The role of the context in countability: a psycholinguistic approach.

Chiara Zanini[1], Giorgio Arcara[2], & Francesca Franzon[3]
[1]University of Zürich, Switzerland; [2]San Camillo Hospital, Venice, Italy; [3]SISSA, Trieste, Italy

1 Introduction

1.1 Theoretical accounts

Traditionally, grammar descriptions trace a clear-cut division between mass nouns (*butter*), that would refer to substances, and count nouns (*mug*), that would refer to objects. Formal semantics accounts (e.g. Chierchia 2010; Jackendoff 1991) state that every noun is lexically marked as either countable or uncountable. Since the use of certain mass nouns in a countable syntactic context is not uncommon, lexicalist accounts assume that a semantic shift operation of *portioning* has taken place. The reverse operation of *grinding* (e.g.count nouns used in a mass context) is also admitted. Syntax-based approaches (e.g. Borer 2005) claim instead 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). These assume that the mass-count 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, the mass reading would be the default. A third way is represented by those theoretical accounts agreeing with the idea of unspecified lexical roots and yet pointing to the role of the linguistic context of occurrence in denoting the countability of any nominal expression (Allen 1980; Pelletier 2012; Rothstein 2010).

1.2 Experimental studies

Despite a long debate in linguistics as well as in philosophy on the the properties that should distinguish mass from count expressions (e.g. Quine 1960), few empirical studies have actually measured the occurrences of nouns in mass and in count contexts within the use of language. Yet, quantitative data collected on language corpora and by means of rating studies point unanimously to the fact that nouns do not massively gather towards two poles with respect to their morpho-syntactic distribution, namely "pure mass-used"and "pure count-used"nouns (Katz & Zamparelli 2012; Kulkarni et al. 2013; Schiehlen & Spranger 2006; Vermote et al. 2017).

The uneven distribution of frequency of use of nouns has possibly driven to some inconsistency in the results reported in psycholinguistic studies, which in fact do not show a clear picture about the representation and processing of mass and count expressions as two different categories. Some lexical decision studies reported longer RTs for mass nouns (Gillon et al. 1999; Mondini et al. 2009), other did not report significant differences (e.g. Mondini et al. 2008). ERP studies found differences in the elicited ERP components (El Yagoubi et al. 2006), notwithstanding a largely overlapping anatomical representation (Semenza 2005). Among these, Chiarelli et al. (2011) observed an overall bias for the count syntactic context, reporting faster RTs for sentences with count nouns independently from the concreteness factor. A similar effect was also described in an eye-tracking study by Frisson & Frazier (2005): when nouns were presented in a context that was unmarked for countability, a longer fixation time was reported for mass nouns with respect to count nouns. Language acquisition studies depict a similar scenario by reporting a preference for count interpretation and count syntax (Barner & Snedeker 2005; Gathercole 1985; Zanini et al. 2016).

1.3 Aims of this study

Albeit some inconsistency in the experimental literature, to our knowledge no study has ever reported longer response times or inferred a more demanding processing cost for count with respect to mass reference. The aim of the present study is threefold. First, we aim to disentangle the role of non-core linguistic properties, such as frequency of occurrence, from the role of core linguistic aspects related to grammar. To do so, in the present study we relied on the actual occurrence of nouns in the different syntactic contexts in assigning them to the mass or to the count experimental list. The occurrence of nouns was measured by means of a corpus study described in paragraph 2.1. Furthermore, we aim to distinguish the effects of lexical features from the effects of syntactic parsing in the processing of mass and count expressions. If differences are observed in the lexical decision task (paragraph 2.2), they can be ascribed to the presence of a lexical feature. If differences

are observed in the self-paced reading task (paragraph 2.3), they can capture the effects of the integration in a syntactic context. Lastly, we aim to verify whether mass or count expressions require higher processing costs with respect to each other and, if any, how the observed trend can fit the theoretical linguistic descriptions.

2 The study

2.1 Corpus investigation

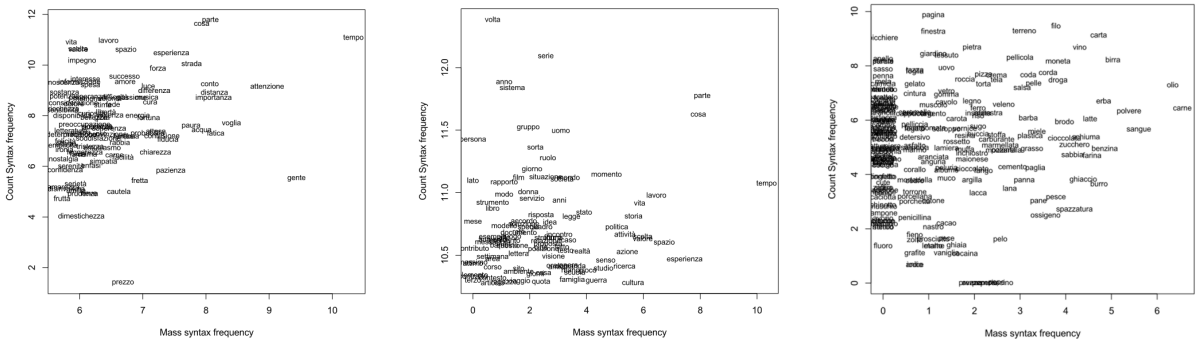
In order to disambiguate the mass use from the count use of the nouns, we performed a corpus analysis. Katz & Zamparelli's (2012) study on English was taken as an example for building appropriate queries for Italian. We designed queries in CQP syntax. One of the major problems in designing the queries 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-SG pizza-SG 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:

Count (singular)]: Indeterminate article (+Adjective(s)) + Noun-SG
Mass Verb (+ Adverb(s)) (+Adjective(s)) + bare Noun-SG
Quantification expression *(molto, poco, troppo, un po' di, . . .)* + Noun-SG

We queried the itWaC corpus (Baroni et al. 2009) in order to check the existence of a correlation between occurrences of a same noun in mass syntax and in count syntax. We collected the 100 nouns that occur most frequently in a mass context and the 100 nouns that occur most frequently in a count context and we measured their distribution with respect to both syntactic contexts. Then, we analyzed data for 224 concrete nouns to be used in the following experiments. We selected these nouns following the mass and count theoretical definitions given in traditional grammars. The list included only nouns with transparent inflection for morphological Number and did not involved compound nouns, derived nouns, loan words. The calculated correlations are summarized in the table below and plotted in Figure 1.

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
Experimental stimuli	r(205)= 0.13	p = 0.07

Table 1: Correlations between frequency of occurrence in mass syntax and in count syntax of the 100 top mass-used nouns, 100 top count-used nouns and experimental nouns in itWaC.



(a) 100 mostly mass-used nouns (b) 100 mostly count-used nouns (c) Experimental stimuli

Figure 1: Distribution of nouns in the itWaC corpus

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, the fact that a noun is frequently used as count does not imply that it is scantly used in mass contexts. Even more surprisingly, the positive correlation found for the top 100 nouns most frequently used as mass suggests 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 in Italian.

2.2 Lexical decision task

2.2.1 Methods

Stimuli. From the 224 stimuli mentioned above, we chose 128 stimuli (i.e. 64 nouns inflected both in the plural and in the singular). To allow a comparison with previous studies (Gillon et al. 1999; Mondini et al. 2009), 64 stimuli were categorized as mass-used or count-used following their occurrence in mass or in count contexts in the corpus. Mass-used nouns were the 16 nouns with the highest frequencies in mass context that were not among the top 16 count-used nouns; count-used nouns were the 16 nouns with the highest frequencies in count context that were not among the top 16 mass-used nouns. 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 32 mass-used nouns and 32 count-used nouns used in the experiment. The mass-used and count-used forms in the singular were matched for subjective frequency (as reported in Franzon et al. 2016), absolute corpus frequency (from itWac, Baroni et al. 2009), and orthographic length (independent t-tests: all ps $>.05$). In the plural, mass-used stimuli had significant lower absolute frequency (independent t-tests: all ps $<.05$), but comparable length (independent t-test: p $<.05$) with respect to count-used nouns. The remaining 64 experimental stimuli (32 nouns, both in the singular and in the plural) were not categorized in such terms since they can appear both in mass and count contexts. The final list included also 80 filler adjectives and 160 phonotactically plausible nonwords.

Procedure. Sixty Italian native speakers (female= 36; mean age= 23.5, SD= 2.37; mean education= 15.16, SD= 1.64) participated in the experiment. Participants saw a series of letter strings presented at the center of the screen one at a time. They had to press a key when they thought that the displayed string was a real Italian word, another key in the converse case. They were asked to answer as quickly and accurately as possible. A fixation point was shown for 500 ms on a dark gray background, followed by the target that remained on the screen until a response or until 2000 ms expired. Stimuli were delivered by means of DMDX software (Forster & Forster 2003).

Data analysis. Data were analyzed by means of the R software for statistical analysis (R core team 2014) We used generalized linear mixed models (Baayen et al. 2008) to investigate the influence of the categorical *Type* of the stimuli (mass-used vs. count-used nouns) as well as other variables such as subjective *Frequency*, sequential number of the trial (*Ntrial*) and response time to the previous trial (*Ntrial-1*). Among the fixed effects, we included also morphological *Number* (singular vs. plural) since mass-used nouns occur less frequently in the plural. Subjects and items were added as crossed random effects.

Data were analyzed via three different models targeting specific theoretical issues. The first model (Model 1) aimed to replicate previous findings reported in the literature. For this purpose, the model included three fixed effects: *Type* of noun, *Number* and the interaction *Type x Number*. In a second analysis, we verified whether eventual categorical differences obtained in Model 1 (and in the previous literature as well) could be better predicted by continuous variables such as frequency. We fitted a second model (Model 2) in which, besides the fixed effects in Model 1, subjective *Frequency*, response time to the previous trial (*Ntrial-1*) and sequential number of the trial (*Ntrial*) were added as predictors. The third model (Model 3) intended to test whether the effects reported in Model 2 could be extended to other nouns or could be due to an artifact of the selected stimuli since these latter are peculiarly distributed when compared to the great majority of the nouns (as highlighted in the corpus study, paragraph 2.1). In fact, the great majority of the nouns is not distributed in a dichotomous way with respect to the context of occurrence. Model 3 included the same predictors of Model 2, but it encompassed all 128 experimental stimuli: 64 mass-used and count-used nouns plus 64 non-categorized nouns.

2.2.2 Results and interim discussion

Responses above 2000 ms and below 200 ms were discarded; only correct responses were considered for response time (RT) analysis. Model 1: The analysis revealed a main effect of *Type*, with a disadvantage for mass nouns. The interaction *Type x Number* also reached significance: mass-used nouns in the plural elicited longer RTs than in the singular.

	Estimate	Std. Error	DF	t value	Pr ($>\|t\|$)
(Intercept)	6.499278	0.037217	94.650000	174.632	<2e-16 ***
Type mass-used	0.146242	0.045525	114.890000	3.212	0.00171 **
Type mass-used x Number pl	-0.165683	0.058104	128.750000	-2.851	0.00507 **

Table 2: Summary of Model 1.

Model 2: No main effect of *Type, Number and Ntrial-1* as well as interaction *Type x Number* were found. *Frequency* and *Ntrial* are significant predictors of the RTs.

	Estimate	Std. Error	DF	t value	Pr (>|t|)
(Intercept)	6.702e+00	3.508e-02	1.370e+02	191.075	<2e-16 ***
Frequency	-5.003e-02	7.177e-03	8.900e+01	-6.971	5.37e-10 ***
Ntrial	-3.060e-03	6.218e-04	1.087e+05	-4.921	8.60e-07 ***

Table 3: Summary of Model 2.

Model 3: The analysis revealed a main effect of *Frequency* and *Ntrial-1* whereas no significant effect of *Number* and *Ntrial* were found.

	Estimate	Std. Error	DF	t value	Pr (>|t|)
(Intercept)	6.775e+00	3.054e-02	1.440e+02	221.815	<2e-16 ***
Frequency	-4.431e-02	8.497e-03	1.540e+02	-5.215	5.84e-07 ***
Ntrial-1	-3.089e-03	6.163e-04	1.102e+05	-5.012	5.39e-07 ***

Table 4: Summary of Model 3.

In Model 1 we found a main effect of the category (*Type*), namely slower RTs for mass-used nouns. However, this effect can be driven by their lower frequency in the plural, as resulting from the interaction *Number x Type*. In fact, when *Frequency* is added in the model as a fixed effect (Model 2), it turns out to predict the RTs while the category effects are no more significant. Notably, these findings hold true even in the case the analysis is extended to other noun stimuli encompassing the range of the mass-count distribution, and not only its subpart concerning the most polarized distribution (Model 3). In sum, even if the stimuli were categorized according to their frequency of occurrence in mass and in count contexts, the possibility for a noun to occur in these different contexts does not predict the RTs of their recognition as a single word in our lexical decision task. Instead, frequency considered as a continuous variable is a better predictor of RTs than a distinction attributed to alleged lexical categories.

2.3 Self-paced reading task

2.3.1 Methods

Stimuli The same categorized nouns used in the previous lexical decision study were chosen: 16 mass-used nouns, 16 count-used nouns and 32 neutral nouns. The experimental conditions resulted from the manipulation of two variables: the *Type* of the noun (mass-used, count-used and neutral) and the syntactic *Context* (mass and count). For each noun, two identical sentences were created: in one the noun appeared in a mass context (i.e. with the partitive article *del/della* lit. 'of the' meaning 'some'), in the other one the same noun appeared in a count context (i.e. with the indeterminate article *un/una* 'a'). Each noun was presented only in the singular. As a result, 32 sentences were not well-formed: in 16 stimuli a mass-used noun occurred in count context, and in 16 stimuli a count-used noun occurred in mass context (incongruent condition). The remaining sentences were well-formed (congruent condition): in 16 sentence stimuli a mass noun occurred in mass context, in other 16 sentence stimuli a count noun occurred in count context. The 32 neutral nouns appeared in 64 sentences, all of them well-formed.

All sentence stimuli (n= 128) were created following the same structure: a subject (determiner + noun), a transitive verb form in the present indicative taking a direct object, that is article + mass/count/neutral noun, and a prepositional phrase (preposition + noun) in the end. Thus, each sentence was made up of seven words. 128 filler sentences were added to prevent participants' focusing on countability contrasts when reading the experimental stimuli. Half of them shared the same structure of the experimental sentences, the other 64 displayed a different one; however, all 128 filler sentences were always made up of seven words. Fillers' grammaticality was balanced to obtain the same number of the respectively correct and incorrect experimental sentences.

Procedure Forty-five Italian native speakers (female= 32; mean age= 22.06, SD= 2.38; mean education= 14.17, SD= 1.76) participated in the experiment. They had to read some sentences, presented at the centre of the screen one word at a time. Participants were instructed to press a key to read the following word: the previous word disappeared and the following one was displayed. After having read the last word, they had to judge the grammaticality of the whole sentence on a Likert scale ranging from 1 to 4 (1= absolutely not acceptable; 4= completely acceptable). The sentence stimuli, delivered by means of DMDX software (Forster & Forster 2003), were presented to each participant in a different random order. The list of 256 sentences was split in four blocks; after each block, participants were encouraged to take a break.

Data analysis Data were analyzed by means of generalized mixed models following the same general guidelines illustrated for the lexical decision study. We fitted one model (Model 4) including the *Type* of noun, the *Context*, the interaction *Type x Context* as well as subjective *Frequency* and RT of the word preceding the target noun (*Target-1*) as fixed effects. Subjects and items were added as crossed random effects.

2.3.2 Results and interim discussion

Grammaticality judgments in the self-paced reading test were reported in Table 5. We considered a mean score of equal or greater than 3 as the threshold for sentence acceptability. Mass-used and count-used nouns in incongruent condition were less accepted than mass-used nouns (correlated t-test: t= 42.09, df= 1286, p<2.2e-16) and count-used nouns (correlated t-test: t= 56.47, df= 1200.7, p<2.2e-16) in congruent condition. Although neutral nouns were judged grammatical in both contexts, they were rated significantly higher in the count context (correlated t-test: t = 1.95, df = 2877.6, p= .05).

	Mass context	Count context
Mass-used noun	*congruent* mean= 3.675, sd= 0.927	*incongruent* mean= 1.673, sd= 1.045
Count-used noun	*incongruent* mean=1.512, sd= 0.927	*congruent* mean= 3.803, sd=0.574
Neutral noun	*congruent* mean=3.425, sd= 0.916	*congruent* mean= 3.492, sd= 0.927

Table 5: Acceptability judgments.

Reading times (RTs) were measured at the chunk displaying the critical noun and at the word immediately preceding it. The analysis revealed a main effect of *Frequency* and of *Target-1*. No main effect of the *Type* of noun and of the *Context* was found, nevertheless the interaction *Type x Context* is significant: count-used nouns in count syntax elicited longer RTs than in mass syntax (t= 2.148, p<.05); similarly, neutral nouns were read faster in count syntax than in mass syntax (t= 2.549, p<.01), whereas no significant difference was found for mass-used nouns.

	Estimate	Std. Error	DF	t value	Pr (>|t|)
(Intercept)	3.71053	0.08881	1200.00	41.779	<2e-16***
Frequency	-0.01242	0.00373	60.00	-3.328	0.0015**
Target-1	0.39789	0.01299	5592.00	30.629	<2e-16***
Type count x Context count	-0.03431	0.01642	5543.00	-2.090	0.0367*
Type neutral x Context count	-0.03209	0.01420	5543.00	-2.260	0.0239*

Table 6: Summary of Model 4.

Grammaticality judgments in the self-paced reading test were consistent with the categorization of the nouns collected on the itWaC corpus. In fact, participants assigned lower scores in the incongruent condition. Yet, a preference for the count syntax emerges: neutral nouns were judged grammatical in both mass and count contexts, nevertheless they were significantly more accepted in sentences involving count syntax. This preference is observed also in RTs across conditions. In fact, whereas count nouns in mass context (incongruent condition) were read significantly slower than in count context (congruent condition), mass nouns were not read faster in mass context (congruent condition) than in count context (incongruent condition). Context alone and category alone do not explain this pattern. If that was the case, a similar performance would have been expected: on neutral nouns in the two contexts and on both mass and count nouns in the respectively incongruent contexts.

3 General discussion

Data from Italian seem to suggest the absence of a clear-cut lexical division between mass and count nouns. As already reported in studies on other languages, the distribution of nouns in corpora shows that most nouns appear frequently both in mass and in count contexts, and do not gather towards two poles of nouns used only in one of the two contexts. From a theoretical perspective, it is thus more likely that the information about countability is contextually encoded, rather than fixed in the lexicon. The lack of a lexical specification for countability seems confirmed by the two behavioral experiments. When nouns are presented in isolation, recognition times are not predicted by mass and count categories, although these latter had been defined on a quantitative basis. When nouns are presented in syntactic contexts marked for countability, differences emerge instead, confirming a general disadvantage in processing uncountability as observed in the previous literature (see paragraph 1.2).

The reported longer RTs with respect to mass syntax are not per se inconsistent with the view of a formally simpler mass syntax. We will discuss the relation of formal analyses with these behavioural data taking into account also possible extra-linguistic factors that may come into play during language processing (as suggested in Zanini et al. 2016).

References

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.

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

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.

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).

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

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.

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

Franzon, F., Arcara, G., & Zanini, C. (2016). Lexical categories or frequency effects? A feedback from quantitative methods applied to psycholinguistic models in two studies on Italian. *Proceedings of the Third Italian Conference on Computational Linguistics* CLiC-it 2016. Torino: Aaccademia University Press, 152-156.

Frisson, S., & Frazier, L. (2005). Carving up word meaning: Portioning and grinding. *Journal of Memory and Language*, 53(2), 277-291.

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.

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

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

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

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.

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.

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.

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

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

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

Schiehlen, M., & Spranger, K. (2006). The Mass–Count Distinction: Acquisition and Disambiguation. *Proceedings of the 5th International Conference on Language Resources and Evaluation* (Vol. 277).

Semenza, C. (2005). The (neuro)psychology of mass and count nouns. *Brain and Language 95*(1), 8889.

Vermote, T., Lauwers, P., & De Cuypere, L. (2017). Transcending the lexical vs. grammatical divide regarding the mass/count distinction. Evidence from corpus studies and acceptability surveys in French and Dutch. *Language Sciences, 62*, 37-51.

Zanini, C., Benavides, S., Lorusso, D., &. Franzon, F. (2016). Mass is more. The conceiving of (un)countability and its encoding into language in five year-old-children. *Psychonomic Bulletin & Review 24*(4) 1330-1340. Doi: 10.3758/s13423-016-1187-2