52.6%) are the most common subclass inside the category of initialisms, whereas 37 disyllabic clippings (76.5%) prevail among other prosodic typologies. For what concerns blending, the two major structural patterns of Italian blends in our list are:

> • semi-complete blends that retain W2 fully, e.g., *apericena* 'a happy hour that can replace dinner' (see note 26) (46/66);
> • fragment blends formed by an initial and a final splinter, e.g., *fantastiliardo* 'zillion' (< *fantastico* 'fantastic' + *miliardo* 'billion') (25/31)[31].

Before commenting our data occurrences inside the PEC corpus (Spina 2014), it could be useful to mention once again two brief considerations: the size of the corpus (see note 11) proved not to be ideal for an analysis which takes into account neologisms and/or lexemes which look rather obsolete or restricted in usage at the present time; at the same time though, the corpus subdivision in different textual genres consented to investigate in which texts such lexemes are more attested or, conversely, disfavoured. The subsequent graph will show in which diaphasic variety blends, clippings

31. Semi-complete blends that retain W1 fully are also quite numerous (20/66). Other minor patterns of fragment blending have been registered as well, but they won't be discussed further as they constitute minor tendencies.

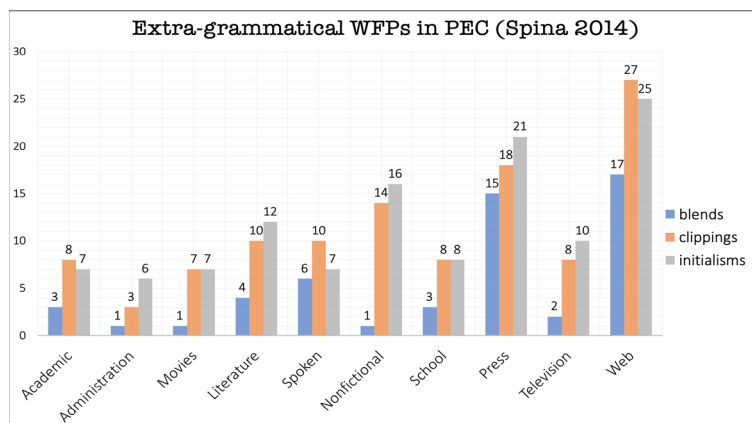and initialisms extracted from the GDLI 2009 are attested (**fig. 5**):

fig. 5

In terms of textual distribution and frequency of occurrence, in the PEC corpus: 29/38 initialisms are attested, with *TV* as the most frequent initialism in terms of global number of occurrences (2919 occ.) and textual distribution (1419 texts); 30/47 clippings are attested, with *auto* as the most frequent clipping in terms of global number of occurrences (2843 occ.) and textual distribution (1123 texts); 30/107 blends are attested, with *cantautore* as the most frequent blend in terms of global number of occurrences (159 occ.) and textual distribution (77 texts)[32].

The usage of lexemes formed by extra-grammatical WFPs on the web clearly emerges as general trend. This is not surprising since the presence of lexemes related to technology or to the web itself. For example, concerning non-translated loanwords, we have listed the following items: *ADSL*, *URL*, *USB* and *IAD* (< *Internet Addiction Disorder*) among initialisms; clippings such as *blog* and *nick* (< *nickname*); blends formed by the *-ware* splinter (*adware*, *freeware*, *pornoware*, *spyware* in our corpus), one fragment blend formed by two final splinters, as *netizen* (< *internet* + *citizen*) and two analogical fragment blends, like *vlog* (< *video* + *blog*) and *webjay*.

As it may be noted, the number of blends attested in most textual genres is very low. Indeed, many of them are neither neologisms, nor of current usage, take, for instance, *merinozio* (see above) or *gaglioffardo* 'scoundrel' (< *gaglioffo* 'clumsy' + *infingardo* 'treacherous'); other items, instead, may have a very restricted usage, e.g.,

---

32. Looking at the global number of occurrences for each lexeme inside the corpus implies that feminine and plural forms (e.g., *cantautori*, *cantautrice*, *cantautrici*) and spelling variants (*tv*, *tivù*, *tivvù*) have been counted as well.

skank < *ska* + *punk*, *ferrotel* 'hotel that host railwaymen on assignment' < *ferrovia* 'railway' + *hotel*). Moreover, technicisms, as *pervaporazione* or *acesulfame* (34/107 – 31.7%), are only scarcely attested (13/107).

The fact that many clippings of our corpus pertain to non-standard varieties (35/47 – 74.4%) could reflect their informality in varieties such as spoken and Internet Italian, in which clippings outweigh other extra-grammatical WFPs[33]. At a closer look though, in the PEC corpus only occurrences of 18 diaphasically-marked clippings are registered, whereas the 12 clippings that pertain to standard Italian are all attested[34]. In facts, the presence of lexicalised items, like *auto*, *bici* and *fax*, together with lexemes of common usage, as *meteo* and *blog*, accounts for the wider distribution of this latter subclass. As expected, initialisms tend to prevail in written varieties of Italian, such as literary, nonfictional texts and newspapers. In comparison to other extra-grammatical WFPs, the presence of initialisms in administrative texts is pervasive in comparison with other extra-grammatical categories, in particular, with clippings, which are probably too colloquial to occur massively in this textual genre[35]. However, initialisms occur more than other categories in the special language of television (spoken language). If this tendency may surprise at first sight, it could be explained by the fact that many items in our corpus refer to social, political or economic issues, often discussed in news reports, such as *c.t.* (see note 10), *OGM* (see above), the French calque *Pacs* (< *PAtto Civile di Solidarietà* 'civil solidarity pact'), the loanword *Nasdaq* (< *National Association of Securities Dealers Automated Quotation*); whereby others are now of common general usage, for example, *sms* (< *Short Message Service*), *tv* (< *TeleVision*), *dj* (< *Disc Jockey*). Initialisms are thus not restricted exclusively to written varieties of Italian as they are attested in the spoken language as well.

---

33. Indeed, 10 clippings are attested in spoken Italian, whereas only 7 initialisms and 6 blends appear, while 27 clippings are registered in the Italian web variety along with 25 initialisms and 17 blends.

34. Among non-standard clippings there are not-attested items pertaining to very sector-specific varieties, like *mono* (< *monolocale* 'one-room apartment'), *bilo* (< *bilocale* 'two-room apartment') and *trilo* (< *trilocale*), typical of estate operators' jargon and only one, e.g., *garga* (< *gargagnano* 'pimp'), which is a clipping of a lexeme pertaining to a diatopic variety, specifically, the Turinese dialect. The diaphasic markedness is sometimes reported on the GDLI 2009, though for some items it was necessary to check the *Nuovo vocabolario di base della lingua italiana* (De Mauro 2016).

35. Among the 10 lexemes attested in bureaucratese we find: 6 initialisms (60%), 3 clippings (30%) and 1 blend (10%).

V. Conclusions

In this study, we have tried to analyse synthetically three extra-grammatical WFPs, namely, initialization, clipping and blending, which all involve *brevitas* on the word level, that is, phonological reduction. Brief, structural and semantic consideration on each WFP has been made, adopting a non-discrete view to categories of word formation. A list of lexemes formed by initialization, clipping and blending has been retrieved from a preliminary survey of the GDLI 2009 and a computation of the occurrences of each lexeme in this list inside the PEC corpus (Spina 2014) has been made. The scrutiny has confirmed that disyllabic clippings as well as semi-complete and fragment blends could be considered nearest to the categorial core of clipping and blending, at least in terms of sheer numbers. For what concerns initialisms, instead, alphabetisms are the most common subclass in the GDLI 2009. Passing on to the textual analysis, the three extra-grammatical WFPs under scrutiny are attested above all in Internet Italian and in newspapers.

This is not surprising, since, for instance, the abbreviatory nature of initialisms and clippings enable editors and authors to save space and characters. In facts, it has been pointed out that such devices respond to a universal principle of linguistic economy, which consent to employ a shortened lexeme maintaining its original referential function at the same time (cfr. Merlini Barbaresi 2007, pp. 42-43; Mattiello 2012, pp. 158-159). As we have seen, clippings and acronyms could be often morphosemantically opaque. In such cases, to disambiguate the referent expressed a common background of encyclopaedical knowledge between encoder and decoder is needed. Blending is employed to form lexemes with an unexpectedly-assembled shape in order to gain the reader/hearer's interest. For this reason, it is not difficult to find them in journal headings, especially in sport dailies. Initialization, clipping and blending are instances of word formation that are not marginal, as such words surround our daily life. English plays certainly a certain role in the spreading of extra-grammatical lexemes in Italian, though more data would be needed to state if this role is crucial or collateral. However, Italian blends and clippings appear to diverge from English ones, at least in structural terms, displaying patterns which are registered, but not predominant in English blends.

In this study the decision to treat extra-grammatical WFPs together inevitably led to a summary analysis, which must be carried further in more detail. GDLI 2009 has proved to be a source of data

which is not sufficient for such an analysis, indeed dictionary and databases of neologisms would suit best to this purpose and may add further evidence of these instances of word formation. From a corpus linguistics perspective, bigger web corpora would guarantee many more data and evidence on such phenomena. Nonetheless, many web corpora may be fuzzy and not always clear-cut as the PEC corpus is, an improvement from this point of view would be a *desideratum*. A suggestion for further improvements of research, then, addresses also the question of instruments currently available for corpus-based enquiries in Italian, which are still neither enough nor ideal.

References

Algeo J., *Blends, a Structural and Systemic View*, in «American Speech», 52(1/2), 1977, pp. 47-64.

Aronoff M., *Word Formation in Generative Grammar*, Cambridge (Mass.), London, MIT Press, 1976.

Battaglia S. B. (a cura di), *Grande Dizionario della Lingua Italiana*, Torino, Utet, 2002.

Bauer L., *English word-formation*, Cambridge, Cambridge University Press, 1983.

Bauer L., *Compounds and Minor Word-formation Types*, in Aarts B. M. (a cura di), *The Handbook of English Linguistics*, Malden, Oxford, Victoria, Blackwell, 2006.

Beliaeva N., *A study of English blends: From structure to meaning and back again*, in «Word Structure», 7(1), 2014, pp. 29-54.

Berman J. M., *Contribution on Blending*, in «Zeitschrift für Anglistik und Amerikanistik», 9, 1961, pp. 278-281.

Bertinetto P. M., *Blends and syllabic structure: A four-fold comparison*, in Lorente M., Alturo N., Boix E., Lloret M.-R., Payrató L. (a cura di), *La gramàtica i la semàntica en l'estudi de la variació*, Barcelona, Promociones y Publicaciones Universitarias, 2001.

Brinton L. J., Traugott E. C., *Lexicalization and Language Change*, Cambridge, Cambridge University Press, 2005.

Cannon G., *Abbreviations and Acronyms in English Word-Formation*, in «American Speech», 64(2), 1989, pp. 99-127.

Castagneto M., Parente E., *Ti va una apericena al ristobar? I blend in italiano come lessicalizzazioni complesse*, in Valenti I. (a cura di), *Lessicalizzazioni "complesse" ricerche e teoresi*, Roma, Aracne, 2020.

De Mauro T., *Nuovo Vocabolario di Base della Lingua Italiana*, 2016. Tratto da Internazionale: dizionario.internazionale.it/nuovovo-cabolariodibase.

Dressler W. U., Mayerthaler W., Panagl O., Wurzel W. U. (a cura di), *Leitmotifs in Natural Morphology*, Amsterdam, Philadelphia, John Benjamins, 1987.

Dressler W. U., Merlini Barbaresi L., *Morphopragmatics*, Berlin, New York, Mouton de Gruyter, 1994.

Dressler W. U., *Extragrammatical vs. marginal morphology*, in Doleschal U. T., Thornton A. M. (a cura di), *Extragrammatical and Marginal Morphology*, München, Lincom Europa, 2000.

Dressler W. U., *Word-formation in Natural Morphology*, in Štekauer P. L. (a cura di), *Handbook of Word-formation*, Dordrecht, Springer, 2005.

Fandrych I. M., *Non-Morphematic Word-Formation Processes: A Multi-Level Approach to Acronyms, Blends, Clippings and Onomatopoeia* (Ph.D. thesis), Bloemfontein, University of the Free State, 2004.

Fandrych I. M., *Submorphemic elements in the formation of acronyms, blends and clippings*, in «Lexis: Lexical Submorphemics», 2, 2008, pp. 103-121.

Grossmann M., Rainer F. (a cura di), *La formazione delle parole in italiano*, Tübingen, Max Niemeyer, 2004.

Iacobini C., *Composizione con elementi neoclassici*, in Grossmann M., Rainer F. (a cura di), *La formazione delle parole in italiano*, Tübingen, Max Niemeyer, 2004.

Iacobini C., *A constructionist account of the emergence of a new lexeme-formation pattern: Italian right-headed verbal compounds*, in *Carnets de Grammaire n°22 - Proceedings of the Décembrettes 8th International conference on morphology*, Toulouse, *CLLE-ERSS*, 2014.

Jakubíček M., Kilgarriff A., Kovář V., Rychlý P., Suchomel V., *The TenTen corpus family*, in Hardie A., Love R. (a cura di), *7th International Corpus Linguistics Conference CL*, Lancaster, UCREL, 2013.

Ježek E., *Lessico. Classi di parole, strutture, combinazioni*, Bologna, Il Mulino, 2005.

Lehrer A., *Blendalicious*, in Munat J. (a cura di), *Lexical Creativity, Texts and Contexts*, Amsterdam, Philadelphia, John Benjamins, 2007.

López Rúa P., *On the structure of acronyms and neighbouring categories: A prototype-based account*, in «English Language and Linguistics», 6(1), 2002, pp. 31-60.

López Rúa P., *Acronyms & Co.: A typology of typologies*, in «Estudios Ingleses de la Universidad Complutense», 12, 2004, pp. 109-129.

Marchand H., *The categories and types of present-day English word-formation*, München, C. H. Beck, 1969.

Mattiello E., *Abbreviations in English and Italian scientific discourse*, in «ESP Across Cultures», 9, Santo Spirito, Edipuglia, 2012, pp. 149-168.

Mattiello E., *Extra-grammatical morphology in English*, Berlin, Boston, de Gruyter, 2013.

Mattiello E., A *corpus-based analysis of new English blends,* in «Lexis: Blending in English», 14, 2019, pp. 1-28. DOI: https://doi.org/10.4000/lexis.3660.

McArthur T., *Code-mixing and Code-switching*, in McArthur T. (a cura di), *Concise Oxford Companion to the English Language*, Oxford, Oxford University Press, 1998. Retrieved on September 3, 2022 from Encyclopedia.com: https://www.encyclopedia.com/humanities/encyclopedias-almanacs-transcripts-and-maps/code-mixing-and-code-switching.

McCarthy J., Prince A., *Prosodic Morphology* (Unpublished manuscript), Amherst, Brandeis University, University of Massachusetts, 1986.

McCarthy J., Prince A., *Foot and word in Prosodic Morphology: The Arabic broken plural*, in «Natural Language and Linguistic Theory», 8, 1990, pp. 209-283.

Merlini Barbaresi L., *Il fenomeno delle sigle: Un primo schema di analisi*, in Garzone G., Salvi R. (a cura di), *Linguistica, linguaggi specialistici, didattica delle lingue: studi in onore di Leo Schena*, Roma, CISU, 2007.

Migliorini B., *Uso ed abuso delle sigle*, in Migliorini B. (a cura di), *Conversazioni sulla lingua italiana,* Firenze, Le Monnier, 1949.

Migliorini B., *I nomi maschili in -a*, in Migliorini B. (a cura di), *Saggi linguistici*, Firenze, Le Monnier, 1957.

Montermini F., *Formazione delle parole e variazione sociolinguistica: il caso dell'apocope in italiano*, in Bauer R., Goebl H. (a cura di), *Parallela IX. Testo – variazione – informatica / Text – Variation – Informatik*, Weinheim, Gottfried Egert, 2002.

Renner V., *An ecosystem view of English word-formation*, in «The Mental Lexicon», 15(1), 2020, pp. 4-20.

Ronneberger-Sibold E., *Lexical Blends: Functionally Tuning the Transparency of Complex Words*, in «Folia Linguistica», 40 (1/2), 2006, pp. 155-181.

Rosch E. H., *Natural Categories*, in «Cognitive Psychology», 4(3),

1973, pp. 328-350.

Sanguineti E. (a cura di), *Grande Dizionario Della Lingua Italiana. Supplemento 2009*, Torino, Utet, 2009.

Scalise S., *Generative morphology*, Dordrecht, Foris, 1984.

Scalise S., Bisetto A., *La struttura delle parole*, Bologna, Il Mulino, 2008.

Spina S., *Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione*, in Basili R., Lenci A., Magnini B. (a cura di), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014*, 1, Pisa, Pisa University Press, 2014.

Thornton A. M., *Italian blends*, in Tonelli L., Dressler W. U. (a cura di), *Natural morphology. Perspectives for the Nineties*, Padova, Unipress, 1993.

Thornton A. M., *On some phenomena of prosodic morphology in Italian: «accorciamenti», hypocoristics and prosodic delimitation*, in «Probus», 8(1), 1996, pp. 81-112.

Thornton A. M., *Riduzione*, in Grossmann M., Rainer F. (a cura di), *La formazione delle parole in italiano*, Tübingen, Max Niemeyer, 2004a.

Thornton A. M., *Parole macedonia*, in Grossmann M., Rainer F. (a cura di), *La formazione delle parole in italiano*, Tübingen, Max Niemeyer, 2004b.

Thornton A. M., *Morfologia*, Roma, Carocci, 2005.