

Stories of Words, Words as Stories.

Some lexico-statistically based Reflections on the Meaning Unit in Spoken Language*

Eleonora Massa (Rome)

Abstract

This paper develops a theoretical discussion about the definition of the unit of content in spoken language. The issue originates from the applicative field of corpus-based lexico-statistical surveys, which are traditionally and prevalently used to optimize and standardize the programs for the vocabulary didactics of foreign languages.

The main critical limitation of lexico-statistical inquiries can be identified in their impossibility to determine a representative threshold of the basic content lexicon of a language or, put otherwise, the most important words referring to the concrete things that are spoken about. Beyond the threshold of the most recurring 1,000 lexemes, in fact, words virtually show low and irregular as well as semi-equivalent probabilities to recur in spoken texts.

The lexico-statistical applications that have followed aim exactly at overcoming this limitation. Albeit through different methodologies, the various approaches conceive in fact the basic content lexicon as made up by the most frequent or used concrete substantives of a language. From time to time, either limited or particularly extensive series of concrete nouns have been compiled: however, these nouns are *de facto* subject to sporadic and irregular trends of recurrence values in spoken texts and are likely to be encountered very rarely by the language user in actual spoken utterances.

The discussion on basic spoken contents simply ends up in a theoretical flaw and in a mere representational paradox, because it investigates and describes exactly what is not constitutive of the examined phenomenon.

The consideration of the very semiotic peculiarity of spoken language constitutes the premise of an alternative definition of its meaning unit. The things that are talked about are in fact expressed only sporadically, because they are embedded in the situational context wherein they are shared – and mostly reiterated – by the conversation partners. More than with a discrete lexical element, the unit of spoken content seems to be identifiable with the holistic conversational practice that is instead regularly carried out by the speakers within likely ordinary frames of experience; consequently it seems to be closer to a basic unit in the “practice of meaning” than to an isolated meaning constituent. As the habitudinary modality of constructing, inhabiting and sharing our everyday form of life, the meaning unit in spoken language rather unveils as a narrative unit, for reasons that this paper explores in details. Such an alter-

* Particular thanks go to Prof. Grazia Basile for her expert advice and encouragement.

native theoretical vision is dealt with in the final part of the contribution, which also outlines further issues related to the possibilities of both its representation and its didactic usability.

1 Introduction

What are the most important things we speak about? The essential, fundamental, basic things we address in our spoken utterances?

By speaking, we surely refer to the things as physical, concrete entities: in short, as objects. Exactly as such, however, the main important contents that are the focus of our speaking seem to be indeterminable: they are in fact the most variable and, consistently with this, the words referring to them are the most irregularly used. Furthermore, spoken utterances generally display a low concentration of concrete substantives, as if things were spoken about by means of a different way than through the regular manifestation of the specific nouns referring to them.

This paper explores precisely such a possibility, that is to say, how we are able to build our meanings not just as a static content but as the result of a linguistically pragmatic negotiation. In carrying out this exploration, the paper discusses the issue of the semantic unit in spoken language.

Every time we speak, we are fundamentally building an ordinary, common and habitudinary sense for our life. This activity is constantly performed in regularly recurring situations and, consistently with this, mostly displays some lexical regularity: the latter is not to be solely and mainly considered in relation to the words that are used but, rather, to the way by which some words are used. By doing this, we are tracing some stable threads of experience, weaving it into stable plots and, thus, narrating our life: this continuous process of sense-making can be understood as the unitary agency in spoken language.

From the procedural and constructivist perspective adopted here, spoken language is in fact considered as a tool used in order to achieve certain goals: to make things rather than name them.

The discussion takes cue from the lexico-statistical field, extends to the one of foreign language didactics and converges in the semantic theoretical reflection. Finally, it results in the consideration of spoken language as a semiotic mode.

In the first part of the paper, the fundamental principles and laws of lexicostatistics are introduced (cf. §2); furthermore, their fertile reception by the basic lexicography, and thus prevalently by the field of foreign language didactics, is pointed out (cf. §3). It is in this same context that two main lexico-statistical studies of spoken language are reviewed (cf. §4).

The second part (cf. §5) stresses the main characters and critical aspects of the word lists that are compiled on the basis of lexico-statistical surveys, in order to clarify the usefulness they can have for the foreign language learner. The following main issues are dealt with: the internal discrepancies of the recurrence values of words (cf. §5.1), the general character of the

words that show a highly recurrence probability and thus provide for the most part some structure textual information (cf. §5.2), the peculiar character of the lexemes that concentrate within low recurring value ranges and thus mainly provide some content textual information (cf. §5.3). Consequently, the hypothesis of the non-quantitative determinability of the basic content vocabulary of a language is formulated.

The third section (cf. §6) deals with the principal methodologies through which lexicostatistics has faced this essential limitation. The results achieved through them are described and discussed (cf. §6.1, §6.2 and §6.3) and the common outlook the different approaches propose on the unit of spoken content is highlighted: in fact, they identify the semantic unit with the form of the concrete substantive and its regular recurrences, thus converging towards its firmly discrete vision (cf. §6.4). At the same time, this perspective is strictly improbable, since the assumption of such procedures is the very low occurrence, if not the sheer absence of nouns referring to concrete things in spoken language. In this way, the discussion concerning the identification of the basic content lexicon of a language seems to strand in a vicious circle, because it turns out to identify and describe something that is not inherent in the investigated object.

The fourth part of the paper aims at a constructivist revision of the hypothesis of the discrete and substantival character of the meaning unit. §7 focuses in fact on the suprasegmental profile of the processes of construction of spoken content and, through this, lets emerge the assumption of a correspondingly holistic interpretation of the semantic unit: more than to a segmental nominal entity, it seems in fact to correspond to a linguistic-pragmatic activity of content configuration. This praxis mostly takes shape by means of the highly recurring general lexicon that is aptly identified by the lexico-statistical surveys and exactly used in highly recurring, situated contexts, to negotiate the objects of our spoken utterances. In other words, the very process of linguistically pragmatic negotiation turns out to be the essential modality through which the most common frames of our daily existence are lived and within which the things are experienced rather than named. In §8 the intrinsic regular feature of a such practice is understood as the sole homogeneous trait of spoken language and, consequently, as the keystone of an alternative definition of its meaning unit. The path of spoken habitudinary content can be in fact understood as the basic manner through which, by speaking, we make a sense of our ordinary life, this in turn emerging as the main thing (or the essential content) that is addressed in speaking. §9 finally unveils the structural identity between this very modality and the narrative mode through which we normally experience and share our life, disclosing the perspective of a narrative semantic unit and of a narrative semantic approach.

In §10 a final consideration of the different issues that have emerged and have been discussed is provided and further inputs of analysis are outlined.

2 Lexical frequency and text coverage

A first systematical description of statistical regularities in historico-natural languages is provided by the work of Zipf (1935). According to the philologist's observations, verbal systems, like any other expression of human activities, are subject to the "principle of the least-effort" (ibd.: passim).

As far as the lexical level of verbal languages is concerned, this assumption implies that frequency is the most significant marker of words (cf. *ibid.*: 30–31). Guiraud (1960: 31) would later observe that lexical unities thus “come true” through their recurring character. Further on, Herdan (1966: 15, italics in original) states that “[...] there is a far-reaching similarity between the members of a speech community, not only in the [...] vocabulary [...] but also in the *frequency of use* of particular [...] lexicon items (words)”.

Frequency appears to be the factor on which different lexical regularities depend. For instance, the rank of a word in a list is inversely proportional to its recurrence: the higher the frequency the lower the position of the word in the list (cf. Zipf 1935: 40–44).¹ The length of the word is as well tied to its recurring character: the connection between the two factors is inversely proportional, since “[...] as the relative frequency of a word increases, it tends to diminish in magnitude” (*ibid.*: 38).²

As to the issue of this paper, “a few words occur with very high frequency while many words occur but rarely” (Zipf 1935: 40–41).

The first investigations in the statistical configuration of vocabulary stress the connection between high recurrence of words and text coverage. A representative profile of these surveys suggests that the first 1,000 most frequent words of a language provide 80% coverage of each text, the most recurrent 2,000 words cover 90% and the most frequent 4,000 cover up to 97,5% of each text (cf. Guiraud 1954: 10).³

Studies on lexical frequency are carried out on text *corpora* that are supposed to be representative of the whole state of the language taken into account. According to the so called “principle of representativeness”, the data collected in the sample have to recur analogously in the whole population. As Leech (2007: 135) points out, “[...] without representativeness, whatever is found to be true of a *corpus* is simply true of that *corpus* – and cannot be extended to anything else”.⁴ Most frequent lexical units in a representative sample are thus supposed to be most recurrent in the majority of texts produced in the language.⁵

The connection between lexical frequency and text coverage is analogously stressed by the most recent studies in applied linguistics: it is still agreed upon that the first 2,000 most frequent words are sufficient to allow a reasonable comprehension of 80% of each text (cf.

¹ The connection had already been established by the stenograph Estoup (1916). It is referred to as the “Zipf-Estoup law”. An exemplification of this law is in Crystal (1987: 87).

² The principle is also known as the “Zipf-Guiraud law”. The relationship between the two quantities was previously pointed out by Kaeding (1898) in his pioneering frequency list of German. The *Häufigkeitwörterbuch der deutschen Sprache* represents the outcome of an attempt to accelerate methods in shorthand writing. As we have observed (cf. fn. n. 1), the first contributions to an analysis of lexical frequency come from research fields that lie outside the domain merely pertaining to linguistics.

³ A first description of the same relationship in the Italian research scene is aptly represented by De Mauro (1961). For one of its more recent formulations cf., among others, Crystal (1987: 87).

⁴ Further essential argumentations of such principle are in Biber, Conrad and Reppen (1998) and Tognini-Bonelli (2001).

⁵ Various terms refer to the whole “text-population”: Oehler and Sörensen (1968), for instance, speak about “normal texts”, whereas according to Kosaras (1980) most frequent words in the sample are useful to communicate in relation to the “majority of daily topics”. Similarly, the German list *Zertifikat DaF* (cf. Steger 1972) refers to “most communicative situations”.

Nation/Waring 1997: 9–10) and the same number of words would provide an even greater coverage (around 90%) of informal spoken texts (cf. Tschirner 2005: 134).⁶

The usefulness of word frequency studies is being stressed with regard to the optimization of vocabulary teaching as well: “[...] lexical frequency should be an important criterion in the selection of words, i. e. in general, words which occur most frequently in the language should be among those taught in the earlier stages of instruction” (Jones 2004: 165).

3 Frequency word lists and foreign language didactics

A remarkable use of word frequency lists for didactic purposes can be traced back in the first decades of the 20th century.⁷ The standardization of vocabulary programs is a consequence of the greater diffusion of modern foreign language acquisition in the school context; furthermore it is often tied to the needs of particular learner categories, such as immigrant workers in the USA and northern European countries.⁸ The usefulness of basic vocabularies, i. e. dictionaries based on frequency word lists, is one of the major issues the 1934 New York Conference on Language Simplification focuses on (cf. Bongers 1947).

A remarkable series of basic dictionaries is compiled between the 20s and 40s. Pioneering works, respectively addressed to Spanish and English learners, are the ones by Keniston (1920) and Thorndike (1921). Knease (1931) completed a first frequency list for Italian learners, whereof a further representative example can be identified in the frequency dictionary for beginners by Migliorini (1943).

A great number of word frequency books is to be found in the field of French didactics: among these, the works by Henmon (1924), Cheydleur (1929), Vander Beke (1929), Tharp et al. (1934) and Haygood (1937). The dictionary compiled by Morgan (1928) can be considered as one of the pioneering word frequency books in German didactics.

Although different word quantities are provided by different vocabularies, 2,000 words can be considered as the average number of basic lexemes: according to the “lexical least-effort hypothesis”, 2,000 words can in fact provide between 80% and 90% coverage of each text (cf. §2). Among others, the dictionaries compiled by Morgan (1928), Knease (1931) and Migliorini (1943) are around this figure.

First word frequency books are mainly based on the results of the investigation of literary sources. The sample examined by Thorndike (1921), for instance, counts around 3,000,000 of 5,000,000 tokens from literary texts.⁹ The frequency list by Knease (1931) is exclusively

⁶ In their pioneering study Schonell, Meddleton and Shaw (1956) raise the coverage threshold to 96%.

⁷ Previous isolated attempts can be found in the handbook for language education of deaf-and-dumb children by Abbé de L’Épée (1776). It provides three series of 1,800 most frequent words.

⁸ Up to the 19th century, foreign language teaching prevalently concerns classical languages, which are understood as a means for the growth of intellectual faculties. Even modern languages like French are taught following the programs of Greek and Latin didactics. For a historical perspective on second language teaching cf. Richards and Rodgers (2000) and Rodgers (2001).

⁹ A *token* is identified as ‘graphical word’: a group of letters separated through punctuation marks and/or blanks from the former and following alphabetical series. A synonym for the term is *running word*, whereas similar tokens constitute a (*word*)*type*. Frequency investigations based on *corpora* proceed from a first identification of tokens to assigning a frequency value to each type. The relationship between the number of types and the num-

based on the investigation of literary sources, which constitute the largest data base of the works by Henmon (1924) and Vander Beke (1929) as well. In short, first generation *corpora* are generally made up of written texts.¹⁰

The following generation of basic dictionaries includes the word books which are based on the frequency parameter and its integrations with so called “user-oriented” or “communicative” criteria.¹¹ Starting from the 60s, foreign language didactics is in fact mainly permeated by theories focusing on “communicative competence”, built on the assumption that language ability doesn’t only consist in producing grammatically correct sentences, but in constructing utterances that are consistent with the socio-cultural context in which language is normally used.¹²

Criteria such as the “general comprehensibility degree” or the “usability” of the word are among the communicative parameters taken into account when determining the most useful words of the language. Such criteria can’t be de facto objectively defined: this is why second generation dictionaries are still mostly based on quantitative, i. e. word frequency investigations. At most they simply organize selected words into thematic fields (e. g. “family”, “school”, “holiday”) or according to communicative situations (e. g. “going to the restaurant”, “buying a ticket at the station”). Basic dictionaries like the ones by Oehler/Sörensen (1968) and Kosaras (1980), as well as the *Zertifikat DaF* (cf. Steger 1972) word list, are based on this model: they collect an average of 2,000 words and are for the most part made up of those lexical items that recur most frequently in written corpora.¹³

In a similar way, later explorations in lexico-statistics are mostly carried out on written samples: in comparison to spoken data these are generally easier, as well as cheaper, to be collected. The recent *Frequency Dictionary of German* (cf. Jones/Tschirner 2006), for instance, counts 4,200,000 tokens: 3,200,000 of them cover literary, journalistic, scientific and directional written texts; only 1,000,000 of them derive from spoken sources.

ber of tokens is known as the type-token ratio (TTR). A first survey of these aspects is offered by Muller (1963: 155–166); for a more recent one cf., among others, McEnery and Hardie (2012: 48–52).

¹⁰ The grammar-translation method, which derived from the classical method of Greek and Latin teaching, was still dominating in the first part of the 20th century. It focused both on learning rules in order to translate sentences from/into the foreign language and on the development of written comprehension skills. Vocabulary teaching was oriented towards the introduction of grammar exceptions, too. A description of a typical lesson unit can be found in Larsen-Freeman (2000). The centrality of spoken language was in fact one of the main assumptions of the subsequent direct method, yet teaching practice still mostly consisted in transmitting grammar structures. A further historical survey of language teaching methods is provided by Zimmermann (1997).

¹¹ For a wider historical reconstruction of basic lexicography see Kühn (1990). A more recent view of the matter is given by Koesters Gensini (2009a: 342–343, 2009b: 198–200).

¹² The term *communicative competence* was formally introduced by D. H.ymes (1966). It implies the re-interpretation and critical enlargement of Chomsky’s (1957) definition of *competence*, which was limited to considering the finite system of mental and innate rules as the only aspect that drives language learning and use; on the other side it overlooks the social and pragmatic character of linguistic behaviour. Further theoretical references for the notion of communicative competence are represented by Austin’s (1962) linguistic-pragmatic and Searle’s (1969) speech-acts theories.

¹³ A theoretical and methodological flaw seems to characterize this phase of basic lexicography. Theories of communicative competence tend in fact to stress the role of the speaker’s spoken ability, yet its basic work books are still based on the results of inquiries into written *corpora*.

4 Spoken frequency lists: *Français Fondamental (Ier degré)* and *Lessico di frequenza dell'italiano parlato*

The first organized attempt to investigate and represent the statistical configuration of spoken language vocabulary is represented by the *Français Fondamental (Ier degré)* (FF1) (cf. Gougenheim et al. 1964). The planning of the work started in 1951 and had specific didactics goals: as a first stage of French it should offer a basic vocabulary knowledge to the autochthonous speakers in the *Union Française* (cf. Gougenheim 1952: 113–114). The territorial-political entity had replaced the ancient French colonies and included, among others, Guadalupe and Martinique Islands, Madagascar, the French coasts of Somalia, Togo, Cameroun, Morocco and Tunisia. In these regions French was both the education language and the *lingua franca* of public utility, business and trade.

The *corpus* on which FF1 is based consists of both preexistent phonograph recordings and, for the most part, conversations purposely collected for the inquiry. These involve 275 speakers (138 men, 126 women and 11 children); for the most part they come from Paris and its outskirts (with the exceptions of few participants who come from other regions like Savoy, Brittany and Normandy). Interviewed people are asked to talk about everyday topics and situations such as “family and friends”, “at the workplace”, “on holiday”, “at home”, “on public transports” etc. (cf. Gougenheim et al. 1964: 63–67). Nevertheless, the number of participants interviewed about each topic is quite unhomogeneous.

From a quantitative point of view the investigation is carried out on the basis of 312,135 transcribed spoken tokens.¹⁴ 7,995 types are identified in 1,090 typed pages, each of them consisting of 300 lexical units. Only those words that recur at least twenty times and in five different texts of the sample are included in the conclusive frequency list: on the whole they amount to 1,063 lexical units (cf. *ibid.*: 66–69).

The *Lessico di frequenza dell'italiano parlato* (LIP) (cf. De Mauro et al. 1993) constitutes the first statistical investigation of spoken Italian. It has been compiled on the basis of a much wider *corpus* than the one investigated by the FF1: it is in fact made up of 500,000 tokens representing the standard *corpus* size between the 70s and the 90s.¹⁵ The sample includes spoken sources recorded in four Italian cities (Milan, Rome, Florence, Naples) and five utterance categories: face-to-face bidirectional conversations, non face-to-face bidirectional conversations (e. g. phone talks), face-to-face bidirectional conversations with non-free turns of speech (e. g. interviews), one-way utterances in the presence of receivers (e. g. lectures) and at distance one-way utterances with absent receivers (e. g. TV and radio programs).

¹⁴ We have already pointed out the theoretical and methodological inconsistency that characterizes basic vocabularies focusing on the learner’s communicative competence (cf. fn. n. 13). It emerges as well later on in statistical investigations that specifically focus on spoken language, since they can’t set aside its written representation, i. e. its segmentation in discrete units. This aspect will be specifically dealt with in relation to the possibility of defining and representing the meaning unit in spoken language (cf. §6.4).

¹⁵ Inquiries based on this *corpus* size are initiated by the *Frequency Dictionary of Spanish Words* (cf. Juilland/Chang-Rodriguez 1964). The same quantity of tokens is then exploited, among others, in the *Frequency Dictionary of Rumanian Words* (cf. Juilland/Edwards/Juilland 1965). Nowadays electronic data processing systems allow to collect samples that largely exceed 500,000 tokens: the project *Das digitale Wörterbuch der deutschen Sprache* (cf. DWDS), for instance, has assembled so far over 1,000,000,000.

The LIP represents a methodological improvement of mere frequency investigations like the *Français Fondamental* (cf. Gougenheim et al. 1964): the frequency index isn't in fact the only one computed by this collection, which also calculated the so called "complex dispersion" of words, i. e. the stability of their frequency in the different parts or *sub-corpora* of the sample.¹⁶

The product of frequency and dispersion indexes represents what is known as the "usage coefficient" of the word.¹⁷ Starting with 500,000 running words and 29,432 types, the LIP usage list identifies a spoken vocabulary core that includes on the whole 15,641 words (cf. De Mauro et al. 1993: 436–530).

Various differences therefore characterize the French frequency (cf. Gougenheim et al. 1964) and the Italian usage list (cf. De Mauro et al. 1993): these concern the temporal context in which the two surveys were carried out, the *corpus* size they respectively investigate, their methodological foundation and the word quantity they finally identify. Nevertheless, a very similar profile of the statistical configuration of spoken vocabulary emerges from them.

5 The statistical configuration of spoken vocabulary: trends and problems

In the following paragraphs a closer description of the quantitative lexicon profile will be developed: the general trends in the distribution of recurring values will be clarified in §5.1, whereas their qualitative characterization, i. e. the illustration of what kind of words correspond to these general trends, will be provided in §5.2 and §5.3. This, in turn, will lead us to circumscribe the primary problem of the statistical representations of spoken vocabulary in §5.4.

5.1 The internal lack of homogeneity in frequency and usage value-bands

The relationship between word recurrence¹⁸ and text coverage constitutes the basic theoretical assumption of the quantitative approaches to the definition of basic vocabularies. As we have already pointed out, the first 2,000 most frequent words of a language are supposed to cover about 90% of each text and the first 4,000 should provide up to 97,5% coverage (cf. §2). The

¹⁶ The *formula* has once again been elaborated and tested by Juilland and his collaborators (cf., among others, Juilland/Brodin/Davidovitch 1970; Juilland/Traversa 1973). Studies like the FF1 (cf. Gougenheim et al. 1964), instead, limit themselves to taking into account a simple dispersion index: this term refers to the whole number of texts of the sample in which the word recurs and not to its frequency in each *sub-corpora*. Clearly, the calculation of complex dispersion presupposes the internal subdivision of the sample itself into sections, the latter often coinciding with the different text typologies included. Given these premises, a lexical unit that recurs five times in a single text typology or part of the *corpus* (e. g. legal texts), for instance, can't be considered as important as a word that appears five times too, each of them though occurring in a different *sub-corpora*: in other words, they display a different degree of systematic centrality.

¹⁷ The fundamental distinction between word frequency and word usage has been precociously recognized by the Italian research field. First examples of its reception are offered by the survey *Lessico di frequenza della lingua italiana contemporanea* (LIF) (cf. Bortolini/Tagliavini/Zampolli 1971) and the study *Vocabolario di base della lingua italiana* (VdB) (cf. De Mauro 1980). In the German research context the issue has been firstly discussed by Kühn (1979) and more recently highlighted by Koesters Gensini (2009a: 343–345, 2009b: 198–200).

¹⁸ On the basis of what has been discussed so far, the term *recurrence* can be understood as both 'word frequency' and/or 'word usage coefficient' (cf. §4).

connection between word frequency and/or usage and text coverage can be understood in terms of a so called “relation of systematical productiveness”. Yet the internal configuration of the above-mentioned threshold levels needs to be investigated more closely.

As we have seen, the *Français Fondamental* (cf. Gougenheim et al. 1964) lists the most frequent 1,063 lexical units of the exploited *corpus* (cf. §3) and thus, according to the principle of representativeness (cf. §2), of spoken French. However, the recurrence coefficients of these words are far from regular: quite on the contrary, substantial disparities do emerge among them. Suffice it to say, for example, that the frequency difference between the first (the verb *être*, ‘to be’) and the fiftieth word (the interjection *oh*) of the list amounts to almost 13,000 recurrences. Similarly, the hundredth word (the verb *trouver*, ‘to find’) recurs about 800 times less than the fiftieth word and, again, more than 13,000 times less than the first item (cf. Gougenheim et al. 1964: 69–71). Further on, the recurrence difference between the three hundredth (the numeral adjective *cinquante*, ‘fifty’) and the hundredth lexeme amounts to over 300 repetitions (cf. *ibid.*: 71–75).

Similar disparities in the distribution of recurring values emerge in the Italian usage list (cf. De Mauro et al. 1993). Some representative examples can be summarized as follows:

RANK	WORD	USAGE COEFFICIENT
1	<i>il</i> (‘the’)	37,076
50	<i>quindi</i> (‘therefore’)	1,399
100	<i>giorno</i> (‘day’)	480
200	<i>situazione</i> (‘situation’)	188
500	<i>alzare</i> (‘to raise’)	58
700	<i>contento</i> (‘happy’)	38
1,000	<i>solito</i> (‘[the] usual’)	24

Table 1: The relationship between rank and usage coefficient in the LIP (cf. *ibid.*: 437–443)

Frequency and usage bands show a significant internal lack of uniformity. Even within limited numbers of rankings like the first 500 or the first 1,000, recurrence values tend to decrease rapidly and substantially.

This trend is uniform in the following coefficient bands. Here, as well, it can be useful to isolate some ranking and value sections to better clarify its extent:

RANK	WORD	USAGE COEFFICIENT
1,200	<i>novantuno</i> (‘ninety one’)	19
1,500	<i>luglio</i> (‘July’)	13
1,600	<i>giallo</i> (‘yellow’)	12
1,900	<i>cura</i> (‘care’)	9

Table 2: Differences in recurrence values after rank 1,000 in the LIP (cf. *ibid.*: 444–448)

This particular distribution of recurring values has been firstly stressed by the French school (cf., among others, Gougenheim 1952). Such a remark can be otherwise understood as the probable reason why the FF1 list (Gougenheim et al. 1964) stops just beyond 1,000 words.¹⁹

An additional trend seems to distinguish the words ranked beyond the thousandth. These are not only much less frequent or used than the previous rank words: they rather manifest very close, or even equipollent, recurrence values. As a consequence, more words share a very similar or even the same probability to occur in texts: a low and broadly equivalent chance to appear therein.

Chart n. 3 provides some examples of this additional trend:

RANKS	TOTAL WORDS	USAGE COEFFICIENT
1,200–1,300	100	19–16
1,500–1,600	100	13–12
1,800–1,900	100	10
1,901–2,000	100	9–8

Table 3: Value concentration after rank 1,000 in the LIP (cf. De Mauro et al. 1993: 444–449)

Chart n. 4 shows in more detail the relation between low-usage coefficients and the number of words that share it:

RANKS	TOTAL WORDS	USAGE COEFFICIENT
1,200–1,300	51	18
	41	17
1,500–1,600	53	13
	47	12

Table 4: Quantity of words sharing the same (low)-recurrence value in the LIP (cf. ibd.: 444–446)

When compared with the ones of the first 100 or 1,000 ranks, the recurring values of these words are generally microscopic.

The decrease of recurrence indexes corresponds to the diminution of the difference in the recurrence of two consecutive terms. The latter difference is so slight that it turns out to be insignificant in actual speech utterances: beyond a macroscopic value band, wide coefficient ranges emerge, wherein each word has practically the same probability to manifest itself as others. There will thus be an irrelevant frequency or usage difference between a word that is included in a basic vocabulary and a word that, although very close to the former one, is excluded from it. Lists tend to stop at an area of recurrence values wherein many successive lexical units have a very similar and, even before, a very low chance to appear in texts. This aspect has been firstly highlighted by the French school as well (cf. Michéa 1949).

Hence the hypothesis according to which the most frequent 2,000 words of the language provide 90% text coverage (and the most recurring 4,000 up to 97,5%) needs to be revised on the basis of what has been discussed so far. The underlying assumption doesn't imply that all

¹⁹ It may be worth stressing once more that value discrepancies actually characterize the rankings 1–1,000, too. For instance, the frequency index of the first word (the verb *être*, 'to be') amounts to 14,083 (cf. ibd.: 69), whereas the value of the five hundredth one (the noun *état*, 'state') is only 55 (cf. ibd.: 79).

these words have the same high and regular chance to become manifest in oral texts, and consequently to cover large text portions. Few words are really subject to this trend: these can be identified with the first 500 - if possible with the first 1,000 – most recurring words of a language; moreover they are themselves characterized by significant, internal value disparities. Many other words, on the other side, appear much more rarely in spoken texts: they only recur, and mostly co-occur, when one or few among many other words don't manifest themselves.²⁰

On the side of the language user, and in particular of the foreign language learner, there will be only few words that recur very often in texts; many others will occur only sporadically, instead.

5.2 Highly recurring character, low-content information: the general lexicon

The significant concentration of grammar words is the first qualitative trait emerging in high-occurrence and text-coverage word bands. Some aspects of their distribution can be schematized as follows:

FF1	TOTAL PERCENTAGE OF GRAMMAR WORDS	HIGH-CONCENTRATION RANKS
	25,3%	1–1500

Table 5: Grammar words in the FF1 (cf. Gougenheim et al. 1964: 69–79)

LIP	RANKS	PERCENTAGE OF GRAMMAR WORDS
	1–500	29,6%

Table 6: Grammar words in the LIP (cf. De Mauro et al. 1993: 437–440)

Apart from few exceptions, they are generally made up of one, two or three syllables. As previously stressed, if the recurring character of a word increases, its length tends to diminish (cf. §2).

For the most part, very frequent or highly used verbs are also bisyllabic or trisyllabic words. Moreover, they let a further characteristic of most occurring words emerge: that is, their polysemy (their tendency to include different, but related senses into their meaning).²¹ The auxiliary verbs *être* and *avoir*, *essere* and *avere* ('to be', 'to have'), represent the most recurring ones in both spoken lists. Further examples of very frequent French verbs are,

²⁰ A similar configuration of vocabulary emerges in statistical investigations of written *corpora*. Among the 2,000 most frequent words of the LIF (cf. Bortolini/Tagliavini/Zampolli 1971), the first 500 lexemes alone provide more than 80% text coverage, whereas the following 1,500 share a total text coverage of about 10% (cf. *ibid.*: V). If the usage coefficient is taken into account, ranks 1–500 cover analogously about 85% of the sample, while ranks 501–2,000 provide altogether less than 9% (cf. *ibid.*: LXXIV). Nevertheless, the exploitations of spoken samples offer a first systematic evidence of these aspects.

²¹ Just consider, for instance, that the third most used verb of the LIP, the verb *fare* ('to do') (cf. De Mauro et al. 1993: 437), has seventeen senses as a transitive verb and ten as an intransitive one. It also has more than hundred senses when appearing in collocations (e. g. *avere da fare*, *fare amicizia*, *fare carte false*) (cf. De Mauro 1999, vol. II: 1033–1039). Similarly to other lexico-statistical regularities, the relationship between word frequency and semantic breadth has been firstly stressed by Zipf (1935).

among others, *faire* ('to do'/to 'make'), *aller* ('to go'), *voir* ('to see'), *pouvoir* ('can'), *vouloir* ('to want'), *venir* ('to come'), *prendre* ('to take'), *devoir* ('must'), *parler* ('to speak') (cf. Gougenheim et al. 1964: 69–71). Similarly, the verbs *fare* ('to do'/to 'make'), *andare* ('to go'), *vedere* ('to see'), *potere* ('can'), *volere* ('to want'), *venire* ('to come'), *prendere* ('to take'), *dovere* ('must'), *parlare* ('to speak'), *trovare* ('to find'), are among the most used in Italian (cf. De Mauro et al. 1993: 437–438). The concentration of verbs significantly and/or firmly characterizes highly recurring word bands: for instance, within the first threshold level of the LIP (cf. ibd.: 437–449), they amount to around 20%.

Highly recurring adjectives also show a tendency towards polysemy. *Grand* ('big'), *plein* ('full'), *vrai* ('true'), *simple* ('simple'), for instance, all have four or five senses (cf. Garzanti 1992: *passim*). They are among the most frequent ones in the FF1 (cf. Gougenheim et al. 1964: 69–76).²² Further examples of highly recurring adjectives can be found among *petit* ('small'), *vieux* ('old'), *seul* ('alone'), *dernier* ('last'), *cher* ('dear'/'expensive'), *certain* ('certain'), *joli* ('pretty'), *chaud* ('warm'), *malade* ('sick'), *nouveau* ('new'), *difficile* ('difficult'). A similar profile of the word category emerges in the list of spoken Italian (cf. De Mauro et al. 1993), in which adjectives like *grande* ('big'), *bello* ('beautiful'), *certo* ('certain'), *vero* ('true'), *importante* ('important'), *buono* ('good'), *nuovo* ('new'), *prossimo* ('next'), *solo* ('alone'), *difficile* ('difficult'), *alto* ('high'/'tall'), *scorso* ('last') are among the most used between rank 1 and rank 500 (cf. ibd.: 437–440).

Finally, the most frequent or used substantives show two qualitative trends. On one side they are constituted by nouns that have to do with the categorization of time: substantives like *heure* ('hour'), *jour* ('day'), *moment* ('moment'), *mois* ('month'), *année* ('year'), *matin* ('morning') (cf. Gougenheim et al. 1964: 69–73) and *anno* ('year'), *volta* ('time'), *giorno* ('day'), *tempo* ('time'), *momento* ('moment'), *ora* ('hour'), *mese* ('month'), *sera* ('evening'), *settimana* ('week'), *pomeriggio* ('afternoon') (cf. De Mauro et al. 1993: 437–440) are ranked at the top of both lists. On the other side, the most recurring substantives are nouns denoting common things and people: *chose* ('thing'), *truc* ('thing'/'whatsit'), *monsieur* ('Mister'/'Sir'), *enfant* ('child'), *femme* ('woman'), *mère* ('mother'), *maman* ('Mum'), *père* ('father'), *mari* ('husband'), *mademoiselle* ('Miss'), *garçon* ('boy'), *place* ('space'), *ville* ('city'), *pays* ('country'), *monde* ('world') (cf. Gougenheim et al. 1964: 70–79) and *cosa* ('thing'), *parte* ('part'), *persona* ('person'), *signora* ('lady'), *bambino* ('child'), *mamma* ('mum'), *famiglia* ('family'), *madre* ('mother'), *ragazza* ('girl'), *gente* ('people'), *casa* ('house'/'home'), *lavoro* ('work'/'job'), *modo* ('manner'/'way'), *fatto* ('fact'), *caso* ('case'/'matter') (cf. De Mauro et al. 1993: 437–440) represent examples thereof.

To conclude, some qualitative constants seem to emerge in high-occurrence and text-coverage word bands. These are: a) the significant amount of grammar words, b) the high concentration of polysemous verbs and adjectives, c) the accumulation of nouns denoting time, common

²² Nevertheless the polysemous character of highly recurring lexemes has been only partially dealt with by the quantitative approaches aimed at the determination of basic dictionaries. Usually the lists don't explicit which one is the most frequent or used sense of a word; the values of the different senses rather converge into a single index of recurrence. Some first critical observations on this issue can be found in Kühn (1979: 49–54). More recently it has been addressed by Russo (2005: 20).

things and people. We will first refer to this cluster of words as “the general lexicon”, meaning they allow the language user to obtain a general, i. e. a limited grammar and/or structure textual information – a kind of language scaffold. They don’t provide, on the contrary, high- or detailed-content information. It may be worth reminding that these words are the most frequent or regularly used ones in spoken texts: that is, the ones a foreign language learner will encounter most often.

5.3 Low recurring character, highly textual information: the content lexicon

The constant decrease in the number of grammar words can be considered as the first trend characterizing low recurring word bands. As far as the LIP (cf. De Mauro et al. 1993) is concerned, this decreasing tendency can be represented as follows:

RANKS	PERCENTAGE OF GRAMMAR WORDS
1–500	29,6%
501–1,000	6%
1,001–1,500	4,8%
1,501/2,000	2%

Table 7: Rankings and distribution of grammar words in the LIP (cf. *ibid.*: 437–449)²³

It is worth noting that the diminution factor turns out to be macroscopic between the first usage band and the following three ones. The difference is therefore not only constant but also significant.

With regard to other lexical categories, the percentage of adjectives isn’t subject to important variations. From rank 1 to rank 2,000 of the LIP (cf. *ibid.*), for instance, it varies from 13,4% to 16,2%. From this point of view, adjectives show a tendency similar to the one we have already highlighted for the category of verbs (cf. §5.2).

What really emerges among low-recurrence and text-coverage lexemes is the increase of substantives: suffice it to say that their quantity almost doubles from rankings 1–500 to rankings 501–1,000 of the LIP (cf. De Mauro et al. 1993: 437–443). Similarly, more than 50% of the words included from rank 1,501 to 2,000 are nouns (cf. *ivi*: 443–449).

Substantives are thus the word category that is mostly subject to low and irregular recurring values (cf. §5.1).

Consequently, they are also the word category that is mostly subject to the trend of close or equipollent values (cf. *ibid.*). In this respect, it might be useful to isolate the sections of substantives following the rank 1,000 in order to verify the range of values within which they concentrate. An example from the LIP (cf. De Mauro et al. 1993) can be summarized as follows:

²³ A few examples are adverbs like *fuori* (‘outside’), *durante* (‘during’), *dietro* (‘behind’), conjunctions as *difatti* (‘in fact’), *perciò* (‘therefore’) and *ebbene* (‘well then’), possessive pronouns like *suo* (‘his’/‘her’/‘its’) and *tuo* (‘your’).

RANKS 1,200–1,300	TOTAL NUMBER OF SUBSTANTIVES	47
RANKS 1,200–1,300	USAGE VALUE RANGE	19–16
RANKS 1,200–1,300	SUBSTANTIVES WITH SAME COEFFICIENT	24

Table 8: Semi-equivalence of substantive-usage values in the LIP (cf. ibd.: 444–445)²⁴

The probability that a word like *colpo* ('stroke') has to appear in spoken texts isn't thus macroscopically dissimilar from the one substantives like *augurio* ('wish', 'greetings'), *riga* ('line'), *papà* ('dad'), *fame* ('hunger'), *riflessione* ('reflection') and *peccato* ('sin') have. The recurrence value of the latter, in turn, is surprisingly close to the coefficient of nouns like *fonte* ('spring', 'source'), *apparecchio* ('device'), *io* ('[the] self') and *memoria* ('memory'). Even further on, the word *teoria* ('theory') has exactly the same occurrence probability as *microfono* ('microphone'), *piatto* ('plate'/'dish'), *concerto* ('concert'), *confine* ('border'), *aiuto* ('help'), *comunicazione* ('communication'), *proprietà* ('property'), *intenzione* ('intention'), *filosofia* ('philosophy') and *sentimento* ('feeling'), just to give a few examples. Besides, all these lexemes are ranked in low and irregular bands of usage values. Finally, they can differently co-occur in the limited text portions that are not covered by the general lexicon.

The low and inconstant frequency of substantives, and more specifically of the *mots concrets* ('concrete words'), has been first pointed out by the authors of the FF1 (cf. Gougenheim et al. 1964: 137–145). Indeed, the scholars had verified the insignificant recurrence values of nouns like *jupe* ('skirt'), *fourchette* ('fork'), *métro* ('underground', 'subway'), *boulangier* ('baker'), *épicier* ('grocer's'), *allumette* ('match'), *autobus* ('autobus'). They had also observed the likely sporadic frequency of substantives like *boulangerie* ('bakery'), *boucher* ('butcher'), *chocolat* ('chocolate'), *cinéma* ('cinema'), *ciseaux* ('scissors'), *film* ('film'), *moto(cycllette)* ('motorcycle'), *radio* ('radio'), *téléphone* ('telephone'), *télévision* ('television'). Such words

²⁴ The series we have examined includes more specifically the following nouns: *Francesca* ('Francesca', proper name), *casino* ('mess'), *colpo* ('stroke'), *teoria* ('theory'), *carico* ('load'), *prospettiva* ('perspective'), *marzo* ('March'), *microfono* ('microphone'), *piatto* ('plate'/'dish'), *concerto* ('concert'), *confine* ('border'), *polemica* ('controversy'), *inglese* ('English'), *banco* ('desk'/'counter'), *aiuto* ('help'), *comunicazione* ('communication'), *proprietà* ('property'), *roccia* ('rock'), *bilancio* ('balance'), *scusa* ('apology'), *intenzione* ('intention'), *filosofia* ('philosophy'), *sentimento* ('feeling'), *definizione* ('definition'), *offerta* ('offer'), *filo* ('thread'), *bello* ('beauty'), *augurio* ('wish', 'greetings'), *riga* ('line'), *papà* ('dad'), *fame* ('hunger'), *riflessione* ('reflection'), *peccato* ('sin'), *venerdì* ('Friday'), *crisi* ('crisis'), *massimo* ('maximum'), *circolazione* ('circulation'), *aprile* ('April'), *quantità* ('quantity'), *pianta* ('plant'), *figlia* ('daughter'), *corrente* ('stream'), *conflitto* ('conflict'), *fonte* ('spring'/'source'), *apparecchio* ('device'), *io* ('[the] self'), *memoria* ('memory') (cf. ibd.). Substantives included between *colpo* ('stroke') and *bello* ('beauty'), for instance, all share the same usage coefficient, i. e. 19 (cf. ibd.: 444). As emerges from the listed series, various substantives can also function as other parts of speech (e. g.: *scusa* as the third singular person in the present and the second in the imperative tense of the infinitive *scusare*/'to excuse', *inglese* as adjective, *io* as the personal pronoun 'I'). Unlike what has been observed in relation to the issue of polysemy (cf. fn. n. 22), the aspect of the functional breadth of vocabulary items is usually dealt with appropriately by the different lists. For instance, the word *bello*, which has been already mentioned within the most recurrent Italian adjectives ('beautiful'), appears again as a noun in the above series ('beauty')

don't even appear in a list: they don't show, in fact, a good frequency coefficient (cf. *ibd.*: 138).²⁵

Substantives referring to concrete things, persons, places, objects, circumstances and conditions – to what can be conceived as a “state of affairs” – can thus be understood as the word category that is mostly subject to the trend of low and semi-equivalently recurring probabilities. They can also be defined as *mots thématiques* (cf. Michéa 1950a, *Id.* 1950b; cf. also Gougenheim et al. 1964: 144–145): as words that provide information about the **themes**, or again the **things**, that are spoken, talked, or even told about.

On the premises given so far, we may conclude by saying that those words which appear rarely and non-systematically in spoken texts, and which the learner will encounter more rarely in his/her experience, are exactly the ones he/she specifically needs in order to gain access to spoken **content**.

5.4 The non-quantitative determinability of basic linguistic contents

The result of lexico-statistical surveys that aim at the identification of the spoken vocabulary nucleus can be profiled as follows.

First, the only lexical items whose probability of recurrence can be scientifically defined – in so far as it is macroscopic and regular – only provide general, i. e. grammar or structure textual information (cf. §5.2). Such words are those that appear in a text independently from its content (the so called *termes athématiques* or non-thematic words): accessory- or grammar words (*mots accessoires*), a large number of verbs and adjectives and a few general and current nouns (cf. Michéa 1950b: 328).

Further on, the outcome of frequency lists is restricted to the identification of those words that

[...] se retrouvent [...] régulièrement [...] dans n'importe quel texte [...] parce qu'il n'existe pas de rapport de dépendance vérifiable entre leur apparition dans le discours et le thème choisi.
[...] On peut ranger dans cette catégorie les mots accessoires, qui ne marquent que de rapports, un grand nombre d'adjectifs et de verbes courants et quelques noms très généraux.

(Michéa 1950a: 188)²⁶

Secondly, those words that are useful or needed to construct a hypothesis of content are ranked beyond the threshold of high-occurrence and text-coverage word bands: they show sporadic and semi-equivalent frequency or usage indexes. In general, the lists don't inform us about concrete words at all and, as observed by the French school (Gougenheim et al. 1964), “[...] il y a dans ce fait quelque chose qui, au premier abord, semble surprenant” (*ibd.*: 138).²⁷

²⁵ The closeness and/or equivalence of values also emerges in the French spoken list. For instance, 50 substantives concentrate from rank 650 to rank 750 (cf. *ibd.*: 81–83): yet they are included in a frequency range that varies from 38 to 31.

²⁶ “[...] regularly [...] appear [...] in no matter which text [...] because there is no relation of verifiable dependence between their appearance in speech and the chosen theme. [...] Accessory-words marking nothing else but (structure) relations, a large quantity of current adjectives and verbs, as well as some general substantives, can be brought under this category”. Unless otherwise specified, all the translations of texts written in a language other than English, are mine.

²⁷ “[...] at first sight there is something astonishing in this fact”.

From another point of view, quantitative inquiries into the vocabulary configuration have to face the fact that

[...] les diverses catégories de mots ne sont pas également [...] justifiables de la statistique. La plupart de noms concrets (*content words*, *Dingwörter*, mots thématiques) échappent pratiquement à ce moyen de sélection. Une grand partie de ceux qui apparaissent dans les résultats d'un dénombrement, même bien conduit, sont en rapport avec le choix, nécessairement, [...] des données de base. Rien d'étonnant à cela: le concret n'est, au fond, que le particulier, et reste par conséquence en dehors du domaine de la statistique.

(Michéa 1950b: 328, italics in original)²⁸

The limits of lexico-statistical methods are not to be intended as practical ones. Studies like the LIP (cf. De Mauro et al. 1993) in fact succeed in determining a far larger vocabulary core than the FF1 (cf. Gougenheim et al. 1964): however, this doesn't imply a qualitative difference in the output.

The limits of lexico-statistical investigations seem rather to be of a theoretical kind: they concern the sense of assigning a coefficient to words that, on their own, occur rarely and have a semi-equipollent probability to appear in texts. Beyond a macroscopic threshold level, word lists only inform the language learner about many words that have a low and even more similar chance to occur in texts: what is, again, the sense of assigning them a recurrence value?²⁹

In general, whereas the overall trends of the statistical configuration of vocabulary are clearly describable, their function in order to understand the general mechanisms of language functioning is quite ambiguous.³⁰

In particular, assigning a coefficient doesn't prove functional in order to determine the **basic contents** of spoken texts: their linguistic form or, more generally, **basic linguistic meanings**.

²⁸ “[...] the different word categories aren't all equally [...] justifiable by statistics. The most part of concrete nouns (*content words*, *Dingwörter*, mots thématiques) practically elude this criterion of selection. A great part of those included in the results of a *corpus* exploitation, even if carried out well, are necessarily related [...] to the selection of the basic data. There is nothing surprising about this: the concrete is, basically, nothing else than the particular and consequently remains out of the domain of statistics”. This second feature of content words emerges in particular in the lists that are compiled on the basis of written *corpora* and precedes then, both theoretically and historically, the problem concerning word lists of spoken language. Frequency dictionaries as the ones by Kaeding (1898) and Morgan (1928), for instance, assign significant coefficients to words that are distinctive of literary and juridical texts, which, in turn, cover a wide percentage of both samples. According to what we have observed so far, the issue of content words in spoken sources emerges rather in terms of low occurring, or even absent, values: the reasons of this particular tendency will be better clarified in the next paragraphs. Anyway, lexico-statistical investigations of both written and spoken samples, do turn out to be partial when it comes to determining the content lexicon: in fact, they identify either a specific part of it or, paradoxically, no one in particular. In general, they fail to provide a description of the basic, i. e. common, standard or cross-texts level of linguistic contents.

²⁹ The author of this paper has carried out such a test with specific reference to written Italian (cf. Massa 2013: 61–72), whereby the relationship between the most used words of the LIF (cf. Bortolini/Tagliavini/Zampolli 1971) and the text coverage of a short story for Italian learners (cf. Felici Puccetti 2010) has been verified. In line with what we have been discussing so far, the words that are useful and/or needed to construct a hypothesis of content are all ranked beyond the first band of highly recurring values.

³⁰ A similar discussion of the issue is being aptly carried out in the most recent Italian research field (cf., among others, Russo 2005).

6 The study of basic meanings in spoken language

In the following paragraphs we are going to clarify the modalities through which the basic lexicography has dealt with the problem of determining content words. The theoretical and empirical perspective of the French school, as well as an examination of its results, will be closer discussed in §6.1 and §6.2. The output of the discussion as it developed in subsequent surveys, like the LIP and more recent investigations of lexical frequency, will be dealt with in §6.3. To conclude, the main critical issue regarding the research on basic spoken meanings will be outlined in §6.4.

6.1 The FF1 and the survey on “available words”

If basic linguistic contents can't be classified through the exploitation of *corpora*, they necessarily have to be investigated elsewhere than in texts: they have to be identified in the “speaker's mind”. This is exactly the spirit that has guided the survey on “available words”.

As we have seen so far, content words are in fact identified with lexemes that occur rarely and irregularly, and yet are useful and usual (cf. Gougenheim et al. 1964: 145). Content words “come to mind” when they are needed (cf. Michéa 1953: 340) and are hence at the “speaker's disposal” (cf. Gougenheim et al. 1964: 145):

[...] que faut-il entendre par «*vocabulaire disponible*»? Un mot disponible est un mot qui, sans être particulièrement fréquent, est [...] toujours prêt à être employé et se présente immédiatement et naturellement à l'esprit au moment, où l'on en a besoin.

(Michéa 1953: 340, italics in original)³¹

The discussion about basic linguistic meanings in spoken language therefore involves an essential theoretical shift towards the mental dimension – the *esprit* – of the speakers.³²

The resort to the psychic dimension leads to a consistent methodological formulation: the investigation of available words is equated with exploring the series of substantives that refer to concrete objects and are most frequently “associated” by the speakers to contiguous thematic fields or “centres of interest” (*centres d'intérêts*).³³ From a procedural point of view, the survey on available words can be described as follows:

³¹ “[...] what is meant by «*available vocabulary*»? An available word is a word that, despite its low frequency, is [...] always ready to be used and comes immediately and naturally to mind at the moment when it is needed”.

³² This aspect is remarked by Michéa (1953) as follows: “[...] beaucoup plus que la fréquence, [...], la disponibilité est en rapport avec notre vie psychique [...], et c'est [...] ce qui en fait la valeur pédagogique” (ibid.: 341). “[...] much more than to frequency, [...], lexical availability is related to our psychic life [...], and this is [...] what makes for its pedagogical value”.

³³ The term refers to the principle of classification of material substances as formulated by the lexicography practice in the Middle Age: it can be understood as a group of objects that are assembled according to the “principle of material similarity” (cf. Quémada 1968: 363), or again as a “nomenclature”.

[...] demandons à un grand nombre de personnes d'écrire une série de noms (20, par exemple) se rapportant à un centre d'intérêt déterminé et classons les mots obtenus par ordre de fréquence décroissante. Ceux qui se trouveront le plus souvent, qui se seront, par conséquent, imposés à l'esprit du plus grand nombre de personnes, pourront être considérés comme plus communément disponible que les autres, si nous entendons par disponibilité la propriété que possède un mot d'être évoqué d'une façon plus ou moins immédiate au cours de l'association des idées.

(Michéa 1953: 341)³⁴

The French survey on available words has been carried out on the basis of a sample that includes 904 speakers, i. e. school pupils aged between 9 and 12 coming from four different areas of France³⁵. The number of centres of interest investigated totals 16: among them are thematic fields such as “body parts”, “clothes”, “parts of the house”, “school”, “food and drink”, “means of transport”, “animals”, “trades and professions”.³⁶

For instance, the analysis of the tests concerning the topic “parts of the house” has led, among others, to the identification of the following content words or substantives the interviewed pupils have most frequently and intersubjectively associated to it: *fenêtre* (‘window’), *porte* (‘door’), *mur* (‘wall’), *cheminée* (‘fireplace’), *chambre* (‘room’), *plafond* (‘ceiling’), *tuile* (‘tile’), *cuisine* (‘kitchen’), *toit* (‘roof’), *salle à manger* (‘dining room’), *escalier* (‘stairs’) (cf. Gougenheim et al. 1964: 158–159). These words are considered as the basic linguistic contents – namely the basic linguistic meanings – of such thematic field.³⁷

6.2 Available nomenclatures: quantitative and qualitative characteristics

The purpose of the survey on the available vocabulary is to overcome the main critical aspects of the investigations of textual *corpora*: that is, the tendency of recurrence values to rapidly and substantially decrease after the rank 1,000 (cf. §5.1) and the consequent difficulty to identify the content lexicon (cf. §5.4).

³⁴ “[...] let’s ask a large number of people to write a series of nouns (20, for instance) sticking to a specific centre of interest, the words are then classified according to the principle of decreasing frequency. Those words that will recur most frequently, having consequently come to the mind of most people, will be considered as more available than the other ones, if availability is understood as the property of a word to be more or less immediately evoked in the course of mental associations”.

³⁵ They include: *Dordogne* (in the South-West of France), *Marne* (North-East), *Eure* (Centre-North) and *Vendée* (North-West). As an addition to their profile, 488 speakers are male pupils whereas 416 are female. A detailed analysis of the French survey on available words has been carried out by Zeidler (1980): among other important aspects, the scholar has highlighted the internal lack of homogeneity characterizing the sample as concerns the diatopic distribution of the participants (500 out of 904 come from *Dordogne*, whereas the remaining 404 are distributed among the other three areas) and the diastratic one (this aspect being exemplarily represented by the case of *Eure*, where 73 male but only 12 female pupils have been interviewed) (cf. *ibid.*: 226).

³⁶ Similarly to what has been observed with regards to the sample composition, the arbitrary choice of the fields has also been pointed out (cf. Zeidler 1980: 242).

³⁷ A further study on English available words (cf. Dimitrijevič 1969) has led to the identification of a similar word series with regards to the same centre of interest. This includes the following substantives: *window*, *floor*, *fireplace*, *door*, *kitchen*, *wall*, *roof*, *bedroom*, *diningroom*, *bathroom*, *livingroom*, *chimney*, *cupboard*, *attic*, *ceiling* and *toilet* (cf. *ibid.*: 129–133). Surveys like the one by Dimitrijevič (1969), the study on German available words by Pfeffer (1964) and the one on the available vocabulary of Acadian French (Mackey/Savard/Arduin 1971), have actually been carried out on the basis of substantially different samples (both if they are compared to one another and to the FF1 sample): despite this, it is worth noting that they eventually provide very similar results. A comparison of the different studies has already been developed elsewhere by the author of this paper (cf. Massa 2013: 93–99).

Nevertheless, the lexical-associative method isn't constitutively different from an inquiry into frequency values. As Michéa (1953) has clearly stated, “[...] les mots [...] qui se trouveront le plus souvent” (ibid.: 341)³⁸ can be considered as at the speakers' disposal: if not as an investigation of the statistical probability of lexical recurrences in texts, the research on available words can then be understood as an investigation of mental occurrences. And in fact it ends up posing again the same problem already showed by the exploitations of *corpora*: like the most frequent or used words in a sample, intersubjective mental associations tend as well to significantly decrease beyond a macroscopic threshold level. Suffice it to say that only 240 content or available words have been added to FF1 list, i. e. to the most frequent 1,063 lexemes (cf. Rivenc 1973: 16). Furthermore, a maximum of only 20 words has been selected for each centre of interest. These numbers appear to be minimal if related to the total quantity of tested speakers (904) and centres of interest (16): evidently, the common associations that follow have turned out to be sporadic and irregular.³⁹

Since paradoxically “[...] d'un individu à un autre, le vocabulaire disponible [...] varie beaucoup plus [...] que le vocabulaire fréquente” (Galisson/Coste 1976: 160),⁴⁰ the method centered on mental associations seems to stop in front of the very problem it should overcome. From a quantitative point of view, then, the series of available substantives restrict themselves to include a limited number of lexical items.⁴¹

The qualitative profile of the available nomenclatures can thus be understood as the direct result of their quantitative character: as limited progressions of conceptual occurrences, the selected series resemble in fact some kind of general associative or mental structure: more than to linguistic contents, they could be compared to “conceptual primitives”.⁴²

³⁸ “[...] the words [...] that will recur most frequently”.

³⁹ We should also note that no information about the minimum occurring value of an available word is given in the FF1. Some interpretational attempts in this respect are offered by Rivenc (1967) and Mackey/Savard/Ardouin (1971). However, the tendency of mental associations to rapidly and consistently decrease is confirmed by other surveys: the study by Pfeffer (1964), for instance, counts a total number of 833,000 associated words, of which only 347 are included in the spoken frequency list (cf. ibid.: 69-74). Furthermore, the survey by Mackey/Savard/Ardouin (1971) leads us to stress another aspect mentally associated words share with lexemes recurring in a *corpus*: their tendency to concentrate in ranges of values that are near or equivalent to one another (cf. §5.1 and §5.3). For example, among over 200 nouns that have been rarely associated by the speakers to the topic “parts of the house”, a hundred items share the coefficient 2 (cf. Mackey/Savard/Ardouin 1971: 41-49).

⁴⁰ “[...] from an individual to another, the available vocabulary [...] varies much more [...] than the frequent vocabulary”.

⁴¹ This aspect leads us to stress once again the similarity of the results achieved by the different inquiries. At first sight, for instance, the series of substantives proposed by Mackey, Savard and Ardouin (1971) seem to include a much wider number of words than the 20 listed by Gougenheim et al. (1964) for each centre of interest. The quantity of words that have been commonly associated by the speakers to the topic “parts of the house”, for example, amounts to 372: nevertheless for 234 of them the coefficient of associative frequency is lower than 10. The nouns that are macroscopically associated by the whole sample to the same thematic field are, again, quantitatively limited.

⁴² The tendency to identify overarching conceptual formats seems to be confirmed by the fact that the diverse surveys actually investigate similar, or even identical, centres of interests. Moreover, as we have already showed (cf. fn. n. 37), they end up identifying overlapping nomenclatures.

Albeit mental and not textual occurrences are investigated here, the outcomes of the research on the available vocabulary are qualitatively limited to a just as general content lexicon as the one identified through statistical surveys (cf. §5.2).⁴³

6.3 Wide-ranging series of nouns: the LIP and other studies

The problem of determining basic content words isn't specifically tackled in the *Lessico di frequenza dell'italiano parlato* (cf. De Mauro et al. 1993): this is why this work doesn't discuss the notion of available vocabulary.

According to what has been observed so far (cf. §4), the LIP (cf. De Mauro et al. 1993) proposes a methodological refinement of the mere frequency parameter whereupon the *Français Fondamental (1er degré)* (cf. Gougenheim et al. 1964) had previously been compiled. The definition of the usage coefficient and, even before, of the complex dispersion index, aims in fact at the identification of a quantitatively wider and qualitatively more stable vocabulary core.

The issue of the non-determinability of basic linguistic meanings (cf. §5.4) is essentially embedded into a mere perfecting of the lexico-statistical frame of discussion: basic content words are to be included within a larger and more stable series of lexical items.

Nevertheless, the identification of wider quantities of words dispersed and used in the text of the *corpus* doesn't entail an actual overcoming of the problems already emerging in frequency lists like the FF1 (cf. Gougenheim et al. 1964). The LIP (cf. De Mauro et al. 1993) lists in fact an extensive string of lexical items – and among them a large number of substantives – that are ranked beyond the threshold of macroscopic usage values, these yet showing sporadic, irregular and semi-equipollent probabilities to recur in spoken texts (cf. §5.1 and §5.3).

Paradoxically, the inadequacy of quantitative methods to be employed to identify the content lexicon results even more clearly in the determination of larger word and substantive sequences than when it comes to more limited lists of simple frequency values. On a larger scale, the Italian usage list (cf. De Mauro et al. 1993) poses the same problem the FF1 (cf. Gougenheim et al. 1964) had already posed and tried to solve through the lexical-associative method: the LIP (cf. De Mauro et al. 1993) offers a different representation of this issue, though doesn't overcome it.

Its methodological frame rather represents the main trend of the discussion concerning content vocabulary in the investigations that have followed the French study (cf. Gougenheim et al. 1964), i. e. its absorption into mere quantitative experimentations: in fact, available words are no longer considered as an issue to be faced.⁴⁴

⁴³ The assumption implies that, after this general range of available words, content lexemes are equally non-determinable through the assignation of a recurring index, even if not statistical but mental. Many more words, in fact, concentrate within bands of low associative values.

⁴⁴ The only exceptions are represented by few studies that propose a pragmatically oriented reinterpretation of the original centres of interest (cf., among others, Raasch 1970; Galisson 1971, 1975) and by the *Vocabolario di base della lingua italiana* (cf. De Mauro 1980: 161–202). Inspired by Austin's (1962) and Searle's (1969) theories of communicative competence, the first surveys however don't succeed in substantially overcoming the nomenclative character of the thematic fields: the series of nouns they select are set in fact within lists of stereotyped, if not radically artificial sentences. On the other hand, the inquiry into the Italian available

The configuration of the recurrence values provided by the more recent lexico-statistical research is thus extremely close to the one observed in the LIP (cf. De Mauro et al. 1993). The series included between rank 1,900 and rank 2,000 of *A Frequency Dictionary of German* (cf. Jones/Tschirner 2006: 74–77), for instance, lists 55 substantives, which show frequency values included between 44 and 41, 21 of them sharing a coefficient of 43 and 20 of them showing a recurrence value of 42. The probability for a word like *Auswahl* ('choice') to appear in spoken texts isn't therefore macroscopically different from the one substantives like *Boot* ('boat'), *Gehirn* ('brain'), *PC* ('computer'), *Strategie* ('strategy') and *Wechsel* ('change') have. The frequency value of the latter, in turn, is extremely similar to the coefficient of nouns like *Bahnhof* ('railway station'), *Erwartung* ('expectation'), *Holz* ('wood') and *Schicksal* ('fate'). In addition, all these lexemes are ranked in low and irregular bands of frequency values.⁴⁵

6.4 Basic meanings in spoken language: on the supremacy of the discrete unit

On one hand, the attempts carried out to determine the basic content words of spoken language lead to the identification of nomenclatures including the few substantives whose lexico-associative frequency is least significant: that is, they stop before they end up representing the same vocabulary configuration lexico-statistical analysis had already showed (i. e. the rapid substantial decrease of the recurrence values and the low and irregular occurrence of concrete nouns, cf. §5.1 and §5.3). This is the case of the quantitatively and qualitatively limited series listed by the FF1 (cf. Gougenheim et al. 1964) and of the studies it has inspired (cf. Pfeffer 1964; Dimitrijevič 1969; Mackey/Savard/Ardouin 1971).

On the other hand, the exploitation of more extended textual samples like the one proposed by the LIP (cf. De Mauro et al. 1993) and more recent investigations like the one by Jones/Tschirner (2006), determine wide-ranging series of nouns showing the usual characteristics of the statistically investigated content lexicon: that is, once again, the sporadic and semi-equivalent recurring feature of concrete substantives (cf. §5.3) beyond the threshold of macroscopic recurring values (cf. §5.1). In the end they also determine quite restricted sequences of content words displaying a representative probability of recurrence.

As we can see, two different attempts of overcoming the same problem, and two different modalities of representing it, don't entail substantially different outcomes: in fact an appreciable threshold of basic spoken contents seems to be ultimately indeterminable from both per-

vocabulary (cf. De Mauro 1980: 161–202) represents an *unicum* from different points of view: for instance, it constitutes an integral part of a basic vocabulary that is essentially addressed to Italian native speakers rather than to foreign language learners; moreover, its investigation is based on a method different than the lexical-associative one (for a more detailed discussion of these aspects cf., among others, Gensini/Vedovelli 1983; De Mauro 1994). The Italian study is thus theoretically and methodologically incomparable with other surveys on available words.

⁴⁵ *A Frequency Dictionary of German* (cf. Jones/Tschirner 2006) actually investigates a sample only partially consisting of spoken sources (1,000,000 tokens out of 4,200,000, cf. *ibid.*: 2–3): so, to a larger extent, it can be considered representative of the more general non-quantitative determinability of content words (i. e. of the diminution and semi-equivalence of their recurrence values after the first 500, if possible after the first 1,000 ranks). As already explained, this tendency firstly emerges in investigations of written *corpora* and is then specifically themed in relation to spoken language (cf. §5.1, in particular fn. n. 20 and §5.4, in particular fn. n. 28 and n. 29).

spectives. When identified with concrete nouns, basic meanings in spoken language appear to elude any attempt of a systematic determination and description.

However, the partiality of these results can't be differently considered than as an unavoidable consequence of the theoretical fallacy characterizing both lines of research.

They converge in fact towards a likely interpretation of the meaning unit that turns out to be essentially segmental. Both research directions identify the unit of spoken content with a single or, even better, with a **discrete unit**: in particular with the occurrence of substantives that refer to concrete things. As a consequence, they end up proposing the hypothesis of a lexical type which is supposed to provide content information by means of its regular recurrences as a token in spoken texts.

Yet, by doing this, the discussion concerning basic meanings *de facto* contradicts the inherent character of the investigated phenomenon, i. e. the intrinsic sporadic nature of substantival occurrences in spoken language. This is actually the premise of the inquiries into the available vocabulary: the verification of the low and irregular recurrence of single lexical units providing content information and, in particular, of concrete substantives (cf. §5.3 and §5.4).

Studies like the FF1 (cf. Gougenheim et al. 1964) and the surveys on word associations (cf. Pfeffer 1964; Dimitrijevič 1969; Mackey/Savard/Arduoin 1971) reveal this theoretical flaw in that they resort to the speaker's psychic or mental dimension as the fundamental context of analysis (cf. §6.1). By asking the participants to associate words to merely potential, and not actualized, thematic fields, they alter – if not contradict – the constitutive character of the examined phenomenon: the situational anchoring of natural spoken language, in which contents only rarely appear as evident occurrences of concrete substantives (cf. §5.3 and §5.4).⁴⁶

Shifting their interest from the syntagmatic speech actualization to a conceptual context of inquiry, the surveys on available words are consequently destined to a representational paradox: the paradigmatic substantive series they identify are intrinsically improbable, both because their chance of effectively recurring in spoken texts is far from regular and because they would never appear in speech utterances in the order presupposed by the lists themselves.⁴⁷

So, as already observed, then, the LIP (cf. De Mauro et al. 1993) and the more recent lexico-statistical approaches (cf., among others, Jones/Tschirner 2006) rather embed the segmental vision of basic meanings into an improvement of the calculated quantitative coefficients (cf. §6.3). And in fact they only represent far more extended series of single words, and among these of substantives, proposing once again the same paradigmatic profile of the limited available nomenclatures: if it were not that such segmental units show surprisingly similar low probabilities of recurrence in speech.

To conclude, the discussion on basic contents ends up in a theoretical flaw and in a representational paradox, because it contrasts with its own theoretical assumption, investigates what

⁴⁶ In this sense, “[...] la notion de centre d'intérêt s'oppose à celle de *situation*” (Galisson/Coste 1976: 81, italics mine)./“[...] the notion of centre of interest is in opposition to that of *situation*”, meaning its mental or potential character contrasts with its actualization in real speech.

⁴⁷ In order to construct a hypothesis of content, the language user has not only to get by through many rare words but, rather, to get by through seldom items mostly co-occurring in limited text portions (i. e. in those sections that aren't covered by the most recurring and general words, cf. §5.3).

isn't constitutive of the analyzed phenomenon and, finally, describes a non-probable unit of spoken content.⁴⁸

Before the possibility of an alternative interpretation of the meaning unit is discussed, it may be thus worth considering more closely which ones are the things that are addressed in speaking: more precisely, how they are talked about and where they are actually to be found.

7 On the situated and shared praxis of lexical negotiation: the construction of spoken meaning

The participants in the dialogue analyzed below are talking about “a baby’s portable cot”: more exactly, they are assembling this object in a child’s bedroom and, while doing this, they are speaking about it. A part of the conversation reads as follows:

<Speaker 1> *It's not difficult as it first seemed*

<Speaker 2> *She says you've got to twist these round and it makes them solid or something*

<Speaker 1> *And all this just for you* [<Speaker 3> *Oh*] (laughs)

<Speaker 2> *There that's solid now*

<Speaker 3> *I think I've made it unsolid sorry I've done it the wrong way round I have I*
(3 secs)

<Speaker 2> *Solid*

(4 secs)

<Speaker 1> (laughs) (inaudible)

<Speaker 2> *Right now it's your end now*

<Speaker 3> *Oh I see right okay*

(4 secs)

<Speaker 3> *Not too much*

<Speaker 2> *There ... what's that in the middle*

(5 secs)

<Speaker 3> *Oh it's*

(2 secs)

<Speaker 1> *Found some more legs*

<Speaker 3> *Mm ... is it legs or is it erm*

(2 secs)

<Speaker 2> *It doesn't tell you what that is*

<Speaker 1> (laughs)

<Speaker 4> *Yeah that looks right surely*

<Speaker 2> *Yeah*

<Speaker 1> *Yeah well done*

<Speaker 3> *D'you like that*

<Speaker 1> *Yeah*

<Speaker 2> *Oh aye*

(cf. McCarthy/Carter 1997: 32–33)

⁴⁸ In general, the origin of such inconsistency can be seen in the very methodological foundation of lexicostatistical surveys: as already observed (cf. §4, in particular fn. n. 14) they investigate spoken language on the basis of its written transcription and of its following segmentation into tokens, these being constitutively discrete constituents.

The total number of tokens that recur in the dialogue section amounts to 107 occurrences. To a large extent they are grammar or structure words as prepositions (*in, for*), conjunctions (*or*), adverbs (*well, not, too, much, now, there, just, surely*), pronouns (*I, it, you, she, them*), flexed forms of auxiliary verbs (*is, have, doesn't*), possessive (*your*), demonstrative (*that*) and indefinite (*some, something*) adjectives and pronouns. Several lexical items within them are discourse markers that affect the attitude of the speakers towards the utterance by expressing, for instance, their greater or lesser degree of certainty (*seemed, think*). More generally these items stress the interpersonal dimension of the speech (*Oh I see right okay, mm, yeah, oh*). Furthermore, diverse deictic words (*you, I, your, that, there, now*) are included within this grammatical structural vocabulary.

The proper content lexicon counts exactly 25 tokens: to a large extent it also comprises quite general words, whereof adjectives like *difficult, wrong*, substantives like *way, end, middle*, verbs like *tell, make* and *do* offer some examples. Adjectives like *solid* and *unsolid*, verbs like *twist* and nouns like *leg* are within the few lexemes that seem to refer to a more circumscribed content: nevertheless they represent only sporadic occurrences in comparison to the more frequent and extended general lexicon (e. g. *solid* occurs four, *legs* recurs two and *twist* only one time). By means of these low and irregular frequency values one wouldn't understand that the participants in the dialogue are talking about a "baby's portable cot": their rare and unsystematic recurrence character wouldn't make clear that this is the **thing** the interlocutors are talking about within the whole utterance.

Even more precisely, the **content** of the spoken interaction wouldn't be made explicit simply because the **meaning** 'baby's portable cot' never recurs within the dialogue. The whole spoken utterance is based on, and concerns, the **object** "bed for a baby to be assembled", and yet this never occurs as an evident substantival token. As already mentioned, the recurrence values of those nouns that would contribute to better circumscribe the content of the conversation, such as *legs (of the cot)*, recur rarely and unhomogeneously too.

From a solely lexico-statistical perspective the meaning of the dialogue appears to be constitutively indefinite, especially if the recurrences of concrete nouns are to be considered. In this regard the spoken interaction turns out to be lexically vague or, put otherwise, characterized by a constituent low lexical density (cf. Halliday 1985).⁴⁹

The partners involved in the conversation don't need to explicit the object they are speaking about by naming it (thus producing substantival occurrences of the corresponding word), because this same object belongs to the very joint context in which the speakers are acting. In other words: the thing they are talking about is a constitutive part of the **situation** they are verbally experiencing.

It is worth noting that here the situational frame can't be understood as the mere space-time context in which the spoken utterance takes place. It must rather be conceived as a complex set of interrelating factors, surely including the above-mentioned time and space of the speech event, yet extending to further elements like the social hierarchy of the involved partners, the

⁴⁹ A previous discussion of this character of spoken utterances is offered by De Mauro (1970). Among others, the issue has been further argued by Vedovelli (1995). Likewise, McCarthy and Carter (1997: 33) resort to the same terms with reference to the dialogue analyzed above.

prosodic, mimic and gestural aspects as well as the pragmatic dimension. Among others, factors like the inference mechanisms, the dynamics of the contextual and co-textual cross-references, the modality of the verbal exchange, as well as the interrelationship existing among them, can be understood as elements configuring the situation as well. In turn, the situational frame constitutes a manifold system by means of which spoken verbal symbols do “anchor” to the speakers’ experienced world: consequently, by means of which they can have a meaning.⁵⁰

On the premises given so far, and recalling again Bühler’s (1934) position, spoken language can be understood as the indissoluble synthesis between a “verbal-symbolic field” and an “indexical field”: as the crucial intersection between the level of lexical occurrences and their frequency on one hand and the set of tools of referring to the extra-symbolic context in which the spoken interaction develops on the other.

It may be worth stressing that by the term *extra-symbolic field* a dimension different – and in no way secondary – from the verbal-symbolic one is meant. As has become evident here, the extra-symbolic field must indeed be understood as an internal, inherent or intrinsic feature of the spoken modality of language: as a datum without which this same modality couldn’t exist or, even more, it wouldn’t mean anything. The concentration of several deictic words in the examined dialogue can be considered as a clear evidence of this aspect: without a situational embodiment, i. e. a “here and now” in which the speakers interact and make a shareable sense of them, these symbols wouldn’t mean anything at all.⁵¹

By virtue of its semiotic peculiarity, spoken language turns out to be “compacted” or “economized”: that is, released from the necessity to name and make explicit the things or the objects that are talked about. In line with Bühler’s (1934) observations, a first discussion of these aspects has been offered by Bally (1952). According to the remarks of the Swiss linguist, the distinctiveness of spoken language is due to the fact that, within it,

[...] l’échange des idées [...] est encadré par une situation que les interlocuteurs trouvent toute faite: entourage matériel, choses connues des intéressés, rapports familiaux ou sociaux, communauté d’intérêts, etc. L’énonciation en est considérablement facilitée et abrégée. Cette économie de l’effort est refusée à la langue écrite; elle doit, dans chaque cas, se créer sa situation par des procédés artificiels, des combinaisons plus ou moins compliquées.

(Bally 1952:

105)⁵²

⁵⁰ The internal complexity of the situational frame has been discussed, among others, by De Mauro (2002: 50–57) and, more recently, by Albano Leoni (2009: 13–16).

⁵¹ It is no coincidence that the terms *deixis* and *field of indication* are used by Bühler (1934) as nearly synonymous terms, since deictic markers are considered as the main lexical category configuring the indexical field. A recent consideration of Bühler’s perspective on spoken language has been offered again by Albano Leoni (2009): the fertility of Bühler’s assumptions has been particularly developed by the scholar in order to understand the phonological vagueness of spoken utterances.

⁵² “[...] the exchange of ideas [...] is framed within a situation that the interlocutors find as already given: material environment, things known by the involved speakers, family or social relations, common interests, etc. This considerably facilitates and abbreviates the utterance. Such economy of effort is denied to written language; this has, in any case, to create its situation by means of artificial procedures, of more or less complicated combinations”.

So if contents aren't given by explicitly naming the objects that are concerned in a conversation, then how are they conveyed and used? The hypothesis that proves to be plausible is the one according to which the transmission of contents in spoken language doesn't exhaust itself in the occurrences of a segmental or a discrete nominal unit but, exactly the opposite way, it extends to a much wider dimension: indeed the transmission of spoken contents unveils as a holistic process, as an activity or, again, as enunciative-conversational **praxis**.

What are the participants in the analyzed dialogue actually doing? They are co-acting and interacting in a linguistically-experienced, and probably common, space of life: in a verbally-**shared** daily situation. Most of this activity takes shape through a general and indexical vocabulary: that is, by means of the continuous **lexical negotiation** that develops through it. Within this practice the things that are spoken about are continually configured and re-defined, evoked and recalled, re-formulated and re-interpreted. The interlocutors aren't actually naming the objects they are talking about: they are rather handling them, shaping them – they are linguistically **constructing** them. Considered in this perspective, spoken contents don't seem to correspond to any entity but, instead, to a process of meaning configuration.

8 Towards an alternative definition of the meaning unit: the “path of habitudinary spoken content”

If the participants in the dialogue analyzed so far (cf. §7) had been talking about something else than a “baby's portable cot” – about a “tie”, for instance – their interaction would have presumably developed in a likely manner. Most probably, the recurrences of the substantive *tie* would have been rare and irregular. The occurrences of concrete nouns that help to clarify the item being spoken about would have presumably been seldom and unsystematic, too.⁵³ On the contrary, the speakers would have seemingly resorted to the situated and shared praxis of lexical negotiation: that is, to the prevalent use of a general and indexical vocabulary and, by means of it, to the constant configuration and definition of the thing being spoken about.⁵⁴

Suppose, then, the participants in the dialogue concerning the “baby's cot” had never talked about such a thing: they would have probably negotiated it once again, namely in a manner similar to the one through which they have negotiated other items in other shared and situated interactions focusing on other objects. In short, they would have resorted for the umpteenth time to the praxis of meaning configuration.

⁵³ It is quite probable that the occurrences of the substantives potentially associable to the centre of interest “clothes”, would have been sporadic and non-homogeneous exceptions.

⁵⁴ Such an outcome has been pointed out somewhere else by the author of this paper (cf. Massa 2013: 82–85). Although entirely focused on the object “Krawatte” (‘tie’), the analyzed conversation didn't include any single occurrence of the corresponding concrete noun *Krawatte* (‘tie’): the highest text coverage was provided in fact by the general and indexical lexicon (e. g. by the verbs *haben* /‘to have’, *sein* /‘to be’, *sehen* /‘to see’, *reden* /‘to talk’ /‘to speak’, by personal pronouns as *ich* /‘I’, *du* /‘you’, *wir* /‘we’, *ihr* /‘you’, by the articles *der* /‘the’, *das* /‘the’, the prepositions *auf* /‘on’ and *von* /‘of’, the averbs *hier* /‘here’ and *auch* /‘also’, the noun *Sache* /‘thing’). The thing referred to in speaking belonged to the situation shared by the partners who, consequently, didn't need to explicitly name it. It is no accident that the substantive *Krawatte* (‘tie’) can be excluded from the lists of the most frequent German words as, for instance, the one by Jones/Tschirner (2006): as has become evident here, its explicit recurrence in spoken texts can be extremely irregular. Nevertheless, this does not mean that a “tie” is something not being spoken about.

If their modality of manifestation is considered, things seem to be the most variable, and consequently the non-determinable part of spoken interactions: such a modality doesn't correspond to a discrete-substantival configuration but rather to an activity of situated and shared construction of the same objects that are addressed in speaking (cf. §6.4 and §7).

Moreover, and in a certain way even more interestingly, things and their corresponding concrete nouns seem to be non-determinable if they are considered as the very entity – i. e. unit – of spoken content. Among the premises to the study of the available vocabulary, the French school (cf. Gougenheim et al. 1964) had already stressed this point:

Prenons le mot *fourchette*. Voilà bien, dira-t-on, un mot qui doit être fréquent: nous manions cet instrument deux fois par jours, un enfant de trois ans, doué d'une intelligence normale, sait ce que désigne ce mot. Mais quand le prononçons-nous? Quand nous disons à un enfant: «Ne laisse pas tomber ta fourchette» et dans telle autre circonstance. Mais nous pouvons rester des jours et des semaines sans le prononcer.

(Gougenheim et al. 1964: 138–139, italics in original)⁵⁵

On the whole, if on one side things in themselves undergo an irregular recurrence and substantival manifestation, on the other something else proves to be far more uniform: what seems to be a stable and consistent activity is in fact the very process of lexical negotiation.

In other words, the procedural modality by means of which the shared situations of our daily space of life are acted and experienced, and within which things are included, appears to be a homogeneous practice.⁵⁶

Therefore, our ordinary way of using spoken language turns out to be the essential one through which our common life is configured and lived: as a general semiotic **mode** (*modus*) of “making sense of our life”. The process of meaning configuration that has been dealt with in this work (cf. §7), has thus to be mainly understood as the construction of the sense – i. e. of the meaning that, while and by speaking – we're constantly attributing to our everyday existence.

The things – the contents – that are talked about, coincide with ourselves and our being in the world⁵⁷.

⁵⁵ “Take the word *fork*. In short, one would say, a word that has to be very frequent: we deal with this tool two times a day, a three-year-old child showing a normal intelligence knows what this word designates. But when do we pronounce it? When we say to a child: “Don't drop your fork”, and in other similar circumstances. But we are likely not to be pronouncing it for several days or weeks”. Further observations concerning this character of concrete objects and words can be found in those according to which, for instance: “[...] nous nous servons tous les jours de nos *dents*, nous n'en parlons que quand nous en souffrons. Les Parisiens utilisent tous les jours le *métro*, mais en parlent-ils-constamment? Peut-être, à la rigueur, prononcent-ils son nom une ou deux fois par jours. [...] Le hasard veut que je rencontre un ami au moment où il sort du cabinet du dentiste, il parlera de ses *dents*. Un incident ou un accident dans le *métro* provoquera soudain l'emploi de ce mot (ibid.: 139, italics in original).” “[...] we use our *teeth* every day, we talk about them only when we have a toothache. Parisians use every day the *underground*, but are they constantly talking about it? At most, maybe, they pronounce it once or twice a day. [...] By chance, I might meet a friend in the moment when he is getting out of the dentist's, he will talk about his *teeth*. An incident or a accident on the *underground* will cause the sudden use of this word”.

⁵⁶ In some preliminary reflections concerning the definition of the spoken semantic unit (cf. Massa, in print), the same process of lexical negotiation has turned out to be “[...] più un modo di esperire e condividere le cose che un denominare le cose stesse”./ “[...] a way of experiencing and sharing things rather than one of naming things themselves”.

Some kind of uniformity in spoken language is therefore likely to be identified in this solely procedural perspective, and it is in this same perspective that an alternative and more plausible definition of the spoken semantic unit is to be formulated: this shall be understood as a unit of the practice of meaning construction, as the basic modality of making sense of our common life, by speaking.

To conclude, the meaning unit in spoken language unveils as the fundamental manner of the current praxis of configuration, experience and re-formulation of the most central threads of content of our subjective and intersubjective world: as the elementary mode of our regular and holistic discourse process.

Such a unit can be understood as a **path of habitudinary spoken content**: ultimately, as a story.⁵⁸

9 The path of meaning as a narrative semantic unit: or why we tell a story every time we are speaking

We can consider spoken language and the paths of meaning as its main modality of fulfillment, as the form by means of which our habitudinary life unfolds.

In turn, our very modality of experiencing and knowing our ordinary subjective and intersubjective life unfolds in a narrative form.

The narrative principle that characterizes the human species consists of weaving our common space of life into stable or customary routes of content: these are habitudinary wefts of life to which new data of experience can be traced back and thereby our knowledge can be integrated. Narrativity unveils as the configuration of a constant plot of experience and knowledge by means of which we can make hypotheses on the course of events and on the roles that single events have for one another.

At the same time, our way of configuring our usual personal and interpersonal dimension is a gestalt modality: indeed it is based on complexes of stereotyped actions that occur in routinized scenes and wherein conventional roles are in action.

Thinking through telling thus consists of assigning a causal order to experienced actions and events, so that they are perceived as interconnected in a meaningful and not in a merely casual way. This manner roughly includes an initial situation, a complication and a solution: briefly, the very framework of a story.

By means of such a narrative mode of thinking we continuously assign a plausible meaning to the world, configure an understandable version of ourselves and our life and imagine further possible worlds.

⁵⁷ As has been observed somewhere else (cf. Massa, in print), spoken language can be understood as “[...] la modalità [...] attraverso cui i contenuti vengono comunemente esperiti e condivisi: più semplicemente, attraverso cui vengono vissuti”/“[...] the modality [...] through which contents are commonly experienced and shared: more simply, through which they are lived”.

⁵⁸ Such a definition of the semantic unit has been firstly sketched (cf. Massa 2013: 159–162) and more recently formulated (Massa, in print) by the author of this paper.

Narrativity can be furthermore understood as the primary tool through which we can avoid “getting lost” in the continuous flow of perceptions and actions but, instead, we can interpret and manage this same flow on the basis of previously experienced sequences of events, plots or, in other terms, of stories. From this perspective, narrativity is an extremely powerful vehicle for constructing our autobiographical and shared memory: for building up the story we are a part of or, put otherwise, for entering and inhabiting the culture we belong to.⁵⁹

The human peculiarity of producing and using verbal symbols accomplishes exactly the task of weaving, i. e. of reiterating, stabilizing and thus structuring the *continuum* of our experience into stable paths, these being in turn experienced every time in the same or, at least, in a similar form. Verbal languages are the fundamental instrument of configuration and experience of our everyday single and shared life: consequently they are the chief tool through which our narrative cognition unfolds.⁶⁰ As Bruner (2002: 3) has underlined, “Siamo così bravi a raccontare che questa facoltà sembra ‘naturale’ quasi quanto il linguaggio”.⁶¹

An essential manifestation of the structural connection between narrativity and language is offered by the process of linguistic ontogenesis: in fact the first words are uttered by the child to map the first sequences of actions and events that articulate his/her *routine*, his/her first stories of life. These are to be understood as constitutively interactive, shared or social events: according to the Soviet psychologist Vygotskij (1960 [1997]), again, as “natural forms”.

Learning to mean turns out to be an intrinsically socio-pragmatic praxis (cf., among others, Bruner 1978, 1983, 1987; Tomasello 2003). Finally, the essential modality through which it unfolds seems to also coincide with the path of linguistically inter-acted experience: that is, with a sequence of spoken habitudinary actions and events.⁶²

Further inputs into a narrative definition and comprehension of spoken meaning systems can be traced in their modality of mental configuration, in the sense that the lexemes that belong to the same plot, or that allow to reconstruct one, are more easily associated and associable to one another. First suggestions in this respect are offered by the notions of “associative relations” by Saussure (1916) and of “associative field” by Bally (1940), which can in fact be understood as paths of meaning that branch off from a sign and arrange themselves around it, equaling the beginning or *incipit* of a story that has a linguistic content (cf. Massa/Simeoni 2014: 84). Finally, they constitutively unveil as a path of stable linguistic content, too.⁶³

⁵⁹ Essential reference studies on the topics that have been introduced here are in Bruner (1986, 1991, 2002, 2004), Dennet (1992), Gazzaniga (2000), Gazzaniga, Russel and Senior (2009), Halbwachs (1952 [1992]), Hinchman and Hinchman (1997), Mitchell (1981), Nelson (1986), Polkinghorne (1988), Ricoeur (1981, 1984), Siegel (1999), Smorti (1994, 2007) and White (1980).

⁶⁰ A discussion concerning the principle of narrativity as a cognitive-experiential, and thus as a linguistic-cultural mode, has been recently carried out by Massa and Simeoni (2014).

⁶¹ “We are so good at telling that this faculty seems almost as ‘natural’ as language”.

⁶² Clear evidence of the narrative entity of verbal systems is provided by the studying of the connection between language acquisition and the emergence of memory, i. e. of the configuration of his/her personal and social identity by the child (cf., among others, Fivush/Nelson 2004; Pasupathi 2001; Pasupathi/Hoyt 2009). Further evidence of the ontogenesis process as a narrative practice has been given by Basile (2010, 2012). On the linguistic-habitudinary ontogenesis cf. also Massa/Simeoni (2014: 80–83).

⁶³ The hypothesis of “local holism” by Kittay and Lehrer (1992) seems also to converge towards a lexical-narrative hypothesis. It corresponds in fact to a re-definition of the traditional notion of semantic field, which can

Early hints at the intrinsic narrative character of verbal systems come again from the reflections by Saussure (1916) who underlines the need for the same systems to be embodied in the habits of the speaking mass in order to make sense. From this point of view,

[...] la langue ne se présente pas comme un ensemble de signes délimités d'avance, dont il suffirait d'étudier les significations et l'agencement; c'est une masse indistincte où l'attention et l'*habitude* peuvent seules nous faire trouver des éléments particuliers.

(Saussure 1916: 150, italics mine)⁶⁴

Besides, the observations of the Swiss linguist are the result of his interest for the radical arbitrariness of the linguistic sign, this coinciding with the definition of linguistic meanings as a necessarily social form (cf. *ibid.*: 99–110, 161–175).⁶⁵ Saussure's (1916) reflections are finally addressed to the definition of spoken language as the very object of linguistics and thus as arbitrary or social value (cf. *ibid.*: 45–48): consequently as a necessarily habitudinary, and thereby narrative system.⁶⁶ As has become evident so far (cf. §8), the principal mode by means of which the spoken social form unfolds can be identified in the path of linguistic habitudinary content.

The way spoken meanings are used isn't thus different from the one they are shaped and configured: insofar as they are considered with reference to their modality of forming, of configuration and use, spoken meanings disclose through the form of sequences, i. e. of paths of linguistically experienced and known consuetudinary life: hence they prove as constitutively narrative tools and products. In this sense the structural connection between narrativity and language can be conceived even in terms of an identity relation.

hence be understood as “[...] a set of lexemes [...] applied to some content domain (a conceptual space, an experiential domain, or a practice)” (*ibid.*: 3), having again the configuration of a scheme, a frame, a narrative script. Similar formulations concerning the mental narrative conformation of linguistic meanings are offered by Violi (1997) and Basile (2001).

⁶⁴ “[...] language does not offer itself as a set of predelimited signs that need only to be studied according to their meaning and arrangement; it is a confused mass, and only attentiveness and *familiarization* will reveal its particular elements” (trans. Baskin 1959: 104, italics mine).

⁶⁵ For instance, “La langue [...] est à chaque moment l'affaire de tout le monde; [...] elle est une chose dont tous les individus se servent toute la journée. [...] Elle fait corps avec la vie de la masse sociale, et celle-ci, [...] apparaît avant tout comme un facteur de conservation. [...]. C'est parce que le signe est arbitraire qu'il ne connaît d'autre loi que celle de la tradition, et c'est parce qu'il se fonde sur la tradition qu'il peut être arbitraire” (*ibid.*: 109–110)./“Language [...] is at every moment everybody's concern [...]; language is something used daily by all. [...] It blends with the life of society, and the latter [...] is a prime conservative force. [...] Because the sign is arbitrary, it follows no law other than that of tradition, and because it is based on tradition, it is arbitrary” (trans. Baskin 1959: 73–74). The inherent social, and consequently consuetudinary character of verbal systems is stressed by Bally (1952) as well: Saussure's disciple highlights in fact the necessity of the linguistic signs of being “unified” in order to carry out their social task (cf. *ibid.*: 46).

⁶⁶ “Langue et écriture sont deux systèmes de signes distincts; [...] l'objet linguistique n'est pas défini par la combinaison du mot écrit et du mot parlé; le dernier constitue à lui seul cet objet” (*ibid.*: 46)./“Language and writing are two distinct systems of signs; [...] The linguistic object is not both the written and the spoken forms of words; the spoken forms alone constitute the object” (trans. Baskin 1959: 23–24). Likely observations can be identified in the reflections by Bally (1952), according to whom “[...] tout d'abord la langue parlée, par opposition à la langue écrite, est considérée comme l'unique objet de l'étude linguistique” (*ibid.*: 151)./“[...] first of all spoken language, in opposition to written language, is considered as the sole object of the linguistics study”.

The definition of the meaning unit in spoken language as a narrative unit must consequently be understood firstly in relation to its unity of identity, being its structural mode of unfolding a procedural, i. e. sequential manner: this can be identified in the path of habitudinary content, unveiling hence as a narrative semantic unit.

As far as the issue of this paper is concerned, and on the basis of what has been discussed so far (cf. §7 and §8), the discussion referring to the basic modality of use of spoken meanings seems to converge once more towards the definition of its holistic, non-discrete, and thus procedural unit: in addition to that, towards its local definition.

Some kind of stability in the use of spoken language doesn't seem to be traceable otherwise than in the habitudinary character of the shared situations of our life, these tending since their earliest phases to recur regularly and to be regularly experienced in and by speaking.⁶⁷ Once more, the semantic spoken unit turns out to coincide with the habitudinary configuration of experience that is carried out in these situations through the constant process of lexical negotiation.

Furthermore, ours shall be a pragmatic definition: in the same way as by the ontogenetic process, in fact, the normal use of spoken language is always embedded in a socio-relational context and related to the achievements of goals.⁶⁸

A plausible definition of the spoken meaning unit shall be also a syntagmatic one, being the local configurations of *parole* the only semantic manifestations we have proof of.⁶⁹ Besides, any attempt of a paradigmatic description ends up in a theoretical flaw and in a representational paradox: either way they are ineffective to deal with the longstanding question of contextual variability.⁷⁰

The unit of signification in spoken language coincides with the speakers' everyday activity of reiterated construction and experience of their most central routes of content, "if not in treading the boards of our [...] inter-subjective life *discourse*" (Massa/Simeoni 2014: 85, italics in original).

⁶⁷ It is no accident that the notion of "shared situation" is firstly used to refer to the relational contexts within which the child speaks his/her first words (cf. Basile 2010, 2012). A precedent can be identified in the definitions of "format" (cf. Bruner 1983) and of "joint attentional frame" (cf. Tomasello 1999, 2003). As has been observed, spoken language is a narrative system firstly if its structural unity is considered.

⁶⁸ Basile (2005) has underlined the continuity relation between the development of the semantic and the pragmatic competence in the process of language acquisition.

⁶⁹ According to Bally (1952), these are the very forms of being of the language: in fact, "[...] *parole désigne le fonctionnement pur et simple de la langue*" (ibid.: 76, italics in original)./"[...] *parole designs the pure and simple functioning of the langue*". Likewise Saussure (1916) underlines that "[...] la langue est nécessaire parce que la parole soit intelligible et produise tous ses effets; mais celle-ci est nécessaire pour que la langue s'établisse" (ibid.: 38)./"[...] language is necessary if speaking is to be intelligible and produce all its effects; but speaking is necessary for the establishment of language" (trans. Baskin 1959: 18).

⁷⁰ For a synthesis of these aspects cf. Violi (2003). Especially if confronted with the so-called scientific "paradigmatic thought", narrative thought seems indeed to have a horizontal and syntagmatic orientation, because it concerns the causal connection and organization among the parts of a whole. An essential confrontation of the two modes of thought and discourse is provided by Lyotard (1979 [1984]), whereas a further discussion is offered, among others, by Bruner (1990). Nevertheless, the compenetration of the two modalities of thought across disciplines is discussed, for example, by Smorti (1994).

As already observed by Wittgenstein (1953: §23, italics in original), “[...] the *speaking* of language is” exactly “part of an activity, or of a form of life”.⁷¹

By means of the paths of spoken content we give an order to our experience, we make it consuetudinary and, to close the loop, we weave it into stories that become the form of our habitudinary life: our *habitus* or our story, our common linguistic story.⁷²

10 The semantic unit in spoken language: some conclusions and further questions

The results achieved by lexico-statistical researches have constituted the starting point for the discussion developed in this paper. In a certain way, it is as we had looked at their limits from another, and quite different, perspective: not focusing, indeed, on what they don't allow to understand of the functioning of spoken language but, oppositely, on what they permit to comprehend as far as this same aspect is concerned. At the same time, our discussion hasn't concentrated on what the lexico-statistical outcome has failed to identify and to describe but, instead, on what this same lack can reveal about the very essence of spoken language.

From a far more practical point of view, that is from the perspective of a language user, the reflections carried out in this paper have focused on what the outcome of lexico-statistics allows to do or, even better, to “construct”: our ordinary way of inhabiting the world or, in other words, our habitudinary form of life. Considered in this perspective, the final results of lexico-statistical investigations prove as particularly productive.

Such an aim of the analysis has clearly coincided with the theoretical discussion concerning the semantic identity and, along with it, the semiotic peculiarity of spoken language: the relevance of such an examination has consequently turned out to be still very urgent. The procedural or constructivist perspective on such issues has disclosed its fecundity: indeed, the alternative definition of the semantic unit in spoken language doesn't coincide with the definition of an entity but, instead, with the definition of a unit of the praxis of meaning. The goal of the discussion developed here hasn't been therefore the identification of the basic words of con-

⁷¹ Within the reflections collected in the *Philosophical Investigations* (Wittgenstein 1953), various references to the habitudinary agency and its role as an element that unifies the same form of linguistic life can be found. In order to produce meaning, in fact, signs and their combinations have to be consuetudinary used. Suppose, for example, “I say the sentence: “The weather is fine”; [...] so let's put “a b c d” in their place. But now when I read this, I can't connect it straight away with the above sense. — I am not used, I might say, to saying “a” instead of “the”, “b” instead of “weather”, etc. . [...] (I have not mastered this language.)” (ibd.: §508). The customary agency is furthermore supposed to be inherent to the functioning of any other semiotic system. If “I am not used to measuring temperatures on the Fahrenheit scale. [...] such a measure of temperature *'says'* nothing to me” (ibd., italics in original).

⁷² The concept of *habitus* is understood here as the “socially embodied order” theorized by Bourdieu (1979 [1984], 1980 [1990]) and, along with such a definition, as a “linguistically embodied order”. As the result of a regular practice that creates and sediments habits, these coinciding with our practical way of inhabiting the world, the notion of *habitus* entails a structural affinity with the concept of story: this is in fact the result of telling conceived as a practice of weaving contents into plots and thus depositing layers of meaning. The *habitus* is consequently a “[...] modo di essere, ossia tutto ciò che noi siamo [...] soliti ad *avere con noi*, a portarci dietro continuamente” (cf. ETIMO, italics in original). “[...] a way of being, that is to say everything we are [...] used to having *with us*, to carrying continuously”. Last but not least the *habitus* can be understood in terms of ‘appearance’, this coinciding in turn with a set of stable characters and physical traits that materially show an individual or a group's way of being and habits.

tent in spoken language but, rather, the clarification and understanding of what it is essentially like to mean through spoken language itself.

Hence, the necessity for the semantic reflection to exceed a merely linguistic, i. e. symbolic level, has emerged. In this respect, the potentialities of a wider cognitive and linguistic perspective on why and how spoken language is used have become evident. The narrative approach adopted here aptly enfoldes this very necessity: as a structurally constructivist and procedural perspective, in fact, it allows to understand the complexity of the spoken modality of language as the fundamental form of human life.

The field of foreign language didactics, and more precisely its still urgent need for streamlined methods for teaching and learning, has actually represented the thread of the discussion carried out above and, in a certain way, the field to which these final considerations would like to look at.

In this regard, further challenges shall concern the usability of the narrative semantic approach and its potentialities in the structuring of the programs for foreign language teachers and learners. From a more concrete perspective, the main issue to be dealt with shall concern the possibility of a procedural didactics that aims to construct a learners' habitudinary space of life in and by speaking the foreign language. At this point, some important inputs come from the approaches that focus on the configuration of the speakers' memory in and by using the foreign verbal system. This should become a crucial field of research for further experiments and achievements concerning the definition and the usability of a narrative-based didactics.

In such a field, the use of *corpus*-based methods of investigation would be desirable as well: the collection of samples of memories told by foreign language learners and the inquiry into salient plots of remembering could represent a first issue to be explored. The analysis should tend towards the identification of macroscopic tendencies displayed by the learners within the processes of identity construction, and thus of telling about their life story, in and through the foreign verbal system. Similarly, outstanding patterns could be investigated in relation to the processes of configuration, and again of telling about their present and future life story: as a consequence, the constitution of corresponding *corpora* would be analogously advantageous.

Apart from being investigated as the process of configuration of the personal identity in a foreign language, the narrative praxis could also be examined as the essential procedural tool used by learners to experience, give a meaning to actions and share an habitudinary space of life: again, as the typical modality of weaving a life story. In connection with this, the identification of prototypical narrative frames and features could benefit from the potentialities offered by corpus-based applications as well.

At the same time, the risk for a narrative approach to end up for the umpteenth time in a mere nomenclature of stories (i. e. in a list or handbook of communicative daily contexts), should be avoided. The modality of a plausible representation of the semantic narrative unit and of its consequent usability clearly constitutes one of the main issues to be dealt with in the future.

That's how it works with stories, since they are never told for themselves: instead, they are part of a story that has already been narrated and entails the beginning of another story to be told.

This is what we have tried to do here.

References

- Albano Leoni, Federico (2009): *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: Il Mulino.
- Austin, John L. (1962): *How to do things with words*. Oxford/New York: Oxford University Press.
- Bally, Charles (1940): « L'arbitraire du signe. Valeur et signification ». *Le Français Moderne* 8: 193–206.
- Bally, Charles (1952): *Le langage et la vie: linguistique générale et stylistique, mécanisme de l'expressivité linguistique, langage transmis et langage acquis, la contrainte sociale dans le langage, l'enseignement de la langue maternelle, F. de Saussure et l'état actuel des études linguistique*. 3. édition augmentée. Genève/Droz/Lille: Giard.
- Basile, Grazia (2001): *Le parole nella mente. Relazioni semantiche e struttura del lessico*. Milano: Franco Angeli.
- Basile, Grazia (2005): « Può darsi una semantica senza pragmatica o viceversa? Cosa accade nel processo di denominazione ». In: Frigerio, Aldo/Raynaud, Savina (eds.): *Significare e comprendere. La semantica del linguaggio verbale. Atti dell'XI Congresso nazionale della Società di Filosofia del linguaggio (Milano, 16–18 settembre 2004)*. Roma, Aracne: 67–84.
- Basile, Grazia (2010): « I percorsi del senso. In che modo il senso prende forma in parole ». *Bollettino di Italianistica VII/I n.s.*: 9–29.
- Basile, Grazia (2012): *La conquista delle parole. Per una storia naturale della denominazione*. Roma: Carocci.
- Biber, Douglas/Conrad, Susan/Reppen, Randi (1998): *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bongers, Herman (1947): *The History and Principles of Vocabulary Control as it affects the Teaching of Foreign Languages in general and of English in particular*. Woerden: Wocopi.
- Bortolini, Umberta/Tagliavini, Carlo/Zampolli, Antonio (1971): *Lessico di frequenza della lingua italiana contemporanea (LIF)*. Milano: Garzanti.
- Bourdieu, Pierre (1979): *La Distinction. Critique sociale du jugement*. Paris: Les Éditions de Minuit. (Trans. Nice, Richard (1984): *Distinction: A Social Critique of the Judgment of Taste*. Cambridge/Mass.: Harvard University Press).
- Bourdieu, Pierre (1980): *Le Sens pratique*. Paris: Les Éditions de Minuit. (Trans. Nice, Richard (1990): *The Logic of Practice*. Stanford: Stanford University Press).
- Bruner, Jerome S. (1978): « Dalla comunicazione al linguaggio: una prospettiva psicologica ». In: Camaioni, Luigia (ed.): *Sviluppo del linguaggio e interazione sociale*. Bologna, Il Mulino: 75–112.
- Bruner, Jerome S. (1983): *Child's Talk. Learning to Use Language*. New York: Northon & Company.
- Bruner, Jerome S. (1986): *Actual Minds, Possible Worlds*. Cambridge/London: Harvard University Press.

- Bruner, Jerome S. (1987): "The Transactional Self". In: Bruner, Jerome S./Haste, Helene (eds.): *Making Sense. The Child's Construction of the World*. London, Methuen & Co.: 81–96.
- Bruner, Jerome S. (1990): *Acts of Meaning*. Cambridge/Mass: Harvard University Press.
- Bruner, Jerome S. (1991): "The Narrative Construction of Reality". *Critical Inquiry* 18/1: 1–21.
- Bruner, Jerome S. (2002): *La fabbrica delle storie. Diritto, letteratura, vita*. Roma/Bari: Laterza.
- Bruner, Jerome S. (2004): "Life as Narrative". *Social Research* 71/3: 691–710.
- Bühler, Karl (1934): *Sprachtheorie: die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Cheydleur, Frédéric D. (1929): *French Idiom List based on a Running Count of 1,183,000 Words*. New York: MacMillan. (= Publications of the American and Canadian Committees on Modern Languages 16).
- Chomsky, Noam (1957): *Syntactic Structures*. The Hague/Paris: Mouton & Co.
- Crystal, David (1987): *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Dennett, Daniel (1992): "The Self as a Center of Narrative Gravity". In: Kessel, Frank S./Cole, Pamela M./Johnson, Dale L. (eds.): *Self and Consciousness: Multiple Perspectives*. Hillsdale, Erlbaum: 103–115.
- De Mauro, Tullio (1961): « Statistica linguistica ». In: *Enciclopedia italiana. III Appendice, Vol. II*. Roma, Istituto della Enciclopedia Italiana: 820–821.
- De Mauro, Tullio (1970): « Tra Thamus e Theuth. Uso scritto e parlato dei segni linguistici ». *Bollettino del Centro di Studi Filologici e Linguistici Siciliani* 11: 167–179.
- De Mauro, Tullio (1980): *Guida all'uso delle parole*. Roma: Editori Riuniti.
- De Mauro, Tullio (1994): « Com'è nato il vocabolario di base ». In: Thornton, Anna M./Iacobini, Claudio/Burani, Cristina (eds.): *Una base di dati sul Vocabolario di Base della lingua italiana*. Roma, Istituto di psicologia del CNR: 51–55.
- De Mauro, Tullio (2002): *Capire le parole*. 2. edizione. Roma/Bari: Laterza.
- De Mauro, Tullio (1999): *Grande dizionario italiano dell'uso*. Torino: Utet.
- De Mauro, Tullio et al. (1993): *Lessico di frequenza dell'italiano parlato (LIP)*. Milano: Etaslibri.
- Dimitrijevič, Naum (1969): *Lexical Availability. A New Aspect of the Lexical Availability of Second School Children*. Heidelberg: Groos.
- DWDS (*Das digitale Wörterbuch der deutschen Sprache*). www.dwds.de [10.04.2015].
- Estoup, Jean-B. (1916): *Les Gammes Sténographiques*. Paris: Institut sténographique.
- ETIMO. www.etimo.it/?term=abito [10.04.2015].
- Felici Puccetti, Alessandra (2010): *Errore fatale*. München: Compact.
- Fivush, Robyn/Nelson, Katherine (2004): "Culture and Language in the Emergence of Autobiographical Memory". *Psychological Science* 15/9: 573–577.
- Galisson, Robert (1971): *Inventaire thématique et syntagmatique du français fondamental*. Paris: Hachette.
- Galisson, Robert (1975): « Thèmes de prédilection et vocabulaires thématiques à charge incentive ». *Étude de linguistique appliquée* 18: 59–114.
- Galisson, Robert/Coste, Daniel (1976): *Dictionnaire de didactique de langues*. Paris: Larousse.

- Garzanti (1992): *Il nuovo dizionario Garzanti di francese: francese-italiano, italiano-francese*. Milano: Garzanti.
- Gazzaniga, Michael S. (2000): "Cerebral Specialization and Interhemispheric Communication: does the Corpus Callosum enable the Human Condition?". *Brain* 123: 1293–1326.
- Gazzaniga, Michael S./Russell, Tamara/Senior, Carl (2009): *Methods in Mind (Cognitive Neuroscience)*. Cambridge: MIT Press.
- Gensini, Stefano/Vedovelli, Massimo (eds.) (1983): *Teoria e pratica del glotto-kit. Una carta d'identità per l'educazione linguistica*. Milano: FrancoAngeli.
- Gougenheim, Georges (1952): « Le français élémentaire ». *ORBIS* 1/1: 113–115.
- Gougenheim, Georges et al. (1964): *L'élaboration du français fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Nouvelle édition refondue et augmentée. Paris: Didier.
- Guiraud, Pierre (1954): *Les caractères statistiques du vocabulaire*. Paris: Presses Universitaires de France.
- Guiraud, Pierre (1960): *Problèmes et méthodes de la statistique linguistique*. Paris: Presses Universitaires de France.
- Halbwachs, Maurice (1952): *Les cadres sociaux de la mémoire*. Paris : Presses Universitaires de France. (Trans. Coser, Lewis A. (1992): *On Collective Memory*. Chicago: University of Chicago Press).
- Halliday, Michael A. K. (1985): *Spoken and written language*. Victoria/Australia: Deakin University.
- Haygood, James D. (1937): *Le vocabulaire fondamental du français: étude pratique sur l'enseignement des langues vivantes*. Paris: Droz.
- Henmon, Vivian A. C. (1924): *A French Word Book based on the Count of 400,000 Running Words*. Madison: University of Wisconsin/Bureau of Educational Research.
- Herdan, Gustav (1966): *The Advanced Theory of Language as Choice and Chance*. Berlin/Heidelberg/New York: Springer.
- Hinchman, Lewis P./Hinchman, Sandra K. (eds.) (1997): *Memory, Identity, Community: The Idea of Narrative in the Human Sciences*. Albany: State University of New York Press.
- Hymes, Dell H. (1966): "Two types of linguistic relativity". In: Bright, William (ed.): *Sociolinguistics*. The Hague, Mouton: 114–158.
- Jones, Randall (2004): "Corpus-based Word Frequency Analysis and the Teaching of German Vocabulary". *Fremdsprachen Lehren und Lernen* 33: 165–175.
- Jones, Randall/Tschirner, Erwin (2006): *A Frequency dictionary of German*. London/New York: Routledge.
- Juilland, Alphonse G./Chang-Rodriguez, Eugenio (1964): *Frequency Dictionary of Spanish Words*. London/The Hague/Paris: Mouton & Co.
- Juilland, Alphonse G./Edwards, Prior M. H./Juilland, Ileana (1965): *Frequency Dictionary of Rumanian Words*. London/The Hague/Paris: Mouton & Co.
- Juilland, Alphonse G./Brodin, Dorothy/Davidovitch, Catherine (1970): *Frequency Dictionary of French Words*. London/The Hague/Paris: Mouton & Co.
- Juilland, Alphonse G./Traversa, Vincenzo (1973). *Frequency Dictionary of Italian Words*. London-The Hague/Paris: Mouton & Co.

- Kaeding, Friedrich W. (1898): *Häufigkeitwörterbuch der deutschen Sprache*. Steglitz bei Berlin: Selbstverlag.
- Keniston, Hayward (1920): "Common Words in Spanish". *Hispania* 3: 85–96.
- Kittay, Eva F./Lehrer, Adrienne (1992): "Introduction". In: Lehrer, Adrienne/ Kittay, Eva F. (eds.): *Frames, Fields and Contrasts. New Essays in Semantic and Lexical Organisation*. Hillsdale, Erlbaum: 1–18.
- Knease, Tacie M. (1931): *An Italian Word List From Literary Sources*. The State University of Iowa/Toronto: University of Toronto Press.
- Koesters Gensini, Sabine E. (2009a): „Grundwortschatz und lexikalischer Spracherwerb: methodologische Überlegungen und erste Untersuchungsergebnisse“. In: Di Meola, Claudio et al. (hrsg.): *Perspektiven Drei. Akten der 3. Tagung Deutsche Sprachwissenschaft in Italien (Rom, 14.–16. Februar 2008)*. Frankfurt/Main, Lang: 339–348.
- Koesters Gensini, Sabine E. (2009b): „Der deutsche Grundwortschatz zwischen Lexikologie und Sprachdidaktik“. *Deutsch als Fremdsprache* 46/4: 195–202.
- Kosaras, István (1980): *Der Grundwortschatz der deutschen Sprache*. Budapest/Berlin: Tankönyvkiadó/ Volk und Wissen Volkseigener Verlag.
- Kühn, Peter (1979): *Der Grundwortschatz: Bestimmung und Systematisierung*. Tübingen: Niemeyer. (= Reihe Germanistische Linguistik 17).
- Kühn, Peter (1990): „Didaktische Spezialwörterbücher“. In: Hausmann, Franz J. et al. (eds.): *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin/New York, Walter de Gruyter: 1353–1362.
- Larsen-Freeman, Diane (2000): *Techniques and Principles in Language Teaching*. 2. edition. Oxford: Oxford University Press.
- Leech, Geoffrey (2007): "New resources, or just better old ones? The Holy Grail of representativeness". In: Hundt, Marianne/Nesselhauf, Nadja/Biewer, Carolin (eds.): *Corpus Linguistics and the Web*. Amsterdam, Rodopi: 133–149.
- Lyotard, Jean-F. (1979): *La condition postmoderne: rapport sur le savoir*. Paris: Les Éditions de Minuit. (Trans. Bennington, Geoffrey/Massumi, Brian (1984): *The Post-Modern Condition. A Report on Knowledge*. Minneapolis: University of Minnesota Press).
- L'Épée, Charles-M. de (1776): *Institution des sourds-muets par la voie des signes méthodiques*. Paris: Nyon l'aîné.
- Mackey, William F./Savard, Jean G./Ardouin, Pierre (1971): *Le vocabulaire disponible du français. Le vocabulaire concret usuel des enfants français et acadiens: étude témoin (tome 1). Vocabulaire disponible des enfants acadiens (tome 2)*. Paris/Bruxelles/Montreal: Didier.
- Massa, Eleonora (2013): *Raccontare, raccontarsi: per un approccio narrativo-abitudinario al problema del vocabolario di base*. Università « Sapienza » Roma: Tesi di dottorato. <http://padis.uniroma1.it/handle/10805/2103> [10.04.2015].
- Massa, Eleonora (in print): « Sull'unità di significato della lingua parlata: alcune osservazioni su base lessico-statistica ». To appear in: De Meo, Anna/Bardel, Camilla (eds.): *Proceedings of the International Conference GSCP (Gruppo di Studi sulla Comunicazione Parlata) at the University of Stockholm and Uppsala 2014 (9–12 April)*. University of Naples « L'Orientale », University Press Unior.

- Massa, Eleonora/Simeoni, Valentina (2014): "Once upon a Tale. On the foundational Role of Narrative in Constructing Linguistic and Social Experience". *Journal of Comparative Research in Anthropology and Sociology (COMPASO)* 5/1: 71–96.
- McCarthy, Michael/Carter, Ronald (1997): "Written and spoken vocabulary". In: Schmitt, Norbert/ McCarthy, Michael (eds.): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge, Cambridge University Press: 20–35.
- McEnery, Tony/Hardie, Andrew (2012): *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Michéa, René (1949): « Introduction pratique à une statistique du langage ». *Les Langues Modernes: la revue trimestrielle de l'Association des Professeurs des Langues Vivantes de l'Enseignement Public* 43/3A: 173–186.
- Michéa, René (1950a): « Vocabulaire et culture ». *Les Langues Modernes: la revue trimestrielle de l'Association des Professeurs des Langues Vivantes de l'Enseignement Public* 44/3: 187–192.
- Michéa, René (1950b): « La culture par la langue ». *Les Langues Modernes: la revue trimestrielle de l'Association des Professeurs des Langues Vivantes de l'Enseignement Public* 44/5: 328–335.
- Michéa, René (1953). « Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage ». *Les Langues Modernes: la revue trimestrielle de l'Association des Professeurs des Langues Vivantes de l'Enseignement Public* 47/4: 338–344.
- Migliorini, Bruno (1943): *Der grundlegende Wortschatz des Italienischen. Die 1500 wesentlichsten Wörter*. Marburg: Elwert.
- Mitchell, William J. T. (ed.) (1981): *On Narrative*. Chicago: University of Chicago Press.
- Morgan, Bayard Q. (1928): *German Frequency Word Book*. New York: MacMillan.
- Muller, Charles (1963): « Le MOT, unité de texte et unité de lexique en statistique lexicologique ». *Travaux de linguistique et de littérature* 1: 155–173.
- Nation, Paul/Waring, Robert (1997): "Vocabulary size, text coverage and word lists". In: Schmitt, Norbert/ McCarthy, Michael (eds.): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge, Cambridge University Press: 6–19.
- Nelson, Katherine (1986): "Event Knowledge and Cognitive Development". In: Id. (ed.): *Event Knowledge: Structure and Function in Development*. Hillsdale, Erlbaum: 1–19.
- Oehler, Heinz/Sörensen, Ingeborg (1968): *Grundwortschatz Deutsch. Allemand fondamental. Vocabolario di base del tedesco*. Stuttgart: Klett.
- Pasupathi, Monisha (2001): "The Social Construction of the Personal Past and Its Implications for Adult Development". *Psychological Bulletin* 127/5: 651–672.
- Pasupathi, Monisha/Hoyt, Timothy (2009): "The Development of Narrative Identity in Late Adolescence and Emergent Adulthood: The Continued Importance of Listeners". *Developmental Psychology* 45: 558–574.
- Pfeffer, Jay A. (1964): *Grunddeutsch. Basic (Spoken) German Word List. Grundstufe*. Englewood Cliffs/NJ: Prentice-Hall. (= Publications of the Institute for Basic German/University of Pittsburgh 1).
- Polkinghorne, Donald E. (1988): *Narrative Knowing and the Human Sciences*. Albany: State University of New York.
- Quémada, Bernard (1968): *Les dictionnaires du français moderne*. Paris: Didier.

- Raasch, Albert (1970): „Centres d'intérêt-Lernfelder. Zur Wortschatzarbeit im Französischunterricht (II)”. *Französisch an VHS: Hinweise und Anregungen für den Fremdsprachenunterricht in Jugend- u. Erwachsenenbildung* 2: 41–44.
- Richards, Jack C./Rodgers, Theodore S. (2000): *Approaches and Methods in Language Teaching*. 2. edition. Cambridge: Cambridge University Press.
- Ricoeur, Paul (1981): “Narrative Time”. In: Mitchell, William J. T. (ed.): *On Narrative*. Chicago, University of Chicago Press: 165–186.
- Ricoeur, Paul (1984): *Time and Narrative*. Chicago: University of Chicago Press.
- Rivenc, Paul (1967): « État actuel des enquêtes sur les langues parlées et les langues de spécialité ». In: *Les théories linguistiques et leurs applications. Actes du 1^{er} congrès international de Linguistique Appliquée (Nancy 1964)*. Strasbourg: AIDELA: 153–176.
- Rivenc, Paul (1973) : « A l'aube de l'ère des corpus ». *Voix et Images du CREDIF* 18: 12–16.
- Rodgers, Theodore S. (2001): *Language Teaching Methodology*. ERIC Issue Paper. Washington/ DC: ERIC Clearinghouse on Languages and Linguistics.
- Russo, Domenico (2005): *LIN. Lessico Italiano di Notorietà 2004: il lessico di alta disponibilità*. Roma: Aracne.
- Saussure, Ferdinand de (1916): *Cours de linguistique générale*. Lausanne/Paris: Payot & Cie. (Trans. Baskin, Wade (1959): *Course in General Linguistics*. New York/Toronto/London: McGraw-Hill).
- Schonell, Fred J./ Meddleton, Ivor G./Shaw, Barnard A. (1956): *A Study of the Oral Vocabulary of Adults*. Brisbane: University of Queensland Press.
- Searle, John R. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Siegel, Daniel J. (1999): *The Developing Mind. How Relationships and the Brain interact to shape Who we are*. New York: The Guilford Press.
- Smorti, Andrea (1994): *Il pensiero narrativo. Costruzione di storie e sviluppo della conoscenza sociale*. Firenze: Giunti.
- Smorti, Andrea (2007): *Narrazioni. Cultura, memorie, formazione del sé*. Firenze: Giunti.
- Steger, Hugo (ed.) (1972): *Das Zertifikat Deutsch als Fremdsprache. Bonn/Bad Godesberg/München: Deutscher Volkshochschul-Verband*.
- Tharp, James B. et al. (1934): “A Basic French Vocabulary”. *Modern Language Journal* 18: 238–274.
- Thorndike, Edward L. (1921): *The teacher's Word Book*. New York: Columbia Teachers College.
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam/Philadelphia: Benjamins.
- Tomasello, Michael (1999): *The Cultural Origins of Human Cognition*. Cambridge/Mass.: Harvard University Press.
- Tomasello, Michael (2003): *Constructing a Language: a Usage-Based Theory of Language Communication*. Cambridge/Mass.: Harvard University Press.
- Tschirner, Erwin (2005): „Korpora, Häufigkeitslisten, Wortschatzerwerb”. In: Heine, Antje/ Hennig, Mathilde/ Tschirner, Erwin (eds.): *Deutsch als Fremdsprache. Konturen und Perspektiven eines Faches*. München, Iudicium: 133–149.
- Vander Beke, George E. (1929): *French Word Book*. New York: Macmillan. (= Volume XV of the Publications of the American and Canadian Committees on Modern Languages).

- Vedovelli, Massimo (1995): *La lingua italiana d'uso: il lessico del parlato*. Genova: IRRSAE Liguria.
- Violi, Patrizia (1997): *Significato ed esperienza*. Milano: Bompiani.
- Violi, Patrizia (2003): « Significati lessicali e pratiche comunicative ». *Rivista di linguistica*, 15/2: 321–342.
- Vygotskij, Lev S. (1960): *Istorija razvitija vysših psichičeskich funkcij*. Moskva: Akademija Pedagogičeskich nauk RSFSR. (Trans. Hall, Marie J. (1997): *The History of the Development of Higher Mental Functions*. In: Rieber, Robert W./Carton, Aaron S. (eds.): *The collected Works of L. S. Vygotskij*, Vol. IV. New York: Plenum Press).
- White, Hayden (1980): “The Value of Narrativity in the Representation of Reality”. *Critical Inquiry* 7/1: 5–27.
- Wittgenstein, Ludwig (1953): *Philosophical Investigations*. Oxford: Basil Blackwell.
- Zeidler, Heidemarie (1980): *Das Français fondamental (1er degré): Entstehung, linguistische Analyse und fremdsprachendidaktischer Standort*. Frankfurt/Main: Lang.
- Zimmermann, Cheryl B. (1997): “Historical trends in second language vocabulary instruction” In: Coady, James/ Huckin, Thomas (eds.): *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Cambridge, Cambridge University Press: 5–19.
- Zipf, George K. (1935): *The Psychobiology of Language*. Boston: Houghton Mifflin.