Sede Amministrativa: Università degli Studi di Padova

Dipartimento di *Studi Linguistici e Letterari DiSLL*

SCUOLA DI DOTTORATO DI RICERCA IN : Studi Linguistici, Filologici e Letterari
INDIRIZZO: Linguistica
XXVIII CICLO

The encoding of countability and numerosity
in nominal morphology

Direttore della Scuola: Ch.mo Prof. Rosanna Benacchio
Supervisore: Ch.mo Prof. Davide Bertocci

Dottoranda: Francesca Franzon

Abstract

The encoding of countability and numerosity in nominal morphology.

The aim of this research was to examine the role of Number morphology for what concerns the encoding of information about the numerosity and countability of referents. The issue was approached both from a theoretical and from an experimental point of view.

Number morphology is a widespread category and only few languages in the world seem to completely lack it (Corbett, 2000). Why is Number such a common feature among natural languages? In general, it can be assumed that language grammaticalises only some of all the possible information present in the referential world. The fact that information about numerosity is grammaticalised in such a widespread way in natural languages may mirror the salient role that such information has from a biological point of view, i.e. the fact that this information stems from cognitive processes that are biologically relevant in order to behave successfully in a given environment (Hauser & Spelke 2004). Language provides the means to communicate salient information readily. Morphology is one of these means in general, and Number morphology is the one specifically set for the encoding of the information about numerosity of referents.

Number morphology is designed to convey salient information expressing numerosities, but this possibility takes place only when the noun is linked to a countable interpretation. Within morphological Number systems, countability plays a crucial role: in fact, in absence of countability, nouns are not inflected but assigned a Number value by default. Although the great amount of interest dedicated to countability both by theoretical and experimental approaches, no account has fully succeeded in explaining countability and its relation with morphological Number.

In the present thesis we propose a formal model and provide empirical data - collected in quantitative morphology, psycholinguistics and language acquisition – in order to support the idea that in encoding countability more than one factor comes into play: namely, core grammar rules, effects of non-strictly grammatical processing of linguistic stimuli, and effects related to non-verbal cognitive processes that deal with the information encoded into language.

Riassunto

La codifica della contabilità e delle numerosità nella morfologia nominale.

Questa ricerca ha lo scopo di esaminare il ruolo della morfologia di Numero per quanto riguarda la codifica della numerosità e della contabilità. La questione è stata affrontata sia dal punto di vista teorico che dal punto di vista sperimentale.

La morfologia di Numero è una categoria tipologicamente molto diffusa e solo poche lingue al mondo sembrano esserne completamente prive (Corbett, 2000). Dove va ricercato il motivo di una tale diffusione? In generale, si può ritenere che la lingua grammaticalizzi solo alcune di tutte le possibili informazioni presenti nel mondo referenziale. Il fatto che le informazioni relative alla numerosità siano grammaticalizzate in modo così diffuso nelle lingue può rispecchiare il ruolo saliente che tali informazioni hanno da un punto di vista biologico, cioè il fatto che tali informazioni derivino da processi cognitivi che sono necessari per comportarsi con successo rispetto all'ambiente (Hauser & Spelke 2004). La lingua fornisce i mezzi per comunicare prontamente le informazioni salienti. La morfologia è uno di questi mezzi, in generale, e la morfologia di Numero è il mezzo specificamente deputato alla codifica delle informazioni sulla numerosità dei referenti.

La flessione nominale, e quindi la codifica di informazioni riguardo alla numerosità, è presente solo quando il nome è legato ad una interpretazione contabile. All'interno dei sistemi morfologici di Numero, la contabilità gioca quindi un ruolo cruciale: infatti, in assenza di contabilità, ai sostantivi è assegnato un valore di Numero per default. Nonostante l'ampio interesse dedicato alla contabilità sia a livello teorico che sperimentale, nessun approccio è riuscito a dare una spiegazione del tutto coerente della contabilità e della sua relazione con la morfologia di Numero.

Nella presente tesi si propone un modello formale e vengono forniti dati empirici - raccolti in due studi di morfologia quantitativa, due di psicolinguistica e uno studio in acquisizione - sostenendo l'idea che nella codifica linguistica della contabilità e delle numerosità entri in gioco più di un fattore: non solo un set di regole della core grammar, ma anche effetti dell'elaborazione cognitiva di stimoli linguistici, ed effetti legati ai processi cognitivi non verbali che si occupano di informazioni codificate nel linguaggio.

The encoding of countability and numerosity in nominal morphology.

Acknowledgements

The encoding
of countability
and numerosity
in nominal morphology.

I. Introduction

Number morphology is a widespread category and only few languages seem to completely lack it (Corbett, 2000). Why is Number such a common feature among natural languages? In general, it can be assumed that language grammaticalises only some of all the possible information present in the referential world. The fact that information about numerosity is present in such a widespread way in the morphology of natural languages may mirror the relevant role that such information has from a biological point of view, i.e. the fact that this information stems from cognitive processes that are biologically relevant in order to behave successfully in a given environment (Hauser & Spelke 2004). This biological relevance is what formal linguistics would traditionally define as "saliency". It may be assumed that language provides the means to communicate salient information readily. Morphology is one of these means in general, and Number morphology is the one specifically set for the encoding of the information about numerosity of referents.

Number morphology thus appears as a suitable point to investigate intersections between language and numerical cognition. The few studies that have explored this possibility so far gave encouraging results, by reporting a link between the processing of numerical information and the processing of formal features of Number morphology (Carreiras et al., 2010).

Crucially, Number morphology is designed to convey salient information concerning numerosities, but this possibility takes place only when the noun is linked to a countable interpretation: countability is in fact necessary in order to dispose of a full paradigm of Number inflection. It is thus particularly interesting to explore countability as a grounding property of the morphological Number systems, and to take into account the possibility that it may be encoded contextually with respect to the referent, as any other feature handled within Number morphology.

Although the great amount of interest dedicated to the issue by formal linguistics, experimental psychology, and philosophy, no account has fully succeeded in explaining countability and its relation with morphological Number. To our opinion, this is due to some methodological shortcomings, that stem from a lack of integration of the different approaches that have examined countability so far. The encoding of countability in fact

may not depend only from a morphological rule, or a lexical tag, nor only from a property of the referent. It is necessary to build an approach that integrates all that concerns countability both within the language and outside its strict domain: in the present thesis we propose that in defining countability more than one factor comes into play: namely, core grammar rules, effects of non-strictly grammatical processing of linguistic stimuli, and effects related to non-verbal cognitive processes that deal with the information encoded into language.

The aim of this work is thus to overcome any partial view of the issue, by creating an interdisciplinary link between formal linguistics, psycholinguistics, cognitive psychology.

The discussion is organised as follows: §II.1 will deal with the biological relevance of numerical information, and its conceiving by means of non-verbal cognitive systems. The information processed by such systems will be compared with the information that can be encoded in Number morphology in natural languages in §II.2. The role of countability as a necessary property for numerical information will be faced in II.3; different theoretical and experimental approaches that have tried to explain it will be presented. In section II.4 a formal model will be proposed: the formalization will describe all the possible information that may be encoded into Number morphology in natural languages, from the basic property of countability, to the Number values that can surface. The model aims to overcome the inconsistent results and approaches found in literature, by suggesting that countability can be better taken into account by within an approach that integrates grammar rules, effects of non-strictly grammatical processing of language, and effects related to non-verbal cognitive processes that deal with the conceinving of countability and numerosity of entities.

In section III, some experimental works that aim to explore the interplay of linguistic and cognitive different factors in the expression of numerosity and countability will be presented. Each one is described in a way that allows it even to be readable independently from the rest of the dissertation. The picture-word matiching task in §III.1 will concern the interplay of referential numerosity and morphological Number value. The quantitative studies in §III.2 will provide a measure of the distribution of Italian nouns in countable and uncountable syntactic contexts. The lexical decision study in III.3 will try to overcome some experimental biases found in literature, and it will show how non strictly grammatical properties of nouns, such as subjective frequency and the frequency of occurrence in some syntactic contexts, are a crucial point with respect to the processing

of countability. The study in language acquisition displayed in §III.4 will explore the role of non-verbal cognitive domains in the encoding of countability.

Conclusions about the role of the different factors that are relevant respect to the encoding of numerosity and countability within language will be drawn in §IV.

II. Numerosity, countability and Number morphology

Reference to quantities and numerosity can be encoded at different levels (syntax, lexicon and morphology) in language. To date, experimental studies examining the link between linguistic processing and its underlying conceptual representations have mainly focused on the lexical level, while the morphological level has not received the same amount of attention. On the one hand, only a few experimental studies have highlighted the connection between numerical cognition and Number morphology (Carreiras et al. 2010). On the other hand, theoretical linguistics has not been overly concerned with considering the fact that information encoded in Number morphology must interface with non-linguistic domains of cognition that provide the information. This dissertation expands on these topics to include the cognition of numerosity and countability, and offers a formalization of its link with language.

II.1 Number morphology and cognition

Number morphology is a grammatical category that expresses semantic oppositions concerning the numerosity of the referents. Morphological Number values, such as singular and plural, are not inherent in nominal lexemes: they are contextually selected with respect to the communicative context in order to encode information about the numerosity of a referent (Thornton, 2005). It may be the case that everything related to assigning a value in Number morphology is a contextual operation based on the reference, such as countability. On the one hand, this contextual encoding concerns operations within the language as a system, on the grammatical level; on the other hand, it concerns language as a cognitive process; and non-verbal cognitive operations such as the conceiving of numerosity and countability.

Thus, we can assume that Number morphology encodes salient information: in fact, it cannot be only by chance that Number morphology is widespread throughout the languages: only few languages seem to lack completely a Number category (Corbett, 2000: 50 – 51). Notably, the opposition present in Number morphology are not limited

to the widespread one (singular) vs. more than one (plural). In fact, some languages have morphological Number values also for expressing exact numerosities up to three/four or for expressing estimations of quantities. Moreover, the values that are present in Number morphology are a limited set and that they can encode a limited set of information concerning the numerosity of the referent.

In addition, the marking of morphological Number can be restricted by some constraints, mostly related to other salient properties, such as animacy and definiteness, not only cross-linguistically but also within a same language (Corbett, 2000). Overall, the fact of displaying morphological Number values is more probable when the referent stands in the higher positions of an animacy hierarchy, such as this one adapted from Dixon (1979): Personal pronouns > Kinship terms > Human nouns > Animate nouns > Inanimate nouns. Some exceptions are however reported by Corbett (1996) and Brown (2013). For example, Malay has number on pronouns but not on nouns, Sarsi has Number only for kinship terms, Manchu has number on pronouns and nouns of humans, Comanche has number for animates, rarely for inanimate referents. Data form WALS reported in the chapters 34A (Haspelmath, 2013) and 35A (Daniel, 2013) provide a measure of this distribution.

Given these observations, morphology appears as a suitable point to investigate intersections between language and numerical cognition.

In what follows we will compare the possible morphological Number values reported in typology with data from cognitive psychology and psychobiology on numerical cognition. We will propose that there is a parallelism between the information that can be expressed within Number morphology throughout the languages and the information processed within non-verbal numerical cognition.

Up to now, cognitive psychology has approached the relation between numerical cognition and language mainly by taking into consideration the words expressing quantities and numbers, for example quantifiers, ordinal and cardinal numbers (i.e. Butterworth et al., 1999; Carey, 2004; Clark & Grossman, 2007; Gelman & Gallistel, 2004; Gordon, 2004; Lipton & Spelke, 2003; Pica et al., 2004; Rath et al., 2015; Semenza, 2008; Troiani et al., 2009), but it has rarely been extended to other domains of language different from the lexicon, such as Number morphology (Carreiras et al., 2010). Particularly, a long-debated issue concerns the existence of numerical thoughts without the words to express them. It has been proposed that number words are necessary to

solve numerical problems (for a review: Gelman & Butterworth, 2004; 2005). However, more recent studies have shown that non-human animals (Agrillo et al., 2014; Rugani, et al., 2013; Vallortigara, 2012; Cantlon & Brannon 2006), pre-verbal infants (deHevia, 2011; McCrink & Wynn, 2007) and adults speakers of languages that have no number words (Butterworth et al., Lloyd, 2008; Pica et al., 2004) do master numerical abilities. Moreover, educated adult humans are able to solve a subset of numerical tasks when, under specific experimental conditions, the use of language is prevented (Cordes et al., 2001). In these conditions, non-verbal numerical skills (i.e. all those calculations that could be solved in absence of symbolic numerical words) are therefore preserved and can be compared with those of other non-linguistic organisms such as preverbal infants and non-human animals (Cordes et al., 2001).

Non-verbal numerical systems

The similarity in performance of different species, qualitatively and quantitatively compared, suggests that we share with other animals a subset of non-verbal numerical skills, available soon after birth, that are considered the evolutionary foundations of more complex numerical reasoning (Dehaene, 1997; Spelke, 2000; Cantlon & Brannon, 2007; Starr et al., 2013; Rugani et al., 2015). The non-verbal number system is supposed to be based on two systems: the Object File System (OFS) and the Analog Magnitude System (AMS).

The OFS is founded on the capability of individuating each new object entering into a scene, to which a new file ("object file") is assigned and stored in the working memory. Spatio-temporal information and property/kind changes are used by the OFS to this aim. The signature of the OFS is a limit to the number (usually 3 or 4) of object-files that can be simultaneously tracked and stored in the working memory (Trick and Pylyshyn, 1994). Differences at the upper limit, 3 in the case of salamanders (*Plethodon cinereus*; Uller, Jaeger et al., 2003) fishes (*Xenotoca eiseni*; Stancher et al., 2013), chicks (*Gallus gallus*; Rugani, et al., 2008; 2009; 2010; 2014), frogs (*Bombina orientalis,* Stancher et al., 2015) and 4 in the case of adult monkeys (*Macaca mulatta*; Hauser et al., 2000), have been attributed to maturational factors (Carey, 2009). Hence such a system is not specific to number representation though number is implicitly represented.

Estimation involving also larger numerousness would be dealt by the AMS. The functioning of that system would be ratio-dependent according to the Weber's law

(Gelman & Butterworth, 2004; 2005). As the ratio between the numbers to be discriminated becomes larger, response times decrease and accuracy increases (Gallistel & Gelman, 1992). The minimum discernible ratio narrows over development from 1/3 for newborns, to 1/2 at six months, 2/3 for nine months and 3/4 for preschool children (Izard et al., 2009; Feigenson et al., 2004; Halberda & Feigenson, 2008). Recently dissociable neural signatures of the non-verbal numerical systems has been demonstrated in preverbal infants (6-7.5 month-old), by recording event-related potentials (ERPs) while they were viewing either small (1-3) or large (8-32) sets of objects in a number alternation paradigm. Small numbers, evoked an earlier peaking occipital-temporal response around 400 ms, that was dependent on the absolute value of the numbers presented in successive sets, regardless of their ratio. On the contrary, large numbers evoked a mid-latency parietal response around 500 milliseconds that was dependent on the ratio between successive large numbers, irrespective of their absolute values (Hyde & Spelke, 2011).

Interestingly enough, the Number values that language can communicate throughout morphology seem to resemble the values that non-human animals and pre-verbal infants can distinguish by non-verbal number systems. While number words can refer precisely to numerosities that are perceptively undistinguishable (such as ninety-nine vs. one hundred), it can be hypothesized that Number morphology encodes only numerosities and estimations that are distinguishable at a perceptive level.

Interesting experimental evidence that suggest the involvement of the non-verbal number system in Number morphology comes from an fMRI study conducted by Carreiras and colleagues (2010). Participants saw noun phrases in three conditions: in one condition, phrases presented a Gender agreement violation (i.e. * *la piano* 'the-FEMININE piano-MASCULINE'), in one condition phrases presented a Number agreement violation (i.e. **los piano* 'the-PLURAL piano-SINGULAR'), in another condition the agreement was correct (*el piano*), both for Number and for Gender. The authors found increased activation of the right superior parietal gyrus and of the right intraparietal sulcus in the condition displaying morphological Number agreement violations with respect to the other two conditions. Significantly, the activation of these areas was found to be associated with non-verbal numerosity processing (e.g., Dehaene et al., 2003; Piazza et al., 2002, 2006; Piazza et al., 2007; Pinel et al., 2004). The authors concluded that the right parietal lobe, which is involved in non-verbal number processing, is also activated during language processing concerning Number morphology. Thus, the processing of

grammatical Number values is not strictly confined within the domain of language processing.

II.2. A typological survey on Number morphology

The non-verbal numerical systems handle the information about perceptively discriminable numerosities. Information about the numerosity of entities of the surrounding environment is highly salient from a biological perspective, in fact the non-verbal numerical systems develop early in the human and are shared between species.

It may be the case that the capability to communicate this information readily may be someway advantageous as a consequence of its saliency. We propose that Number morphology is shaped with the purpose to readily and efficiently encode information about numerosity of entities. This seems to be partially supported by the observation that the information about numerosity processed by the non-verbal numerical systems match the possible sets of morphological Number values attested in the grammar of the languages spoken in the world, as will be discussed in this chapter.

Data set

The possible Number values and systems found in languages will be described. Since an important part of our survey concerns the relationship between the numerical cognition and the language, all the parts of speech that can refer to entities (i.e. nouns and pronouns) will be taken into consideration.

The mere attestation of a Number value is sufficient to admit its existence. However, the diffusion of some values with respect to other allows to measure what phenomena are typologically marginal, and what are frequently attested instead; moreover, quantitative data on typology allows to know how the distribution of values in systems is constrained, which is very relevant both for linguistic and for cognitive theory.

This survey is based on data collected on a typological sample of 210 languages from 50 language families plus 8 language isolates. Morphological Number values for each of the 218 languages are reported for the first time in a synoptic table available in the Appendix A.

It is not possible to report data about all the languages in the world, first of all because accurate morphological descriptions are not available for most languages. Moreover, the Ethnologue catalogue, updated to the year 2015, lists 7102 living languages distributed in 141 families (Lewis et al., 2015). However, this number is constantly shifting due to the extinction of some languages and to the fact that linguists sometimes disagree what are distinct languages and what are dialects of the same language (i.e. Lewis & Simons, 2010).

In addition, it is not possible to collect a statistically unbiased sample of natural languages even when considering all the languages (Rijkhoff & Bakker, 1998). In fact, some language families include more languages than others. For example, according to the Ethnologue catalogue, the most represented families are the Austronesian family (with 1257 languages) and the Niger-Congo family (with 1538 languages), while the Indo-European family lists only 444 languages. This disproportion is found even when the distribution of world languages by area of origin is considered. For example, 1313 living languages (i.e. the 18.5% of the world languages) originate in the Pacific area and they are spoken by the 0.1% of the total number of speakers; in Europe, instead, only 286 living, native languages (i.e. the 4% of the world languages) are counted, but they are spoken by the 26% of the total speakers (data from the Ethnologue catalogue).

This study will mostly discuss the data from the languages reported in the most exhaustive data collection about the typology of Number (Corbett, 2000). This sample is quite representative as far as language size and language family are concerned, and is consistent to the most used methods of typological sampling (i.e. Dahl, 2008; Rijkhoff et al., 1993; Rijkhoff & Bakker, 1998).

Possible morphological Number values

The basic opposition for Number: singular vs. plural

Singular vs. plural is the basic opposition in Number morphology (214/218 languages reported in the Appendix A show this opposition). If a language marks morphological Number, it shows at least the opposition in which a numerosity equal to one is encoded into a value of singular, while numerosities different than one are encoded into a value of plural. In the absence of further specification about numerosity, the plural conveys the meaning of a numerosity that is interpretable as "larger than one".

The values for exact numerosities

Some languages show the possibility to denote precise numerosities up to four by means of morphological Number values of dual, trial, quadral (Corbett, 2000). At the moment, there is no evidence of languages that display morphology for numerosities greater than four.

Dual

A precise numerosity of two is expressed by means of dual, which is quite common and attested across language families (84/218 languages belonging to 26/50 families have a dual). See this example from Sikuani language adapted from Aikhenvald (2014): emairibü 'a yam' - emairibü-nü 'yams'- emairibü-behe 'two yams'.

Trial

A precise numerosity of three is encoded by a trial. Its diffusion seems to be limited to the languages spoken in different families geographically distributed in the Oceanic area (20/218 languages in 4/50 families) and its occurrence is restricted to pronoun inflection, or is either ruled by constraints concerning animacy of the referents. Lihir provides an example: wa 'you' (singular) – go 'you' (plural) - gol 'you two' - gotol 'you three'.

Quadral

The possibility to encode a numerosity of four is under debate. The corresponding value of quadral seems to be found only in two Austronesian languages, namely Marshallese and Sursurunga. Moreover, its use seems to be confined to personal pronoun inflection and kinship terms (Corbett, 2000) as in Sursurunga: -i/on/ái 'he/she/it' - di 'they' - diar 'they two' - ditul 'they three' - dihat 'they four'.

The values for approximate numerosities

The morphological values that map exact numerosity do not vary across communicative conditions nor with respect to the referents: a value of dual encodes in any case a numerosity of two. Instead, the value that refer to approximate numerosity do not carry

information of numerosity expressed in absolute terms: they refer to an evaluation of the magnitude of a set of units, instead. Paucals are values that refer to an estimation of a small set of entities whose numerosity is not precisely defined; in English a concept of paucal could be expressed with the quantifier a few. For this reason the value of paucal can be used to refer to sets of different numerosities evaluated as small by the speaker.

Paucal

Paucal is present in languages of families variously distributed from a geographical point of view (30/218 languages from 9/50 families) and, differently from trial and quadral, it is marked mostly on nouns, as in Bayso: lubán-titi 'a lion' - luban-jool 'lions' - luban-jaa 'a few lions'.

Greater paucal

The greater paucal is very rare, being found only in two Austronesian languages, Tangga and Sursurunga. It can occur only in a language that display also a paucal, and it is used to designate small numerosities, greater than the ones that a paucal would refer to.

The other values

General

A Number value of general is used to refer to a referent without expressing any information about its numerosity. An example is given in Fula: besides the singular vs. plural opposition *nyaarii-ru – nyaarii-ji* 'cat-cats', the value for general number *nyaari* 'cat(s)' is also available when the speaker wants to refer to cat(s) without specifying if one or more. Usually, such a value surfaces as a zero-morpheme in a paradigm in which other values are phonologically realized (Corbett 2000).

Collective

A number value of collective is used to refer to a group of items considered together rather than individually. In English the collective meaning is marked in the lexicon (i.e. 'fleet' designates a group of ships), in other languages it can be expressed by

morphological inflection; see for example the (residual) opposition of *frutto* 'fruit' – *frutti* 'fruits' – *frutta* 'fruit-collective' in Italian (Corbett, 1996).

Greater plural

The value of greater plural, sometimes called plural of abundance (Corbett 2000: 30), is used to refer to an excessive, unlimited, overwhelming number. An example of this is taken from Banyum: the form *i-sumɔl* means 'snakes', while the form *ti-sumɔl* refers to 'an unlimited number of snakes'.

Possible morphological Number systems

The term morphological Number system refers to a set of at least two morphological Number values that display a regular opposition between form and meaning in order to systematically denote different numerosities. The basic morphological Number system is singular vs. plural, which is also the most attested in all language families. There is no possibility for other morphological Number values to emerge in a system if there is no singular vs. plural opposition: the presence of the other morphological Number values (such as dual, trial, paucal, …) is possible only given the basic opposition singular vs. plural.

Further than this basic condition, some constraints regarding the occurrence of values leave some Number systems unattested. The constraints concerning the values mapping exact numerosities have been known at least since Greenberg's (1963) work. In fact, Universal #34 states that "No language has a trial Number unless it has a dual. No language has a dual unless it has a plural." Likewise, the presence of quadral is constrained to the presence of trial. Thus, a morphological Number system can be: singular – plural – dual (e.g. Slovene, Sanskrit); singular – plural – dual – trial (e.g. Ngan'gityemerri, Larike) ; singular – plural – dual – trial – quadral (e.g. Sursurunga and Tangga, but see above for remarks concerning the quadral); but morphological Number systems as *singular – dual or *singular – plural – trial have never been attested so far.

The presence of a paucal is constrained only by the presence of the plural and it concurs in building morphological Number systems as singular – plural – paucal (e.g. Avar, Kayapò).

The surfacing of values of general, collective, greater plural is rather unconstrained by implications, but even in this case the total values of the number system cannot be more than five. For example, Tigre has singular – plural – dual – paucal- general and Katyeye has singular – plural – dual – greater plural – general.

Systems such as singular – plural – dual – paucal (e.g. Longgu, Pilagà), in which values referring to precise and to approximate numerosities coexist, are not rare. Notably, morphological Number systems like these can show phenomena of syncretism of values. Moreover, in some languages the morphological Number values mapping exact numerosities can assume a value of paucal; in other words, for example, the trial value may be used to refer to few entities, and not strictly to three. As a consequence, morphological Number values of a system singular – plural – dual – trial can alternatively be interpreted as singular – plural – dual – paucal (e.g. Larike, Murrinh - Patha); similarly, the morphological Number values of a system like the one of Sursurunga can shift their interpretation to singular – plural – dual – paucal – greater paucal (the situation of Sursurunga and Tangga is still under debate). Crucially, this syncretism of morphological Number values does not take place with values of dual.

The syncretism as a property of morphological Number systems will be discussed in §II.4. Up to now, it is important to notice that the values that are present in Number morphology are a limited set and that they can encode a limited set of information concerning the numerosity of the referent.

Interestingly enough, the Number values that language can communicate throughout morphology seem to resemble the values that non-human animals and pre-verbal infants can distinguish by non-verbal number systems (see above in §II.1). It can be hypothesized that Number morphology encodes the numerosities and estimations processed by non-verbal number systems.

In summary, the non-verbal numerical systems handle the information about perceptively discriminable numerosities. Information about the numerosity of entities of the surrounding environment is highly salient from a biological perspective, in fact the non-verbal numerical systems develop early in the human and are shared between species. It may be the case that the capability to communicate this information readily may be someway advantageous as a consequence of its saliency. We proposed that Number morphology is shaped with the purpose to readily and efficiently encode information about numerosity of entities. This seems to be partially supported by the observation that

the information about numerosity processed by the non-verbal numerical systems match the possible sets of morphological Number values attested in the grammar of the languages spoken in the world. From the discussion above it can be inferred that some cognitive operation takes place when selecting a pertinent morphological Number value for nouns. It may be the case that selecting the pertinent Number value implies the encoding another grounding property of morphological Number systems: countability.

II.3 Number and countability

So far, it has been highlighted that the function of Number morphology is encoding a set of semantic features that denote the numerosity of the reference. However, this possibility takes place only in the case that the reference is countable. In fact, in the absence of countability, it is not possible to inflect nouns for all of the values present in a Number paradigm (usually, in this case, a value of singular is assigned as a default).

Countability is thus a crucial feature with respect to Number morphology. However, it is still unclear how its encoding takes place within language and how it is related to properties of the referents. Some approaches assume that countability is a feature of lexemes, and that it depends on certain properties of the entities in the referential world; some others state that lexemes are not marked for countability and point to the role of the context in defining it. Up to now, no account has fully succeeded in explaining countability and its relation with morphological Number.

With respect to this point, we will show that the issue is better dealt with by considering language as more than a set of grammatical rules and lexical items. Namely, in order to better understand countability, it is necessary to take into account also the effects related to the processing of non strictly grammatical features of the linguistic stimuli, as well as the effects related to cognitive domains that provide the information that is encoded into the language. Finally it would be desirable to formalize how they relate to the core grammar.

In what follows, we will propose that the countability of a referent is encoded on the basis of the communicative context, as any other semantic feature pertaining the domain of Number morphology. In the next paragraphs, the theoretical and the experimental

literature on countability will be briefly reviewed. In chapter II.4 a new account about countability and morphological Number will be proposed.

Theoretical works on countability

The issue of countability has been faced by plenty of studies pertaining the linguistic side, the strictly cognitive side, and the neuropsychological side as well (for a review: Fieder, 2014). Traditional grammar descriptions (at least since Cheng 1973) trace a clear-cut division (with exceptions, as is to be expected) between words that refer to "stuff", that are defined as "mass nouns" (*blood*, *butter*, *iron*, …), and words that refer to "things", i.e. countable entities, that are defined as "count nouns" (*bullet*, *chair*, *apple*, …). In turn, this division between things and stuff in lexicon is reflected in distinct syntactic properties between the former and the latter. A basic but widespread explanation for this division consists in a one-to-one mapping of an alleged physical property of an entity into lexical properties. For example, a substance such as *blood* would have no boundaries *per se,* and every part of the substance *blood* shares the properties of the whole substance. According to this traditional interpretation, the alleged physical mass property of the substance *blood* would be encoded in lexicon as an inherent property that would be inextricably linked to the noun *blood*.

Traditionally, grammars provide evidence in favour of such a dichotomic division on the basis of specific morpho-syntactic constraints, the most basic of which is that mass nouns cannot have a plural form. However, the use of a plural form of certain mass nouns is not uncommon. Such phenomena is seen by traditional grammar as some kind of semantic shift (for Italian: Marcantonio & Pretto, 2001), without explaining when and why this possibility can take place. In literature, such phenomena are referred to as a "portioning" operation. There is also the case of count nouns used in a mass context, called "grinding"[1]. Moreover, no convincing account is given for the fact that some mass nouns can bear a kind interpretation in the plural form (e.g. *three wines* can mean three types of wine), while others can bear both a kind interpretation and a count interpretation

---

[1] The notion of *Universal Grinder* and *Universal Packager* have been widely used in literature on the issues of mass and count nouns (including: Cheng 1973; Pelletier 1975; Frisson & Frazier 2005); it is unclear when the terms were first formulated - Jackendoff (1991) attributes their first use to Victor Yngve, during his lectures in the 1960s.

(e.g. *three beers* can mean both three types of beer and three pints of beer, depending on the context).

Other formal linguistic approaches have tried to give a detailed account of such phenomena, starting from semantic or syntactic perspectives. As far as formal Semantics is concerned, Jackendoff (1991) and Chierchia (1998; 2010) start from the idea that every noun is marked as either count or mass *a priori* in lexicon on the basis of the physical properties of the referent. Alternatives to this intrinsic, lexical interpretation are derived by means of type-shifting operators and/or features that would not change the feature values of the lexical entry, but "add operators around the base" (Jackendoff 1991: 21).

Formal syntactic approaches like the one proposed by Borer (2005) assume that roots are not categorized as count or mass in lexicon. Instead, this distinction is syntactically derived: once a root enters syntactic computation it merges with the syntactic head Div° (Dividing Head) and yields a count reading. If Div° is not present in the syntactic computation, it yields a mass reading. In other words, a mass reading is the default. In order to derive a kind reading as well, De Belder (2011) postulates an additional syntactic head Size°. The presence of both Div° and Size° in syntactic computation drives a countable reading; the absence of both drives a mass reading; and the presence of the solely Div° results in a kind reading. Theories that take into account also typological data from languages displaying a classifier system describe "mass" as a basic feature of nouns, and "count" as derived (Krifka, 1995). The idea that mass and count properties are not codified in lexicon but in the context itself was first proposed by Allan (1980) and Pelletier (1975); a recent formalization is found in Rothstein (2010). These approaches enable a nominal expression to be labelled as mass or count depending on the syntactic context. Pelletier (2012a; b) showed that there are pairs of words such as garlic vs. onions "where one is mass and the other is count and yet the items in the world that they describe seem to have no obvious difference that would account for this" (Pelletier 2012a: 8), thus shifting the attention to the referential context as well. Concerning the extra-linguistic, referential level, also De Belder importantly (2011) notes that "linguistics is probably not the appropriate science to account for the salience of the unit for certain concepts and the oddness of the 'There is dog on the wall' example which results from it" (De Belder 2011: 34). Analyses like the one proposed by De Belder (2011) are based on the observation that the capability to recognize a unit is not exclusive to the human species, and thus does not require language abilities. To this

regard, De Belder (2011) quotes a great number of studies - Hirai & Jitsumori (2009), Irie-Sugimoto (2009), Pepperberg (2006) - demonstrating that animals are able to count units; more recent data is also discussed by Rugani et al. (2013) and Vallortigara et al. (2010). If it is true that the capability to recognize a unit does not stem from the faculty of language, there would be no link between the perception of an entity as a mass or count and its encoding in lexicon as mass or count. In addition, and most importantly, the same entity, although perceived in the same way, may be conceived as being either stuff or a thing depending on one's thinking about the structure of that entity. Prasada et al. (2002) investigated the cognitive processes underpinning the conceiving of an entity as either a thing or as stuff, suggesting that their conception is driven by various characteristics of that entity (i.e. the regularity of the structure, its repetition, and the existence of structure-dependent functions), whose relevance depends on the context.

On the one hand it seems that countability is not an ontological property of entities, and thus cannot be directly mapped in lexicon. On the other hand, however, it seems equally reasonable to think that, at least to some extent, language could interface with the cognitive elaboration of the perception of an entity. There is no doubt that certain nouns are mainly used with a mass interpretation and others with a count interpretation. Some of the former sometimes allow a countable reading. For example, it can be argued that a noun such as *wine,* which is normally used as mass, receives a kind interpretation when it occurs in a countable context, since the existence of different types of wine is culturally relevant. Conversely, another noun such as *blood* will hardly ever surface in a countable context, since, for various reasons, the fact that there are different types of blood is, in general, not pertinent to human culture; one exception could be represented by a blood donation centre. Another good example is given by the metaphor about the vampire bar (Chierchia 1998). In a vampire culture, for example, use of the word *bloods* could be considered to be normal. This latter example demonstrates that exclusively syntactic approaches cannot explain why some mass nouns and not others can be inflected in the plural; in other words, there would be no syntactic constraints which rule out the form *bloods* but not the form *wines*.

Experimental works on countability

From a psycholinguistic point of view, Gillon (1992) tries to overcome the issue of semantic shift by classifying nouns that can appear in both mass and count syntax as

"dual". Gillon (1989) assumes that being a "dual" noun is a feature codified in mental lexicon, and that these nouns would be learned as being exceptions to the rule (Bale & Barner 2005). However, experimental literature does not have a clear direction concerning this point. In fact, the majority of psycholinguistic and neurolinguistic studies exploring mass vs. count processing starts from a clear-cut dichotomy between two classes, either trying to avoid "dual" nouns in experimental lists or ignoring one of their two possible readings. Nevertheless, nouns with multiple readings are by no means simply a marginal phenomenon in lexicon (Katz & Zamparelli 2012). It follows that, before conducting an experimental study on mass vs. count processing, it would be desirable to measure the distribution of such nouns, for example by means of analysing language corpora in order to have a better estimate of what and how often nouns effectively occur in mass or in count contexts. Despite recent attempts in this direction (Kulkarni et al., 2013; Katz & Zamparelli 2012 and the study that we will present in III.2), the majority of experimental studies do not take this into consideration and in experimental literature the classification of stimuli as mass or count nouns has mostly been based on the experimenters' intuition. This is partially a consequence of the fact that linguistic theory is not always concerned with providing a firm ground for constructing testing hypotheses, let alone for the selection of experimental items and protocols. As a result, experimental studies on the mass vs. count distinction do not present a totally coherent picture.

For example, many of the experiments conducted in behavioural studies (Mondini et al. 2008; Bisiacchi et al. 2005; El Yagoubi et al. 2006) have failed to observe significant RT differences between the two types of noun in lexical decision tasks, even though these studies were particularly concerned with balancing the two types of items for difficulty. In other studies, however, Mondini et al. 2009 *did* find significant RT differences: mass nouns require longer RTs than singular count nouns; a similar result was found by Gillon et al. 1999. ERP studies do not provide a clear picture either. In an implicitly lexical decision task, El Yagoubi et al. (2006) found a very early detection of the differences between the two noun types (120-160 ms time interval after word onset). In a similar task, Mondini et al. (2008) obtained a comparable result: a very early distinction between mass and count nouns at 160 ms after word onset (N 150), with mass nouns eliciting a stronger negativity in left anterior brain regions than singular count nouns. Bisiacchi et al. (2005) set up a semantic categorization task to directly address the semantic processing level; the authors found that singular count nouns elicited a stronger early negative

component (N150) in comparison with mass nouns, which showed the opposite pattern. Other works study mass vs. count processing in sentence context. In a judgement task, Steinhauer et al. (2001) report that count with respect to mass nouns elicited a frontal negativity independent of the N400 marker for conceptual semantic processing, but resembling anterior negativities related to the processing of grammar. While Steinhauer et al. (2001) conclude that the mass vs. count distinction is primarily related to syntactic processing, Bisiacchi et al. (2005) and Mondini et al. (2008) showed that there are differences in semantic processing as well, and that results of experiments on the mass vs. count distinction can be strongly task-dependent.

To sum up, distribution of ERP components are generally interpreted as suggesting the recruitment of different brain regions. However, a consistent overlapping of structures involved in mass and count processing is found. This result converges with neuropsychological evidence, that has often failed to find patients with clear dissociation in mass vs. count processing (Semenza et al. 1997; 2000).

II.4 The core proposal

So far, it has been shown that morphological Number values resemble the numerosities encoded by the non-verbal number systems. However, the literature in cognitive psychology has rarely taken into account Number morphology; when dealing with countability it suffered from the possible shortcomings of theoretical approaches considering language as more than a set of grammatical rules and do not weight processing effect of non-strictly grammatical properties of linguistic stimuli (see §III.2 and III.3).

An first step in overcoming the theoretical entanglement and the consequent experimental biases can be made by developing a descriptive linguistic model that provides an account of the actual possibilities and constraints of language concerning the issue of countability, namely i) its encoding into the core grammar within Number morphology, ii) the effects related to the processing of the linguistic stimuli iii) the interface with other cognitive domains. An exhaustive explanation of the possibilities of

language may not be found if not by looking outside the mere domain of language, if by language we intend a mere set of lexical words connected by rules.

In fact, these possibilities are dependent on the referential context, and language needs to interface with other cognitive domains in order to express information. It follows that the model itself must act as an interface between language and the other cognitive abilities that come into play, providing a common ground in which theoretical linguistics and experimental psychology can be mutually implemented.

In other words, in order to provide an account of the possibilities actually available within the domain of language, it is necessary to look outside the domain itself. As opposed to the semantic approaches previously mentioned in II.3, it will henceforth be assumed that:

(i)   language does not encode information on the alleged ontological properties of entities (for example as having "naturally definite boundaries"); instead, it encodes the contextual properties of a reference. We define a reference as being an instance of a referent.

(ii)   The properties of a reference are processed within other cognitive domains

(iii)   and set available for linguistic encoding.

(iv)   Some of the properties pertaining to a reference, such as numerosity, are more salient than others (such as for biological reasons (§II.1), and are more likely to be grammaticalised, in this case into morphological Number.

Starting from these assumptions, the following points will be formalised:

(v)   how the contextual properties of each reference (such as numerosity) are encoded into language by means of *configurations* of semantic and morphological features; feature configurations surface into formal values (e.g. singular, plural, dual, ...): the presence of the feature grants the semantic interpretability of the formal values with respect to the reference.

(vi)   The possibility of *underdetermining* the expression of such properties is also entailed in language: in this case, no features are encoded and, as a consequence, a default value surfaces. This value cannot encode one of the semantic interpretations linked to those allowed by the feature.

Overall, it can be assumed that language grammaticalises only some of all the possible information present in the referential world. This fact may mirror the relevant role that

such information has (and has had) from a biological point of view, i.e. the fact that this information stems from cognitive processes that are biologically relevant in order to behave successfully in a given environment. (see §II.1) Language provides the means to communicate this information in the most immediate way for the context.

Moreover, language entails the possibility of building a reference to an instance that underdetermines such contextual properties; in this way, the reference refers to that instance in a more abstract way. This possibility may be more costly in terms of processing resources, as the indication pointing to a grounding property of the system (not of the referent!) must be inhibited. However, such a possibility means that unnecessary information can be cut out, which is a major advantage in terms of communicative efficiency. More generally, it can be concluded that communicative efficiency tends to head towards the encoding of the optimal range of information (both in quantitative and qualitative terms) in respect to the salient properties of the reference. The points outlined above will be clarified in the consequent chapters.

Countability and Number morphology

Inflectional morphology displays the following peculiarities: (i) its expression is obligatory, and (ii) with respect to other lexical and functional categories, its constituent elements, i.e. inflectional morphemes, are smaller in terms of iconicity of quantity (i.a. Dressler 1994; for a critique of the notion of iconicity see Haspelmath 2006,2008). These peculiarities make morphology particularly suitable for the encoding of salient properties whose expression may turn out to be somewhat biologically advantageous (§II.1). In light of this, numerosity and *ab origine* countability seem to be especially salient properties; in fact, they are encoded into Number morphology in the great majority of natural languages, as discussed in §II.2.

Given that the majority of natural languages display morphological values (see above XXXX) to express numerosity, it can be assumed that (i) numerosity is a salient property, and (ii) morphology is the proper place in which salient properties are encoded into grammar for the sake of necessity and the immediacy they require. The starting point of the present analysis of the interface between Number morphology and the cognition of numerosity will be an exploration of countability and Number. The term *interface* is used here to designate the point of contact between two or more levels of analysis; in other words, an interface is a "trading place" that can integrate information from different

domains, and make this information available for encoding and decoding. These domains concern both language and non-linguistic levels of cognition.

On the basis of the approach described above a coherent analysis on countability and Number cannot share the view that nouns are inherently marked as either countable or uncountable, as is stated in most formal semantic assumptions. In this regard, it is proposed that there is no *a priori* label attached to lexemes to define the latter as countable or not. The choice of one of the two uses stems from a contextual analysis operated by the speaker on various properties of the referent, i.e. the fact that it is salient to describe the referent as an object with definite boundaries or, for example, as a substance (for the cognitive processes underpinning such choices, see Prasada et al., 2002, and §III.4). In other words, countability (and uncountability) is not a property of the lexical referent but of the contextual reference.

We believe that the encoding of non-countability vs. countability (just as the expression of numerosity when countability is present) is a peculiarity of morphology. It can be observed that the choice of a particular Number value (i.e. between plural vs. singular) is only available when the noun is linked to a countable interpretation. Where a noun bears an uncountable interpretation, the morphological value of singular is assigned as the default. In this case, singular does not entail a numerosity equal to one: it does not encode any semantic connection at all to a numerosity.

Consider, the following examples from Italian[2], (1a) vs. (1b): in both cases the noun *pizza* is inflected at the morphological Number value of singular; however while in (1a) the value of singular stands for a numerosity equal to one, in (1b) the value of singular does not make any reference to a numerosity equal to one. In (1b), the underdetermination of countability allows the reference of the noun *pizza* to be interpreted as a substance rather than a unit. Note that in (1c) it is not possible to disambiguate whether the reference of the noun *pizza* is countable or uncountable. Notably, examples such as that in (1c) seem to point to the fact that the link between the countability and the reference of a noun cannot be an inherently lexical property, otherwise the (un)countability of (1c) could only be retrieved on the basis of the alleged inherently lexical property of the noun *pizza*, irrespective of the syntactic and communicative context.

---

[2] For the sake of convenience, we will mainly consider examples from Italian, however, *mutatis mutandis*, the same analysis holds for other languages

(1)   a.   *Vorrei   una pizza   subito*.

             I would like   a   pizza-SG right now

    a.   *Mangio   pizza   tutti i giorni*.

             I eat   pizza-SG   every day

    c.   *Vorrei   la   pizza*.

             I would like   the pizza-SG

A morphological value must be expressed in order to be available for syntactic operations, such as agreement. Thus, morphology requires the mandatory assignment of a formal value, even in cases in which a semantic interpretation linked to that value is absent.

Such phenomena may be better understood by assuming the presence of the feature [unit] - from now on [u] - emerging at the interface level between morphology and the cognition of numerosity. The morphological values assigned by [u] are singular vs. plural in Italian. Note that the presence of [u] does not entail *per se* one of these two morphological Number values; the feature [u] stands for the encoding of the properties of countability into morphology.

As already illustrated, the countability of a reference does not concern only the numerosity of the unit(s) *per se*, but can also refer to the numerosity of the different kinds related to the unit(s). This means that [u] can be linked to either the semantic feature [UNIT] or the semantic feature of [KIND]. To sum up, the interpretation of a numerosity equal to one or different from one stems from an encoded [u]; whereas the fact that numerosity refers to a number of units or a number of kinds depends on the semantic feature, [UNIT] or [KIND], linked to [u]. In the present discussion, the term "semantic feature" is used to designate the output of the elaboration of a salient property by a cognitive, non-linguistic, domain.

When the semantic features [UNIT] or [KIND] are not present, since the countability of the referent is not relevant, [u] is also absent, and a morphological Number value of singular (as is the case in Italian) is assigned as the default. In this case, the morphological Number value of singular does not entail any numerosity.

Languages that dispose of a Number value of general (see §II.2 and Appendix A) entail the possibility of expressing the countability of the reference without specifying its numerosity. In these cases, the morphological Number value is linked to the feature [u] and thereby introduces a reference to countability, but is underdetermined with respect of numerosity.

At this point in the discussion, it is very important to repeat the fact that the semantic feature [UNIT] must not be confused with any physical property of the entity to which the noun refers. In fact, depending on the reference, it may or may not be relevant to describe the boundaries of an entity (see Prasada et al. 2002; Chesney & Gelman, 2015; and the discussion in §III.4). For example, the noun *birra* 'beer' can denote the substance: in this case, there is no possible [UNIT] interpretation. Rather, the same noun *birra* can denote a reference whose boundaries are delimited (e.g. a glass or a bottle), and thus the interpretation [UNIT] is available. The noun *birra* 'beer' can be related to the [KIND] interpretation as well: in this case, the interpretation relies on the feature [KIND], and the morphological value is selected from those made available by [u].

To sum up (table 1 in §II.4), if it is salient to refer to an entity as an individual with boundaries, the corresponding noun will be linked either to a [UNIT] or [KIND] interpretation. The presence of [UNIT] and the presence of [KIND] are mutually exclusive: [u] can imply a countability of either units or kinds, but not both at the same time. However, where an entity is referred to without particular reference to its boundaries, the corresponding noun is not linked to either a [UNIT] or a [KIND] interpretation.

CONSTRAINT no.1   Each reference cannot be assigned more than one indication
            regarding its countability.

Table 1: The encoding of countability.

COGNITION OF INDIVIDUAL	SEMANTIC FEATURE	MORPHOLOGICAL FEATURE	MORPHOLOGICAL VALUES (IN ITALIAN)
SALIENT	[UNIT]	[u]	SG vs. PL
	[KIND]		
NOT SALIENT	[NOT DEFINED]	[not defined]	Default (SG)

In the present study we considered a language, Italian, that bears an overtly marked distinction of singular vs. plural in Number morphology. This analysis is based on a feature system that, at least in theory, should be applicable cross-linguistically, even in the case of languages that present other systems for Number morphologies, e.g. with classifiers instead of inflectional morphology. More precisely, since the interface features are responsible for the encoding of countability (and thus its interpretability), it may be the case that they are universally present at the core of grammar, even in languages in which the Number value opposition is not overtly expressed in nominal morphology.

More precisely, since these features are responsible for the encoding of countability (and therefore its interpretability), they may be universally present at the core of human language, even in languages in which Number value opposition is not overtly expressed in nominal morphology. Conversely, the set of morphological Number values that is made available by the features *is* language-specific, and confined to the space admitted by typological variation (see II.2 for a deeper discussion on this point).

Partial conclusions

The discussion seems to reshape the roles of lexicon and morphology concerning the issue of what is referred to in literature as "mass and count". In lexicon, each noun is not *a priori* labelled for countability: countability and non-countability are instead interpretations assigned on a contextual basis with respect to the reference, and encoded into morphology by means of a configuration made up of semantic features and morphological interface features. Such features allow for an interpretable morphological Number value to be assigned, while their absence results in a default value being assigned. In the first case, the reference of a noun is countable; in the second case, the reference of that noun is underdetermined with respect to countability.

Potentially, every noun could be linked to every feature configuration, since each of them is equally well-formed if placed in the appropriate context, i.e. if it is salient to refer to the instance of that entity within its boundaries or not (see example 1 in §2). The fact that a particular noun has traditionally been considered as lexically countable or lexically uncountable may depend on the frequency with which it is associated with a specific feature configuration. In other words, a noun such as *sangue* 'blood' has been traditionally classified as a "mass noun", since the contexts in which it is linked to the configuration [UNIT][u] or [KIND][u] are less frequent than those in which it carries no encoded reference to countability. But, at least in theory, this does not mean that such a noun cannot be linked to each possible configuration, i.e. that it cannot be a countable noun at all.

The selection of a configuration basically depends on the referential and communicative context. As some contexts are much more plausible than others, it follows that some configurations are so frequently associated with a lexeme that they are commonly understood to be lexical properties inherent to that lexeme. Nominal expressions such as *i sangui* 'the bloods' are not formally agrammatical, but are simply less plausible from a

referential point of view. Since there is no *a priori* label assigned in lexicon, there is no need to postulate semantic shift operations from an original lexical category to another; equally, there is no need to assume a lexicon including full lists of exceptions (i.e. the "dual nouns", Glllon 1992), as instead proposed by formal semantic approaches (§II.3). The unpredictability of the application of rules is a concern for both the formal semantic and the formal syntactic approaches (i.e. there would be no syntactic constraint to rule out the form *bloods* but not the form *wines*; see §Ii.3). These latter reasonably claim that it is not possible to solve the problem of countability within the core linguistic domain. The present work tries to offer a broader perspective, and to move the problem of predictability from the strictly linguistic side towards the interface between language and other cognitive domains. In the light of this discussion, it may be useful to look at a broader perspective of the phenomenon in order to ideate a formal model that can be used as a testing ground for experimental questions, and shed light on the puzzling results obtained in experimental literature (§II.3). Such an approach allows the extent and role of each cognitive domain, and of the interfaces between them, to be understood; it follows that experimental questions can be advanced in a more precise way with respect to the processes involved in the domain of interest. As a consequence, experimental results can receive a clearer interpretation. For example, some approaches cited above (Borer, 2005; Chierchia, 1998; 2010; De Belder 2011; Krifka 1995) state that in order to derive a "mass" noun, less syntactic operations are required than in deriving a "count" noun. This theoretical description does not seem to agree with previous results in psycholinguistic studies (i.a. Gillon 1999; Mondini et al. 2009) that point to the fact that the processing of "mass" nouns is more demanding; moreover in acquisition studies, children show preference for the count syntax (see discussion on §III.4): that does not seem consistent with the assumption that the "mass" condition is simpler.

The present work tries to form a coherent picture of these different perspectives. Information concerning numerosity of a referent is so relevant that it is grammaticalised in morphology. Each value of the paradigm is interpretable, as it refers to a referential numerosity. When the countability of the reference is not relevant, it is not pertinent to encode a numerosity: in this case a specific value of the inflectional paradigm works as a default, and is not interpretable with respect to numerosity. To recall an example given earlier (1a vs. 1b, §2), the singular value of a noun surfacing in an uncountable context does not refer to a numerosity equal to one. It may be the case that the surfacing of a

default value is not straightforward, and does not consist in a mere absence of operations but in an operation of inhibition with respect to the encoding of properties that are otherwise relevant in a cognitive sense. It may be the case that such an operation of inhibition is costly, since it is a means to express uncountability within a system that is built up around countability.

In the next few paragraphs the question of countability and quantification will be further investigated, and the relevant operations that can take place in interfaces will be formalized.

Connecting interfaces

Chapter II.3 and II.4 provided an outline of the problem concerning the predictability of the surfacing of a noun in a context of countability or uncountability. Since the present approach distinguishes Number values (such as singular and plural) from morphological features (such as [u]), the issue of predictability must shift from the surfacing of values to the surfacing of features. The features are not univocally associated with a single lexeme, and their surfacing depends on the link that is contextually established between them and other domains of cognition. Such cognitive domains analyse the relevant information from the reference, in this case regarding quantification. As outlined above in this chapter, it follows that every noun can potentially be linked to each configuration of features, and each of them is equally well-formed if placed in the appropriate communicative context. The choice of features takes place among a set of properties that are relevant for biological reasons (see §II.2)

The choice of features is performed by virtue of Saliency. *Saliency Choice* consists in discriminating whether a semantic feature of the properties set (in this case [UNIT] or [KIND]) must be encoded into a morphological feature (in this case [u]) or not. When this link is present, the morphological value is interpretable, in this case with respect to the type of countability it expresses (of course within the language-specific range of values of nominal inflection).

For example, in Italian when the semantic feature [UNIT] (or [KIND]) is encoded into the morphological feature [u] the Number value of singular stands for a numerosity of the reference equal to one, where the plural stands for a numerosity different from one. In this regard, it is important to note that the feature [u] allows for the presence of other

semantic features such as [ONE], [≠1] and [∞], that come into play in respect to the assignment of Number values (fig.1).

In the case that a value is not linked to a morphological feature [u], it may only be interpreted *in absentia* as an underdetermination of the properties encoded by that feature within that particular inflectional paradigm. For countability, when [u] is not present, a reference cannot be interpreted as a unit nor as a kind. In fact, referring to an entity is independent from the encoding of the boundaries of that entity. This latter, that is always present at the perceptual level, is underdetermined in language by omitting the features that refer to it. In this sense, this reference is interpreted in a more "abstract" way. In this case, referring to a bounded entity is not salient, and countability is underdetermined with respect to the reference; as a consequence, it is not possible to encode a numerosity and to select the corresponding value from the inflectional paradigm. Thus the noun surfaces with a default value, that is language specific. As with most languages, the default value in Italian is singular, which, in this case, does not stand for a numerosity equal to one, but means that countability is not salient.

In the flow chart below (figure 1 in §II.4), the surfacing of features into values of general Number, singular, plural and greater plural is illustrated. This latter surfaces as being linked to the semantic feature [∞] in addition to those proper of the plural. The semantic feature [∞] stands for a numerosity so huge that cannot be imagined.

Figure 1: From features to values, part I
In the flow chart below, when a question box is present, lines heading down stand for a "YES" answer, while lines heading right stand for a "NO" answer. In the boxes on the bottom, the configurations of features are reported under their dedicated values. Their formal annotation will be explained in §4. For the sake of brevity, only configurations with [UNIT] are written, however [UNIT] can be replaced with [KIND].
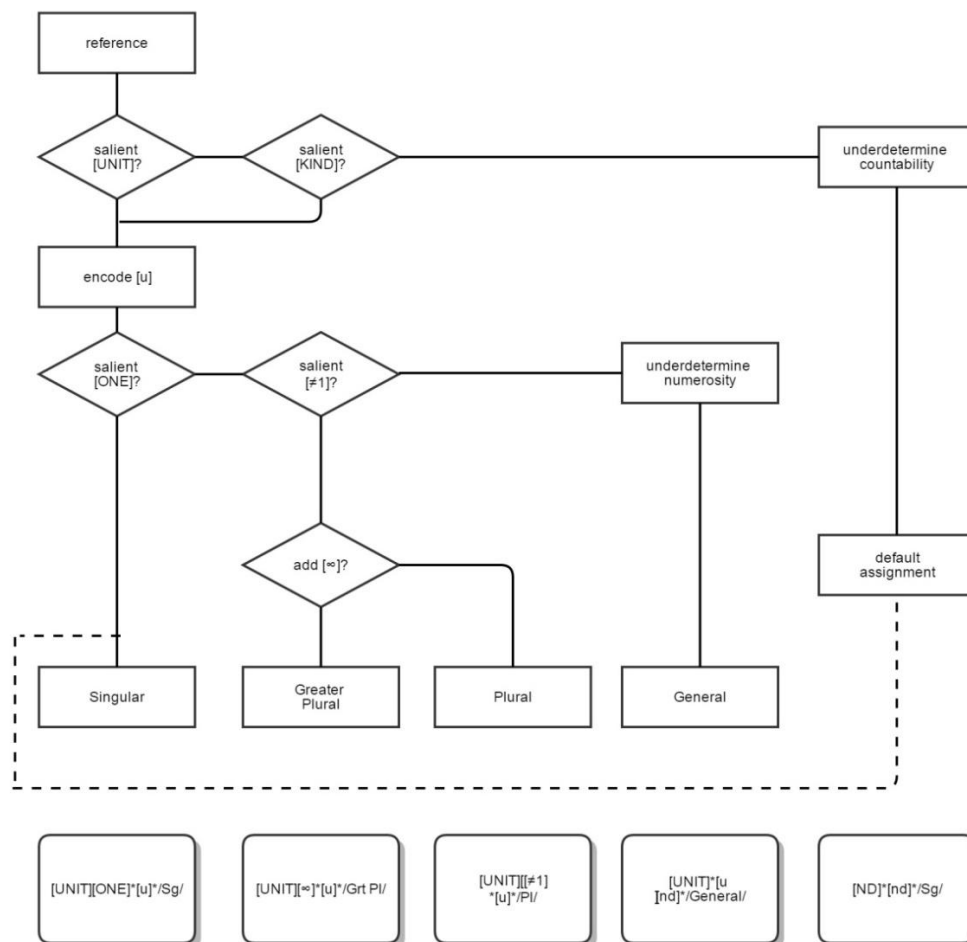
Saliency encodes various relevant semantic properties into morphological features. Recognizing those relevant properties is not a matter of language, but of other cognitive domains. Saliency works as a bridge between these cognitive domains and language. It follows that the predictability of countability, i.e. of the emergence of features, cannot be detected by considering language alone. The present model attempts to formalise the linking interface between quantification in the language system and the cognitive abilities of quantification that provide the information to be encoded. Countability is processed on-line with respect to each reference; some contexts are much more plausible than others, and even the same entity can be intended as either an object or as stuff depending on the context (Prasada et al. 2002). The potential outcome of this is that the same entity can be described by different sets of features in language. For some entities, one of these sets of features may be encoded much more frequently than others.

Countability with respect to a noun is not an inherent lexical property, and can only be measured in terms of the frequency of the use of that noun in correspondence to the emergence or the absence of its salient features. The following paragraphs will consider three other features that add specification to numerosity and quantity.

The emergence of [b]

The encoding of a property of individuality does not pertain strictly to [u], nor to morphology. The semantic feature [BOUND] grants the possibility of referring to an individual entity whose limit is not that traced by [u]. [BOUND] is encoded into language by means of the feature [bound], from now on [b], and it can emerge in syntax as well. In fact, [b] is a feature that sets a new limit for the reference, which is different to the limit eventually encoded by [u]. An example is given below (2), where [b] emerges in correspondence of the quantification expression (QE) *un sacco di* 'a great amount of'. Semantic features are written in capital letters between squared brackets; morphological (and eventually syntactic) features are indicated in lower case between squared brackets; and values are individuated between slashes.

(2)   a.   *È stato versato <u>un sacco di sangue</u>.*   [BOUND][ND]*[b][nd]*/Sg/
                '<u>a great amount of blood</u> has been spilt'
    b.   *Il sommelier ci ha descritto <u>un sacco di vini</u>.*   [BOUND][UNIT]*[b][u]*/Pl/
                'the sommelier described <u>a great number of wines</u>'

As it will be demonstrated, this feature may surface not only in syntax but within the domain of inflectional morphology as well. For the sake of convenience, the surfacing of [b] in syntax will be described first. At both levels, however, there is a fundamental constraint in respect to the use of [b]: when [b] is present, it traces a new boundary with respect to a reference, and thus determines the interpretation of the morphological value (singular or plural in Italian). If [b] surfaces in the configuration [b][u]*/Pl/, as in (2b), the morphological Number value of plural is interpreted as referring to more than one unit whose boundaries are determined by the feature [b]. When [b] surfaces in correspondence to a Number value of singular, it is not possible to interpret that singular number as being linked to a [u]. In fact, in a case like this, the boundaries determined by [b] would overlap that traced by [u]; this overlap is constrained, since the boundary set by [b] and that traced by [u]*/Sg/ occur in complementary distribution. In other words, the

37

limit set by [u] cannot coincide with that traced by [b]. This complementary distribution stems from a semantic constraint: a reference cannot be semantically described by the co-occurrence of the feature [BOUND] and the feature [UNIT], both referring to a numerosity equals to one. To sum up, this seems to suggests that as [BOUND] (and, as a consequence, [b]) is *per se* a reference to a singularity, it can only set the limit for entities whose reference is that of a non-singularity, namely a reference to more than one unit (2b) or the reference of non-countability linked to [ND] (2a).

CONSTRAINT no.2   Each reference cannot be assigned more than one specification of
                   singular individuality.

A formalization of the hypothesis stated above may consist in the impossibility of linking [ONE] with [UNIT] in presence of [BOUND]. It is important to note the consequence of this clash of semantic features: when [b] surfaces in correspondence to a Number value of singular, [u] cannot be present and the value of singular cannot be interpreted as referring to a numerosity equal to one; in such cases, there can only be an uncountable interpretation (2a). Another consequence of this is that [BOUND][ND] is the sufficient semantic feature configuration for assigning a singular value, and [BOUND][UNIT] is the sufficient semantic feature configuration for assigning a Collective value. If a language does not entail a dedicated morphological value of Collective, this feature configuration parasites the value of plural (on *parasiting* see §II.2 and further in §III.3).

Table 2: The constraints of singular individuality.

SINGULAR INDIVIDUALITY		
[UNIT][ONE]	[BOUND]	
	[UNIT]	[ND]
[UNIT][ONE]*[u]*/Sg/ *An apple* on the table	[BOUND][UNIT]*[b][u]*/Pl/ *A kilo of apples* in the bag	[BOUND][ND]*[b][nd]*/Sg/ *A kilo of apple* in the cake

The observed constraint, as well as the discussion conducted so far, seems to lead to an advantageous consequence: operations such as "packaging", "grinding", "portioning" and similar tools which often recur in literature (see §II.3) are better considered in terms of the interaction between the two features [b] and [u]. It is worth considering that [b] and [u] are not only related to the explanation of phenomena concerning countability, but

are also involved in the encoding of the output of cognitive processes related to quantification and numerosity into language.

In fact, the introduction of [b], [u], and the Saliency Choice allows the interaction of numerosity and quantification within the domain of the nominal phrase to be formalised. In (3) an example of this will be shown.

(3)   a.   *Un chilo di anguria*   [BOUND][ND]*[b][nd]*/Sg/
                a kilo of watermelon-SG

    III.   *Venti chili di angurie*   [BOUND][UNIT]*[b][u]*/Pl/
                twenty kilos of watermelon-PL

    IV.   *Un chilo di angurie*   [BOUND][UNIT]*[b][u]*/Pl/
                a kilo of watermelon-PL

    d.   *Un chilo di mela*   [BOUND][ND]*[b][nd]*/Sg/
                a kilo of apple-SG

In these phrases, the boundary is expressed by the QEs *un chilo di/venti chili di* 'a kilo of/twenty kilos of'. In (3a) the QE traces a limit to a quantity of an entity described as an uncountable reference. In (3b), the QE traces a limit to a quantity of entities described as a countable reference.

The surfacing of a value of singular on the noun *anguria* in (3a) is pertinent with respect to the QE *un chilo di* that encodes [b], since it is plausible that a watermelon weighs more than one kilo. Therefore, the limit set by [BOUND][ND] in the feature configuration [b][nd]*/Sg/ is in this case smaller than (or equal to) the limit that would be set by the features configuration [UNIT][ONE]*[u]*/Sg/. It may also be the case that the boundary set by [BOUND][ND] may refer to a limit greater than that which could be set by [UNIT][ONE] (3d). In this sense, the choice of [BOUND][ND] is a matter of Saliency. There is no pure linguistic constraint to rule out (3c), it is just implausible that, for the total weight of one kilo, there could be more than one watermelon. The same holds for (3b).

Another remark on the interaction of [b], [u] and Saliency Choice concerns the similarity of meaning between bare singular nouns with an uncountable reference and bare plural nouns. Consider the following example:

(4)   a.   *Bere birre*   [UNIT][≠1]*[u]*/Pl/
                drinking beer-PL

        b.   *Bere birra*   [ND]*[nd]*/Sg/

                drinking beer-Sɢ

In literature, these expressions are classified as atelic since the event they describe does not entail either an endpoint nor definite boundaries (Jackendoff 1991; Vendler 1957; Verkuyl 1972). With respect to the present approach, in both (4a) and (4b) no boundary is specified in relation to the reference. In (4b) not only the expression of a boundary is absent, but the feature [u] is not encoded either. Ultimately, the atelic reading stems from the fact that there are no definite and/or bound individuals that can work as a *telos* for the infinitive verb.

The emergence of [b] in syntax and the different types of QE

As stated in §4, the possibility of referring to an individual entity whose limits are those traced by [u] is given by the feature [b]. In addition, language allows more specific information to be encoded regarding the numerosity or quantity relative to a particular reference.

With respect to this, the surfacing of [b] in syntax is linked in its most evident way to the presence of Quantification Expressions (QEs). In (5) two different kinds of QEs are illustrated. In pseudo-partitive constructions (5a, b) the boundary is defined by the sole feature [b]. The other kind of QE emerges as an output of the presence of the feature [e], which will be examined in the following paragraphs (§6). It is worth noting here that the QE in (5c) introduces a [b] and encodes an evaluation of extension by means of the feature [e].

(5)   a.   *Ho bevuto della birra.*   [BOUND][ND]*[nd]*/Sg/

              I drank OF-THE beer-Sɢ ('some beer')

        b.   *Ho bevuto delle birre.*   [BOUND][UNIT]*[b][u]*/Pl/

              I drank OF-THE beer-Pʟ ('some beers')

        c.   *Ho mangiato molte mele.*   [BOUND][LARGE][UNIT]*[b][e][u]*/Pl/

                I ate a lot of apple-Pʟ

The emergence of [b] in morphology

The presence (or absence) of [b] and [u], dedicated to the encoding of semantic properties, respectively [BOUND] and [UNIT] (or [KIND]), generates a range of possible combinations, that also offer an explanation for the nouns classified as exceptions in

traditional grammar. These particular cases are described as the output of a combination of features made available by the language system, and not as oddities.

Inflectional morphology

The case of *i muri* (/Masc; Pl/) 'walls' – *le mura* (/Femm; Pl/) 'surrounding wall'.

In Italian, the encoding of [b] can emerge at the level of inflectional morphology and surface in a dedicated value in the inflectional paradigm. This particular value is syncretic with a value of Gender morphology, but, given its systematic opposition with the Masculine plural morpheme, is interpreted as a Number value. Consider the systematic alternations *il muro* (/Masc; Sg/) 'wall', *i muri* (/Masc; Pl/) 'walls', *le mura* (/Femm; Pl/), 'surrounding wall'. *Il muro – i muri* contains the regular opposition of singular vs. plural morphemes (-*o* vs. -*i*) for the declension of inherently masculine nouns. Plurals in –*a*, like *le mura,* contain an Inflectional morpheme which is homophonous to the regular morpheme for the large majority of singular Feminine nouns, but it requires an agreement with plural[3]. This type of inflection is no more productive in Italian, but such nouns show a systematic opposition with plurals that do not entail a collective meaning. From a referential point of view, *muri* relates to a plurality of entities, while *mura* introduces a collective value, since it refers to either the city walls or to the totality of the domestic walls. This collective value is mapped at the level of Inflectional morphology by the feature [b].

A restricted number of nouns that shows a plural in –*a* do not seem to present all three forms: surprisingly, their inflection seem to show just the forms in -*o* for singular vs. –*a* for plural, and not the regular plural in –*i*. It is not to say that this latter form is not used at all, it is just much less common than the form ending in -*a*. This difference in frequency may be due to a Saliency effect. For example, *il dito - le dita* 'finger – fingers' seems to omit the regular plural in –*i*: *i diti*. Furthermore, the form *le dita* seems to appear in contexts that are distributionally covered by plurals in other nominal expressions *tre dita /#tre diti della mano* 'three fingers of one hand', as reported in Acquaviva (2002) and Corbett (1996). However, *tre diti mignoli* ' three little fingers' is perfectly fine because in this case the collective value reference is not computed. On the contrary, due to a Saliency effect, it is preferential to compute a collective value reference in the case of

---

[3] Except in the case of *la frutta* 'fruit-COLLECTIVE', that shares the same collective meaning, but is singular. However, in a previous diachronic phase of Italian *le frutta* (/Femm; Pl/) is attested.

dita. Even in the case of *tre dita* 'three fingers', the collective value - and thus the [b] feature - is available, since the referring point is always represented by the presence of all of the fingers on a hand. This alternation can be better explained by starting from *osso* 'bone', another noun of this type. The nominal expression *tre ossa* refers to three bones possibly belonging to the same skeleton, while the converse condition holds for *tre ossi*. The case of *la forbice* (/Sg/) – *le forbici* (/Pl/) 'scissors'.

Another case is represented by nouns like *le forbici* 'scissors', *i pantaloni* 'trousers' and *gli occhiali* 'glasses'. According to normative grammar, these nouns designate singular objects but do not surface as singular; as they refer to entities made up of two or more similar subparts, each of these latter can be perceived in a very clear-cut way but cannot be considered to be an autonomous entity. These nouns can also occur as inflected at the singular (*la forbice* 'scissor-Sg') with the same meaning. This ambiguity must be traced back to the encoding of the semantic properties; in the case described by normative grammar systems (*le forbici*-Pl), the encoded semantic features are [UNIT][TWO][4] surfacing in a plural value. The latter case (*la forbice*-Sg) is the result of the encoding of [UNIT][ONE]. As a consequence, when these nouns surface at the plural, the interpretation is ambiguous even for a native speaker. In this case, these nouns can be intended as either a single object or as a plurality of objects. To conclude, these case are different from those represented by *muro-muri-mura* since no [b] is involved. It is not a matter of collective meaning; instead, the variation stems from Saliency Choice, depending on the relevance of the object or its subparts.

Derivational morphology

In the present analysis, the term derivational morphology is intended to describe a non-mandatory step of the computation in which features can be inserted. Some occurrences hint at the fact that the semantic feature [BOUND] can be encoded in derivational morphology by means of dedicated derivational morphemes, such as the Italian -*eto* in *quercia – querceto* 'oak' – 'wood of oaks': while *quercia*, which is singular, refers to a single tree, *querceto*, which is also singular, refers to a group of trees. The plural *querceti* refers to more than one group of oak trees. The collective meaning ("group of") is present in the dedicated derivational morpheme that entails the semantic feature [BOUND].

---

[4] The feature [m], responsible for the encoding of precise numerosities, like one and two, will be introduced immediately after [b].

However, this derivational morpheme does not insert the feature [b] at morphological level; it does not result in the surfacing of a Number value of plural. As a consequence, numerosity is encoded by the morphological feature [u], as usual.

Contrarily to what happens with [BOUND] inserted at the lexical level (§4.3), it is worth noting that [BOUND] introduced in derivational morphology implies that all units of the group share the same semantic reference described by the root (for example, compare bosco 'wood' vs. querceto 'wood of oaks').

Similarly to –eto, the suffix -ame introduces [BOUND] at the level of derivational morphology: while foglia 'leaf' has no [BOUND] in its computation, fogliame 'foliage' introduces [BOUND], allowing a collective interpretation.

The emergence of [b] at the lexical level

So far, what happens when [b] is computed at the syntactic and morphological level has been explained; however, this feature can surface in lexicon as well. This is the case for nouns such as flotta, gregge, bosco 'fleet', 'herd', 'wood', so-called collective nouns that indicate a group of units. In this case, [BOUND] is present at the lexical level and it belongs to the bundle of necessary features that define the referent (not reference!) of these collective nouns. For example, a fleet cannot be intended as such if it were not a group of boats defined by some kind of boundary. As a consequence, the collective meaning ("group of") is present in lexicon, and the feature [BOUND] has no encoding into inflectional morphology and it cannot assign Number value. Numerosity is determined by the morphological feature [u], as usual. It follows that, in Italian, when these nouns have singular Number morphology, they refer to a single group of units; and when they have plural Number morphology, they refer to more than one group.

It should be noted that some languages entail the possibility of expressing the collective meaning encoding the semantic feature [BOUND] into morphology: the outcome is the feature configuration [BOUND][UNIT]*[b][u] that, as illustrated in §4, is a sufficient feature configuration for assigning a plural value (in the case that there is no dedicated morphological value for Collective in the inflectional paradigm). This choice is driven by Saliency (possibly triggered by the animacy of the reference), and is much more evident in a language like English where there are alternations with respect to the Number agreement, such as this furniture is cheap vs. the police are coming. In the first case, the agreement takes place at the singular since it is performed on [UNIT][ONE]*[u]; whereas

in the case of the noun *police*, the agreement takes place at the plural since it is performed on [BOUND][UNIT]*[b][u].

The emergence of [m]

Besides the features of individuality described so far, [u] and [b], there is a feature of measurement responsible for the encoding of exact numerosities referring to units: it will be called [m], and it refers to mathematical numerosities smaller than five. This feature relates to the ability to set a link between a unit and a precise numerosity encoded in a semantic feature that may be [TWO], [THREE] or [FOUR], that may in turn surface into morphological Number values of dual, trial, quadral . Such distribution of Number values may be the linguistic counterpart of the arithmetical system for keeping precise track of small numbers, Object File System (OFS; see above in II.1) An important remark must be made regarding the semantic feature [ONE] (as well as its counterpart [≠1]): this feature does not need to be linked to an [m] in order to be interpretable. In fact, the semantic feature [ONE] is linked to a [u], since it is pivotal in defining the difference between singularity and plurality: plurality is defined as a non-singularity, and not vice-versa. As a consequence, it is only possible to interpret a numerosity equal to one linked to a singular value.

[ONE] does not need to be mapped into [m], differently from [TWO], [THREE]and [FOUR]; however, [ONE] refers to a precise numerosity just as [TWO], [THREE] and [FOUR] do. Given this common property among these features, it can follow that if [UNIT][ONE] occurs in complementary distribution with [BOUND], then even [UNIT][TWO], [UNIT][THREE] and [UNIT][FOUR] cannot co-occur with [BOUND]. As illustrated in §4, [UNIT][ONE] cannot co-occur with [BOUND], since it is not possible to contemporaneously encode the precise numerosity of the unit(s) and refer to the magnitude of the boundary that ties the unit(s) together. The same holds for [UNIT][TWO], [UNIT][THREE] and [UNIT][FOUR]: in other words, [m] is in complementary distribution with [b], since they stand for the morphological (and/or syntactic) encoding of two different types of measurement (see figure 2 in II.4). Each reference can be quantified no more than once.

CONSTRAINT no.3   Each reference cannot be assigned more than one indication of its quantity.

Typology and [m]

As stated so far, [m] allows the encoding information on a precise numerosity different from one. In this sense, the encoding of [m] is not mandatory from a cross-linguistic point of view, nor eventually within the same language: unlike [u], [m] is not a primary feature. In fact, the presence of [u] is necessary and sufficient for the encoding of countability with respect to the reference. The presence of [m] qualifies further information on the numerosity set available from the previous encoding of [u]. The ranking of [m] with respect to [u] mirrors the fact that many languages in the world do not display morphological values to express precise numerosities. Obviously, [m] is universal, while the set of values of [m] that can be encoded into morphological Number values is language specific.

It should be noted that if a language encodes [m] into morphology, it does not follow that the whole possible set of values are encoded into that language (see II.2). In (6) examples of dual (6b) vs. plural (6c) are given along with the corresponding feature configurations. Examples from the Guahibo language Sikuani are adapted from Aikhenvald (2014).

(6)   a.   *emairibü*   [UNIT][ONE]*[u]*/Sg/

              'a yam'

        b.   *emairibü-behe*   [UNIT][TWO] * [u][m]*/Dual/

              yam-DUAL

              'two yams'

        c.   *emairibü-nü*   [UNIT][≠1]*[u]*/Pl/

              yam-PL

              'many yams'

Up to this point, the case of languages that display dedicated inflectional values for dual, trial and Quadral have been described. If a language, does not have dedicated inflectional values, [TWO], [THREE] and [FOUR] surface by *parasiting* other values, namely the plural (as in Italian and in many other languages), or the paucal and greater paucal (§6 and §6.1). The parasiting of values is defined as the surfacing of more than one feature configuration in one same value. This syncretism results in a decrease in informativity. The informativity of a value is defined here as the possibility of maximising the encoding of the features with respect to the perceived features of the reference. In this sense, the value of dual is more informative than the value of plural, as the surfacing

of dual entails a direct, bi-univocal mapping between a feature configuration and a morphological value. For example, in Italian, the configuration [UNIT][TWO] *[u][m] and the configuration [UNIT][≠1]*[u] surface with the value of plural. In this case, a configuration which is richer in information parasites the morphological value dedicated to the expression of a configuration which is poorer in information.

This parasiting proceeds in the opposite direction with respect to that illustrated for uncountable references. In that case, the less informative configuration, i.e. the one dedicated to the underdetermination of countability, [ND]*[nd], parasites a value that would be dedicated to the expression of singular, [UNIT][ONE]*[u], that is richer in information and thus necessarily more informative. This occurs since the morphological system of Number is shaped for the expression of numerosities, therefore the underdetermination of numerosities cannot be bi-univocally linked to a dedicated value. Such observations are part of a deeper consideration on the syncretism of values within the morphological system of Number, that will be drawn in the last paragraphs of this chapter.

The parasiting of values is not confined to Italian, but constitutes a grounding property of morphological Number systems of natural languages, since there is no language that would map each possible feature configuration into a dedicated Number value (as seen in §II.2). Moreover, [m] is a feature encoded into morphology, but this does not imply that [TWO], [THREE] and [FOUR] must be encoded at this level only: they can surface as Numerals; of course, in this case their presence is entirely optional.

The emergence of [e]

There is a feature responsible for the encoding an evaluation of quantities into morphology and syntax: it will be called [evaluation], from now on [e]. This feature must be linked to either the semantic feature [SMALL] or the semantic feature [LARGE]. Such evaluations need to modify a feature of individuality, [u] or [b]; as a consequence, [e] cannot be linked to a reference [ND] whose countability is underdetermined. Secondly, the semantic features [SMALL] and [LARGE] linked to [e] result in a slightly different interpretation, dependent on the semantic feature defining the individual entity they are referring to:

    i)   when they refer to a [UNIT], they express an evaluation of the physical

                magnitude of that unit (7a);

ii) when they refer to a [KIND], they cannot qualify an extension referring to a perceptual limit: they consist in an evaluation of the properties that are relevant to individuate [KIND] (7b);

iii) when they refer to a [BOUND], they evaluate the extension of the boundary (7c, 7d and 7e); for example, in the case [SMALL] modifies [b][u], a paucal interpretation is available (7d).

Since the interpretation of [e] is not properly linked to a measurement but to an evaluation, this feature can emerge recursively if it is not linked more than once to the same feature of individuality (7e).

(7) a. *Una piccola torta per te*   [UNIT][SMALL][ONE]*[u][e]*/Sg
         a small-SG cake-SG for you

     b. *I grandi vini della Toscana*   [KIND][LARGE][≠1]*[u][e]*/Pl/
         the great-PL wine-PL of Tuscany

     c. *Poco formaggio in frigo*   [BOUND][SMALL][ND]*[b][e][nd]*/Sg/
         a little-SG cheese-SG in the fridge

     d. *Pochi formaggi in frigo*   [BOUND][SMALL][KIND]*[b][e][u]*/Pl/
         few-PI cheese-PL in the fridge

     e. *Molti grossi calamari*   [BOUND][LARGE][UNIT][LARGE]*[b][e][u][e]*/Pl/
         many-PL big-PL squid-PL

Parallel to what stated for [m], it can be supposed that the evaluation of the magnitude of individuals linked to a [b] by means of [e] is the linguistic counterpart of the non-verbal numerical system for approximate quantities, namely the Analog Magnitude System (AMS). The values of trial and quadral can be parasited by a feature configuration [b][e][u] that would have its dedicated value in a paucal: the morpheme that would be dedicated to a trial ([UNIT][THREE]*[u][m]*/Trial/) could therefore be interpreted as a paucal ([BOUND][SMALL][UNIT]*[b][e][u]*/Trial/). It is not possible for a syncretism of values to result in an increase in informativity (§). In this sense, [b][e][u] codifies features that are less informative than those mapped by [u][m]. In fact, [SMALL] (or [LARGE]) may refer to a range of possible numerosities. Conversely, the configuration [u][m] entails the possibility of encoding exact numerosities: it is more informative because an exact numerosity linked to [m] is encoded and decoded precisely and bi-univocally.

A higher level of informativity is demanding: [m] can only map numerosities up to four. The approximation of the magnitude encoded by [b][e] refers to quantities corresponding

to ranges of numerosities, obviously leaving the precise number as implicit. In other words, the numerosity linked to [m] remains constant across speakers and does not vary across references: the numerosity encoded by [UNIT][TWO]* [u][m] is always decoded as equal to two. Conversely, [b][e] denotes an evaluation on a quantity, and the numerosity expressed by this evaluation is not univocally interpreted. As an outcome, variability is observed in the use of paucal between speakers, and even for the same speaker referring to different entities. For example, "elephant-PAUCAL may typically refer to fewer real word entities than ant-PAUCAL" (Corbett 2000: 40).

The emergence of [e] in morphology

As stated above, the feature [e] can emerge in syntax as well as in morphology. This section will be dedicated to additional remarks concerning the emergence of [e] in morphology.

The feature [e] can surface in inflectional morphology linked to a [b]. In this case, the dedicated values in inflection are respectively paucal if [e] is linked to the semantic feature [SMALL], and greater paucal if it is linked to [LARGE]. It is important to note that, from a typological point of view, languages that display a dedicated value for greater paucal also have a value for paucal. In addition, typological data reported i.a. by Corbett (2000) seems to point to the fact that morphemes dedicated to the expression of exact numerosities can sometimes be interpreted as paucals. It seems that such phenomena take place because the feature configuration [b][e][u], that is less informative than [u][m], parasites the morphological value dedicated to the encoding of this latter.

So far, the surfacing of [e] linked to [b] in morphology has been shown. This feature also can appear as being linked to [u]. In examples (7a) and (7b), [e] is encoded in syntax by means of dedicated adjectives. Adjectives introducing evaluations of extensions display peculiarities and functional properties (in some languages, such as Yoruba, adjectives are a closed class) in the sense intended by Cinque (2010). However, [e] can also emerge as linked to [u] in inflectional morphology and derivational morphology. Italian is illustrative with respect to this emergence, and also concerning the relationship between inflection and derivation. As for inflection, a set of Italian nouns of inanimate referents appear in oppositions as *buco* (/Masc/) – *buca* (/Femm/) 'hole – large hole', *fosso* (/Masc/) – *fossa* (/Femm/) 'ditch – large ditch'. In this case, the value of Feminine in Gender Inflection encodes [UNIT][LARGE]*[u][e], thus defining a unit of a relatively larger dimension. Even if this kind of opposition is widely attested, it seems to be no more

productive from a synchronic point of view. Conversely, in Italian, the encoding of [e] is productive within the domain of derivational morphology by means of morphemes such as –ino (*guanto* – *guantino* 'glove – little glove'), -etto (*casa* – *casetta* 'house – little house'), and -one (*naso* – *nasone* 'nose – big nose').

Given that [e] can emerge in both inflectional and derivational morphology, something must be said on the interaction with respect to [e] between these two levels. As a starting point, it is useful to remember that, at least in Italian, every noun must receive a morphological value of Gender (similarly to what happens for Number), even when the value of Gender lacks any semantic interpretability. The insertion of derivational morphemes is not mandatory, but they are always semantically interpreted when inserted. It follows that when [e] is encoded at the derivational level, the interpretability of the semantic feature linked to [e] is underdetermined at the inflectional level.

In general, when a morpheme encodes a semantic feature, its formal value cannot be assigned by default. Conversely, when the feature needs to be underdetermined, the morpheme must be assigned the default value. When the morphological value of Gender bears another semantic feature, different to [e], i.e. the feature that encodes sex in the case of an animate referent, [e] must surface in morphology by means of derivation: *cavallo* (/Masc/) – *cavallino* (/Masc/) 'horse – little horse'; *cavalla* (/Femm/) – *cavallina* (/Femm/) 'mare - little mare'. When the morphological value of Gender does not bear another feature interpretable at a semantic level, it may surface with a default value, which in Italian is Masculine. This latter case is exemplified by alternations such as *faccia* (/Femm/) – *faccino* (/Masc/) – *faccina* (/Femm/) 'face – little face – little face'. The same holds even in nouns of animate referent, if the sex of the referent is not encoded in inflectional morphology but by means of a lexical feature: *donna* (/Femm/) – *donnone* (/Masc/) – *donnona* (/Femm/) 'woman – big woman – big woman'.

Since it is no one of the aims of this thesis to provide an exhaustive description of the phenomena concerning Gender, the discussion will not be explored further. However, one last remark must be made: the feature [e] is properly involved in Number morphology only when it co-occurs with [b]. If there is no [b], and therefore no [BOUND], [e] may appear as being linked to [u] and does not properly refer to numerosity. The encoding of [e] as linked to [u] is useful for analysing the phenomenon of so-called singulative morphemes (i.a. Treis 2014). It has been noted that the function of such morphemes depends on the noun they modify: from a lexicalist point of view, if the base form of the

noun has a mass interpretation (8a), the derived noun will become countable. However, this countable reading is always connected to an evaluation on the extension. If the base form of the noun has a count interpretation (8b), the derived noun will often receive a meaning of endearment. Examples in (8) are from Kambaata, a Cushitic language (data from Treis, 2014):

(8)   a.   *hix-í-ta*   'grass'   vs.   *hix-ichch-ú-ta*   'blade of grass'

        b.   *mesel-ée-ta*   'girl'   vs.   *mesel-éechch-o*   'o dear girl'

It seems that the morpheme *–ichch-* (and its allomorph *–éechch-*) always encodes [e], and is necessarily linked to [u]. It can be stated that so-called singulative morphemes do not properly encode numerosities, but they introduce an evaluation of the features linked to [u]. If [e] evaluates [UNIT][u], the resulting meaning is a diminutive; if [e] evaluates [KIND][u], the result is a term of endearment. A similar situation is found in Italian: occurrences like *gessetto* 'piece of chalk', Lit: 'small chalk' and *legnetto* 'stick of wood', Lit: 'small wood', which are morphologically derived from lexical nouns frequently associated to [ND]*[nd]*/Sg/ (*gesso*, 'chalk' and *legno*, 'wood'), must be necessarily linked to the configuration [UNIT][SMALL]*[u][e].

The geometry of features

So far, the features responsible for the encoding of the countability and quantification properties of nominal expressions have been described. The following table provides a summary (table 3 in §II.4):

Table 3: The features responsible for the encoding of countability and quantification properties.

INDIVIDUALITY			MEASUREMENT	
SEMANTIC	MORPHOLOGIC			SEMANTIC
[UNIT]	[u]			[ONE]
				[≠1]
				[∞]
[KIND]		[m]		[TWO]
				[THREE]
				[FOUR]
[BOUND]	[b]	[e]		[SMALL]
				[LARGE]

On the interaction of features

Individual features are responsible for the encoding of countability, and measurement features are responsible for the encoding of numerosities and quantities. The semantic features listed in table 3 represent every feature that comes into play when dealing with numerosity and countability; their selection is driven by a choice performed by saliency on the referential context. The morphological features map the interpretability of such semantic features into morphological values. The encoding of the features in morphology grants that salient properties (in this case salient properties of countability and numerosity) are encoded as immediately as possible.

As for semantic measurement features, [ONE] and [≠1] are responsible for the split between numerosities equal to one and numerosities different to one. This mirrors the split between the encoding of singularity and plurality within Number morphology that is basic in all Number systems. Defining plurality as a non-singularity is grounding for understanding that every further indication of quantity or numerosity is an added specification within the domain of [≠1]. This is why no language can display values such as dual, paucal (and so on), without displaying the value of plural, as discussed above in §II.2. This formalization can explain the differences in iconicity stated in Universal 35: "There is no language in which the plural does not have some nonzero allomorphs, whereas there are languages in which the singular is expressed only by zero. The dual and the trial are almost never expressed by zero" (Greenberg, 1963). Following this, in the notations of features configurations [≠1] is omitted in presence of further specifications concerning plurality.

The interaction between the features is ruled by the following general constraints as reported above.

CONSTRAINT no.1   Each reference cannot be assigned more than one indication regarding its countability.

CONSTRAINT no.2   Each reference cannot be assigned more than one specification of singular individuality.

CONSTRAINT no.3   Each reference cannot be assigned more than one indication of its quantity.

A possible formalisation of these constraints consists in the fact that:

(i)   [UNIT] and [KIND] are in complementary distribution with respect to the emergence of [u];

(ii)   [BOUND] and [UNIT][ONE] cannot be simultaneously linked to a same reference;

(iii)   [BOUND] cannot co-occur with [UNIT][TWO], [UNIT][THREE] or [UNIT][FOUR].

The flow chart below (fig.2) illustrates what will be an inflectional paradigm if, *ad absurdum*, each typologically possible feature configuration surfaced in a dedicated morphological Number value. Note that if a semantic feature is linked to a morphological feature, it will be interpretable. The morphological features [b], [m] and [e] can emerge in syntax by means of functional categories such as quantifiers and numerals. Even in this case, though, the constraints remain valid. It is important to note that the same feature configuration must be interpretable in the same way, irrespective of whether it emerges solely in morphology or also in syntax.

Figure 2: From features to values, part II

In the flow chart below, when a question box is present, lines heading down stand for a "YES" answer, while lines heading right stand for a "NO" answer.

Properties of morphological Number systems

Syncretism of values

The flow chart in figure 2 in §II.4 takes into account all the possible values typologically present in all the languages across the world. However, no natural language displays an inflectional paradigm in which each configuration of salient features is mapped one-to-one into a dedicated morphological Number value (see §II.2 and Appendix A)

If the information concerning numerosity is so relevant, why has a fully one-to-one inflectional paradigm never been attested? Moreover, why do morphological Number paradigms rarely display more than four values in the same language (Corbett 2000)?

A preliminary observation must be made before attempting to answer these questions. The absence of a dedicated value in a language does not imply that that language cannot encode the feature configuration corresponding to the value in question. Other syntactic or lexical means will be exploited to express all the salient information if this latter is lost to a value that is linked to more than one feature configuration. In this regard, examples in English are compared in (9), with examples from Bora, a Witotoan language (data from Aikhenvald 2014). Bora displays a dedicated value to express [UNIT][TWO]*[u][m] (9a vs. 9b); in English, the feature configurations [UNIT][TWO]*[u][m]   and [UNIT][≠1]*[u] are syncretic in the same value of plural (9a' vs. 9b'), and the semantic feature [TWO] can be made explicit by adding the numeral 'two' (9a').

(9)   a.   *mútsiítsiba-acu*   [UNIT][TWO]*[u][m]*/Dual/
                pear apple fruit-DUAL

         a'. *two pears*   [UNIT][TWO]*[u][m]*/Pl/

         b.   *mútsiítsiba-ane*   [UNIT][≠1]*[u]*/Pl/
                pear apple fruit-PL

         b'. *pears*   [UNIT][≠1]*[u]*/Pl/

In the case of English, the configuration of [UNIT][TWO]*[u][m] that would have dual as a dedicated value parasites the value of plural instead: as a consequence, the value of plural can be the surfacing value of more than one feature configuration. In general, for all languages, the number of values is lower than the number of possible feature configurations due to the aforementioned phenomenon of parasiting.

Parasiting follows two trends, and results in a syncretism of more than one feature configuration into a same morphological value. In the first case, a feature configuration rich in information about numerosity does not surface in a dedicated value for that numerosity, but instead selects a less informative value (syncretism type A). The feature configuration [UNIT][THREE]*[u][m] surfacing into plural instead of trial is an example of this. In the second case, a feature configuration surfaces into a value that would be dedicated to feature configurations which are richer in information (syncretism type B). An example reported by Corbett (2000) is represented by the case of a value of trial interpreted as [BOUND][SMALL][UNIT]*[b][e][u].

Ultimately, both trends of parasiting lead to a loss of informativity, the major consequence of which is the impossibility of a univocal decoding. The confluence of more than one configuration into a same value gives rise to the impossibility of retrieving the feature configuration richer in information and thus a unique semantic interpretation. Unless other (syntactic or lexical) means add further specifications, only the feature configuration which is poorer in information is retrieved. Syncretisms are sometimes optional within a same language (as, for example, in case of languages that do not mandatorily encode[UNIT][TWO]*[u][m] into dual, like in Maltese, a Semitic language). The syncretisms of type A are present in each and every language, and are pivotal in the building of an inflectional Number system. In a nutshell, parasiting reduces the range of information in order to avoid the encoding (or decoding) of too much information

Optimization of a range of informativity

As a side effect of the syncretism of values, no morphological Number system shows more than five values (Corbett, 2000): all of the possible morphological Number values never occur all together as a system in a same language, but each morphological Number system has only a few of them. The amount of information within morphological systems is not maximised, rather it is optimised by means of syncretism of values, i.e. different information may surface into a same morphological value (i.e. Ackerman & Malouf, 2013; Carstairs, 1987; Loporcaro, 2011; Muller, 2007; Pirrelli & Battista, 2000; Stump, 1991; 2006; 2010).

Grammars are built on a finite and discrete sets of elements combing into a potentially infinite number of sentences. Crucially, the set of combing elements must be finite in order to be learnt (as noted in mathematical learning theory, e.g. Malouf et al., 2015; Nowak et al., 2001) and interpreted (as noted in information theory since the pivotal

work by Shannon, 1948). The bigger the amount of information, the more morphological Number values are present in a system. A morphological Number system needs to balance the overall number of morphological Number values and the amount of information encoded. In the observed languages, the optimized set ranges up to five values.

As a consequence of syncretism, the amount of information encoded in Number morphology is always lower with respect to all of the possible information about numerosity. For example, in languages that have a morphological Number system of singular- dual-plural, the numerosity of three cannot be encoded into a dedicated value. Thus, it surfaces covertly in the value of plural and cannot be decoded unambiguously only by relying on Number morphology. Similarly, in a system of singular vs plural, the information about numerosities of two or three shall surface in the plural, thus becoming covert.

Of course, it is advantageous to dispose of a morphological device to encode the properties concerning precise numerosities or estimated quantities about references. In fact, values of an inflectional paradigm are systematically arranged and mandatorily expressed, thus they are suitable for encoding immediately salient properties of the reference. However, in the particular case of one-to-one mapping between a feature configuration rich in information and its dedicated value, it is not possible for the value to surface without encoding the linked configuration; the same holds for decoding. Secondly, the encoding (or decoding) of such features is costly: it may need to involve other cognitive domains interfacing with language, specifically those related to quantification (§II.1; Carreiras et al., 2010). Last but not least, not all the properties of the reference are always salient with respect to the particular communicative context. For all these reasons, it follows that a system that provides a one-to-one mapping of each possible feature configuration into a dedicated Number value, thus maximizing the encoding of information into Number morphology, would not be the best one. In fact, such an inflectional system would even mandatorily encode information about properties of the references that may be not relevant with respect to the particular communicative context. As a consequence, each inflectional paradigm must optimize the ratio between costs and advantages concerning the encoding and decoding of information. In order to do so, the set of values must be confined within a range of optimal informativity, which is reached by eliminating a part of the information. The result of this operation is a reduced

set of values, whose informativity is reduced since they may be hosting more than one feature configuration. To better exemplify this point, imagine that a customer wanted to buy four bottles of beer in a market; unfortunately, beer is only sold in six-packs in that market. Obviously, six bottles cost more than four. So the customer ends up spending more, and buying two beers he did not need. As opposed to the market in the example, language is shaped in order to avoid this double waste.

Underdetermination of features

So far, it has been pointed out that syncretism (II.2) is the effect of the optimization of informativity concerning the inflectional Number paradigm. Finally, a last grounding property of Number morphology must be described. In §2, §2.1 and §3, another particular type of syncretism was referred to, i.e. that involved in the underdetermination of the most salient feature of the Number paradigm [u]. The lack of encoding countability is probably very costly in terms of cognitive resources (see the literature cited in §1.1.1 and §1.1.2). The reason is that such underdetermination needs to take place in a system that is shaped for the immediate encoding of some properties related to its exact opposite, i.e. countability. This is also the reason why a configuration [ND]*[nd] cannot have a dedicated value inside any paradigm: as it eliminates any indication of countability and numerability, not only it does not fit in the optimal range of informativity (as occurs frequently for [UNIT][FOUR]*[u][m]), but also it does not fit in the whole maximal range. Since it entails no information about countability, the only possibility to surface for [ND]*[nd] is to surface within a default value by parasiting a value that is poor in information with respect to numerosity. The eligible candidates for default values are therefore generaland singular. In a nutshell, the elimination of information on countability is costly (uncountable reference), whereas the elimination of the number of values is advantageous in the economy of the system (optimised informativity).

To sum up, there are different types of parasiting: the offspring of all of them is a syncretism of values that builds an economic inflectional system, optimised for the immediate expression of salient features as well as for the underdetermination of non-salient information.

II.5 Conclusions

In this section, it has been proposed that:

i) Morphology plays a crucial role in language since it works as an interface between language and other cognitive domains.

ii) Information about numerosity is very salient from a biological point of view, in fact number cognition is based on cognitive systems that are available soon after birth and shared with other animal species

iii) Number morphology allows highly relevant information about the numerosity of the reference to be expressed in the most immediate way.

iv) In particular, contextual properties concerning the numerosity or quantity of the reference are encoded into semantic features; the latter are in turn linked to morphological features, and surface as formal values (e.g. singular, plural). The presence of the morphological features grants the interpretability of the formal values with respect to the semantic features.

v) The sufficient morphological interface features responsible for the encoding of countability, numerosity and quantification are: two features denoting individuals, [u] and [b], and two other features of measurement, [e] and [m].

vi) The feature system also entails the possibility of eliminating any indication to countability, however the expression of uncountability is costly since it must take place in a system properly designed for the expression of countability and numerosity.

vii) In addition, this feature system optimises the amount of the encoded information on numerosity. In fact, language tends to avoid the encoding of features rich in information but low in saliency by conveying them into morphological values that would be properly dedicated to feature configurations which are poorer in information.

viii) As a consequence, inflectional paradigms are made up of a small number of syncretic values set around the sufficient informativity. This is an advantage since it avoids a waste of resources in the mandatory encoding and decoding of even non-salient information. The quantity of information that may be present in an inflectional paradigm is not to be maximised, but optimised.

ix)   Ultimately, human language displays two advantages for Number morphology: firstly, the ready expression of salient properties concerning numerosity or quantities and, secondly, the possibility of underdetermining some properties of the instances of the reference. This latter property allows informativity to be optimised; as a side effect, less informative references are more abstract.

x)   The information concerning a reference may be underdetermined to the extent that not even its grounding property, i.e. countability, is expressed. The expression of a reference irrespective of its limits is, in this sense, an abstraction. Such a possibility is possibly a peculiarity of human language.

Throughout the dissertation, a model has been illustrated that aims to describe the interplay between language and other cognitive domains underpinning quantification and its expression. The goal of such a formalization was to provide a means for investigating how the different levels interface with each other.

In the next section we will present some experimental works, with the aim of measuring the role of core grammar, non-strictly grammatical processing of linguistic stimuli, and of non-verbal cognitive processing in the encoding of countability.

III. Empirical Evidence

In the present section we will present some experimental works that aim to explore the interplay of different factors in the expression of numerosity and countability; namely core grammar rules, effects of non-strictly grammatical processing of linguistic stimuli, and effects related to non-verbal cognitive processes that deal with the information encoded into language.

Each of the experiments presented in this section is described in a way that allows it even to be readable independently from the rest.

The picture-word matiching task in §III.1 will deal with the effect of the underdetermination of a semantic feature in the interpretation of a morphological Number value.

The quantitative studies in §III.2, i.e. a subjective frequency rating study and a corpus analysis will provide a measure of the occurrence of Italian nouns as countable and of their distribution with respect to countable and uncountable syntactic contexts.

The lexical decision study in III.3 will show how non strictly grammatical properties of nouns, such as subjective frequency and the frequency of occurrence in some syntactic contexts, are a crucial point with respect to the processing of countability

Finally, the study in language acquisition displayed in III.4 will explore the role of non-verbal cognitive domains in the encoding of countability.

III.1 A picture – word matching task

This study will deal with the underdetermination of a semantic feature with respect to a morphological Number value. The Number morphology of Italian, a singular or plural value is mandatorily expressed and it usually encodes the referential numerosity. It is known from the literature that the processing of morphological Number value on nouns should involve a cognitive elaboration of numerosity as well (i.a.: Carreiras et al., 2010, see above in II.1).

Italian allows to test the mismatch between morphological Number value expressed on a noun and the referential numerosity by comparing two quantification expressions (QEs),

qualche and alcuni. Both QEs refer to a plural numerosity, meaning 'some' (Zamparelli, 2007). While in presence of the QE alcuni the noun congruently surfaces at the plural (N-pl), in presence of the QE qualche the noun displays a value of singular (N-sg). Since both conditions are grammatical, they represent a privileged testing ground to observe the relationship between morphological Number value and numerosity.

Methods

Participants

34 students of the university of Padova participated to the study (age 21-35; 18 males and 16 females). All participants were native speakers of Italian, reported no neurological pathologies, had normal or corrected-to-normal vision, and took part to the experiment on a voluntary basis.

Materials

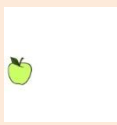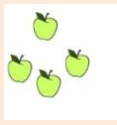30 nouns referring to countable, concrete objects were chosen. All nouns were controlled for

i)   frequency by means of the it-WaC corpus (Baroni et al., 2009),

ii)   subjective frequency by means of a dedicated rating study (Zanini et al., 2014), and

iii)   ortographic length.

 The length and the frequency of the whole phrase qualche/alcuni + N were also controlled. No significant differences were found between the condition with qualche and the codition with alcuni.

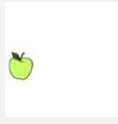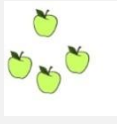The list of the phrases is reported in Appendix B.

Two pictures were created for each noun: one representing one object, and one representing the same object repeated four times. Each picture was presented twice: once with qualche and once with alcuni (table 1 in §III.1). Thus, the experimental conditions were 4, for a total of 120 experimental stimuli.

Table 1- experimental stimuli

PICTURE	Q-EXPRESSION	NUMEROSITY ENCODED BY THE QE	MORPHOLOGICAL NUMBER VALUE	TRUTH VALUE	CONDITION
	qualche + N	plural	singular	F	A
	alcuni/e + N	plural	plural	F	B
	qualche + N	plural	singular	T	C
	alcuni/e + N	plural	plural	T	D

180 filler stimuli were added in order to counterbalance each experimental condition (table 2 in §III.1). The presentation of the stimuli was randomized for each participant.

Table 2 – filler conditions

| PICTURE	Q-EXPRESSION	NUMEROSITY ENCODED BY THE QE	MORPHOLOGICAL NUMBER VALUE	TRUTH VALUE	CONDITION
	un/uno/una + N	singular	singular	T	filler
	plural bare noun	plural	plural	F	filler
	singular bare noun	singular	singular	T	filler
	un/uno/una + N	singular	singular	F	filler
	plural bare noun	plural	plural	T	filler
	singular bare noun	singular	singular	F	filler

Procedure

Stimuli were delivered by means of DMDX software (Forster & Forster, 2003). A picture–phrase matching paradigm was developed. The participant was asked to press one key if the phrase matched with the picture (true), and another key in the opposite case (false). The keys were counterbalanced across the participants.

The structure of each trial was the following: fixation point (800ms), picture presentation (1000ms), blank (200ms), phrase presentation (1000ms) and another blank (3000ms). Participants could answer after the presentation of the phrase. RTs measurement was

triggered at the onset of the phrase presentation. The trial structure Is exemplified in figure 1 in §III.1

Figure 1 - structure of the experimental trial.
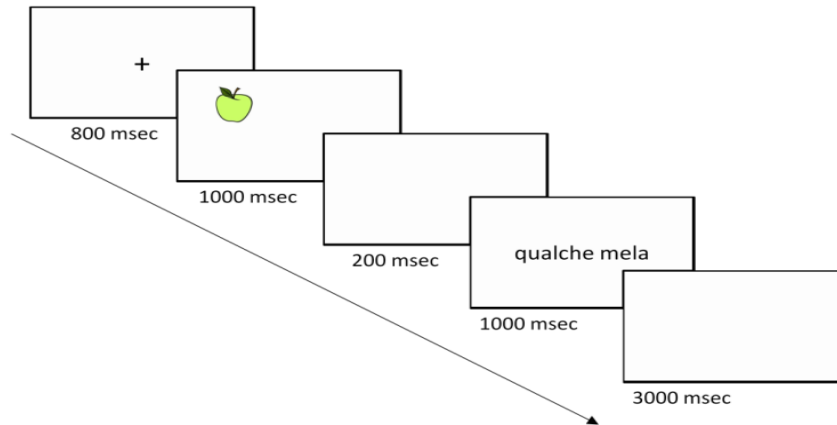

Results

An ANOVA was carried out both by subject and by item. A significant main effect $F(1,33) = 57.81$, $p < .001$ was found for QE: the RTs for conditions with *qualche + N-sg* (mean = 868.75 ms; sd = 224.69) were longer than for conditions with *alcuni + N-pl* (mean = 816.88 ms; sd = 227.15). This means that when the numerosity encoded by the QE matched the morphological Number value as in conditions B and D (*alcuni + N-pl*) faster RTs are recorded, irrespectively of the truth value. Conversely, when the numerosity encoded by the QE did not match the morphological Number value as in conditions A and C (*qualche + N-sg*) RTs are slower. Results are reported in table 3 ad plotted in figure 2 in §III.1.

Table 3 – Results

	TRUE	FALSE	MEAN	SD
qualche+N SG	863.32 (97%)	874.17 (94%)	868.75	224.70
alcuni+N PL	798.51 (98%)	835.26 (97%)	816.88	227.15

Figure 2 – Results for RTs
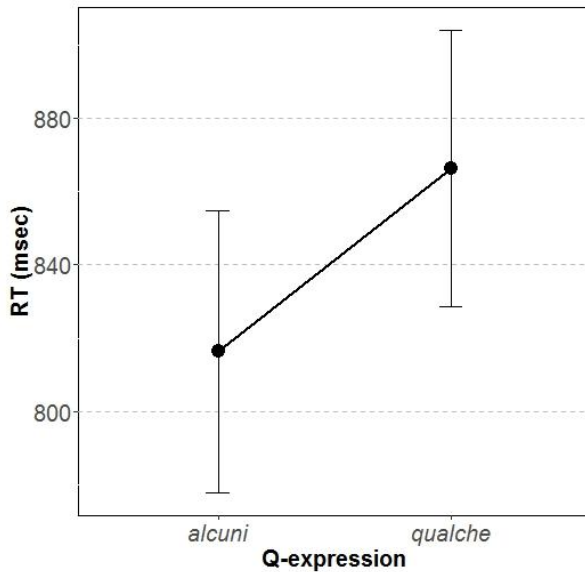
Discussion

Conditions with *qualche* + *N-SG* may elicit longer reaction times because of the mismatch between the numerosity of the reference and the value expressed on the noun by Number morphology. In Italian each occurrence of a noun must bear a morphological value of singular or plural. Usually, the value singular is linked to a semantic feature that implies a numerosity of one, while the value plural implies that the numerosity is different from one. When this link between a morphological value and a semantic feature is not present, the morphological value is "empty" and non-interpretable. In this case the numerosity is encoded by the quantifier and the value on the noun is non-interpretable. Non-interpretability with respect to an inflectional morphological value may be the reason why a larger amount of cognitive resources are required, similarly to what has been observed for Gender morphology (Franzon et al., 2014).

Morphological values must be always present -even when they do not surface at the phonological level- in order to become available for syntactic operations, such as agreement, (see above in II.4). Morphological inflection may be conceived to provide the quick expression of salient semantic features such as numerosity (§II.1, §II.2, §II.4). At the same time morphology entails the possibility to underdetermine such salient semantic features, depending on the syntactic or communicative context (§II.4). With respect to the formal model proposed above, the feature [ONE] that is usually encoded in

the configuration of the morphological Number value of singular, is not interpreted. Such operation of underdetermining consists in an inhibition of the semantic interpretation of the morpheme: this may result in greater processing costs.

III.2 Contextuality of countability: two quantitative studies

The next study will focus on countability. Quantitative data collected from a subjective frequency rating study and a study performed on corpora have been collected in order to provide an actual measure of the occurrence of nouns as countable and as uncountable. The is measure is useful for two reasons: first, the distribution of nouns can endorse some theories better than others, and second, knowing what nouns occur mostly in mass or in count syntactic contexts can be useful for the selection of experimental stimuli for psycholinguistic studies, as we will report in (§XXX)

In contrast with the huge amount of literature dedicated to the issue from the point of view of philosophy, formal linguistics, neuro- and psycholinguistics and cognitive sciences, relatively few attention has been paid to the effective extent of the actual distribution of nouns in the use of language.

Few studies have dealt with the measurement of the distribution of the mass and count nouns by relying on their use. For example, a recent rating study by Kulkarni, Rothstein & Treves (2013) collected a database relative to the mass and count usage of 1,434 nouns in six different languages involving overall 16 informants. A set of metalinguistic questions were asked with respect to specific syntactic contexts and semantic conceptualizations. The authors found that "syntactic classes do not map onto each other, nor do they reflect beyond weak correlations, semantic attributes of the concepts" (Kulkarni, Rothstein & Treves, 2013: 132). Notably, although the metalinguistic questions required a polar answer (y/n), the results gave as an output a non-polar distribution of the items, more similar to a *gradatum* instead.

Rating studies such as the one described above can measure the speakers' acceptance of certain occurrences of nouns. As a complementary measure, the distribution of a noun in a morpho-syntactic context can be better measured by means of corpus analysis. Schiehlen & Spranger's (2006) explored a 200-million-token corpus of German starting from a lexicalist mass vs count distinction of nouns. However, a huge number of words occurrences remained ambiguous. Once the syntactic contexts of the occurrences were taken into consideration, the nouns did not dispose in two dychotomic groups, instead on a *gradatum,* i.e. on an ideal trajectory whose poles are represented by "pure mass" and "pure count" nouns.

Katz & Zamparelli (2012) browsed the ukWaC corpus (Baroni et al., 2009) setting queries based on syntactic contexts to distinguish mass form count use of nouns. Their main investigation aimed at measuring the semantic distribution with respect to the kind reading of nouns, thus providing insightful data concerning the distribution of nouns with respect to countability and uncountability in general. The authors reported that the nouns mostly found in count contexts are found in mass contexts as well, a finding consistent with the perspective that each noun can appear both in countable and uncountable contexts.

In the following paragraphs, we report the results of a rating study and a corpus linguistics analysis conducted on Italian in order to provide a measure of the countability of a sample of nouns. On this purpose, we measured

(i)   the subjective frequency of the nouns listed in Appendix B, inflected both in the singular and in the plural form, by means of a rating questionnaire and

(ii)   the distribution of the occurrences of these nouns in mass or count contexts by means of an analysis on corpora.

We considered a complementary approach, i.e. comparing subjective frequency rating data with the corpus frequency of nouns, a good way to capture information about countability. In fact, from the one side, corpus analysis provides a measure of the distribution of nouns with respect to the contexts of occurrence. However, it may be the case that a corpus does not list all of the occurrences effectively encountered by the native speakers. In this sense, a subjective frequency rating study can work as a useful counterpart of the analysis on corpora since it returns frequency estimations closer to the actual possibility for a noun to occur (i.e. Balota, Pilotti, & Cortese, 2001; Kuperman & Van Dyke, 2013; Williams & Morris, 2004).

If nouns are lexically marked as mass or count, mass will occur mainly in mass contexts and count nouns in count contexts. A space of variation is expected for what concerns the eventual operations of *packaging* and *grinding*, however it must be possible for the great majority of the nouns to trace their basic interpretation of (un)countability. More specifically, we expect that (i) mass nouns in the plural will be scored zero in the subjective frequency rating; (ii) mass nouns will not occur frequently in count contexts (and *viceversa*) in the corpora, i.e. there will be an inverse correlation between the mass use and the count use of a same noun.

Conversely, nouns are not lexically marked as mass or count, the great majority of nouns will occur both in mass and count contexts with a similar frequency and, as a consequence, it will not be possible to trace the basic interpretation of noun with respect to its countability. By assuming that the countability or uncountability of a noun is not a lexical feature but is contextually assigned, we predict that (i) no noun in the plural will be scored zero in the rating; (ii) no inverse correlation between the mass use and the count use of a same noun will be observed.

III.2.1 A subjective frequency rating

A questionnaire was designed in order to evaluate the subjective frequency relative to the occurrence of some nouns. The methodology is based on the previous literature on rating studies (Ferrand et al., 2008).

Materials

The questionnaire listed 448 concrete nouns, i.e. 224 concrete nouns inflected both in the singular and in the plural. The nouns were selected following the theoretical definitions given in traditional grammars in order to encompass the range of mass and count use. Crucially, the list included even the plural of 45 nouns for which only singular occurrences would be expected on a normative basis ("pure mass" nouns such as *sangue* 'blood' - **sangui* 'bloods'). The complete list of nouns is reported in Appendix B.

Procedure

126 informants took part to this study. Participants varied widely in age (range: 22 - 76 years; mean = 36.2, SD = 12.46) and education (1% license primary school; 35% secondary school; 33% bachelor degree; 36% master degree; 8% PhD / medical specialization).

Participants were asked to rate on a 7-point Likert scale the subjective frequency of the nouns. The administered question was: "How frequently have you read or heard these words?". Participants were instructed not to express normative judgments, but to focus on the frequency they have heard or read the words. The possible answers were: 0 = never heard or seen; 1 = one/ few times in my whole life; 2 = once a year: 3 = once a month; 4 = once a week; 5 = once a day; 6 = more than once a day.

The questionnaire was divided into two parts so that a participant never saw the singular and the plural of the same noun. The task started with 15 trial nouns that were not included in the experimental list. The questionnaires were administered online by means of SurveyMonkey platform (surveymonkey.com); the nouns in the questionnaires were presented to each participant in a different random order.

Results

The results of the rating for each occurrence is reported in appendix B. The distribution of the subjective frequency scores for singular and plural across participants is summarized in table 1 and plotted in figure 1 in§III.2.1.

Table 1. Summary of the subjective frequency rating scores.

SCORE MEAN	SINGULAR	PLURAL
n = 0	0	0
0 < n ≤ 1	0	7
1 ≤ n ≤ 2	3	47
2 ≤ n ≤ 3	45	60
3 ≤ n ≤ 4	88	63
4 ≤ n ≤ 5	70	36
n > 5	14	7

Figure 1. Distribution of the subjective frequency rating scores.
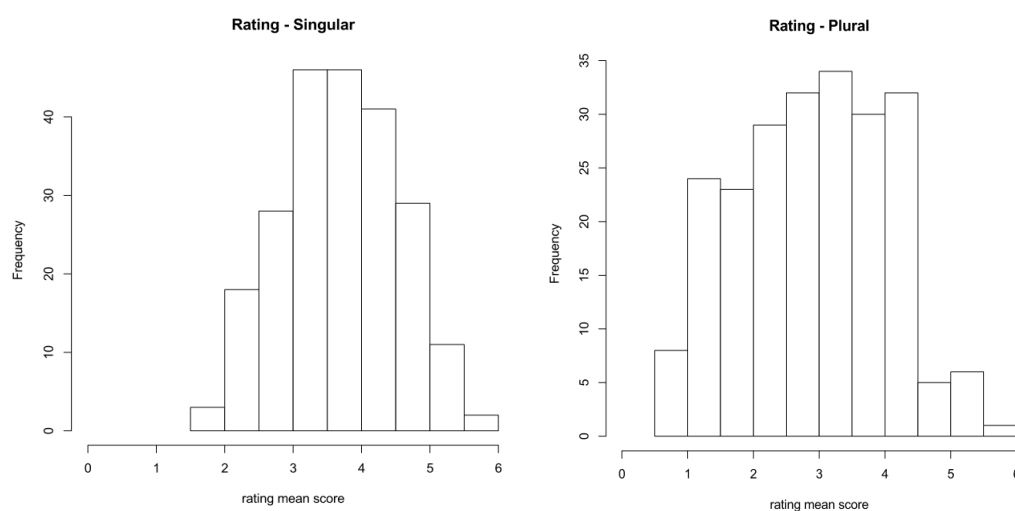
The participants assigned in general lower scores to plural nouns; however no experimental stimulus had a mean of zero. In other words, every noun of the list had been heard or read by the participants at least once in their life, even those plural forms that a normative grammar would consider agrammatical. This result is even more unexpected when considering that only 7 out of 224 plural items were scored <1, i.e. *sangue* 'blood', *cacao* 'cocoa', *cute* 'skin', *pongo* 'play dough', *iodio* 'iodine', *latte* 'milk', *pece* 'pitch' (see the scoring in appendix B). Among these, only *sangue* and *latte* stand as outliers displaying a high score for the singular, while the other five nouns received low scores both in the singular and in the plural.

The survey described above assessed that all nouns in the list can receive a count reading, since they are attested in their plural form. In Italian, in fact, plural is *per se* a hallmark of countability. Different subjective frequency scores for the plural get along with the findings reported in Katz & Zamparelli (2012); Kulkarni, Rothstein & Treves (2013); Schiehlen & Spranger's (2006) as they highlighted a non-polar distribution of nouns with respect to the mass-count ideal trajectory.

It is worth to notice that the plural occurrences of nouns considered as mass by normative grammar are well represented. It follows that every noun of our rating can have a countable interpretation, even if these are attested with a different range of subjective frequency. In this sense, theories that postulate a polar dichotomy between two categories do not predict economically the observed distribution. In fact, the grinding and portioning operations can explain these data, however it does not seem economic to postulate that these operation take place for such a huge portion of the lexicon.

III.2.2. A corpus analysis

The results of the rating study allow to draw inferences for what concerns the plural, but not for what concerns the singular. In fact, from the one side a plural occurrence implies a count reading, from the other side occurrences in the singular may be interpreted either as countable or as uncountable.

The distribution of the singular occurrences with respect to their mass or count use can be better captured by an analysis on corpora. We chose a corpus among the ones available for Italian. Thus, we first measured correlations between subjective frequency scores and absolute frequency values in corpora of all the experimental stimuli. We chose the corpus that showed the best correlation with the subjective frequency scores, i.e. itWaC. We considered that the subjective frequency was a better estimate of the actual possibility of occurrence of nouns, as suggested in the literature (i.e. Balota, Pilotti, & Cortese, 2001; Kuperman & Van Dyke, 2013; Williams & Morris, 2004). Then, we performed queries on the itWaC corpus in order to measure the frequency of occurrence in mass and count contexts relative to each noun.

Selection of the corpus

As a first step, we collected the absolute frequency values of all the experimental stimuli included in the rating questionnaire from various Italian corpora. Following Brysbaert & New's (2009), in the present analysis we considered only POS-tagged corpora of Italian sized 16M tokens as a minimum: La Repubblica (Baroni et al., 2004), ~380M tokens; Subtlex-It (Crepaldi et al., 2013) ~130M tokens; itWaC (Baroni et al., 2009) ~2 billion tokens. Colfis (Bertinetto et al., 2005), although its size is ~3.8M tokens, was included as well since a long tradition of linguistic, psycholinguistic and neurolinguistic studies have used it as a reference for lexical frequency.

Results on estimated subjective frequency and on corpora frequency were compared by means of correlations. In fact, we considered that the subjective frequency was a better estimate of the actual possibility of occurrence of nouns (i.e. Balota, Pilotti, & Cortese, 2001; Kuperman & Van Dyke, 2013; Williams & Morris, 2004). Observed correlations with the set of experimental nouns ranged from 0.70 to 0.75: itWaC showed the best correlation, whereas Repubblica showed the lowest correlation. The observed

correlations are reported in table 1 in §III.1.1; the related plots are reported in figure 1 in §III.1.1.

Table 1. Observed correlations between subjective frequency results and corpora absolute frequency.

REPUBBLICA	r(446) = 0.70, p < 0.001
COLFIS	r(446) = 0.71, p < 0.001
ITWAC	r(446) = 0.75, p < 0.001
SUBTLEX-IT	r(446) = 0.74, p < 0.001

Figure 1. Correlations between subjective frequency results and absolute corpus frequency.
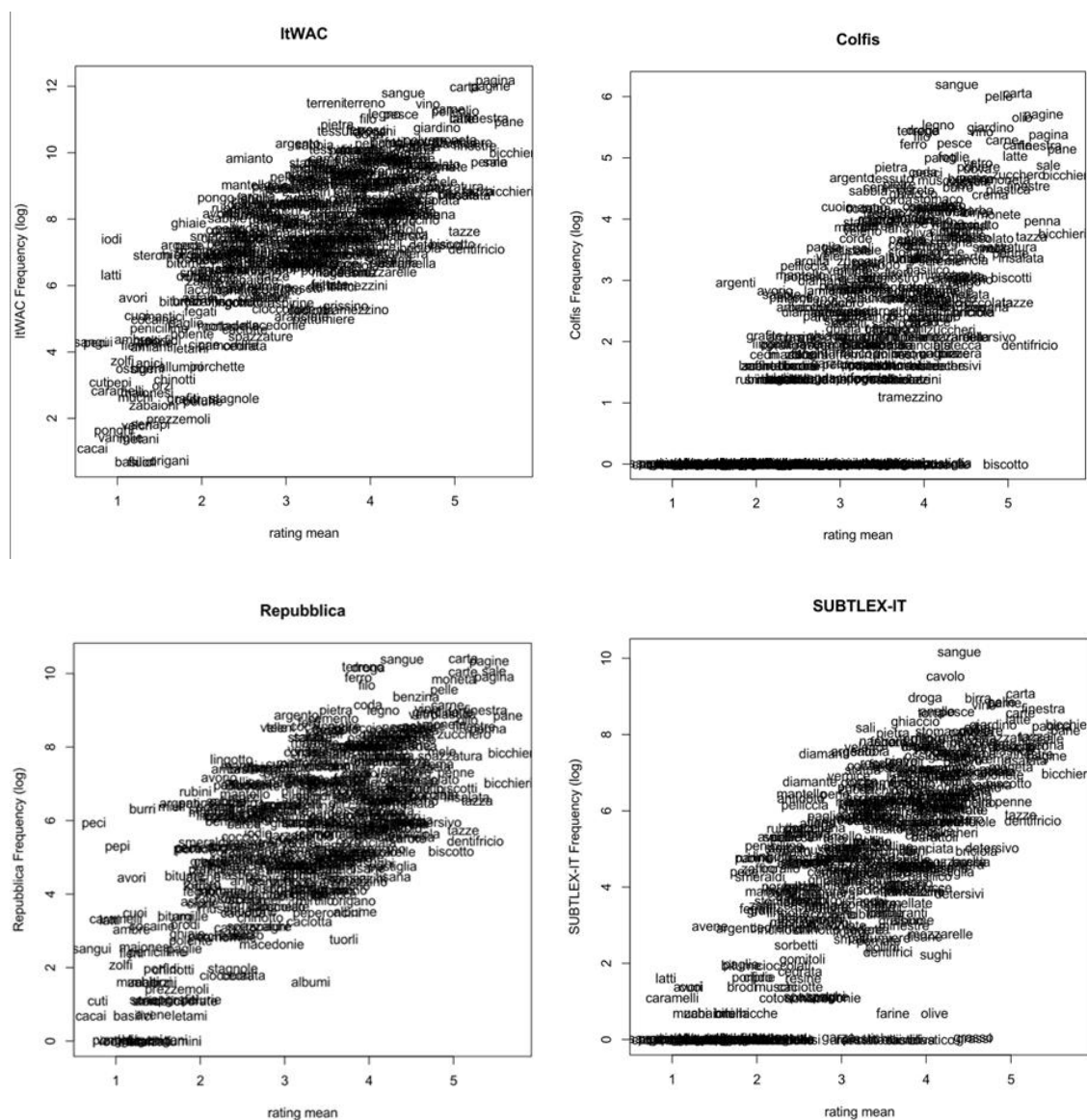
We considered a mean ≥ 2 in the subjective frequency rating as a threshold of a consistent judgment of plausibility across the participants in the questionnaire. Even

when performing correlations only nouns whose subjective frequency was ≥ 2 , two corpora still returned zero occurrences (Colfis: n = 70, Subtlex-IT: n = 23).

From the analysis illustrated in this paragraph, it emerged that some corpora seem to capture better than others the distribution of possibilities that are largely accepted by the native speakers. ItWaC shows the best correlation with the rating scores, thus it was selected for a closer examination on mass and count contexts.

Comparing mass and count occurrences of nouns

The subjective frequency rating suggested that even alleged mass nouns unexpectedly occur in the plural, thus revealing a widespread possibility to surface with a countable reading. The plural is *per se* a hallmark of countability. Conversely, the singular Number value covers both the case of singular count occurrences and the case of mass occurrences. One possibility to disambiguate these latter consists in considering the syntactic contexts of occurrence performing queries on corpora. On this purpose, we performed queries targeting mass context and count singular context. Plural occurrences were queried within syntactic contexts as well, in order to estimate the possible dispersion due to query limits in general.

Methods

In order to disambiguate within itWaC the mass use from the count use of the nouns presented in the rating questionnaire, we designed queries in CQP syntax. Katz & Zamparelli's (2012) study was taken as an example for building appropriate queries for Italian. One of the major problems in designing the queries consists in the fact that not all the syntactic contexts entail the possibility to disambiguate. For example, phrases like *la pizza* 'the pizza' (i.e. definite article + noun) do not provide sufficient information on this purpose. Enlarging the environment up to the sentence level may not provide additional clues as well. In the sentence *La pizza che ho mangiato era fatta in casa* 'the-Sɢ pizza-Sɢ I ate was home-baked', the noun *pizza* may have a count reading (one single pizza) or a mass reading (the substance the pizza is made). Given so, we considered the syntactic contexts unambiguously linked to only one interpretation:

(1)   **Mass:**
        Verb (+ Adverb(s)) (+Adjective(s)) + bare N-Sɢ

Quantification expression (*molto, poco, troppo, un po' di*, …) + N-S<small>G</small>

(2) **Count S<small>G</small>:**

Indeterminate article (+Adjective(s)) + N-S<small>G</small>

(3) **Count P<small>L</small>:**

Verb (+ Adverb(s)) (+Adjective(s)) + bare N- P<small>L</small>

Numeral/quantification expression (*molto, poco, troppo, un po' di*, ….) + N-P<small>L</small>

In appendix B we report the queries designed to match these purposes, and the queries actually performed. Some words must be spent on this point. Due to computational limits it was not possible to perform queries that would catch a phrase such as *un buon vino* 'one-S<small>G</small> good-S<small>G</small> wine-S<small>G</small>', but only phrases where no modifiers are present between the determiner and the noun (*un vino* 'one-S<small>G</small> wine-S<small>G</small>'). Of course, cases such as *un buon vino* are probably not marginal in the corpus. Since it was not possible to perform queries that included cases in which the nouns was preceded by modifiers, the countability could not be traced for all occurrences, as it will be explained in the next paragraph.

Dispersion of occurrences

The results returned by the queries were consistent with the trend of the rating scores with respect for the countability in general: as it will be discussed below, nouns that should be mass from a normative point of view can frequently occur in count context as well. Before illustrating the results in detail, it must be noted that it was not possible to capture all the occurrences. We take as an example the case of *vino-vini* 'wine-wines'(table 2, in §III.1.1). Taking the plural as an undeniable hallmark of countability, it is out of question that for what concerns the plural item *vini* the count occurrences are 42405. The plural contexts returned by the queries are only 198 instead. In this sense, queries performed on the plurals work as an indicator of the size of the possible dispersion. Likewise, at the singular, the dispersion may have equally huge dimension and variable proportions, since singular entails both mass and count readings. For this reason it is not possible to compare directly the total occurrences with respect to the results of the queries.

Table 2. Returned occurrences for the noun *vino* 'wine'.

|	FREQ_itWaC	COUNT_PL.	COUNT_SG.	MASS
vini	42405	198		
vino	96649		5954	73

This dispersion of occurrences may be due to several reasons. Firstly, it was not possible to perform queries based on more complex contexts due to computational limits. In addition, many syntactic contexts, such as the ones with the definite articles, were excluded from the queries for their ambiguity with respect to countability. At the moment, there are no tools to trace the countability of one occurrence with respect to the discourse.

Even if it was not possible to measure the distribution of all occurrences in the mass vs. count contexts, it was possible to check the reliability of the measured occurrences. Thus, correlations of either context with respect to the absolute frequency were performed. The observed correlations are reported in the table 3 and represented in figure 2 in §III.1.1. All correlations were significantly positive. As a consequence of this, it was possible to consider our data, although partial, as reliable for what concerns the occurrence of nouns in mass and count context.

Table 3. Correlations between absolute frequency and syntactic context of occurrence.

ABS_FREQ / MASS	$r(126) = 0.61$, $p < 0.001$
ABS_FREQ / COUNT SG.	$r(201) = 0.52$, $p < 0.001$
ABS_FREQ / COUNT PL.	$r(152) = 0.81$, $p < 0.001$

Figure 2. Correlations between absolute frequency and syntactic context of occurrence.
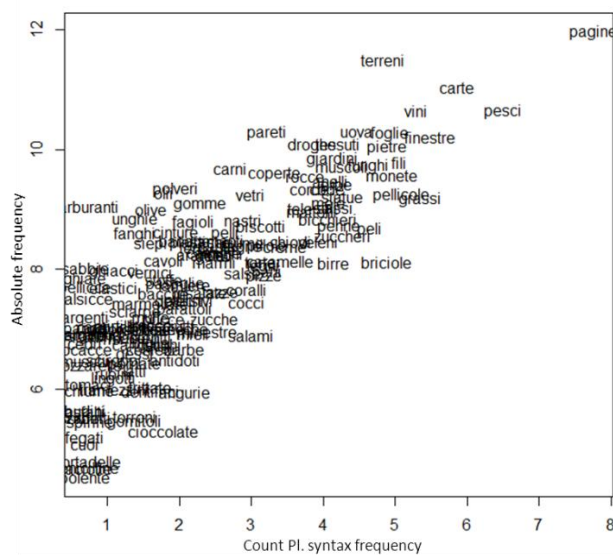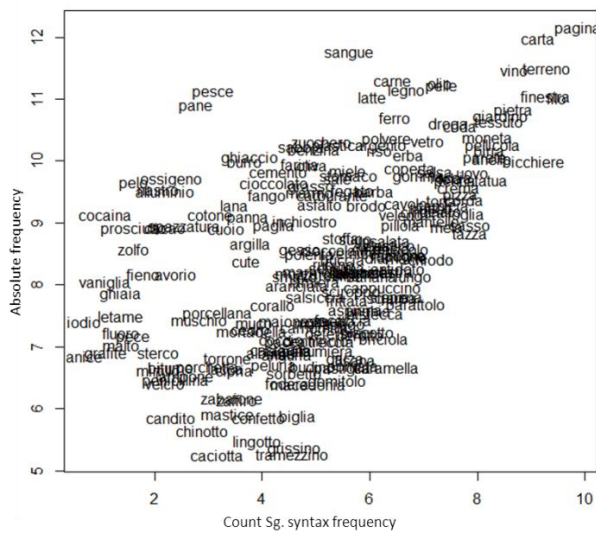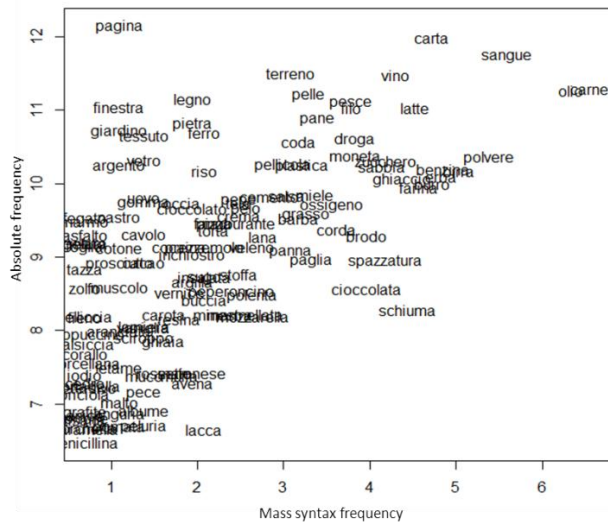
Results and interim discussion

We performed a corpus analysis in order to check the existence of a correlation between occurrences of a same noun in mass syntax and in count syntax. The calculated correlations are summarized in table 4 and plotted in figure 3 in §III.1.1. Firstly, we analysed data for our selected stimuli. Then, in order to avoid possible biases linked to the selection of items, we performed an analysis considering the top 100 nouns that appear more frequently in a mass context as correlated with their appearance in count context; similarly we report the correlation of the top 100 nouns that appear more frequently in a count context and their appearance in mass context.

Table 4. Correlations between frequency of occurrence in mass syntax and in count syntax.

EXPERIMENTAL STIMULI (NOUNS IN THE RATING)	r(205)= 0.13, p = 0.07
100 NOUNS MOST FREQUENTLY USED AS MASS	r(98) = 0.26, p = 0.009*
100 NOUNS MOST FREQUENTLY USED AS COUNT	r(98) = 0.07, p = 0.47

Figure 3. Correlations between frequency of occurrence in mass syntax and in count syntax.

100 most frequent words in Count syntax

100 most frequent words in Mass syntax
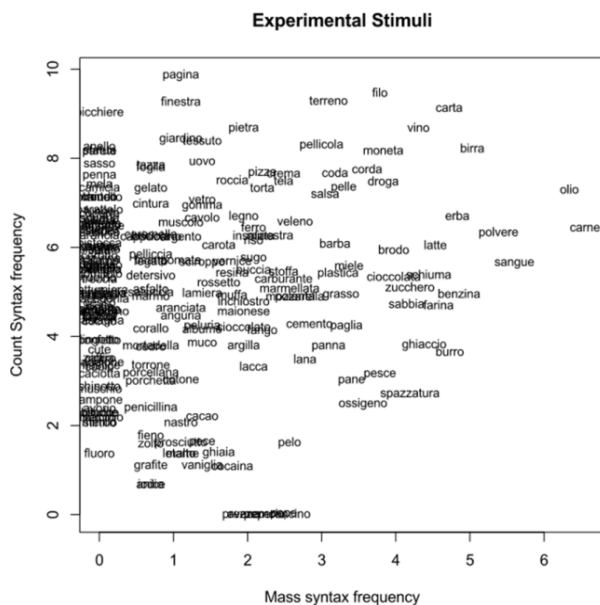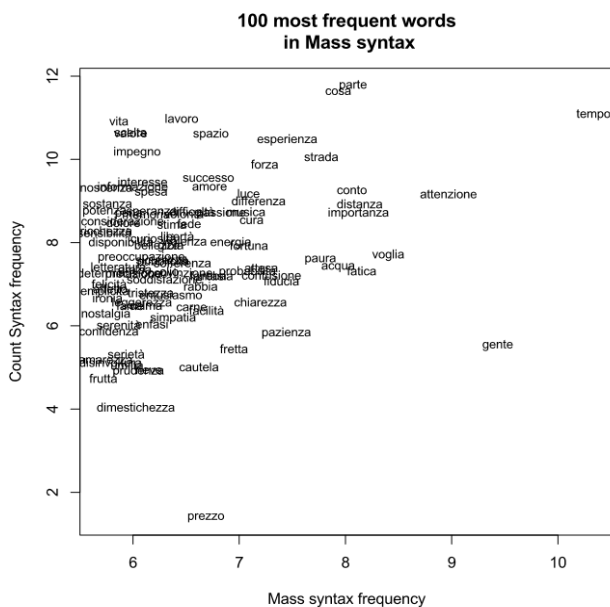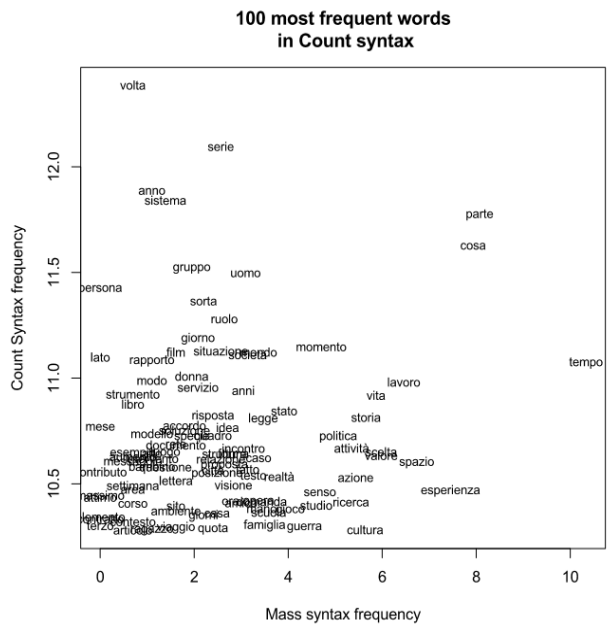
Notably, we did not find negative correlations with respect to our stimuli and even when considering all nouns in the ItWaC corpus; indeed, we found null correlations and, in the case of the top 100 nouns most frequently used as mass, a positive correlation. In other words, a word frequently used as count does not necessarily occur with a low frequency in mass use. Even more surprisingly, the positive correlation found for the top 100 nouns most frequently used as mass implies that the more a nouns is used as mass the more the same noun will be used as count.

The lack of observed negative correlations does not allow to detect the starting point for the operations of packaging and grinding. The data from the corpus analysis do not

consent to disambiguate whether a noun that is found in both contexts is in origin marked as mass or count at the lexical level. As this impossibility occurred in the majority of the cases, there is no empirical evidence for a clear-cut division in the lexicon between mass and count nouns.

Discussion

From the integration of subjective frequency rating data with the corpus frequency of nouns, it emerges that even nouns that traditional grammars consider as "pure mass" are attested in the corpora with a count reading and judged as possible in the plural form by native speakers as well. Crucially, alleged pure mass nouns appear with a count reading much more frequently than what would be expected in literature.

These results may point to the fact that nouns are not lexically specified in relation to their countability. With respect to such findings, from a lexicalist point of view it could be argue that the operation of packaging can potentially be applied to any mass noun. If this was the case, it would be expected that a considerably high number of nouns could occur in the singular much more frequently than they occur in the plural. In fact, if a noun was lexically mass, one would expect that its use in an uncountable context would be much more frequent with respect to its use in a countable context. The data reported seem not to show such a distribution. As a consequence, it would be impossible to recognise which operation took place and from which starting point; more precisely, there are no data that allow us to discriminate unambiguously a grinding operation applied on a count noun from a packaging operation applied on a mass noun.

For such reasons, the empirical data illustrated above better fit with formal approaches that do not trace predictions about being mass or count on the base of alleged lexical properties of nouns and assume the mass and count properties to be contextually assigned. Moreover, such approaches explain empirical data form corpus analyses more economically, without postulating semantic shift operations.

On this regard, it must be kept in mind that linguistic signs do not refer directly to things, but to mental representations of them. Thus, language does not have to mirror all the physical properties of the reference *tout court*. In other words, there are no reasons intrinsic to language for nouns referring to substances to occur only with a mass reading, nor there are for nouns referring to objects to occur only with a count reading.

The rating study and the corpus analysis have revealed that in Italian almost every noun can appear in count contexts. Moreover, negative correlations were never found between the count use and the mass use of nouns. From such results, it emerges that countability is better explained as a contextually assigned property rather than as a feature set in the lexicon.

The predictability for nouns to occur mostly with a mass or a count reading is not an issue that can only be solved within linguistic theory; instead, it must be dealt with also by taking quantitative measurement, that can, in turn be useful to endorse theories that economically explain data. In particular, this integrate approach of rating and corpus analysis performed on other languages may provide insightful data on the issue of countability, especially when considering languages of a different morphological type.

III.3 A lexical decision study

Quantitative data as the one presented in §III.2 about the distribution of the nouns could also lead to a methodological improvements in psycholinguistic research on countability. So far, data from experimental literature are far from being clear and consistent (see above in §II.3). Linguistic theories that describe uncountability as formally simpler (Krifka 1995; Borer 2005; De Belder, 2011) seem not confirmed by experimental literature, as most empirical data report the processing of mass nouns as equally or more demanding if compared to count nouns (see above in §II.3; for a review: Semenza et al., 2012). Some role of the context is eventually reckoned even in studies starting from a lexicalist theory (Frisson & Frazier, 2005), however the complete picture is far from homogeneous. Even within the specific field of lexical decision studies, results are not distinct: some studies reported longer RTs for mass nouns as compared to count nouns (Gillon et al., 1999; Mondini et al., 2009); other studies did not replicate this finding (Mondini et al., 2008; El Yagoubi et al., 2006). It is possible that the discrepancies of empirical data stem from suboptimal choices in the selection of experimental stimuli. Up to now, in experimental literature the classification of stimuli as mass or count nouns has been based mostly on the experimenters' intuition. The quantitative data presented in §III.2

provided a better estimate of the frequency of use of nouns as countable or uncountable. This measure was be used as a proxy to classify a noun as mostly mass-used or mostly count-used in order to build unbiased experimental lists.

Methods

In this study we categorized experimental stimuli as *mass-used* or *count-used.* For this classification, rather than using a theoretical approach, grounded on normative grammar rules or on experimenter choices, stimuli were classified by means of a corpus-based approach (see§ III.2.2)

We classified as *mass* those nouns with high mass frequency and low count frequency, and *count* nouns those with high *count frequency* and low *mass frequency*.

Materials

80 concrete nouns were selected from the questionnaire on subjective frequency described in §III.2.1. The nouns were selected to span as uniformly as possible across the range of possible values of subjective frequency (of singular and plural).

From the set of 80 we identify we classified 18 *mass-used* nouns and 18 *count-used* nouns as follows:

- the *mass-used* nouns were the top 18 nouns with highest mass frequencies and values of count frequencies that were not among the top 18;
- the *count-used* nouns were the top 18 nouns with highest count frequencies and values of mass frequencies that were not among the top 18.

Since the *mass-used nouns* and *count-used nouns* were included both in the singular and in the plural this lead to a total of:

- 36 *mass-used* nouns (18, inflected in the singular and in the plural)
- 36 *count-used* nouns (18, inflected in the singular and in the plural)
- 44 nouns that could not be classified as mass-used or count-used

All stimuli were included in singular and plural form, leading to a total of 160 stimuli.

The *mass-used* and *count-used* in the singular form were matched for length, corpus frequency, and subjective frequency (in independent *t-test* all ps > 0.05). In the plural (as

expected) *mass-used* stimuli had significant lower corpus frequency and subjective frequency (in independent *t-tests* ps < 0.05), but comparable length (in independent *t-test p* < 0.05).

The collection of data on corpus frequency and the data on frequency of occurrence in contexts in the table (and used in the analyses) are described in III.2.2. A complete list of the experimental stimuli, with the associated psycholinguistic properties, is reported in Appendix B. Descriptive statistics for main psycholinguistic properties of the stimuli are reported in Table 1.

The experimental list was split in two versions in a split-plot design, such as each participant saw each stimulus either in the singular or in the plural form of a given word. The final list included also 80 adjectives, included as fillers that were matched for frequency and length with experimental stimuli. Half of the adjectives were in the singular form and half in the plural form. Fillers were matched with experimental stimuli for frequency [Wilcoxon W = 6592, p = 0.72].

For the purpose of the lexical decision task 160 nonword were created from the experimental stimuli and the fillers by replacing or transposing 2-4 letters. All nonwords respected phonotactic rules of Italian. Half of the nonwords ended with a letter that in Italian typically marks the singular form, whereas half of the nonwords had a plural-like ending letters.

To summarize, each participant saw half of the experimental stimuli (80 nouns on a total of 160 stimuli), 80 fillers and 160 nonwords. For each participant half of the stimuli were in the singular, and the other half in the plural.

Table 1 – Experimental items

	NUMBER OF STIMULI	CORPUS FREQUENCY	SUBJECTIVE FREQUENCY	LENGTH
ALL STIMULI	160	11850.32 (27239.65)	3.29 (1.18)	6.41 (1.66)
MASS-USED NOUNS – SINGULAR	18	26204.88 (28831.43)	4.36 (0.57)	6.22 (1.89)
MASS-USED NOUNS – PLURAL	18	824 (1187.38)	1.95 (0.72)	6.28 (1.96)
COUNT-USED NOUNS – SINGULAR	18	38570.05 (54194.95)	4.09 (0.84)	5.78 (1.31)
COUNT-USED NOUNS – PLURAL	18	24365 (36455)	4.07 (0.80)	5.89 (1.27)

Participants

60 students from the University of Padova, participated in the experiment (mean age 23.5 years, SD = 2.37; mean education 15.16 years, SD = 1.64; 36 female, 24 males) All participants were Italian native speakers, reported no neurological pathologies, had normal or corrected-to-normal vision, and took part to the experiment on a voluntary basis.

Procedure

Stimuli were delievered by means of DMDX software (Forster & Forster, 2003). Participants saw a series of letter strings presented at the center of the screen one at a time. They were instructed to decide as quickly and accurately as possible whether or not each string was a real word, by pressing two buttons with the index fingers of both hands. Buttons for word and nonword response were counterbalanced across participants.

A fixation point was shown for 500 ms on a dark gray background, followed by the target presented in white 12-point uppercase Courier New letters. The targets remained on the screen until a response or until 2000 ms expired; an inter-stimulus-interval (blank screen) of 830 ms followed. Participants received no response feedback. Six practice trials were administered before the beginning of the experiment to familiarize with the procedure.

Data analysis

Data were analyzed by means of a mixed model effect analysis (Baayen at al., 2008). The advantage of such method Is the possibility to consider more types of effect in the same analysis, both fixed and random. Specifically, it is possible to include at the same time different variables as random effects (such as items or subjects), and different predictors (continuous and categorical) as fixed effects.

Model 1 – Replicating previous lexical decision studies - 72 mass and count nouns

This analysis aimed at replicating the results in previous lexical decision studies. This model included three fixed effects, namely *Category* (mass-used vs. count used), *Number* (singular vs. plural) and the interaction (*Category x Number*). Subjects and items were considered as random effects.

Two models were calculated: one considered the Accuracy as a dependent variable (Table 2, §III.3) and the other considered Response Times (Table 3, §III.3).

Table 2 – Accuracy.

Effect	Coefficient	Standard Error	z	p-value
Intercept (categoria lessicale = contabile, numero = singolare)	4.90	0.56	8.73	< 0.001*
Categoria lessicale = massa	0.75	0.78	0.96	0.34
Number = plural	0.36	0.75	0.48	0.63
Interazione categoria = massa * numero = plurale	-3.99	1.05	-3.78	<0.001*

Table 3 – RTs

Effect	Coefficient	Standard Error	df	t	p-value
Intercept (Category= count, number = singular)	6.47	0.25	110.79		< 0.001
Category = mass	-0.002	0.02	64.54	-0.095	0.92
Number = plural	0.03	0.02	64.65	1.16	0.25
Interaction Category = mass * number = plural	0.10	0.03	67.63	3.02	0.003*

A significant interaction (*Category x Number)* was found. Specifically, mass nouns inflected in the plural elicit less accurate responses and longer reaction times (see Figure 1, §III.3). No significant difference was found between mass and count nouns in the singular.

Figure 1. Model 1 – RTs. 72 stimuli; predictors: Category, Number, Category X Number
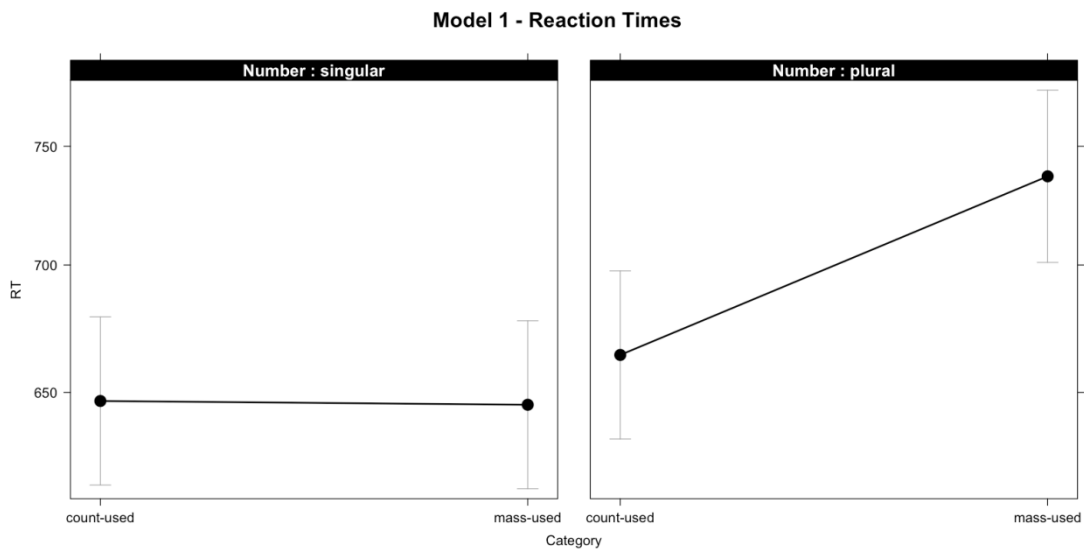
Model 2 – Measuring the effects of Subjective Frequency, Corpus Frequency, Length and Number in lexical decision - 72 mass and count nouns

The aim of Model 2 was to understand if effects observed in Model 1 could be explained as frequency effects.

In the analysis we considered the predictors of Model 1, C*ategory* and *Number*, and added *Subjective frequency, Corpus frequency* and orthographic *Length* as continuous variables. Results for accuracy and RTs are reported in Table 4 and Table 5.

Table 4. Model 2 – RTs

Effect	Coefficient	Standard Error	df	t	p-value
Intercept	6.56	0.05	95.18	130.53	<0.001*
Numero = plurale	0.37	0.02	64.33	2.04	0.04*
Subjective Frequency	-0.04	0.007	74.09	-4.27	<0.001*
Length	0.009	0.004	65.86	2.077	0.04*

Results show show a significant effect of *Length* (longer RTs for longer items), of Number (longer RTs for plurals) and for *Subjective Frequency* (longer RTs for low subjective frequency.

The other variables included in the analysis - *Corpus Frequency, Category, Number* and the interaction *Category x Number* - are not reported as their effects are not statistically significant in this model. It is important to notice that when including the effect of other variables (*Length* and *Subjective Frequency*) in the analysis, the difference reported for plural mass and plural count nouns reported in Model 1. This points to the fact that the effect of longer RTs for plural mass nouns is explainable as a difference in *Subjective Frequency* (or *Length*) of the items.
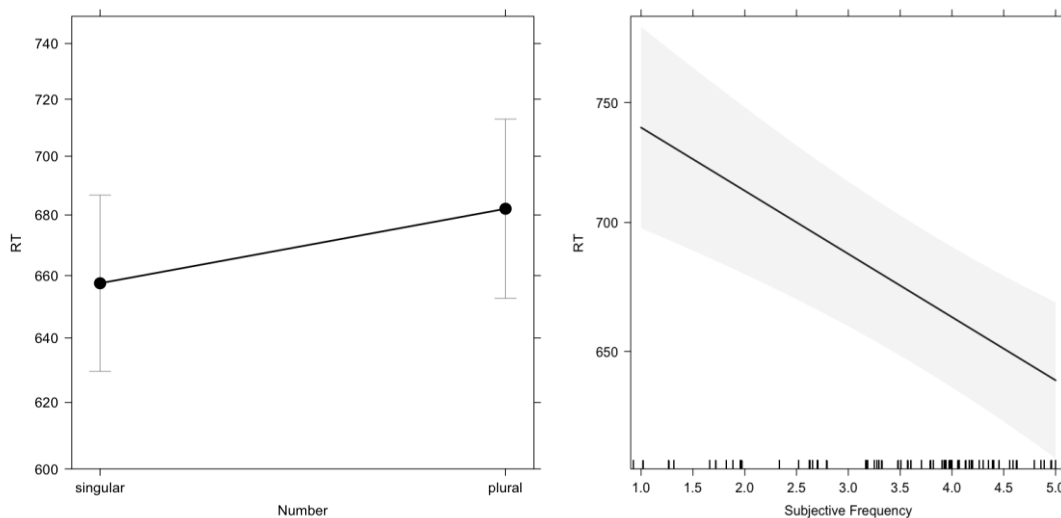
Importantly, the only significant effect for Accuracy is Subjective Frequency (see Table 5, §III.3).

Table 5. Model 2 – Accuracy

Effect	Coefficient	Standard Error	z	p-value
Intercept	-0.34	0.60	-0.57	0.57
Subjective Frequency	1.30	0.18	7.16	<0.001*

The effects are plotted in Figure 2, §III.3.

Figure 2 Model 2 - 72 stimuli; Category, <u>Number</u>, Corpus frequency, <u>Subjective Frequency</u>, <u>Length</u>

Model 3 – Measuring the effects of Subjective Frequency, Corpus Frequency, Length and Number in lexical decision – 160 concrete nouns.

In this last model the analysis is extended to all the 160 experimental items. In the analysis we considered *Number, Subjective frequency, Corpus Frequency* and *Length* as variables. Results on RTs show significant effects of *Length, Corpus Frequency, Subjective Frequency*. Differenty from Model 2, no significant effect of *Number* was found. Results are reported in Table 6 and plotted in Figure 3.

Table 6. Model 3 – RTs

Effect	Coefficient	Standard Error	df	t	p-value
Intercept	6.73	0.04	219.42	172.38	<0.001*
Corpus Frequency	-0.009	0.004	155.55	-2.16	0.03*
Subjective Frequency	-0.05	0.008	152.19	-5.37	< 0.001*
Length	0.008	0.004	2.47	2.11	0.04*

No significant effect was found in the analysis on Accuracy. This difference may be due tu the fact that the added stimuli have improved a ceiling performance, thus not allowing to identify significant effects.

Figure 3 – Model 3 - 160 stimuli; Number, <u>Corpus frequency</u>, <u>Subjective Frequency</u>, <u>Length</u>
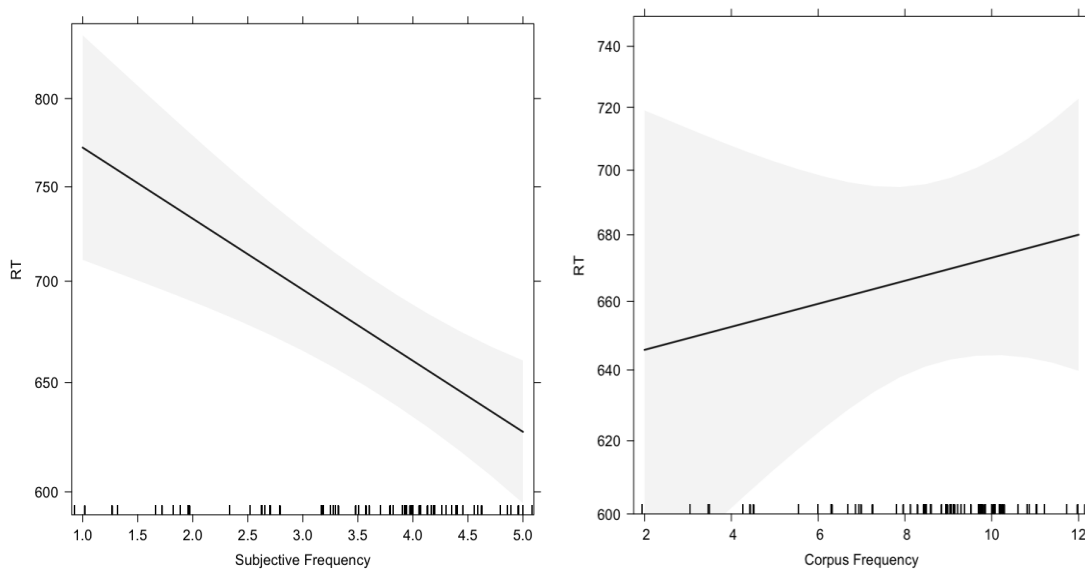
Discussion

Results from this study seem to challenge the conclusions on mass/count distinction as studied in lexical decision tasks in the previous literature. By performing a stricter control on the selection of the stimuli, we did not replicate the results of the previous studies, although the employment of the state-of-the art statistical analyses and in presence of an excellent statistical power.

Frequency effect may be taken into wider consideration in designing experimental tasks on this topic. In sum, some discrepancies with preceding results are easily explainable in terms of suboptimal choices in the experimental designs previously adopted.

Diifferences in lexical access between mass-used and count-used nouns are better explained in terms of frequency. The lack of specific effects countability in lexical access of single words seem to confirm that countability is not a lexical feature and is rather dependent from the syntactic context, as suggested by quantitative studies.

III.4 A study in acquisition

This last study was designed in order to measure the role of non-verbal cognitive abilities concerning the encoding of countability. In other words, the purpose was to test experimentally if the mass-count distinction is merely a linguistic issue or is it coded in representations other than language.

We hypothesised that differences in countability should be observed even neutralising any linguistic distinction driven by frequency of the occurrence of nouns in a syntactic context. To evaluate this hypothesis, we tested the 5-6 year-old children's ability to judge sentences with mass nouns (i.e. those that appear more frequently in mass morpho-syntax, see §III.2, quantitative studies), count nouns (i.e. those that appear more frequently in count morpho-syntax), and neutral nouns (i.e. those that appear in mass and count contexts with similar frequency).

In the present study we hypothesised that even though mass morphosyntax might imply fewer operations at the linguistic level, its processing might be more demanding from a cognitive point of view, because mass reference entails additional abstraction abilities. The extra operations needed to overcome these bias should delay the processing of mass reference with respect to count reference. More critically, because abstract thinking mature in children only throughout development, their gradual acquisition might delay the use of mass morpho-syntax with respect to count morpho-syntax.

Our hypothesis finds support on developmental studies that have explored the mass and count issue not only from a linguistic point of view, but also from a more cognitive perspective, in particular concerning the children's knowledge of objects and substances, typically associated to count and mass nouns respectively. Soja, Carey & Spelke (1991) showed that, before mastering count or mass morpho-syntax, English-speaking children at the age of two know the conceptual distinction between objects and substances. Therefore, the ability to recognize and distinguish these two categories is pre-linguistic and independent from the spoken language (Imai & Gentner, 1997). However, some specific cognitive biases in early word learning seem to favour objects before substances. While acquiring the lexicon, children are likely to assume a new word to refer to a whole-object, not to the substance or parts of the object (for a review see Bloom & Kelemen, 1995; Markman, 1990). This predilection for objects might in turn disfavour the most prototypical mass nouns. In fact, when a peculiar class of mass nouns, namely "object – mass" nouns referring to objects are presented (e.g., *furniture*, *luggage*) young children overextend the count morpho-syntax to refer to them (Barner & Snedeker, 2005), suggesting a strong bias in favour of the count nouns (or the count morpho-syntax). This also suggests that fully conceiving mass nouns and mass morpho-syntax might be associated with the ability to conceive an entity (e.g. milk) independently of its perceived boundaries (e.g. a glass of milk, a bottle of milk).

As reported before with respect to the lexical decision study, frequency is another factor that might also explain the apparent facilitation in the processing of count nouns with respect to mass nouns. In fact, frequency plays a crucial role for the acquisition of nouns. Children learn the mass and count distinction either through semantic exposure, namely on the basis of the frequency with which a noun occurs as referring to either individuals or non-individuals (Bloom, 1999) or through the co-occurrence of the various nouns in a specific morpho-syntactic context (Gordon, 1985; Barner & Snedeker, 2005; 2006). Unfortunately, to our knowledge, no previous study has measured and controlled for the frequency with which the stimuli chosen in experimental settings are used in mass contexts with respect to count context in the spoken language. This leaves unresolved the question of whether, after controlling for frequency, the formal simplicity of the mass nouns (and mass morpho-syntax) in terms of linguistics actually has its parallel in cognition.

Besides these two factors, it is also possible that some suboptimal choices in the experimental designs have influenced the results in previous literature. In particular, the

visual modality of presentation of stimuli in linguistic tasks could per se be a bias in favour of a count interpretation. More specifically, an unbiased representation of mass nouns should not explicitly depict the boundaries of the referent in order to respect the linguistic distinction of this category. On the one side it is true that being an object does not imply an individual reference (i.e. Bloom, 1999). On the other side, however, it is equally reasonable to think that when something is displayed as having physical boundaries, the perception of such boundaries influences the conceiving it as an object (i.e. Prasada et al., 2002; Chesney & Gelman, 2015), thus favouring a count interpretation. A study aimed at attesting mass and count issue from a strictly linguistic perspective could be better designed avoiding every representation of the words at issue.

Methods

We developed an orally-administered experimental task in which the mass-count morpho-syntax was manipulated and the frequency of occurrence of nouns as mass or count was controlled. The aim was two folded. First and foremost the study was designed to explore whether the use of mass morpho-syntax relates to the development of abstract thinking in 5-6 year-olds. Secondly, we aimed at investigating whether the mass vs. count distinction is observable once (i) the frequency with which nouns occur in morpho-syntactic contexts is controlled and (ii) there is no influence of visual presentation of stimuli.

The so designed Mass And Count Test (MACT) assesses the children's ability to judge sentences with mass and count nouns. Mass and count nouns were presented in congruent or incongruent contexts. In congruent contexts a noun that is most frequently used as mass appeared in a mass morpho-syntax and in incongruent contexts a noun that is most frequently used as mass appeared in a count morpho-syntax; the same holds for count nouns.

Since at this age children have completed the acquisition of basic grammar structures (i.e. Tomasello, 2003), we should anticipate no differences in the performance with mass and count nouns in congruent contexts in this test. However, based on previous literature, we expect participants to perform better in the count incongruent condition than in the mass incongruent condition because of their susceptibility to overextend the count morpho-syntax (Gordon, 1985; Barner & Snedeker, 2005). Importantly, the above hypothesis might not hold true if previous results were driven by differences in the

relative frequency of the nouns chosen or biased by the visual presentation of the stimuli, which were eliminated in the present study.

Finally, if the hypothesis that the mass interpretation refers to more abstract properties of an entity was true, we should expect that abstract thinking in children correlates with the participants' performance with mass nouns but not count nouns in the MACT test.

Participants

Parental informed consent was obtained for 63 children attending the last year of kindergarten to participate in the study. Data from 58 children (26 males, 32 females) were included in the analysis. Children were distributed in ten different classes of the "A. Rosmini" school in Andria (Barletta – Andria – Trani), Italy. None of the 58 participants had a clinical history of specific language impairment or intellectual disability. All the children were native speakers of Italian and were at least five years old (age range: 62-76 months; mean = 69,7; SD = 3,1).

Procedure

An experimenter who was blind to the goals of the study, tested each participant individually in a silent room in two testing sessions. Participants completed the Mass And Count Test (MACT) along with the Logical Operations and Conservation test (LOC; Vianello & Marin, 1997) that assesses children's abstract thinking. Additionally, the test of Grammatical Comprehension (TCGB; Chilosi & Cipriani, 1995) was used as a screening measure of morpho-syntactic comprehension.

Materials

Mass And Count Test - MACT

We developed an experimental task in order to test the children's competence with respect to occurrences of nouns in mass or in count morpho-syntactic context. Different from previous studies, items were carefully selected for frequency with respect to their occurrence in mass and count morpho-syntax.

We chose nouns inflected only in the singular, and not in the plural, because plurals are only related to a count interpretation. Conversely singulars can be linked to either a count or a mass interpretation depending on the syntactic context, thus providing an unbiased testing ground. Moreover, we chose syntactic contexts (determiner + noun inflected in

the singular) in which mass or count interpretation is unambiguous: in the count context the determiner was the indeterminate article), in the mass context the determiner was a quantifier (e.g., *tanta sabbia* 'much sand' vs. **tanto anello* 'much ring').

In order to balance the frequency and the frequency of occurrence we relied on the data reported in the quantitative studies reported above in §III.2.

Only high frequency nouns denoting concrete referents were selected. The age of acquisition of the experimental nouns were controlled by means of a dedicated study. All nouns were learnt at the age of five/six by at least the 95% of the participants. The age of acquisition was balanced across types of nouns (count nouns: 97.73%; mass nouns: 98.8%; neutral nouns: 96.48%). Nouns were also controlled for syllabic length (count nouns: mean=2.5; mass nouns: mean=2.3; neutral nouns: mean=2.6).

The items chosen for the MACT were

(i)   "mass": 10 nouns that appear more frequently in a mass context and were not frequent in a count context;

(ii)   (ii) "count": 10 nouns that appear more frequently in a count context and were not frequent in a mass context.

For each noun, two identical sentences were created: in one the noun appeared in a mass context, in the other one, in a count context (table 1 in §III.4). As a result, twenty sentences were not well-formed (incongruent condition): in half of them a mass noun occurred in count context, and in the other half a count noun occurred in mass context. Twenty sentences were well-formed (congruent condition): in half of them a mass noun occurred in mass context, and in the other half a count noun occurred in count context.

Additionally we chose 20 "neutral" nouns that appear in mass and count contexts with similar frequency. Children's responses to neutral nouns are not influenced by the frequency with which they have listened to a noun in a particular context, as such they allow us to measure the participants' response in the absence of this cue. The neutral nouns appeared in 40 sentences. Based on the frequency of occurrence of these nouns in the corpus, the sentences should be considered congruent both in mass and in count context (table 1). The experimental material was therefore made up of 80 sentences (see Appendix B).

Table 1. Experimental conditions.

	Mass context	Count Context
Mass noun	CONGRUENT Leo ha tanta <u>sabbia</u> nelle scarpe. *Leo has much <u>sand</u> in his shoes*	INCONGRUENT Leo ha una <u>sabbia</u> nelle scarpe. *Leo has a <u>sand</u> in his shoes*
Count noun	INCONGRUENT La principessa ha un po' di <u>anello</u> al dito. *The princess has a bit of <u>ring</u> on her finger*	CONGRUENT La principessa ha un <u>anello</u> al dito. *The princess has a <u>ring</u> on her finger*
Neutral noun	CONGRUENT Sul tavolo c'era ancora tanta <u>torta</u>. Lit. *There was still much <u>cake</u> on the table*	CONGRUENT Sul tavolo c'era ancora una <u>torta</u>. *There was still a <u>cake</u> on the table*

The sentences were read aloud by a male native Italian speaker, audio recorded, and presented offline to the participants one at the time. The sentences were grouped in two blocks, with each noun occurring only once a block. Each block was administered in a different testing session. The order of presentation of the sentences was pseudo-randomized, with no more than four incongruent sentences in a row. Four different pseudorandom orders were produced.

Participants received the following instructions: "*Simone is a boy who is learning Italian. Could you please help him? You will listen to some sentences uttered by Simone: some are right and some are wrong. After each sentence, you have to tell me if the sentence was right or wrong*".

Each congruent sentence accepted and each incongruent sentence refused scored 1 point; each congruent sentence refused and incongruent sentence accepted scored 0 points. Each answer of the participants was transcribed by the experimenter. When the participant refused a sentence, he/she was asked to produce what he/she considered to be the correct version of the sentence.

A brief training session was administered before the actual test to make sure that the participants had correctly understood the task.

The Logical Operations and Conservation test - LOC (Vianello & Marin, 1997)

This standardized test assesses logical and abstract thinking in children. It consist of 24 tasks of increasing difficulty divided into four areas: (i) *seriation* which consists in ordering objects of different size; (ii) *numeration* which requires children to recognize the equal numerosity of two sets of objects; (iii) *classification* which consists in grouping together objects by size, colour, or form; (iv) *conservation* which consists in recognizing the identity between two samples of liquid or dough, presented in different containers and shapes. The scores were within the normal range for the age of the participants (M= 13.46, SD= 4.02).

Test of Grammatical Comprehension – TCGB (Chilosi & Cipriani, 1995)

This test requires participants to match 76 sentences of increasing complexity with one out of four pictures. It evaluates the comprehension of the sentence and the generalization ability of linguistic concepts. We chose to use TCGB as a screening test in order to make sure that participants master syntactic elaborations in line with their age. The TCGB scores were within the normal range for all the participants (M= 15.83, SD= 6.53).

Results

The children's performance on the MACT was evaluated using a 2x2 ANOVA with condition (congruent and incongruent) and noun (mass and count) as within-subject factors.
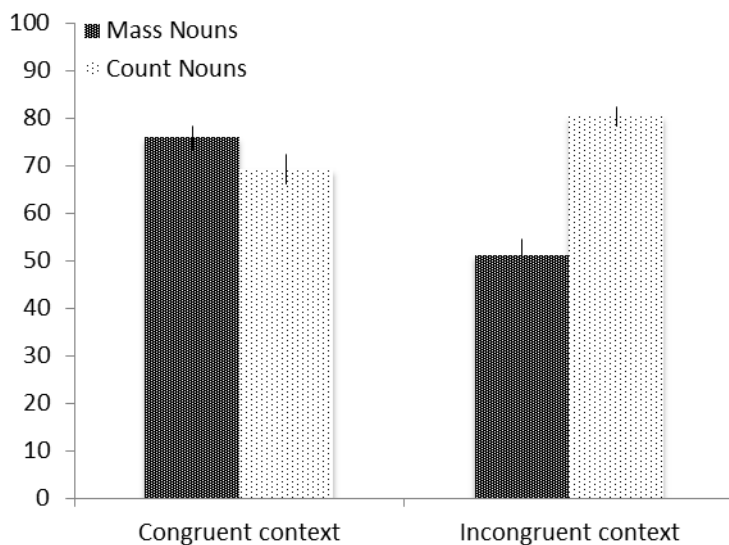
The comparison of the children's ratings on neutral nouns in mass versus count context was carried out separately, using paired t-tests.

Pearson's correlations between the children's performance on each noun-condition of the MACT (e.g. mass congruent, count incongruent, etc.) and the LOC were carried out in order to determine the implication of children's abstract thinking. To make sure that the relationships between those measures were not mediated by a general demographic

factor, we computed partial correlations in which children's age, gender, and class were controlled for.

The ANOVA analysis showed a main effect of context $F(1, 232) = 5.30$, $p<.03$; a main effect of noun $F(1, 232) = 15.99$, $p<.001$; and a significant interaction between both factors $F(1, 232) = 39.71$, $p<.0001$. There was a significant difference between mass nouns (M=51.2%, SD=25.2) and count nouns (M=80.0%, SD=16.1) in the incongruent condition ($p<.0001$). No significant difference was found in the congruent condition (mass nouns M= 75.2%, SD =19.4; count nouns M= 68.8%, SD =24.0; $p>.05$). Results are displayed in figure 2 in §III.4.

Figure 2. Children's performance on the MACT. The y-axis shows the percentage of correctly accepted sentences in the congruent condition and correctly rejected sentences in the incongruent condition for each noun type.

The comparison of the children's performance on neutral nouns showed that children were significantly more accurate in the count context (M= 78.44%, SD =18.54) than in the mass context (M= 63.96%, SD =17.71), t(58) = 5.47, p<.0001 (figure 3 in §III.4).

Figure 3. Percentage of correct responses with neutral nouns.
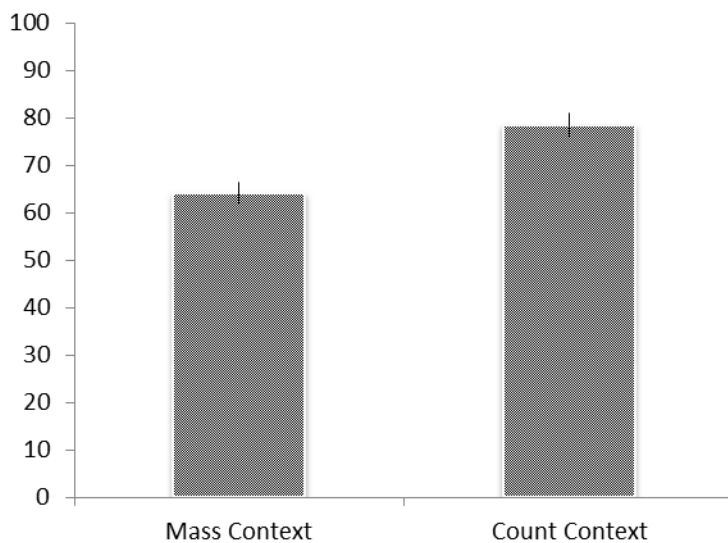
Finally, the results of the Pearson's correlation analysis showed that participants' performance on mass contexts and the children's abstract thinking -as measured by the LOC- were intercorrelated (mass nouns in mass contexts r=.27; p<.05; count nouns in mass contexts showed a similar tendency r=.24; p=.07). By contrast, participants' performance on count contexts did not correlate with the measures of abstraction (mass nouns in count contexts r=.13; p=.35; count nouns in count contexts r=.12; p=.36).

Discussion

The goal of this experiment was to investigate the relation between the development of abstract thinking and the use of nouns in mass and count syntactic contexts in 5-6 years-old children. Moreover the study was designed to evaluate the effects of frequency and verbal presentation over the mass and count distinction.

The results showed that having controlled for frequency and prevented the possible bias of a visual presentation of the stimuli, the distinction between mass and count nouns still emerges in certain conditions. From the one hand, there is no difference between mass nouns and count nouns in congruent syntactic contexts, thus suggesting that neither of the two conditions is per se more difficult than the other at this age. From the other hand, a predilection for count context emerges in the incongruent condition. Even though a similar result has been previously reported in the literature (Barner & Snedeker, 2005), for the first time our study shows that the performance in the incongruent mass

condition, i.e. children accepting mass nouns in count context, cannot be explained assuming a frequency effect. If that was the case an equally bad performance with count nouns in mass context would have occurred. In fact, in the case of low frequency of occurrence with respect to one particular context, we found that mass nouns in count context were widely more accepted with respect to the converse case, even if the sentence was not well-formed. A frequency effect is not explicative of the performance with neutral nouns either since they occur as frequently in both syntactic contexts. Even in this case, nouns in a mass context are accepted less than in count context.

Overall, these results seem to suggest an effect of the frequency of occurrence in judging correctly a noun in its proper context, both for mass and for count syntactic context. In all the other conditions, there is a strong preference for count nouns and count morpho-syntax. These trends are consistent with results in the previous literature and cannot be explained in terms of frequency, nor in terms of a specific linguistic difficulty of the mass nouns or morpho-syntax.

Thus, a possible explanation for this pattern of results seems not to be related with a mere linguistic feature.

We explored the hypothesis that, in absence of frequency effects, the mass reference is less preferred because it relays on a more demanding, extra-linguistic, cognitive process of abstraction. The literature reports the strong role played by the perceptual (or shape) bias in the acquisition and categorization of novel words (i.e. Landau, Smith, & Jones, 1988). This means that the perceived boundaries of an entity are particularly salient in recalling a noun associated to that entity. This may suggest that parsing a (concrete) noun in a phrase is easier when the reference to the boundaries is explicit, as in the case of count morpho-syntax. Conversely, the parsing a noun in a phrase may be more demanding when the reference to the boundary is suppressed, as in the case of mass morpho-syntax.

In the present study we found that measures of abstract thinking (LOC) correlated with the performance in mass contexts, both in congruent and incongruent conditions. In other words, the better the children performed in the LOC the most likely they accepted mass nouns in mass context and refused count nouns in mass context.

These data are consistent with the hypothesis illustrated above. The results obtained in the LOC can offer the basis for a possible explanation of the children's disfavour of mass morpho-syntax in incongruent condition. Since the LOC assesses the capability of dealing with abstract and logical thinking and since there is a positive correlation between the

LOC score and the performance on mass contexts, it may be the case that abstract thinking is implied in the processing of an abstract reference as the mass one. Thus, children preferred count morpho-syntax in incongruent condition since a mass interpretation refers to an abstract conceiving of the entity and the cognitive abilities underpinning this conception are not yet fully disposable at the age of five.

Noteworthy, even if children did not complete the development of the abstract thinking, they performed well with mass morpho-syntax in congruent condition. This is likely because they could base their judgements on the frequency of occurrence of the nouns in the surrounding language. Conversely, in the incongruent condition, they had no cue of any type and, as a result, they performed worse in mass context.

Remarkably, the LOC scores do not correlate with children's performance in count context. This result may point to the fact that parsing nouns in count context does not require the abstract thinking involved in parsing nouns in mass context.

Based on the current findings we argue that the differentiation between these two categories emerges also from the interaction between language and extra-linguistic abilities. Our study also indicates that when the relative frequency of use is controlled, such distinction -apparently intrinsic of language- is no longer so clear-cut, suggesting that linguistic features alone are not sufficient to define it. The distinction at the cognitive level, however, appears associated to more (and not less) processing of mass properties with respect to count properties. These results suggest that the representation of uncountability is effortful because it requires additional abilities of abstract thinking with respect to countability, for what concerns the cognitive level.

III.5 Conclusions

The experiments described in this chapter have explored the relation between countability, numerosity and Number morphology taking into account different levels of analysis. Overall, the results point to the relevance of factors that do not pertain strictly to core grammar properties in what is encoded within Number morphology.

Measuring the distribution of nouns in different syntactic contexts respect to countability demonstrated that countability is not a fixed property of the lexeme. In fact, nouns are

distributed along a *gradatum* between surfacing mostly in 'mass' or in 'count' contexts (§III.2). Countability is thus encoded contextually, as other features encoded into Number morphology.

Moreover, the frequency of occurrence mostly as mass or mostly as count is not a predictor of the response times in lexical decision, as well as a mere feature such as surfacing in the singular or in the plural form (§III.3).

The choice to refer to an entity as countable or uncountable depends from the communicative context, and it is linked to non-verbal cognitive processes. The conceiving of uncountability requires abstraction abilities: in fact, referring to uncountability is more difficult in populations in which the development of abstraction abilities is not complete yet, such as 5-year-old children (§III.4).

The contextual assignment of a Number value of singular or plural was already undoubted in theoretical linguistic literature; here the link with the processing of numerosity was taken into account and experimentally explored (§III.1).

Overall, these experimental results suggest that in order to understand Number morphology it is necessary to take into account other factors complementary to the core grammar.

IV. General conclusions

The information about the numerosity of entities is very salient from a biological point of view in fact number cognition is based on cognitive systems that are available soon after birth in humans and shared with other animal species (§II.2). Language allows to express information about numerosity of entities by exploiting many means, as number words, however this information is systematically encoded in morphological Number paradigms. Its saliency is confirmed by the broad diffusion of morphological Number systems in typology; most importantly a link between Number morphology and non-verbal systems dedicated to the cognition of numerosity seems confirmed by the parallelism of the information they encode.

The present work highlighted that the encoding of morphological number value happens on the basis of the referential context and has to be linked to non-verbal cognitive processing. It has also been proposed that also countability is encoded on the basis of the context, as any other feature that is dealt with within Number morphology.

In particular, the experimental results in section III point to the fact the encoding of countability and numerosity is not merely confined within the core grammar of language, but other factors come into play in defining them, namely non-strictly linguistic properties of words (see III.2 and III.3) and abilities related to other cognitive domains (III.4).

More generally, morphology could play a crucial role in research aimed at understanding the link between language and cognition, since it works at the interface with both of them; other morphological categories could be interestingly explored by relying on the proposed interdisciplinary methods, namely Gender, the verbal Aspect, and more.

This study suggests that language is not an independent system of lexical words and rules: it is instead a complex system continuously interfacing with information processed by other cognitive domains. Further research conducted on this topic needs to be interdisciplinary: formal linguistics, psycholinguistics and cognitive psychology need to work together in order avoid the biases derived from partial approaches reported so far.

References

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language* 89(3), 429-464.

Acquaviva, P. (2002). Il plurale in *–a* come derivazione lessicale. *Lingue e linguaggio* I, 295-326.

Acquaviva, P., & Pahagiotidis, P. (2012). Lexical decomposition meets conceptual atomism. *Lingue e linguaggio* XI, 165-180.

Acquaviva, P. (2013). *Il nome*. Roma: Carocci.

Agrillo, C., Dadda, M., & Bisazza, A. (2007). Quantity discrimination in female mosquitofish. *Animal Cognition 10*, 63–70. doi: 10.1007/s10071-006-0036-5.

Agrillo, C., Petrazzini, M.E.M., & Bisazza, A. (2014). Numerical acuity of fish is improved in the presence of moving targets, but only in the subitizing range. *Animal Cognition* 17(2), 307-316.

Aikhenvald, A. Y. (2014). Number and noun categorisation: A view from north-west Amazonia. A. Storch & G. J. Dimmendaal (eds.). *Number - Constructions and Semantics*. Studies in language companion 151. Amsterdam: John Benjamins.

Allan, K. (1980). Nouns and countability. *Language* 56, 541-567.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390-412.

Bale, A., & Barner, D. (2009). The Interpretation of Functional Heads: Using Comparatives to Explore the Mass/Count Distinction. *Journal of Semantics* 26, 217-252.

Bale, A., & Barner, D. (2011). *Mass-Count Distinction*. Oxford Bibliographies Online, http://ladlab.ucsd.edu./pdfs/BB@Oxford.pdf.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition* 29(4), 639-647.

Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97, 41-66.

Barner D., & Snedeker, J. (2006). Children's Early Understanding of Mass-Count Syntax: Individuation, Lexical Content, and the Number Asymmetry Hypothesis. *Language Learning and development* 2(3), 163-194.

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., & Mazzoleni, M.

(2004). *Introducing the La Repubblica corpus: A large, annotated, TEI (XML)-compliant corpus of newspaper Italian*. Issues, 2, 5-163. Available on http://sslmit.unibo.it/repubblica.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43 (3), 209-226.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. M. (2005). *CoLFIS (Corpus e Lessico di Frequenza dell'Italiano Scritto)*. Available on http://www. istc. cnr. it/material/database.

Bisiacchi, P., Mondini, S., Angrilli, A., Marinelli, K., & Semenza, C. (2005). Mass and count nouns show distinct EEG cortical processes during an explicit semantic task. *Brain and Language* 95, 98-99.

Borer, H. (2005). *In name only*. Oxford: OUP.

Bloom, P., (1999). The role of semantics in solving the bootstrapping problem. In R. Jackendoff, P. Bloom, & K. Wynn (eds*.). Language, logic and concepts: Essays in memory of John Macnamara*. Cambridge, MA: MIT Press.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition* 56, 1-30.

Borer, H. (2005). *In name only*. Oxford: OUP.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41 (4), 977-990.

Butterworth, B., Granà, A., Piazza, M., Girelli, L., Price, C., & Skuse, D. (1999). Language and the origins of number skills: Karyotypic differences in Turner's syndrome. *Brain and Language* 69, 486–488.

Butterworth, B., Reeve, R., Reynolds, F., & Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences 105*(35), 13179-13184.

Cantlon, J.F., & Brannon, E.M. (2007). Basic math in monkeys and college students. *PLoS Biology* 5(12): 2912-2919.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus* 133(1), 59-68.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Carreiras, M., Carr, L., Barber, H., & Hernández., A. (2010). Where syntax meets math: Right Intraparietal Sulcus activation in response to grammatical number agreement violations. *NeuroImage* 49, 1741-1749.

Carstairs, A. (1987). *Allomorphy in Inflexion*. London: Croom Helm.

Cheng, C.Y. (1973). Response to Moravcsik. J. Hintikka, J.M.E. Moravcsik, & P. Suppes (eds.). *Approaches to Natural Language*. Dordrecht: Reidel, 286-288.

Chesney, D.L., & Gelman, R. (2015). What counts? Visual and verbal cues interact to influence what is considered a countable thing. *Memory & Cognition* 43(5),798-810.

Chiarelli, V., El Yagoubi, R., Mondini, S., Bisiacchi, P., & Semenza, C. (2011). The Syntactic and Semantic Processing of Mass and Count Nouns: An ERP Study. *PLoS ONE* 6(10), e25885.

Chierchia, G. (1998). Plurality of mass nouns and the notion of 'semantic parameter'. S. Rothstein (ed.). *Events and Grammar*. Kluwer: Dordrecht, 53-103.

Chierchia, G. (2010). Mass nouns, vagueness and semantic variation. *Synthèse* 174, 99-149.

Chilosi, A.M., & Cipriani, P. (1995). *Test di Comprensione Grammaticale per Bambini*. Pisa, Italia: Edizioni Del Cerro.

Cinque, G. (2010). *The Syntax of Adjectives. A Comparative Study* . Cambridge, MA: MIT Press.

Clark, R., & Grossman, M. (2007). Number sense and quantifier interpretation. *Topoi* 26, 51–62.

Corbett, G. (2000). *Number*. Cambridge: CUP.

Cordes, S., Gelman, R., Gallistel, C.R., & Whalen, J. (2001). Variability signatures distinguish verbal from non-verbal counting for both large and small numbers. *Psychonomic Bulletin & Review* 8(4), 698-707.

Crepaldi, D., Keuleers, E., Mandera, P., & Brysbaert, M. (2013). *SUBTLEX-IT*. Retrieved from

http://crr.ugent.be/subtlex-it/.

Dahl, O. (2008). An exercise in a posteriori language sampling. *STUF*, Berlin 61 (3), 208–220.

Daniel., M. (2013). Plurality in Independent Personal Pronouns. M.S. Dryer, & M. Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/35.

De Belder, M. (2011). A morphosyntactic decomposition of countability in Germanic. *The Journal of Comparative Germanic Linguistics* 14(3), 173-202.

Dehaene, S. (1997). *The Number Sense*. Oxford University Press. New York.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol*. 20, 487–506.

de Hevia, M. D. (2011). Sensitivity to number: Reply to Gebuis and Gevers. *Cognition 121*(2), 253-255.

Dixon, R. M. W. (1979). Ergativity. *Language* 55, 59-138.

Dixon, R. W. (2012). *Basic Linguistic Theory*. Vol.3. Oxford: OUP.

Dressler, W. U. (1989). Prototypical Differences between Inflection and Derivation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 42*(1), 3-10.

Dressler, W. U. (1994). Interactions between Iconicity and Other Semiotic Parameters in Language. Simone, R. (ed.). *Iconicity in Language, CILT 110*. Amsterdam/Philadelphia: John Benjamins, 21-37.

Dryer, M.S. (2013). Coding of Nominal plurality. In: M.S. Dryer, & M. Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/33.

Dryer, M. S. & Haspelmath, M. (eds.). (2013). *The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology*. Available online at http://wals.info.

El Yagoubi, R., Mondini, S., Bisiacchi, P., Chiarelli, V., Angrilli, A., & Semenza, C. (2006). The elctrophysiological basis of mass and count nouns. *Brain and Language* 99, 187-188.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences 8*(7), 307-314.

Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods* 40 (4), 1049-1054.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*,*35*(1), 116-124.

Franzon, F., Peressotti, F., Arcara, G., & Semenza , C. (2014). Semantic interpretability speeds up the processing of morphological features. A psycholinguistic experiment on Gender Agreement. *Stem-, Spraak- en Taalpathologie* 19, Supplement 4: 189-91.

Frisson, S., & Frazier, L. (2005). Carving up word meanings: Portioning and grinding. *Journal of Memory and Language* 53, 277-291.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1), 43-74.

Gathercole, V.C. (1985). He has too much hard questions: The acquisition of the linguistic mass-count distinction in *much* and *many*. *Journal of Child Language* 12(2), 395-415.

Gelman, R., & Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science* 306(5695), 441-443.

Gelman, R., & Butterworth, B. (2005). Number and language: how are they related? *TRENDS in Cognitive Sciences* 9(1), 6-10.

Gillon, B. (1992). Towards a common semantics for English count and mass nouns. *Linguistic and Philosophy* 15, 597-639.

Gillon, B., Kehayia, E., & Taler, V. (1999). The mass/count distinction: Evidence from on-line psycholinguistic performance. *Brain and Language* 68, 205-211.

Gordon, P. (1985). Evaluating the semantic categories hypothesis: The case of the count/mass distinction. *Cognition* 20(3), 209-242.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science* 306(5695), 496-499.

Greenberg, J.H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. Greenberg, J.H. (ed.). *Universals of Language*, 73-113.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology 44*(5), 1457.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics* 42(1), 25-70.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1), 1-33.

Haspelmath, M. (2013). Occurrence of Nominal plurality. M. S. Dryer, & M. Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info/chapter/34.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi–free–ranging rhesus monkeys. *Proceedings of the Royal Society of London B: Biological Sciences* 267(1445), 829-833.

Hirai, S., & Jitsumori, M. (2009). Counting absolute numbers of items, from 1 to 8, in pigeons. *Learning & Behavior* 37, 365-379.

Hyde, D.C., Spelke, E.S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Dev Sci.* 14(2), 360–371.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition* 62(2), 169-200.

Irie-Sugimoto, N., Kobayashi, T., Sato, T., & Hasegawa, T. (2009). Relative quantity judgment by Asian elephants (Elephas maximus). *Animal Cognition* 12(1), 193-9.

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences 106*(25), 10382-10385.

Jackendoff, R., (1991). Parts and Boundaries. *Cognition* 41, 9-45.

Jarema, G. & G. Libben (2006). Disorders of morphology. In H. A. Whitaker (ed.), *Concise encyclopedia of brain and language*, 358-366. Amsterdam: Elsevier.

Katz, G., & Zamparelli, R. (2012). Quantifying Count/Mass Elasticity. J. Choi et al. (eds). *Proceedings of the 29th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 371-379.

Krifka, M. (1995). Common nouns: A contrastive analysis of English and Chinese. G. N. Carlson & F. J. Pelletier (Eds.), *The generic book*: 398–411. Chicago: University Press

Kulkarni, R., Rothstein S., & Treves, A., (2013) A Statistical Investigation into the Cross-Linguistic Distribution of Mass and Count Nouns: Morphosyntactic and Semantic Perspectives. *Biolinguistics* 7, 132-168.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance* 39(3), 802.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R.H.B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-6. http://CRAN.R-project.org/package=lmerTest

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development* 3(3), 299-321.

Lewis, M. P., & Simons, G.F. (2010). Assessing endangerment: expanding Fishman's GIDS. *Revue Roumaine de Linguistique* 55 (2), 103-120.

Lewis, M.P., Simons, G.F., & Fennig, C.D. (eds.). (2015). *Ethnologue: Languages of the World, Eighteenth edition*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense large-number discrimination in human infants. *Psychological Science 14*(5), 396-401.

Longobardi, G. (1988). I quantificatori. L., Renzi, G., Salvi and A., Cardinaletti (eds*.) Grande grammatica italiana di consultazione*. Bologna, il Mulino: 645-696.

Longobardi, G. (1991). Proper names and the theory of N-movement in syntax and Logical Form. *University of Venice Working Papers in Linguistics 9*.

Loporcaro, M. (2011). Syncretism and Neutralization in the Marking of Romance Object Agreement. Maiden, M., & Goldbach, M. (eds.). *Morphological autonomy: perspectives from romance inflectional morphology*. Oxford University Press.

Malouf, R., Ackerman, F., & Seyfarth, S. (2015). *Explaining the Number Hierarchy*. Ms.

Marcantonio, A., & Pretto, A.M. (2001). Il nome. In Renzi, L., Salvi, G., & Cardinaletti, A. (eds.). *Grande grammatica italiana di consultazione*. Bologna: Il Mulino. 329-346.

Markman, E. M. (1990). Constraints Children Place on Word Meanings. *Cognitive Science* 14, 57–77.

Massam, D. (ed.). (2012). *Count and Mass Across Languages*. Oxford Studies in Theoretical Linguistics 42. Oxford: OUP.

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological science 18*(8), 740-745.

Mondini, S., Angrilli, A., Bisiacchi, P., Spironelli, C., Marinelli, K., & Semenza, C. (2008). Mass and Count nouns activate different brain regions: An ERP study on early components. *Neuroscience Letters* 430, 48-53.

Mondini, S., Kehaya, E., Gillon, B., Arcara, G., & Jarema, G. (2009). Lexical access of mass and count nouns. How word recognition reaction times correlate with lexical and morpho-syntactic processing. *The Mental Lexicon* 4, 354-379.

Moro, A. (2006). *I confini di Babele*. Longanesi.

Müller, G. (2007). Notes on paradigm economy. *Morphology* 17(1), 1-38.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, 291(5501), 114-118.

Pelletier, F. J. (1975). Non-singular reference: some preliminaries. *Philosophia* 5, 451-465.

Pelletier, F. J. (2012). Lexical Nouns are Neither Mass nor Count, but they are Both Mass and Count. D. Massam (ed.). *A Cross-Linguistic Exploration of the Count-Mass Distinction*. Oxford: OUP, 9-26.

Pepperberg, I. M. (2006). Grey parrot (Psittacus erithacus) numerical abilities: Addition and further experiments on a zero-like concept. *Journal of Comparative Psychology* 120(1), 1-11.

Piazza, M., Mechelli, A., Butterworth, B., & Price, C.J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage* 15, 435–446.

Piazza, M., Mechelli, A., Price, C., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: an fMRI study. *Brain Res*. 1106, 177–188.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53, 293–305.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306(5695), 499-503.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* 41, 983–993.

Pirrelli, V., & Battista, M. (2000). The paradigmatic dimension of stem allomorphy in Italian verb inflection. *Italian Journal of Linguistics* 12, 307-379.

Prasada, S., Ferenz, K., & Haskell, T. (2002). Conceiving of entities as objects and as stuff. *Cognition* 83(2), 141-165.

Quine W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rath, D., Domahs, F., Dressel, K., Claros-Salinas, D., Klein, E., Willmes, K., & Krinzinger, H. (2015). Patterns of linguistic and numerical performance in aphasia. *Behavioral and Brain Functions 11*(1), 2.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rijkhoff, J., Bakker, D., Hengeveld, K., & Kahrel, P. (1993). A Method of Language Sampling. *Studies in Language* 17(1), 169-203.

Rijkhoff, J., & Bakker, D. (1998). Language sampling. *Linguistic Typology* 2(3), 263-314.

Rothstein, S. (2010). Counting and the Mass/Count Distinction. Journal of Semantics 27(3), 343-397.

Rugani, R., Vallortigara, G., & Regolin, L. (2013). Numerical Abstraction in Young Domestic Chicks (Gallus gallus). *Plos One* 8(6), e65262. doi:10.1371/journal.pone.0065262.

Rugani R., Cavazzana A., Vallortigara G., & Regolin L. (2013). One, two, three, four, or is there something more? Numerical discrimination in day-old domestic chicks. *Animal Cognition* 16: 557–564.

Rugani, R., Vallortigara, G., & Regolin, L. (2014). From small to large. Numerical discrimination by young domestic chicks. *Journal of Comparative Psychology* 128(2),163-71.

Rugani, R., Vallortigara, G., & Regolin, L. (2015). At the root of the left–right asymmetries in spatial–numerical processing: from domestic chicks to human subjects. *Journal of Cognitive Psychology* 27 (4), 388-399.

Schiehlen, M., & Spranger, K. (2006). The Mass - Count Distinction: Acquisition and Disambiguation. *Proceedings of the 5th LREC*. Genova.

Semenza, C., Mondini, S., & Cappelletti, M. (1997). The grammatical properties of mass nouns: an aphasia case study. *Neuropsychologia* 35, 669-675.

Semenza, C. (2008a). Number processing. B. Stemmer, & H. Withaker (eds.). *Handbook of the Neuroscience of Language*. Oxford: Elsevier, 219-227.

Semenza, C., El Yagoubi, R., Mondini, S., Chiarelli, V., & Venneri, A. (2008b). Processing of mass and count nouns in the brain. Bambini, V., Ricci, I., & Bertinetto, P.M. (eds.). *Language and the brain – Semantics. Atti della XLII Congresso Internazionale di Studi della Società di Linguistica Italiana*, Vol. 2 I.C.3. Pisa: Bulzoni.

Semenza, C., & Mondini, S. (2015). Word-formation in aphasia. *Word-Formation: An International Handbook of the Languages of Europe. Berlin: DeGruyter*.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition* 38, 179-211.

Spelke, E.S. (2000). Core knowledge. *American Psychologist* 55: 1233-1243.

Stancher, G., Rugani, R., Regolin, L., & Vallortigara, G. (2015). Numerical discrimination by frogs (Bombina orientalis). *Animal cognition*, *18*(1), 219-229.

Stancher, G., Sovrano, V. A., Potrich, D., & Vallortigara, G. (2013). Discrimination of small quantities by fish (redtail splitfin, Xenotoca eiseni). *Animal Cognition* 16, 307–312.

Starr, A., Libertus, M.E., & Brannon, E.M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *PNAS* 110(45), 18116-20.

Steinhauer, K., Pancheva, R., Newman, A. J., Gennari, S., & Ullman, M. T. (2001). How the mass counts: An electrophysiological approach to the processing of lexical features. *Cognitive Neuroscience and Neuropsychology* 12(5), 999-1005.

Storch, A., & Dimmendaal, G. J. (2014). *Number - Constructions and Semantics*. Studies in language companion 151. Amsterdam: John Benjamins.

Stump, G. T. (1991). A paradigm-based theory of morphosemantic mismatches. *Language*, 675-725.

Stump, G. T. (2006). Heteroclisis and paradigm linkage. *Language* 82, 279-322.

Stump, G.T. (2010). Interactions between defectiveness and syncretism. Baerman, M., Corbett, G. G., & Brown, D. P. (eds.). *Defective Paradigms: Missing forms and what they tell us* (Vol. 163). Oxford University Press, 181–210.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge MA: Harvard University Press.

Treis, Y. (2014). Number in Kambaata. A. Storch & G. J. Dimmendaal (eds.). *Number - Constructions and Semantics*. Studies in language companion 151. Amsterdam: John Benjamins.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review* 101, 80 –102.

Troiani, V., Peelle, J. E., Clark, R., & Grossman, M. (2009). Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers. *Neuropsychologia* 47(1), 104-111.

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: Rudiments of number in an amphibian. *Animal Cognition* 6, 105–112.

Vallortigara, G., Chiandetti, C., Sovrano, V.A., Rugani, R., & Regolin, L. (2010). Animal Cognition. *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 882–893.

Vendler, Z. (1957). Verbs and times. *Philosophical Review* 56, 143-160.

Verkuyl, H. (1972*). On the compositional nature of the aspects*. Dordrecht: Reidel.

Vianello, R., & Marin, M. L. (1997). *OLC. Dal pensiero intuitivo al pensiero operatorio concreto: prove per la valutazione del livello di sviluppo*. Bergamo: edizioni junior.

Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology* 16(1/2), 312–339.

Zamparelli, R. (2007). On singular existential quantifiers in Italian. I., Comorovski and K., von Heusinger (eds.) *Existence: Syntax and Semantics*. Springer: 293-328.

Zanini C., Arcara G., & Franzon F. (2014) Measuring the distribution of mass and count nouns. A comparison between a rating study and a corpus based analysis. *Palc 14, Łódź*, Poland. DOI: 10.13140/RG.2.1.3986.5448

Appendixes