

MASARYK UNIVERSITY
FACULTY OF ARTS



Semantics of Classifiers

An Experimental Study

BACHELOR'S THESIS

Mariia Onoeva

Brno, Fall 2018

MASARYK UNIVERSITY
FACULTY OF ARTS
DEPARTMENT OF LINGUISTICS AND BALTIC LANGUAGES

Field of study: General Linguistics



Semantics of Classifiers An Experimental Study

BACHELOR'S THESIS

Mariia Onoeva

Advisor: doc. PhDr. Mojmír Dočekal, Ph.D

Brno, Fall 2018

Declaration

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Mariia Onoeva

Acknowledgements

I would like to express my deep gratitude to doc. PhDr. Mojmír Dočekal, Ph.D, for his patient guidance, enthusiastic encouragement and useful critiques of this thesis. And I would like to offer my special thanks to Bc. Jitka Koudelková and Mgr. Iveta Šafratová for helping me out with the Czech version of the experiment. Of course, I am very grateful to all of those who participated in both versions of the experiment. I am also grateful to all who were around me during my studying. Finally, I must express my very profound gratitude to my parents, my little sister and my boyfriend for providing me with unfailing love and support.

Contents

Introduction	1
1 Mass and count nouns	2
1.1 <i>Properties of mass nouns</i>	3
1.1.1 The signature property	4
1.1.2 The mapping property	5
1.1.3 The elasticity property	6
1.2 <i>Three ways of encoding mass/count</i>	8
1.2.1 Classifier languages	9
1.2.2 Number neutral languages	10
1.2.3 Number marking languages	11
1.2.4 Fake mass nouns	13
2 Formal theories of mass/count distinction	16
2.1 <i>The lattice theory</i>	16
2.2 <i>Mass nouns and vagueness</i>	21
2.3 <i>The object unit</i>	22
3 Classifiers <i>štuka</i>, <i>čelovek</i> and <i>golova</i>	27
3.1 <i>Semantics of counting and measuring</i>	27
3.2 <i>Types of classifiers</i>	28
3.3 <i>Štuka, čelovek and golova are functional expressions</i>	32
3.4 <i>Štuka, čelovek and golova do not take kind-denoting terms as complements</i>	33
3.5 <i>Štuka, čelovek and golova are measure unit classifiers</i>	34
4 The experiment	38
4.1 <i>Is štuka a translation of kus?</i>	38
4.2 <i>Respondents and distribution</i>	41
4.3 <i>Method</i>	41
4.4 <i>Czech questionnaire</i>	44
4.4.1 Introduction	44
4.4.2 Practice section	46
4.4.3 The experimental part	47
4.5 <i>Russian questionnaire</i>	50
4.5.1 Introduction	50

4.5.2	Practice section	51
4.5.3	The experimental part	53
4.6	<i>Results</i>	54
4.6.1	Czech	54
4.6.2	Russian	60
4.6.3	Discussion	63
5	Summary	66
	Appendices	71
A	Czech items	72
B	Czech fillers	81
C	Russian items	85
D	Russian fillers	94

Introduction

There is a general perception that classifiers are present only in a special group of languages. There are according to Allan (1977) fifty one *classifier languages* which use classifiers as a special closed set of words to denote some salient meanings. In these languages they occur as morphemes in surface structures under certain conditions. Classifiers characterize perceived features such as *shape, material, size, location, quanta, etc.* Nouns by themselves cannot refer to entities in this type of languages, together with the associated nouns classifiers can fully refer to entities (Allan, 1977). However, it is not a whole picture and *non-classifier* languages use them.

The purpose of the present work is to find out how different languages deal with classifiers. The thesis consists of two main parts: first three chapters are theoretical, the forth chapter is a practical application of the given theory on Czech and Russian languages.

It is well-known that languages like English, Czech, or Russian usually use classifiers to measure or count things. Even if it is not necessary to insert a classifier in every occasion of counting, how it happens in Mandarin Chinese or Japanese, some terms in non-classifier languages cannot be felicitously combined with numerals without classifiers. These are *mass nouns*. They are opposed to *count nouns* which are fine with or without classifiers. The first chapter will introduce mass/count distinction and show how some languages represent it.

In the second chapter there will be described three theories which formally explain why mass nouns differ from count; the third chapter is the application one of them onto Russian. The forth chapter will discuss the experiment; during the experiment were compared two at first sight identical classifiers from Czech and Russian, and then will follow the analysis of the results from the experiment. The fifth chapter will conclude the thesis.

1 Mass and count nouns

The grammatical singular/plural contrast in languages generally corresponds to the number in real word but such a contrast is not always relevant. Fluids, minerals, pastes, assorted materials are associated with **mass nouns**, i.e. material aggregates that may be joined and split without changing their nature (Chierchia, 2010). Substances like *water*, *gold*, *dough* do not represent discrete, bounded objects which can be counted. They cannot be multiplied on the model “*one cat plus one cat equals two cats*”, hence they cannot be ‘singular’ or ‘plural’ in real word, but grammatically mass nouns are only singular¹.

However, this definition involves ‘canonical’ mass nouns. The nouns *furniture* and *footwear* are mass in opposite to *table* and *shoe*, abstract nouns like *beauty*, *knowledge*, *advise*, etc. are mass as well, so are *jumping* or *running* which denote nominals (Chierchia, 2010). For the purposes of the present work a circle of examined nouns is restricted by concrete nouns. The difficulties with ‘superordinate’ nouns like *furniture* and *footwear* will be discussed further.

Mass nouns are distinguished from **count nouns**. Prototypically they denote objects with boundaries in real word, grammatically can be either singular or plural. In case of count nouns, a singular form refers to a single element, a plural form usually represents quantity of more than one element (or more than two or three in presence of dual or trial). There are some obvious parallels between mass nouns and plural count nouns, e.g. cumulative reference² or an ability to appear without any determiner (at least in English), but on the other hand there are differences because of which semanticists cannot treat them as something equal (Lasnik, 2011).

It is a good question, whether mass/count distinction is a linguistic or extralinguistic issue, whether it is a grammar relation, or it is incorporated as a given into reality and languages must adjust. At

1. Unless they are shifted to count reading (Doetjes, 2012), for more about it further.

2. If there are two entities, e.g. x and y , to which the predicate in case of mass nouns *water* applies, this predicate applies to their collection as well, so $x \oplus y$ is *water*. Consider the same for count, if the predicate *books* applies to x and to y , the predicate *books* applies to $x \oplus y$ as well (Krifka, 1989).

$\forall P[CUM_S] \leftrightarrow \forall x \forall y [P(x) \wedge P(y) \rightarrow P(x \cup_S y)]$
(P has cumulative reference)

some points, the issue was tested by several researchers³ and it is clear that substance/object contrast is language independent but matches mass/count distinction in language (Chierchia, 2010). Moreover, such a distinction is not a feature of one particular language, many of them (if not all) show it but differently (Chierchia, 2010; Lasersohn, 2011; Doetjes, 2012).

Count nouns had been named *thing-words* (Jespersen, 1913), *bounded nouns* (Bloomfield, 1933), *individual nouns* (Whorf, 1941) before they became 'count' (No Author, 1952). The first notion of mass nouns in its technical sense was made by Otto Jespersen but he detected such 'immaterial' words like *progress*, *admiration*, or *safety* as mass-words, while Bloomfield separated them into abstract nouns (Lasersohn, 2011). Linguists have been studying mass and count terms for a while and the next section will tell about general features of them.

There will be described three properties of mass nouns which seem to be common through languages. But if mass nouns behave similar, languages do not encode them identically. There are three ways of encoding such nouns, they will be analyzed below as well. In the end of the chapter will be described an interesting phenomenon of non-canonical superordinate mass nouns like *furniture* and *cattery*.

1.1 Properties of mass nouns

Some tests, e.g. pluralization or quantification, can help to distinct between mass and counts nouns. But such tests are indicative only in a special group of languages, though some properties of mass nouns seem to be universal in many (if not all) languages. They are **the signature property**, **the mapping property**, and **the elasticity property** which are described below.

These universal properties of mass nouns are mostly based on English examples and were adopted from Chierchia (2010).

3. Quine argued that pre-linguistic child really has no notion of 'object' as opposed to 'substance' then Carley and Spelke challenged it using different experiments and showed that at a few months of age the child has structured expectations as to how objects differ from substances (Chierchia, 2010).

1.1.1 The signature property

In general, the direct combination of a mass noun and a numeral expression is ungrammatical or needs to be seen under the special circumstances to be felicitous (1b), while count nouns (1a) are fine:

- (1) a. Thirty three tables/stars/pieces of that pizza
b. *Thirty three bloods/waters/golds

Majority of English nouns has plural marking ‘-s’ in combinations with numerals other than one, i.e. shows agreement. But even if a numeral is in the predicate position where it does not provoke number plural agreement, the combination of the mass noun and the bare numeral is still impossible (2b):

- (2) a. Those boys are at least thirty.
b. *That gold is at least thirty.
c. That gold is at least thirty pounds.

Unlike English, in Finnish numerals expect singular agreement from count nouns (3a), but numerals still cannot appear with mass nouns directly (3b):

- (3) a. *Yhdeksän omena-a*
Nine-NOM apple-part-SG
‘nine apples’
b. **Yhdeksän vesi-a*
Nine-NOM water-part-SG
‘nine water’

There are some ways how to combine mass nouns and numerical expressions felicitously. The first way is to ‘coerce’ them to be count, the elasticity property is about it. The second way is to enter a measure phrase or a classifier. Consider Mandarin Chinese (4), the combination of mass noun *rou* ‘meat’ and the numeral *san* ‘three’ is grammatical just with the special classifier *bang* alongside (4b):

- (4) a. **san rou*
 'three meat'
- b. *san bang rou*
 three CL meat
 'three pounds of meat'

The same principle applies to different languages, e.g. an Athapaskan language Dëne Sųline (north-western Canada), an ungrammatical expression in (5a) is fine with a measure word in (5b):

- (5) a. **solaghe bër*
 five meat
 'five meats'
- b. *solaghe nedadhi bër*
 five pound meat
 'five pounds of meat'

And it works for English, all examples in (1b) can be fixed by the measure phrases (6). The measure phrase takes number agreement and the combination is felicitous.

- (6) Thirty three litres of blood/bottles of water/kilos of gold.

The inability of mass nouns to occur with numerical expressions is perhaps the steadiest grammatical property associated with these terms across languages. Moreover, it seems to be that all languages solve it by the insertion of an additional grammatical construction, a measure phrase or a classifier, or shift them to be count (more about in *section 1.1.3 The elasticity property*)

1.1.2 The mapping property

The second universal property of mass nouns says that there is no language in which the basic words for *blood* or *air* will come out as count. The idea is that languages have specific morphosyntactic generalizations for this sort of distinction (just like every language has criteria to tease subjects from objects). Substances are coded as mass

by the tests prevailing in a language and the mass/count embedded in grammar.

Perhaps language as a human device had developed the algorithms of mass/count encoding based on reality and its substance/object contrast. Since the perceptual abilities of our species are the same regardless of one's language, people do not have access to minimal components of fluids or minerals and do not treat them as separated objects.

However, for clearly bounded and distinguishable objects there is no such confidence of their countability. Nouns like *furniture*, *footwear*, *jewellery*, *etc.* are mass morphosyntactically by the tests in English and other languages, but they are not substances and have all rights to be count. The researchers call them **fake mass nouns**, **count mass nouns** or **individual mass nouns**, more about them in *section 1.2.4 Fake mass nouns*.

1.1.3 The elasticity property

In some contexts, mass nouns can appear in count use (7):

- (7) a. I need three ropes rock, tape ...
b. I need a lot of rope
c. I drank three beers tea, coffees ...
d. I drank a lot of beer

Rock, *rope*, and *beer* in (7b, c) are naturally mass nouns but in (7a, b) used as count nouns without any misunderstanding. Such substances occur in lumps and pieces or in standard servings/amounts: count *rocks* are lumps of rock, count *beers* are bottles or glasses of beer, count *ropes* might be in coils. The elasticity property says that mass nouns can change their nature in favour of count use.

In (8) there are more examples of elasticity of mass nouns:

- (8) a. I drank three waters.
b. I drank three bloods.

- c. I ate three breads.
- d. I bought three golds.

A sentence in (8a) is reasonably natural as in (7c), for (8b) imagine an exotic vampire bar where it could happen. The sentences in (8c, d) are not as natural as the others but still can be used.

Such a shift from mass to count is called a **universal packager**, i.e. a function that turns mass nouns into count. Like in (7) – (8) they can be forced to count use because of their standardized or expected individual representations. The universal packager does not apply on every mass noun with the same level of naturalness, some shifted meanings are less acceptable than the others.

The second way of turning mass nouns into count use is to make them to be interpreted as kinds, sorts, or types. In (9a) naturally mass noun *wine* has the kind interpretation. However, this shift covers not just mass nouns, the same operation applies to the count noun *dog* in (9b), where *three dogs* stands for three different breeds, not individual dogs.

- (9) a. I like only three wines: chardonnay, pinot, and chianti.
- b. I like only three dogs: Irish setters, golden retrievers, and collies.

As well as mass nouns can change their basic uncountable nature, count nouns can be shifted to mass use. This shift has connections with a **universal grinder**. Pelletier (1975) introduces the concept of such a grinder machine where he suggests taking an apparently count noun and putting it through it. If the result is a homogeneous mass, i.e. it can be joined and split without changing its nature, therefore it can be spread ‘all over the floor’ (10a). But the universal machine grinds only physical objects, count abstract nouns like *kilometer* or *aspect* (10b) are not imaginable as the homogeneous mass (Doetjes, 2012).

- (10) a. There was table/bicycle all over the floor.
- b. There was *kilometer/*aspect all over the floor.

Besides the floor such a grinder can be used in different contexts. In (11a) *apples*, *chicken*, and *rabbit* are in their usual count use, but in (11b) their singular counterparts are shifted to mass:

- (11) a. I ate three apples/chickens/rabbits.
b. There is apple/chicken/rabbit in the soup.

The universal grinder shift exists in many languages, e.g. Mandarin Chinese. In (12a) a noun *ji* ‘chicken’ is count because of a classifier *zhi* which is usually restricted to whole objects. In (12b) a classifier *pan* ‘dish’ is not restricted as the first one and denotes *ji* ‘chicken’ as mass same as in (11b).

- (12) a. *san zhi ji*
three CL chicken
‘three chickens’
b. *san pan ji*
three CL chicken
‘three portions of chicken’

Conclusion It was already mentioned, the signature, mapping, and elasticity properties are generally common across languages and depend on the morphology of a particular languages. But if mass nouns behave alike, languages, which have these properties, do not interpret mass/count distinction in the same manner. There are three types of languages which encode mass/count nouns differently, the following section will describe them all.

1.2 Three ways of encoding mass/count

Languages do not represent mass/count distinction in the same way grammatically mostly because of the category of number (Doetjes, 2012). There are three nominal systems within languages – **classifier languages**, **number neutral languages**, **number making languages**.

In the first group are languages like Mandarin Chinese, Japanese, Mokilese (Austronesian), Nivkh (Siberia), etc. The second group consists of languages Dëne Sųline, Tagalog (Austronesian), etc. (Doetjes, 2012). The third group is rather familiar, here are English, Czech, Russian and other not only Indo-European languages.

The typology and examples were applied from Chierchia (2010).

1.2.1 Classifier languages

Classifier languages are named so because they require **a classifier** to count⁴ any sort of nouns. If count nouns in English, Czech, or Russian can directly combine with numerals languages, no noun in Mandarin Chinese can (13):

- (13) a. *san ge nanhai*
 three CL boy
 'tree boys'
- b. *yi ben shu*
 one CL book
 'one book'

A classifier is a word that denotes something like a measure, a container, a unit. Classifiers presuppose certain properties of nouns they occur with. It is a separate functional category, a closed set of words (Khrizman, 2016). Some classifiers are usually restricted like *zhi* in (12a) and will not combine with mass terms, others could combine and can force mass terms to count use. For example, the classifier *ge* in (13) is unspecific, so if mass nouns go with it, they have two options, they either will not grammatically combine with the classifier or do the shift (14):

- (14) ?*san ge xue*
 three CL blood
 'three portions of blood'

4. There are different types of classifiers in classifier languages which are used for different purposes (Allan, 1977), but we are interested in their mass/count representations, hence in the present work classifiers are used in the quantity expressions.

Languages like Mandarin do not obligatory mark number on nouns. Moreover, it might be that such languages have the macro syntax of each noun like mass nouns in English. But still they do distinguish between mass and count terms. This distinction is encoded in the grammar of languages through the syntax and semantics of classifiers. So, classifiers not only help to count bounded and discrete objects, i.e. count nouns in English, they also have the function similar to the measure phrases as in (6), i.e. felicitously combine mass nouns with numerals.

1.2.2 Number neutral languages

Unlike other types, number neutral languages do not mark number and do not obligatory have classifier systems. Because their nouns do not have any visible singular/plural changes, they are named number neutral. Analyzing sentences in (15) it is clear that Dëne Sųline uses the same form of a noun in singular and in plural, the noun *tth'ay* 'dish' stays the same:

- (15) a. *tth'ay thitsı* *sı*
 dish perf-1SG-make SG
 'I made one dish'
- b. *tth'ay ghighı* *sı*
 dish perf-1SG-make PL
 'I made several dishes'

In (16a, b) count nouns do not require plural agreement or any classifier combining with a numeral, while the combination of the mass noun *bër* and the numeral is ungrammatical (16c):

- (16) a. *solaghe dzot*
 'five ball'
- b. *solaghe k'asba*
 'five chicken'
- c. **solaghe bër*
 'five meat'

Number neutral languages as well as classifier and number marking languages have mass/count distinction. But unlike number marking languages they do not show number on nouns and this makes them similar to classifier languages. However, unlike classifier languages number neutral languages do not use classifiers and encode mass/count distinction in another way. It seems to be that through the syntax of numerals languages like Dëne Sųline show the contrast between mass and count.

1.2.3 Number marking languages

Languages of this group have basic singular and plural distinction⁵, so they mark number morphologically or demonstrate it through agreement. They do not use classifiers in sense of classifier languages, it means that numerals can directly combine with count nouns.

Mass nouns in number making languages fail simple tests. The first is the pluralization test which is not relevant in languages from two other groups because of their number relations. Plural forms in e.g. English or Czech are usually morphologically overt, whether on nouns themselves or through agreement. Mass nouns because of their nature do not have plural forms (17a), while count nouns do (17b):

- (17) a. water - *waters, mud - *muds, blood - *bloods
voda – *vody, bláto – *bláta, krev – *krve
- b. dog - dogs, book – books, table – tables
pes – psi, kniha – knihy, stůl – stoly

The second test runs through a determiner system including articles and quantifiers. In English some of determiners like *a(n)*, *the*, *some*, *every*, *most*, *all* are sensitive to mass/count distinction and do not choose both. In (18) there are count singular *boy*, count plural *boys*, and mass *water*. Some determiners are ungrammatical with the mass nouns and plural count nouns (the indefinite article *a(n)* or *every*), but others do not combine with the singular count nouns but are fine with mass and plural count nouns (*most* and *all*), the definite article *the* and *some* combines with all of them.

5. Moreover, languages from this group can have richer number systems with dual, trial, paucal, greater plural, etc.

- (18) a. a boy, the boy, some boy, every boy, *most boy, *all boy
 b. *a boys, the boys, some boys, *every boys, most boys, all boys
 c. *a water, the water, some water, *every water, most water, all water

A large class of determiners combines with mass and plural nouns, including *a lot*, *more* and *less*. Most of them can be used as adverbs. A small class of determiners is restricted to mass nouns – *a bit*, *much*, *little*. In English they are in complementary distribution with the plural selecting determiners – *many*, (*a*) *few*. They are used as adverbial as well (Doetjes, 2012). But it was observed that e.g. *wine* can be used with *much* in mass interpretation and with *many* in shifted to count interpretation⁶ (Grimm, 2012).

Same sensibility for a certain type of nouns show Czech quantifiers, an individual universal quantifier *všechno* ‘all’ (19a, b) and a plural existential quantifier *několik* ‘some, several, few’ (19 c, d). *Bláto* ‘mud’ and *zlato* ‘gold’ are prototypical mass nouns, *kniha* ‘book’ and *pes* ‘dog’ are count singular nouns, *knihy* ‘books’ and *psi* ‘dogs’ are count plural nouns (Dočekal, 2017):

- (19) a. všechno bláto, všechno zlato, *všechna kniha, *všechn pes, všechny knihy, všichni psi
 b. *několik blát, *několik zlat, několik knih, několik psů, *několik kniha, *několik pes

So, mass nouns in their natural sense fail simple tests in number marking languages but if such a shift as a universal packager takes place they may have ‘count’ determiners and quantifiers alongside (Grimm, 2012).

It seems to be that because of marking the number, only languages from this group have fake mass nouns. They sort of behave as mass nouns but are bounded and discrete objects in real life. The following section will tell more about the phenomenon.

6. section 1.1.3 The elasticity property

1.2.4 Fake mass nouns

As it was already mentioned in the section *section 1.1.2 The mapping property* substances in languages are coded as mass nouns and in no language substances will be represented by count nouns, though bounded objects can be encoded as mass. In Chierchia (2010) nouns like *furniture, footwear, mail, cutlery, etc.* are named **fake mass nouns**.

They are 'mass' because of their syntactic behaviour of prototypical mass nouns. Fake mass nouns cannot be pluralized, do not combine with numerals directly (20a) and are sensible to 'count' determiners (20b):

- (20) a. *I bought three furnitures.
b. I bought three pieces of furniture.
c. *I don't have many furnitures.
d. I don't have much furniture.

This type of mass nouns is 'fake' because: i) they do not denote substances, and ii) according to some tests they more plural than mass. Schwarzchild has found a class of predicates and named it **stubbornly distributive predicates**. The class can differentiate between the prototypical mass nouns and the fake ones (Chierchia, 2010).

Predicates in natural languages are of three types – collective, distributive, cumulative. A collective predicate applies to a plural entity as a whole. A distributive predicate applies to each member of the plural entity. A cumulative predicate is similar to the distributive predicate, it involves two plural entities. The basic examples of these three types are in (21):

- (21) a. The men gathered.
b. The ten girls smiled.
c. Three boys saw two girls.

In (21a) the predicate *gathered* has collective reading and applies to several men simultaneously because one man could not gather by

himself. In (21b) a situation is inverted because ten girls cannot smile as a whole, it is clear that each girl smiled by herself, the predicate *smiled* has distributive reading. In (21c) two boys could have seen one girl and one boy could have seen one girl, it depends on a situation but all *three boys* are under the predicate *saw* and it has cumulative reading (Champollion, 2014).

So back to the fake and prototypical mass nouns and stubbornly distributive predicates. It seems to be that with plurals such predicates as *small*, *big*, *cubical*, *large*, *etc.* have just distributive reading. In (22a) an example definitely has distributive reading. In (22b) the predicate may have the same reading as in (22a) but as well it may have collective reading. However, in (22c) there are together fake mass noun *furniture* and stubbornly distributive predicate *small*, the sentence is well-formed. But the same predicate does not go with the prototypical mass noun *snow* in (22d).

- (22) a. Those violets are small.
b. Those violets occupy little space.
c. That furniture is small.
d. ?That snow is small.

The test shows that fake mass nouns in English are more plural and count than mass. It is interesting that such a phenomenon is restricted only to number marking languages. Moreover, some languages, e.g. French and Italian, have count counterparts of their fake mass nouns, or others may not even have them at all, e.g. Greek (Chierchia, 2010).

The existence of fake mass nouns brings many questions. Even if classifier and number neutral languages do not have fake mass nouns in sense of number marking, their count nouns behave exactly the same as fake ones, i.e. have morphosyntax of prototypical mass nouns but denote objects. It is clear that more careful studying of them will help to test mass/count distinction even better.

Conclusion The mass/count distinction was studied in different periods of time by the numerous researchers. It is clear that substance/object contrast is reflected in languages through the mass/count distinc-

tion. And even if the contrast is basically the same because of human perception, it is curious how differently languages all over the world realize it.

It was important to explain briefly some properties of mass and count nouns because their distinction goes hand in hand with classifiers and measure phrases. Mass nouns require them almost in every case unless, of course, they are shifted to count interpretation. Moreover, count nouns in number marking languages can occur with classifiers. In some languages classifiers play a very essential role, e.g. in Mandarin no noun can go without it. In this chapter it was said that classifiers are obligatory in many cases, but the following chapter will describe formal theories of mass/count distinction and explain why they are obligatory.

2 Formal theories of mass/count distinction

This chapter will show some ways how semanticists explain the mass/count distinction formally. The first is **the lattice theory** which is about singularities, pluralities and mass nouns, by Godehard Link. The second says that the mass/count distinction is a matter of **vagueness**, by Gennaro Chierchia. The third is about **an object unit**, the embedded classifier, by Manfred Krifka.

2.1 The lattice theory

Godehard Link in Link (1983) presented a theory of plural and mass noun phrases. He proposed that the set of individuals associated with a given noun has an internal structure which is relevant for the way one talks about those individuals. For example, the noun *horse*. The meaning of *horse* is a property, i.e. the meaning of a predicate, and the property indicates for each world those entities which are described by the predicate in that world. In (23) there are associations between possible worlds and the sets of horses (Portner, 2005).

- (23) $w_1 \rightarrow \{A, B\}$
 $w_2 \rightarrow \{B, C\}$
 $w_3 \rightarrow \{A, B, C\}$
 $w_4 \rightarrow \emptyset$

In world 1, only A and B hold the property HORSE, so there are two horses, the same in world 2, so there are two horses as well. In world 3, A, B and C hold the property HORSE, in world 4 there are no such entities as horses. So, the predicate *to be a horse* applies to the entities A, B, and C and then the property HORSE, which is the meaning of the predicate, sort of marks entities A, B, C as horses. In world 4 there are no entities which hold the property HORSE, so no predicate describes them as horses.

Paul Portner in (Portner, 2005) focuses on world 3. The meaning of a singular *horse* is modeled as a set in world 3, but the plural *horses* of world 3 has to be modeled as in (24):

Diagram illustrating the decomposition of a 3-qubit state:

Top part (labeled *horses*):

$$A + B + C$$

Branching into:

$$A+B \quad A+C \quad B+C$$

Bottom part (labeled *horse*):

$$A \quad B \quad C$$

The diagram illustrates the part/whole relations between singular *horse* and plural *horses*. A, B and C in the bottom are **atoms** and they make a **lattice** of the part/whole structure, the rest are **plural individuals** or **groups**. The plural individual, e.g. A+B is called **the join** of A and B; A and B are parts of A+B. The parts of A+B+C are harder to decompose because it depends of how one looks at this plural individual: if there are three atoms as parts {A+B+C}, or two parts, one atom and one plural individual, something like {A+B, C}, {A, B+C} or {A+C, B} (Portner, 2005).

17

can be analyzed as an adjective¹ and it develops the predicate to a larger predicate. When *three* combines with *horses*, the result describes any plural individual which is described by horses and which has three atomic parts (Portner, 2005).

There is a table 2.1 to sum up all said:

Table 2.1:

<i>horse</i>	<i>horses</i>
a property which describes any individual horse	a property which describes any plural individual consisting of horses
{A, B, C}	{A+B, B+C, A+C, A+B+C}
<i>three horses</i>	<i>the three horses</i>
a property which describes any plural individual described by horses with three atomic parts	refers to the unique most salient thing described by <i>three horses</i>
{A+B+C}	A+B+C

The discussion above was obviously about count nouns, but how does it relate to mass nouns? For an answer to this question consider Mandarin Chinese. The language does not mark number, so it does not distinguish between singular and plural, as it was noted in the previous chapter. The noun *ma* can be translated as English *horse* and *horses* at once, and in world 3 *ma* would be the set {A, B, C, A+B, B+C, A+C, A+B+C}, as it does not separate atomic parts from plural individuals and numerals have nothing to count. And how it was mentioned in section 1.2.1 *Classifier languages*, Chinese and other languages deal with that by use of classifiers (or measure phrases). Classifiers indicate the units that numerals can apply to, like in (25) classifier *pi* ‘creates’ counting context (large-animal-sized) between the numeral *san* and the nouns *ma* (Portner, 2005).

1. Numerals and adjectives are of the same type $\langle e, t \rangle$ (Portner, 2005).

- (25) *san pi ma*
 three CL horses
 ‘three horses’

And the lattice structure for Chinese nouns is in figure 2.2.

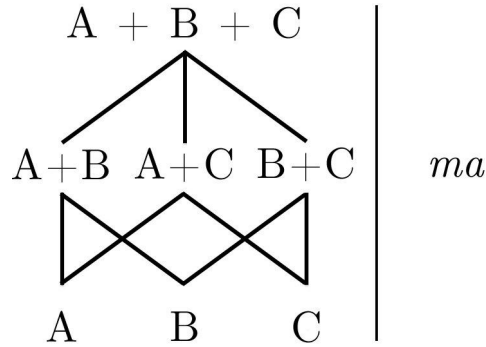


Figure 2.2: The part/whole structure of the Chinese noun *ma* in world 3

In the section 1.2.1 was also said that every noun in Chinese behaves like mass nouns in English, e.g. *gold*. A certain amount of gold can be splitted into two smaller amounts of gold and so on without limit. This is **divisive reference**²: if a mass noun holds of P, and S is a part of P, then the mass noun holds of S as well. The mass noun *gold* holds a property GOLD, and as far *a piece of gold* was divided from *gold*, it is *a part of gold*, then *gold* is also *the piece of gold*. And this division has no end, so mass nouns have no bottom of the part/whole structure, they do not have atoms or any basic units³, same as every Chinese noun (Portner, 2005; Lasnik, 2011).

To be more precise, Chinese count nouns have minimal parts but do not categorize them. This fact makes them more like fake mass nouns in number marking languages (section 1.2.4 *Fake mass nouns*).

2. $\forall P [DIV_S P \leftrightarrow \forall x \forall y [P(x) \wedge y \subseteq_S x \rightarrow P(y)]]$
 (P has divisive reference) (Krifka, 1989)

3. Laws of physics are omitted, languages do not reflect them because of humans' limited perception.

These nouns clearly denote things that come in minimal parts: *furniture* is a pack of chairs, tables, etc., *change* is a pack of coins. Chinese nouns and fake mass nouns just do not care about the atoms, atoms are not relevant for their meaning. As well as Chinese nouns, fake mass nouns cannot felicitously combine with numerals and require a classifier (Portner, 2005).

Talking more about canonical mass count and fake ones, Link in (Link, 1983) comes up with the next example:

- (26) The gold in Smith's ring is old, but Smith's ring is not old.

From (26) follows that the ring is gold and the ring is new, but the gold (the ring was made from) is old. The noun *ring* is surely count and the noun *gold* is mass. And even if the ring was made from the gold, they have different properties, therefore it is necessary to distinguish between them two to avoid the paradox in (30). So, Link in his paper claims that mass nouns hold of the portions of materials, while count nouns hold of more abstract objects constituted of that material (Lasersohn, 2011).

However, this proposal concerns the canonical mass nouns only. It is the question whether the thing like *the ring* is the portion of material like *gold*, but the thing as *a chair* (a count noun) is *a piece of furniture* (a fake mass noun). The chair is not made from furniture in the sense that the ring is made from gold. Moreover, the chair cannot be new while the furniture is old and vice versa (Lasersohn, 2011). There is again the evidence of how canonical mass nouns are distinct from fake ones.

Count plural and mass nouns may have some common qualities, but the lattice theory explains how they differ. Basically, mass nouns do not have the bottom atomic layer in the part/whole structure while count nouns do and they refer by it to singulars. Plurals cannot be divided downward without the end, there is always the minimal point where they became singulars. Substances like *water*, *air*, *mud* and *gold* and objects like *book*, *boy*, *table* and *coin* have the different part/whole structures, thus they have different properties.

2.2 Mass nouns and vagueness

Gennaro Chierchia in Chierchia (2010) introduces the theory of mass/count distinction based on vagueness. He argues that through the concept of supervaluation vagueness can be integrated into singular/plural structures, and it would not be possible to miss where the mass/count distinction comes from. Unlike Link he claims that canonical mass nouns have minimal atoms, but the atoms are vague.

Starting with vagueness it is necessary to explain what it means. Chierchia says that almost all natural concepts are vague. Consider a simple request to count objects in a room. One points at a table and says “There is a table, this is one. The table has a left leg, this is two. The left leg of the table has an upper part, this is three...” So, this example shows that it does matter on which level one counts and there is no right and single way to count. But if he or she uses plurals, it is clear that the person counts the smallest things to which the noun applies, i.e. atoms. If there are two books, no one would ever count the spines of two books or each pages of the books as a minimal part of two books, the combination *two books* has its minimal parts, they are well defined, it means they are not too vague.

In case of mass nouns, minimal elements are not exactly specified, so they are vague. *Gold, water, dough* and other canonical mass nouns do not come in standardized forms and because of it they cannot be pluralized, there is nothing stable to make them plural, and they do not combine with numerals because there is nothing atomic to count (*section 2.1 The lattice theory*). Mass nouns are born too vague.

However, both mass and count nouns are of **inherent vagueness**. The definition of inherent vagueness is that even an expert will be unable to provide uncontroversial criteria according to which a property applies or not to a certain individual or a group. Again, if one is counting three cats, he or she is counting individuals, i.e. cat-atoms, the smallest things on which the property CAT applies. But does the property CAT have specific boundaries? Does a cat embryo already hold the CAT property? Or in case of an unfortunate situation a cat lost a tail, is it still a cat in a sense that the property CAT applies to it, or actual cats have to own tails obligatory? The answers for the questions are hard to find.

The inherent vagueness affects how one counts: the inherent vagueness of the nouns which have defined atoms, i.e. count nouns, it determines how many atomic parts there are; the inherent vagueness of mass nouns prevents from counting in mass predicates. Because even if it is not really possible to define where the property *CAT* begins and where it ends, people still have a clear idea what the cat is. There are plenty of things that fall under the cat concept. The boundary of the property *CAT* is such that there are definitely *x*'s that fall under it, such that no proper part of *x* does. So, the count nouns are vague in the sense that boundaries of their properties are hard to find out, and simultaneously the count nouns are not too vague because of their atomic structure.

And now consider the mass nouns *rice*. The spoon of rice falls under the concept of rice, a half of the spoon of rice is rice, the two kilos of rice fall under the concept of rice, one grain of rice is rice as well. From it follows that the half of grain of rice is rice, and the quarter and etc., it recalls divisive reference. The point is that there is no systematic basis for deciding which rice amount is qualify as rice atoms. The minimal parts are not specified because of the inherent vagueness.

Gennaro Chierchia has all formal proofs of the theory in his paper, for their discussion see Chierchia (2010).

2.3 The object unit

Manfred Krifka in Krifka (1995) reveals an appealing theory which may help to explain mass and count distinction. In the paper he investigates Chinese and English and introduces **the object union** operator. Krifka makes parallels between two languages and claims that classifiers either embedded or overt distinguish count nouns from mass.

He starts with Chinese and shows that the noun *xíong* 'bear' can refer to a kind *Ursus*, to some specimens of the kind *Ursus*, to a specified number of realizations of the kind *Ursus*, to a specified number of individual specimens of the kind *Ursus*, and to the subspecies of the kind *Ursus* with the use of an aspect (27a, b), a measure construction (27c) or classifiers (27d, e) ⁴.

4. ASP stands for 'aspect', CL for 'classifier'

- (27) a. *xióng júe zhǒng le*
 bear vanish kind ASP
 'The bear is extinct'
- b. *wǒ kànjiàn xióng le*
 I see bear ASP
 'I saw (some) bears'
- c. *sān qún xióng*
 three herds bear
 'three herds of bear'
- d. *sān zhī xióng*
 three CL bear
 'three bears' (objects)
- e. *sān zhǒng xióng*
 three CL bear
 'three bears' (species)

In case of English the noun *bear* does not need classifiers or particular grammatical constructions for getting kind, specimens, or other sorts of interpretations (Krifka, 1995). The pluralized noun (bare or modified) can be interpreted as kind referring (28a), as an unspecified number of individuals (28b), may have either individual and specified number interpretation (28c), or denotes subspecies (28d).

- (28) a. Bears are not ferocious.
- b. Children have seen (some) bears in the zoo.
- c. Goldilocks visited three bears.
- d. Mary likes three bears: the panda bear, the grizzly bear and the polar bear.

Krifka suggests that there are two types of relation **the realization relation R** and **the taxonomic subkind relation T**. In general, $R(x, y)$ where x is an instance of y , i.e. R relates kinds to their specimens, e.g.

$R(Xinxin, Ailuropoda melanoleuca)$ simply says that *Xinxin is a panda*. Analogously, $T(dolphin, whale)$ means *a dolphin is a whale*, i.e. $T(x, y)$ where x is subkind of y (Krifka, 1995).

The R and T relations apply to the nouns in the following way. If k is a kind, then the property $\lambda x.R_i(x, k)$ applies to specimens or individual sums of specimens of k in world i , while the property $\lambda x.T_i(x, k)$ applies to subspecies or individual sums of subspecies of k in world i . The relations are used together in a unit relation RT as $RT_i(x, y) \leftrightarrow R_i(x, y) \vee T_i(x, y)$, so R or/and T applies. Thus, back to Chinese examples in (27) and English in (28), the property $\lambda i \lambda x.RT_i(x, Ursus)$ applies to a single bears or collections consisting of bears, and to a single bear individual or collections consisting of bear individuals (Krifka, 1995).

Considering (27d) where $zh\bar{i}$ is a classifier, Krifka introduces the operator OU , i.e. **the object unit**. The operator for each possible world i , takes a kind and yields a measure function that measures the number of specimens of that kind. In (27d) semantically there is $\lambda i \lambda x [RT_i(x, Ursus) \wedge OU_i(Ursus)(x) = 3]$. For the reason that the OU_i applies to the objects only, just the R is realized, so it is from the kind *Ursus* to the objects *bears*. The three specimens were picked from the kind by the OU_i operator, in other words by the classifier $zh\bar{i}$ and the numeral $s\bar{a}n$ ‘three’.

Even if the T relation is not very helpful in (27d) because the talk is about objects not subkinds, it is important to put it there anyway. The RT relation distinguishes for example cats from bears. The OU alone picks specimens of all kinds, so $OU_i(Ursus) = OU_i(Felis) = OU_i(animal)$. And it is the RT what ranks the result of the OU as an individual or a group of individuals to the kind.

The difference between (27d) and (27e) is in classifiers morphologically and in the operator KU , i.e. **the kind unit**, semantically. The KU applies to the kind *Ursus* and yields the subspecies of the kind - the polar, the grizzly, the panda, almost in the same way as the OU does. The difference is that now that the T stands for the RT relation, from the kind to the subkinds. The formal entry is $\lambda i \lambda x [RT_i(x, Ursus) \wedge KU_i(Ursus)(x) = 3]$ (Krifka, 1995).

However, the OU and KU work differently in non-classifier languages. First, as far as there is no different classifiers for different sorts of relation and as in (28), one bare noun can stand for different

meanings, the OU and KU operators are one whole the OKU operator, i.e. **the object or kind unit**. Second, in Chinese any classifier is visible and overt, while the OKU in English is hidden into the structure of count nouns themselves or the structure of numerals.

Even if in (28c) and in (28d) *three bears* refers to different sorts of things (individual bears and the subkinds), its inner semantical structure in these two examples is the same $\lambda i \lambda x [RT_i(x, Ursus) \wedge OKU_i(Ursus)(x) = 3]$. The OKU operator applies to the kind *Ursus* and yields either three individuals from the kind or three subkinds.

So, Krifka proposes to highlight the OU, the KU and the OUK operators in the structure of the nouns and the classifiers, but how does the proposal apply to mass/count distinction? It does in a very elegant way. Consider again the Chinese case, the OU/KU operator is in the classifier, without it nouns cannot refer to kinds, to individuals or cannot be counted. So, in general, it is the classifier and its structure make them count, without it they are like mass nouns English. And if the OKU is embodied in nouns in English, the OKU makes them count. The OU/KU and the OUK do the same job, they create for numerals a countable context. The difference is that Chinese use the OU/KU as a separate functional category, but in English the OKU is hidden in the structure of count nouns (Krifka, 1995).

Krifka argues that mass nouns cannot be count directly with numerals exactly because in their structure there is no the OKU operator. The formal entry of *mud* is $\lambda i \lambda x [RT_i(x, mud)]$, whereas for *bears* it is $\lambda i \lambda x \exists n [RT_i(x, Ursus) \wedge OKU_i(Ursus)(x) = n]$. However, mass nouns can denote individuals with the measure phrases, same as Chinese nouns can be counted with the classifiers alongside. The measure phrases usually consist of the count nouns, i.e. have the OKU operator in their inner structure. The measure phrases also involve measure units like *liter, kilogram, pound, etc.* which are well defined.

Krifka's proposal is crucial for the next chapter and for the experiment.

Conclusion Three theories above in their own way explain mass/-count distinction. In summary, Link claims that the reason why mass nouns cannot be count or pluralized is because of their lattice structure, there is no bottom atomic layer, thus there is nothing to count or

pluralized. Chierchia explains it through the concepts of vagueness. Krifka analyses classifier Chinese and non-classifier English and introduces the OU/KU and the OKU operators because of which mass and count nouns have differences.

All three have their pros and cons but the next chapter uses Krifka's proposal and focuses on Russian classifiers.

3 Classifiers *štuka*, *čelovek* and *golova*

Keren Khrizman in Khrizman (2016) investigates three Russian classifiers, *štuka* ‘item’, *čelovek* ‘person’ and *golova* ‘head’. She argues that these classifiers make together a closed set of measure words, although look very much like Mandarin Chinese sortal classifiers. Khrizman analyses data and concludes that they are not measure nouns but words which measure mass denotations in terms of natural units in the sense of (Krifka, (1989); Krifka, (1995)).

3.1 Semantics of counting and measuring

In (29) are numeral NPs of two subtypes, counting or measuring:

- (29) a. five boys
b. five items of furniture
c. five liters of milk
d. five kilos of potatoes

Counting expressions (29a, b) are the count predicates denoting sets of sums of atomic entities, while measuring NPs (29c, d) are the mass predicates denoting sets of non-individuated quantities of entities/stuff. In other words, in (29a, b) there are the pluralities boys/items of furniture and a cardinal number five which indicates how many discrete parts each of the sets has (30).

- (30) a. $\| \text{five boys} \| = \lambda x. \text{PL}(\text{BOY})(x) \wedge |x| = 5$
(The set of pluralities of boys such that each of pluralities has 5 atomic parts)
b. $\| \text{five items of furniture} \| = \lambda x. \text{PL}(\text{FURNITURE ITEM})(x) \wedge |x| = 5$
(The set of pluralities of furniture item with 5 atomic parts)

The difference between measuring and counting is that in the end of measuring operation there are sets of pluralities in which atomic

parts are not specified, so are mass. Even morphologically count plural complements in measure expressions, according to Rothstein (2011), are shifted to mass interpretation¹. In (29c, d) *milk* and *potatoes* are the quantities, *liters* and *kilos* are predicates calibrated in certain units, liters and kilograms themselves, on a dimensional scale, for this case volume and weight. A remaining cardinal *five* restricts number of certain units to five liters and kilos (31) (Khrizman, 2016).

- (31) a. $\| \text{five litres of milk} \| = \lambda x. \text{MILK}(x) \wedge \text{MEAS}^{\text{VOLUME}}(x) = \langle 5, \text{LITER UNIT} \rangle$
(The set of quantities of milk which measure 5 liters in volume)
- b. $\| \text{five kilos of potatoes} \| = \lambda x. \text{PL}(\text{POTATOE})(x) \wedge \text{MEAS}^{\text{WEIGHT}}(x) = \langle 5, \text{KILOGRAM UNIT} \rangle$
(The set of quantities of potatoes which measure 5 kilos in weight)

This interpretation is very convenient for the paper, Khrizman uses it to distinguish between different types of classifiers in English and Russian.

3.2 Types of classifiers

Traditionally, languages are divided into classifier and non-classifier (number marking and number neutral), e.g. Mandarin Chinese and English. If a language (English) has a category of count nouns, then it does not need a functional category to count. On the contrary, if a language (Chinese) has only mass nouns, it depends on a separate syntactic category of individuator, i.e. classifiers. With all due respect, this terminology is misleading.

It was already mentioned above, and it is necessary to repeat it here: all nouns in Mandarin Chinese even clearly distinguishable individuals, are mass, so they cannot be count directly and require classifiers (32):

1. Much/#many of the five kilos of potatoes remained unused.
 Five kilos of potatoes require much, not many, then the measure expression is mass.

- (32) *yi zhi dou*
one Cl dog
‘a dog’

The classifier *zhi* is a functional head because:

- does not add any new information, just presupposes animal property of a dog;
- cannot be used as a noun;
- is a part of closed class of classifiers.

Languages use various classifiers to categorize different sorts of things (Allan, 1977). Their type so far is $\langle k, \langle e, t \rangle \rangle$, where *k* is a noun with kind denoting mass interpretation and $\langle e, t \rangle$ is a produced predicate denoting the set of atomic instantiations of that kind and can be already counted. This is the basic and primary meaning of **sortal classifiers**.

However, the situation is not so clear with non-classifier languages. Two deceptive points in the interpretation above: i) Hungarian is a language which has both count nouns and sortal classifiers; ii) non-classifier languages use individuating classifiers to count and this is highly important for the further discussion.

English as a non-classifier language has mass and count categories of nouns, both can appear with classifiers. But count nouns do not need a classifier to be count and can be modified by numerals directly, what is ungrammatical for mass nouns. In (33a) *glasses* is the classifier together with a cardinal *five* denoting from *milk* which is a mass noun, the sets of pluralities of individual *glasses* with milk and in each there are five atomic parts. In (33b) the situation is rather similar, the difference is that a word *berries* is not mass but a count noun.

- (33) a. five glasses of milk
b. five glasses of berries

Classifiers in English are count relational nouns at type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$, so instead of kinds (Chinese case) they take predicates and denote predicates. Another difference is an ability of the classifier to be used

as a lexical noun, *glass* is a regular count noun in English. These are the reasons why English and other non-classifier languages do not have ‘proper’ classifiers, i.e. sortal classifiers. From now on these classifiers are named **measure nouns**.

There is the third type of classifiers named **measure words**². This type includes such words as *liter*, *kilo*, *meter*, etc. They are functions at type $\langle n, \langle e, t \rangle \rangle$ from numbers to measure predicates denoting sets of sums of entities which measure *n* number of units on a dimensional scale. The scale is a triple $\langle D, U, N \rangle$:

- *D* is a dimension (volume, weight, etc.);
- *U* is the unit in terms of which the scale is calibrated (kilos, liters, etc.);
- *N* is a set of numbers (the natural numbers, the real numbers, etc.)

So, e.g. *liter* is associated with a volume scale calibrated in liter units (Khrizman, 2016):

$$(34) \text{ Scale: } \langle \text{VOL}, \text{LITER}, N \rangle \\ \parallel \text{ liter } \parallel = \lambda n \lambda x. \text{MEAS}^{\text{VOLUME}}(x) = \langle n, \text{LITER UNIT} \rangle$$

In the section 3.1 *Semantics of counting and measuring* this type of classifiers was already mentioned in the relation to measuring in (31). The measure word is the sort of the OKU operator in Krifka’s proposal, but instead of objects or kinds it denotes certain calibrated units *litres*, *kilos*, *meters*, etc. (section 2.3 *The object unit*.)

Khrizman applies Krifka’s proposal to Russian and analyses *štuka* ‘item’, *čelovek* ‘person’ and *golova* ‘head’ as functional classifiers in Chinese. In Russian, count nouns can be count directly (35a, c) and used as measure nouns (35b, d).

2. In previous chapters terms *measure word* and *measure noun* were described by one term **measure phrase**.

- (35) a. *pjat' predmetov*
five item-GEN.PL
'five items'
- b. *pjat' predmetov mebeli*
five item-GEN.PL furniture-GEN.SG
'five items of furniture'
- c. *pjat' butylok*
five bottle-GEN.PL
'five bottles'
- d. *pjat' butylok vina*
five bottle-GEN.PL wine-GEN.MASS
'five bottles of wine'

But there is a small class of classifiers which does not fit this scheme. These in (36) optionally occur in the counting constructions and have interesting properties.

- (36) a. *pjať štuk jaic*
five item-GEN.PL egg-GEN.PL
'five eggs'
- b. *pjať čelovek rabočix*
five person-GEN.PL/SG?? worker-GEN.PL
'five workers'
- c. *pjať golov korov*
five head-GEN.PL cow-GEN.PL
'five cows'

Khrizman claims that: i) *štuka*, *čelovek* and *golova* are functional expressions, like sortal classifiers in Mandarin, while *predmet* and *butylka* are measure nouns; ii) English and Russian classifiers in contrast with Mandarin take predicates and not kind-denoting terms as complements; iii) Russian classifiers are not measure nouns but are a closed set of functional measure expressions at type $\langle n, \langle e, t \rangle \rangle$ analogous to *liter* (Khrizman, 2016).

All these statements are explained in the following sections below.

3.3 *Štuka*, *čelovek* and *golova* are functional expressions

The nouns like *predmet* ‘object’, *kusok* ‘piece’ or *butylka* ‘bottle’ are indeed lexical words and counting classifiers, i.e. measure nouns. Classifiers *štuka*, *čelovek* and *golova* behave like functional expressions, not like nominals (Khrizman, 2016).

1. *Štuka*, *čelovek* and *golova* do not add any new content to expressions they appear with;
2. Similar as Mandarin classifiers, Russian require certain type of nouns:
 - (a) *štuka* – inanimate objects;
 - (b) *golova* – farm animals;
 - (c) *čelovek* – human beings;
3. Classifiers when they stand alone could not be used or will have different meaning:
 - (a) *štuka* does not have any encyclopaedical definition and restrict meaning;
 - (b) *golova* is a body part but with a certain type of nouns (farm animals) is connected to livestock;
 - (c) although *čelovek* has ‘a human’ meaning when it is discrete, in classifier constructions *čelovek* has morphologically ungrammatical form; the word is suppletive and *ljudi* should go for a plural form;
4. Adjectival modification is not possible with these terms;
5. Require a complement, cannot appear without a number, they are syntactically dependent;
6. *štuka*, *čelovek* and *golova* could not be replaced by words with similar meaning, e.g. *vešč* ‘thing’, *mužčina* ‘man’ etc.

Other classifiers (*butylka*, *kusok*, *predmet*) do not fit for the category of functional expressions and that makes *štuka*, *čelovek* and *golova* a closed set of functional classifiers like Mandarin sortal classifiers.

3.4 *Štuka*, *čelovek* and *golova* do not take kind-denoting terms as complements

Štuka, *golova* and *čelovek* occur mostly with plural count nouns. Russian plural count nouns can be interpreted as referring to kinds as well as predicates denoting sets of individuals. If Russian classifiers as Chinese are functional expressions, they ought to take kind-denoting terms and make them predicates. Khrizman proves that it is not true, the complements of *štuka*, *čelovek* and *golova* are plural predicates. A closer look to these follows.

1. Complements of *štuka*, *čelovek* and *golova* can be modified by temporal (37a) and stage-level (37b, c) modifiers.

(37) a. *pjať štuk včerašnjx kotlet*
 five item-GEN.PL yesterday meatball-GEN.PL
 'five yesterday's meatballs'

b. *pjať čelovek opytnyx oficerov*
 five person-GEN.PL skilled officer-GEN.PL
 'five skilled officers'

c. *pjať golov molofoho skota*
 five head-GEN.PL young cattle-GEN.MASS
 'five head of young cattle'

2. Superlative (38a) and possessive (38b) nominal constructions can be complements of the investigated classifiers.

(38) a. *pjať štuk samyx krasivyx žemčužin*
 five item-GEN.PL most beautiful pearl-GEN.PL
 'five the most beautiful pearls'

b. *10 čelovek našix oficerov*
 10 person-GEN.PL our-GEN.PL officer-GEN.PL
 '10 persons of our officers'

3. If *štuka*, *čelovek* and *golova* took singular count nouns as complements same as *podvid* 'subtype', they would appear with kind-denoting terms. They do not (40) while *podvid* does (39).

- (39) *vsego vydeleno 9 podvidov tигра*
in-all distinguished 9 subtype-GEN.PL tiger-GEN.
in total 9 subtypes of tigers have been distinguished'

- (40) a. **pjat štuk perepelingo jaica*
five item-GEN.PL quail-GEN.SG egg-GEN.SG
'five quail eggs'

- b. **pjat čelovek kamenščika*
five person-GEN.PL brick-layer-GEN.PL
'five brick layers'

4. In Russian it is possible to make from kind-denoting terms singularive nominals by using the suffixes '-inka/-ina'. *Štuka*, *čelovek* and *golova* are fine with them (41b), but ungrammatical with the unaffixed nouns (41a).

- (41) a. **100 000 štuk pyli*
100 000 item-GEN.PL dust-GEN.MASS
'100 000 dust grains'

- b. *100 000 štuk pylinok*
100 000 item-GEN.PL dust-inka-GEN.PL
'100 000 dust grains'

The examples above (37) – (41) have proved *štuka*, *čelovek* and *golova* do not take kind-denoting terms as complements but plural count predicates.

3.5 *Štuka*, *čelovek* and *golova* are measure unit classifiers

Russian classifiers do not make from kind-denoting terms countable predicates like Chinese individual classifiers. Moreover, they do not behave like measure nouns such as *predmet* or *butylka*. Khrizman argues that these classifiers do not have properties of genuine count predicates like *five eggs* but properties identical to measure predicates

like *five liters of water*. She applies the theory of counting and measuring outlined above and calls them measure words.

While the explicit classifier *liter* measures in certain units (42a), *štuka*, *čelovek* and *golova* measure pluralities in natural units (42b).

- (42) a. *pjať litrov N* → the set of pluralities of *N* which measure five litres
 b. *pjať štuk/čelovek/golov N* → the set of pluralities of *N* which measure five natural units

Or if to apply the entry from *section 3.2 Types of classifiers of liter*, follows (43):

- (43) Scale: $\langle \perp, \text{NATURAL UNIT}, N \rangle$
 $\parallel \textit{štuka, čelovek or golova} \parallel = \lambda n \lambda x. \text{MEAS}^{\text{ARBITRARY}}(x) = \langle n, \text{NATURAL UNIT} \rangle$

The dimension is arbitrary, calibration is in terms of **natural units**, *N* is the set of natural numbers (Khrizman, 2016).

Two predictions come with this analysis. First, the given classifiers should take predicates which refers to naturally distinguishable units. Second, constructions with *štuka*, *čelovek* and *golova* should have properties of measure predicates and not of count predicates.

It has been already said that Russian classifiers *štuka*, *čelovek* and *golova* take count predicates as their complements but this is not a full picture. Usually the classifiers take plural count predicates and do not admit kind-denoting terms. More precisely, they take predicates with distinguishable atomic entities, so fake mass nouns could fit it, see (44):

- (44) a. *pjať štuk* *pečenja/kuragi*
 five item-GEN.PL biscuit-GEN.
 ‘five biscuits/dried apricots’
 b. *pjať golov* *rogatogo skota*
 five head-GEN.PL horned cattle-GEN.MASS
 ‘five head of cattle’

- c. *pjať čelovek* *narodu*
five person-GEN.PL people-GEN.MASS
‘five people’

More common usage of these classifiers is where a classifier comes before a numeral, so a construction is inversed. It creates an approximative phrase which notifies about measuring. In many cases with strict word order given classifiers are redundant because do not change the meaning (45), inversed word order is more suitable (46).

- (45) a. *pjať (štuk)* *jaic*
five (item-GEN.PL) egg-GEN.PL
‘five eggs’
- b. *pjať (čelovek)* *rafočix*
five (person-GEN.PL) worker-GEN.PL
‘five workers’
- c. *pjať (golov)* *svinej*
five (head-GEN.PL) pig-GEN.PL
‘five pigs’
- (46) a. *štuk* *pjať jaic*
item-GEN.PL five egg-GEN.PL
‘about five eggs’
- b. *čelovek* *pjaťrafočix*
person-GEN.PL five worker-GEN.PL
‘about five workers’
- c. *golov* *pjaťsvinej*
head-GEN.PL five pig-GEN.PL
‘about five pigs’

It was pointed before, *štuka*, *čelovek* and *golova* syntactically depend on numerals. In measuring constructions with *liter*, *kilo* or *meter* is not easy to drop off a numeral as well because these are calibrated by cardinals.

Conclusion Russian as a non-classifier language does not have sortal classifiers, but uses measure nouns (*butylka*, *predmet*, *kusok*) and measure words (*litr*, *metr*, *kilogram*). In addition, there is a closed class of measure words *štuka*, *čelovek* and *golova* which should to fall under the definition of measure nouns but it does not. In the next chapter will be described the experiment which i) ought to prove Khrizman's theory, and ii) investigate whether the Czech classifier *kus* behaves like *štuka*.

4 The experiment

The experiment focuses on Czech *kus* and Russian *štuka* which are translations of each other in these languages. It is well-known that Russian and Czech come from one group of languages, Slavic, though they belong to different branches Eastern and Western respectfully. The first prediction would be that *kus* and *štuka* behave alike, so a purpose of the experiment is to compare two at first sight identical classifiers from related languages. I use Khrizman's analysis and compare *kus* and *štuka* in the various contexts. The experiment has shown that they have different semantic structure and are not the exact translations of each other.

4.1 Is *štuka* a translation of *kus*?

Kus is a Czech counterpart of the Russian *štuka* in two points. First, the Russian – Czech dictionary translates word *štuka* as *kus* in its first meaning (Vlček, 1974). Second, *štuka* and *kus* are used in the same context in Russian and Czech, e.g. *a pack of wet-wipes* (47) or *a box of eggs* (48)¹:

- (47) a. *Vlhčené hygienické ubrousky s vůní vodního melounu*
wet-PL hygienic-PL wipes with aroma water melon-GEN
v balení 100 kusů.
in pack 100 items
Hygienic wet-wipes with watermelon aroma in a pack of 100 pieces
- b. *Vlažnyje salfetki osvežajuščije 60 štuk v upakovke.*
wet wipes refreshing 60 items in pack
Refreshing wet-wipes in a pack of 60 pieces

1. Czech examples were found in the Czech National Corpus. Russian examples were either found on the Internet or translated from Czech.

- (48) a. *Balení obsahuje 6 kusů čerstvých vajec.*
 pack has 6 items fresh eggs-GEN
 There are 6 eggs in a box.
- b. *V korobke naxoditsa 6 štuk svežix jaic.*
 in pack are 6 items fresh-GEN.PL eggs-GEN
 There are 6 eggs in a box.

Perhaps two points above are all that relate these two words, *kus* and *štuka* are used as individuating classifiers in Czech and Russian. It is possible to find *kus* in different contexts but for *štuka* according to Khrizman it is not. Observing e.g. the compatibility of these words with canonical mass nouns, it is clear that *kus* and *štuka* are totally different terms. Czech classifier is felicitous in (49a) with the mass noun *maso* 'meat', when the combination of Russian *štuka* and the mass noun *mjaso* 'meat' is ungrammatical (49b), *štuka* requires a different noun *kusok* 'piece' (49b):

- (49) a. *Přinesl jsem kus masa.*
 brought be-1SG.PST piece meat-GEN.MASS
 I have brought a piece of meat.
- b. **Ja prines štku mjasu.*
 I brought item meat-GEN.MASS
 I have brought an item of meat.
- c. *Ja prines kusok mjasu.*
 I brought piece meat-GEN.MASS
 I have brought a piece of meat.

Moreover, *kus* is fine with fake mass nouns (50) – (51), a situation in Russian is difficult. Khrizman claims that fake mass nouns could felicitously combine with *štuka* (section 3.5 in (44)), but *mebel* 'furniture' and *odežda* 'clothing' demand another word *predmet*.

- (50) a. *Zřejmě chtěl přemístit nějaký kus nábytku*
 Apparently wanted move some piece furniture-GEN
 Apparently, he wanted to move some
podle vlastního, mužského vkusu.
 according own male taste
 piece of furniture in
- b. *Očividno, chofel peremestiť kakoj-nibud' předmět/*
 Apparently wanted move some piece/*some
**kakuju-nibud' štuku mebeli po svojemu*
 item furniture.GEN according own male
mužskomu vkusu.
 taste
 Apparently, he wanted to move some piece/item of furniture
 in his own male way.
- (51) a. *Otevřela starou masivní šatní skříň a vyndala několik*
 opened old massive clothes closet and take-out several
kusů oblečení
 items clothes
 She opened an old massive clothes closet and took out several
 pieces of clothing.
- b. *Ona odkryla straryj massivnyj garderob i dostala neskolk*
 she opened old massive wardrobe and take-out several
*predmetov/*štuk odeždy.*
 pieces/items clothes
 She opened an old massive wardrobe and took out pieces/items
 of clothing.

It follows from the discussion above that *kus* and *štuka* are different types of classifiers. The goal of the experiment is to confirm the hypothesis:

- (52) *Štuka* is a measure word analogous to *liter* of type $\langle n, \langle e, t \rangle \rangle$ (Khrizman 2016), while the Czech *kus* is a measure noun of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$, so they are different types of classifiers (section 3.2 *Types of classifiers*).

I have prepared two questionnaires to check this hypothesis. As far as Czech *kus* behaves like a measure nouns, it was compared with proper measure words, i.e. *litr*, *kilogram*, *metr*, etc. In case of Russian where *štuka* ought to be a measure word, it was opposed to measure nouns, i.e. *kusok*, *butylka*, *čашka*, *pačka*, etc. Following sections will describe how the experiment was going.

4.2 Respondents and distribution

The total number of respondents is 55, 33 speakers of Russian and 22 speakers of Czech. Unfortunately, I am not aware of the participants age range because all that I required from them was a name, but I can assume² that Russian participants were between 17 – 50, while Czech between 20 – 35. Respondents took part in the experiment on a voluntary basis without any reward or compensation.

The experiment was created using Ibex Farm, an online platform for experiments. I contacted Czech respondents via social media website Facebook and Russian ones through VKontakte. I send them a link to the experiment and a short message with some basic information. Some of them forwarded it to their friends.

To check whether the participants followed the instruction and did not choose the answers chaotically, I created so-called **filler sentences**. Subjects had to rate them as well as tested sentences. I used data from all Czech subjects and all Russian because all of them passed the filler sentences and their answers were reliable. In the next section there are two histograms of participants which show how respondents handled fillers.

4.3 Method

In Russian and Czech versions of the experiment participants had seen 50 sentences, 25 were tested items and 25 fillers. Respondents had to assess them on a Likert scale from 1 to 5, where 1 is *not acceptable* and 5 is *completely acceptable*. This method is called **truth value judgment task**, subjects had to judge well-formedness and naturalness of the

2. My assumption is based on the feedback I have and the known age of my respondents which had the links directly from me.

sentence in a given context. Each version had an introduction where subjects wrote their names, a practice section and an experimental part.

Items and fillers in the experimental parts were mixed, therefore respondents could not have guessed what I tested. In each experiment 12 fillers were well-formed and natural in the given contexts, i.e. completely acceptable, and 13 were either not true in the context or ungrammatical, i.e. not acceptable.

It was checked whether the average of each participant's responses to unacceptable fillers was lower than the average of their responses to acceptable fillers. The histograms follow.

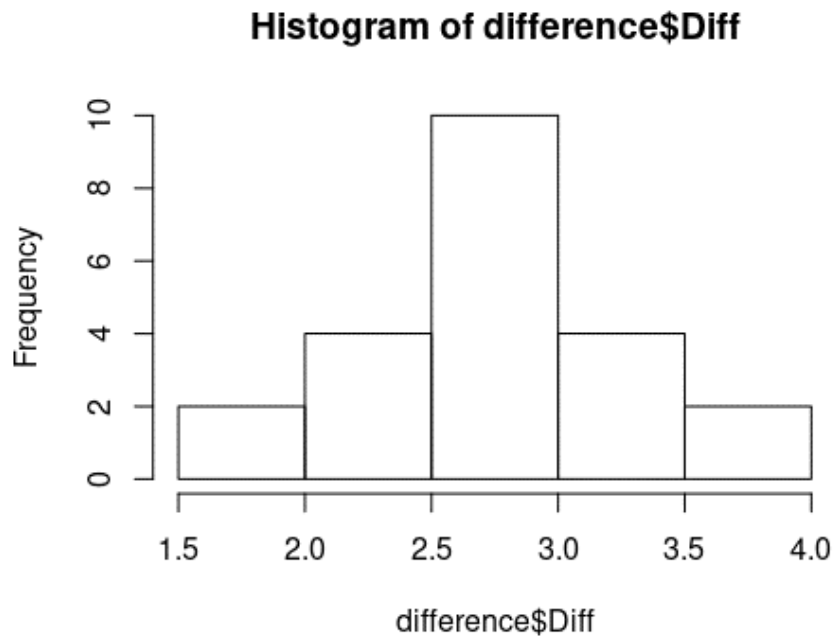


Figure 4.1: The histogram for Czech

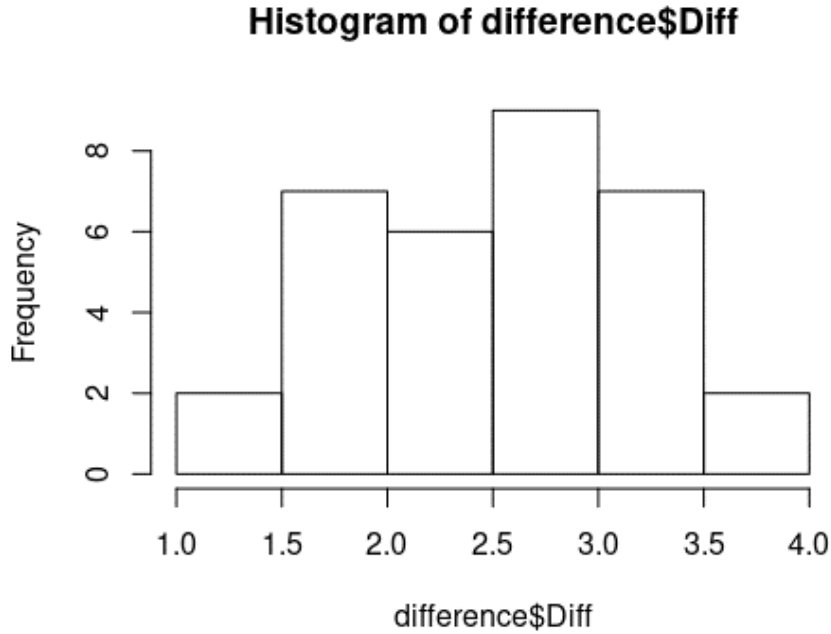


Figure 4.2: The histogram for Russian

The filler sentences were the same for each participant but the tested sentences were picked from 150 – 25 contexts multiplied by 3 conditions multiplied by 2 classifier sentences. The purpose of the Russian experiment was to confirm that *štuka* is a measure word, so there were three sentences with *štuka* and three sentences with measure nouns *kusok*, *butylka*, *čашka*, *pačka*, etc. The prediction was that subjects would mark sentences with *štuka* as not acceptable in the given contexts, and in the same contexts but with the measure nouns they would mark them as completely acceptable. Every respondent had seen only one sentence from six options.

Khrizman in Khrizman (2016) claims that it is not possible to use *štuka* without numeral, A) and B) options were a baseline with a numeral, so A) and B) ought to be acceptable. Options C) and D) should have proved Khrizman’s point 3 in section 3.3. In E) and F) classifiers were modified by adjectives and according to point 4 in section 3.3, *štuka* had to fail this test.

The Czech experiment was designed in the same way, however, its purpose was a bit different. Even if the conditions were the same, *kus* as a measure noun was opposed to measure words *litr*, *kilogram*, *metr*, etc. The prediction was that subjects would mark sentences with *kus* as completely acceptable in the given contexts, and in the same context sentences with measure words would mark as not acceptable.

The structure of Czech and Russian items is following:

Context: *here is a given context*

option A) baseline with the measure noun – *kusok*, *butylka*, *čашka*, *pačka*, etc./*kus*

option B) baseline with the measure word – *štuka*/ *litr*, *kilogram*, *metr*, etc.

option C) lonely word usage (measure noun) - *kusok*, *butylka*, *čашka*, *pačka*, etc./*kus*

option D) lonely word usage (measure word) - *štuka*/ *litr*, *kilogram*, *metr*, etc.

option E) adjectival modification (measure noun) - *kusok*, *butylka*, *čашka*, *pačka*, etc./*kus*

option F) adjectival modification (measure word) - *štuka*/ *litr*, *kilogram*, *metr*, etc.

4.4 Czech questionnaire

4.4.1 Introduction

*Vážení a milí přátelé,
účelem tohoto experimentu je zjistit přijatelnost některých českých vět rodilými mluvčími. Všechny věty jsou uvedeny v kontextu, kde by mohly být proneseny. Klíčové je, abyste při posuzování gramatičnosti a vhodnosti vět vůči kontextu spoléhali pouze na své intuice. Představte si, že se vás ptá např. cizinec, který umí obsojně česky, ale není si jistý, jestli věta, kterou vytvořil, je pro rodilého mluvčího v daném kontextu přijatelná. Vše bude*

vysvětleno v následující cvičné fázi. Žádný z příkladů neinterpretujte jako psaný ironicky.

Dále, i když po vás ve formuláři pod tímto textem požadujeme jméno, celý experiment je zcela anonymní.

Testování začne krátkou fází, ve které je na příkladech vysvětleno, co máte v experimentu dělat. Po dvou příkladech začne samotný experiment.

Translated as following:

Dear friends,
the goal of the present experiment is to test acceptability of several Czech sentences by native speakers. All sentences are given in the context where could be said. The point is that when you judge well-formedness and naturalness of sentences in contexts rely only on your intuition. Imagine that these sentences were made by a foreigner who speaks Czech fluently but not perfect. He or she is not sure whether the native speaker can accept the sentence in the given context. All instructions will be in a practice section. No sentence is written ironically.

Next, even if there is need to put your name below, the experiment is fully anonymous.

The testing will start with the short practicing section. There are two examples which will explain you what you have to do. The experiment itself will follow after them.

4.4.2 Practice section

The first practice sentence:

- Context: *Včera šly Bára a Klára nakupovat oblečení.*
- Sentence: Bára si koupila tři oblečení, Klára si koupila jen jedno oblečení.
- Comment: *Tato věta je pro mluvčího češtiny negramatická, a tím pádem neadekvátní v jakémkoliv (i v tomto) kontextu. Pokud souhlasíte, vyberte jednu z možností v levém konci škály (nejspíš 1).*

Translation:

- Context: *Bára and Klára went shopping yesterday.*
- Sentence: Bára bought three clothes and Klára bought just one clothes.
- Comment: *This sentence is ungrammatical for the native speaker, hence it is inadequate in any context (here as well). If you agree, choose the option in the left end of scale (probably 1).*

The second practice sentence:

- Context: *Marie si včera dala tři sklenice vína.*
- Sentence: Marie pila včera víno.
- Comment: *Tato věta je v daném kontextu pravdivá a je také gramaticky v pořádku. Pokud souhlasíte, vyberte jednu z možností v pravém konci škály (nejspíš 5).*

Translation:

Context: *Marie drank three glasses of wine.*

Sentence: Marie was drinking wine yesterday.

This sentence is true in this context and it is grammatical.

Comment: *If you agree, choose the option in the right end of scale (probably 5).*

After these two examples the experiment starts.

4.4.3 The experimental part

Now I will present two Czech fillers, one is not acceptable and another one is. Then will follow the Czech item with all 6 options.

Filler (bad)

Context: *Eva zaplatila 500 korun za pět stejných triček.*

Sentence: Eva zaplatila 150 korun za jeden kus.

Translation:

Context: *Eva has paid 500 crowns for five identical T-shirts.*

Sentence: Eva has paid 150 crowns for one piece.

The filler is grammatical, but it is not true in the given context, so subjects choose the option in the left end of the scale.

Filler (good)

Context: *Iceberg se kvůli tání odděluje od ledového šelfu
a svobodně plave v oceánu.*

Sentence: Iceberg je obrovský kus ledového šelfu.

Translation:

Context: *Because of melting an iceberg detaches from ice sheet
and swims free in the ocean.*

Sentence: An iceberg is a huge piece of ice sheet.

The sentence is grammatical and true in the given context, so subjects choose the option in the right end.

The full list of fillers is in *chapter B Czech fillers*

Item

- Context: *Každou neděli Lenka své rodině připraví palačinky s jahodami.*
- option A) Lenka potřebuje 10 až 15 kusů jahod, aby ozdobila palačinky.
- option B) Lenka potřebuje 300 gramů jahod, aby ozdobila palačinky.
- option C) Lenka krájí jahody na kusy a zdobí jimi palačinky.
- option D) Lenka krájí jahody na gramy a zdobí jimi palačinky.
- option E) Lenka pro svou rodinu vybírá jenom nejkvalitnější kusy jahod.
- option F) Lenka pro svou rodinu vybírá jenom nejkvalitnější gramy jahod.

Translation:

- Context: *Lenka makes pancakes with strawberries to her family every Sunday.*
- option A) Lenka uses 10 to 15 pieces of strawberry for decorating pancakes.
- option B) Lenka uses 300 grams of strawberry for decorating pancakes.
- option C) Lenka cuts strawberries in pieces and decorates pancakes.
- option D) Lenka cuts strawberries in grams and decorates pancakes.
- option E) Lenka chooses the finest pieces of strawberries for her family.
- option F) Lenka chooses the finest grams of strawberries for her family.

The full list of items is in *chapter A Czech items*

4.5 Russian questionnaire

4.5.1 Introduction

*Дорогие друзья,
в этом эксперименте я тестирую некоторое количество русских предложений на их приемлемость в определённых контекстах. От вас не требуются глубокие познания русского языка, орфографии или пунктуации, полагайтесь только на свою интуицию и чувство языка. Представьте, что эти предложения сочинил иностранец. Этот иностранец довольно неплохо разговаривает на русском, но он не уверен, если составил предложение правильно, и может ли носитель русского языка в данном контексте произнести его. Критерии оценки будут объяснены далее в тренировочной части. Ни одно из предложений не несёт ироничный или саркастический характер.
Ниже, пожалуйста, напишите своё имя для того, чтобы я могла обработать данные. Целый эксперимент полностью анонимный, его результаты будут использованы для моей бакалаврской работы.
Тестирование начнётся с короткой тренировочной части, в которой на примерах будет показано, что от вас требуется. После трёх тренировочных предложений начнётся сам эксперимент.
Огромное спасибо!*

The introduction of the Russian experiment corresponds to the Czech one, except I had changed the number of the practice sentences from two to three.

4.5.2 Practice section

The first practice sentence:

Context: Светлана купила себе золотые украшения.
Она купила золотой браслет и золотое кольцо.

Sentence: Светлана купила несколько зóлот.

Comment: *Носитель русского языка не сказал бы это предложение, так как оно неграмматичное. В этом и в любом другом контексте данное предложение произнести невозможно. Если вы согласны, нажмите на кнопку в левой части шкалы (вероятнее всего 1)*

Translation:

Context: *Svetlana has bought golden jewelry. She has bought a golden bangle and a golden ring.*

Sentence: Svetlana has bought some golds.

Comment: *This sentence is ungrammatical for the native speaker of Russian. It is inadequate here and in any context. If you agree, choose the option in the left end of scale (probably 1).*

The second practice sentence:

Context: Антон выпил вчера три кружки пива.

Sentence: Антон вчера пил пиво.

Comment: *Это предложение истинно в данном контексте и грамматически правильно составлено. Если вы согласны, нажмите кнопку в правой части шкалы (вероятнее всего 5).*

Translation:

Context: *Anton drank three glasses of beer yesterday.*

Sentence: Anton drank beer yesterday.

Comment: *This sentence is true in this context and it is grammatical. If you agree, choose the option in the right end of scale (probably 5).*

The third practice sentence:

Context *Токио населяет более 9 млн. человек, население Нью-Йорка составляет 8,6 млн.*

Sentence: В Нью-Йорке живёт больше людей, чем в Токио.

Comment: *Не смотря на то, что предложение построено грамматически правильно, в данном контексте оно не истинно. Если вы согласны, нажмите на кнопку в левой части шкалы (вероятнее всего 1).*

Translation:

Context: *The population of Tokyo is over 9 million people, the population of New York is 8.6 million.*

Sentence: Far more people live in New Your than in Tokyo.

Comment: *Even if the sentence is grammatically correct, it is false in the given context. If you agree, choose the option in the left end of scale (probably 1).*

The end of the practice section.

4.5.3 The experimental part

Now I will present two fillers and one item from Russian experiment.

Filler (bad)

Context: *Швея получила заказ на отделку платья лентами разных размеров и цветов.*

Sentence: Швея нарезала средние сантиметры от красной ленты и пришила к платью.

Translation:

Context: *The seamstress has received an order for finishing the dress with ribbons of different sizes and colors.*

Sentence: The seamstress has cut middle centimeters from the red ribbon and sewed to the dress.

This sentence is a nonsense, so subjects choose the option in the left end of the scale.

Filler (good)

Context: *Андрей пьёт чай только с сахаром.*

Sentence: Андрей кладёт две ложки сахара в одну чашку чая.

Translation:

Context: *Andrej drinks tea with sugar only.*

Sentence: Andrej puts two spoons of sugar in one cup of tea.

This sentence is fine and true in the given context, so subjects choose the option in the right.

The full list of Russian fillers is in *chapter D Russian fillers*

Item

Контекст:	<i>Мама принесла из магазина сыр разных сортов.</i>
option A)	Мама принесла три кусочка сыра.
option B)	Мама принесла три штуки сыра.
option C)	Мама нарезала сыр, который принесла, на кусочки.
option D)	Мама нарезала сыр, который принесла, на штуки.
option E)	Мама принесла три небольших кусочка сыра.
option F)	Мама принесла три небольших штуки сыра.

Translation:

Context:	<i>Mom has brought different types of cheese.</i>
option A)	Mom has brought three pieces of cheese.
option B)	Mom has brought three items of cheese.
option C)	Mom has cut cheese in pieces.
option D)	Mom has cut cheese in items.
option E)	Mom has brought three small pieces of cheese.
option F)	Mom has brought three small items of cheese.

The full list of Russian items is in *chapter C Russian items*

4.6 Results

The experiment has proved the hypothesis (52) from section 4.1, *kus* is a lexical noun, i.e. when it is used as a classifier, it is of 'a measure noun' type, while *štuka* is a measure word analogous to *liter*. Next will follow the summaries of data sets, i.e. **descriptive statistics**, of Czech and Russian versions of the experiment.

4.6.1 Czech

Starting with the numbers it is important to explain their meaning³. **The mean** is the simple average, the central value. For the Czech case the means are in figure 4.3.

3. basic statistical terminology from Cumming and Calin-Jageman (2016)

##	Scenario	Means
## 1	adj_mn	2.921348
## 2	adj_mw	1.648352
## 3	baseline_mn	3.166667
## 4	baseline_mw	2.855556
## 5	lonely_mn	3.163043
## 6	lonely_mw	1.684783

Figure 4.3: The means of the Czech conditions

So, the mean of the condition `adj_mn` (the adjectival modification of the measure noun *kus*) is 2.921348. It is higher than the mean of the same condition but with the measure word, `adj_mw`, its mean is 1.648352. From figure 4.3 follows that all three conditions with the measure noun *kus* have higher means than the conditions with the measure words.

The next useful number is **the median**, i.e. the value below and above which the half of the data points lie. For the Czech medians see figure 4.4.

##	Scenario	Medians
## 1	adj_mn	3
## 2	adj_mw	1
## 3	baseline_mn	3
## 4	baseline_mw	3
## 5	lonely_mn	3
## 6	lonely_mw	1

Figure 4.4: The medians of the Czech conditions

It was expected that conditions `baseline_mn` and `baseline_mw` would have higher medians, in other words that subjects would mark them as *completely acceptable* but they did not. Still they have higher value than `adj_mw` and `lonely_mw`, so it is clear that subjects liked `baseline_mn` and `baseline_mw` more.

To visualize the given numbers there is **a box plot graph** in figure 4.5 which shows subjects' answers. On the x axis there are six options of three conditions, on the y axis there is the Likert scale. The

boxes are vertical, so it is possible to compare conditions to each other. The dots inside the boxes are the means, the thick lines are the medians. The boxes themselves show the lowest given answer (down part of the box before the median) and the highest (up part of the box after the median).

The vertical lines are so-called **whiskers** and they report that some subjects had selected these answers but still the majority had chosen something from the box. The dots outside the boxes are **outliers**, i.e. answers which are distant from the main data. The outliers simply say that there was a subject which assessed the sentence as *completely acceptable* while the majority did not like it at all.

Analyzing e.g. the condition `adj_mw` which is the second box we can see the following data. The median is 1, the mean is a little higher than 1.5. The majority of subjects marked sentences with the condition as *not acceptable* which is 1 or 2 on the scale. But there were respondents which gave 3 to these sentences and some outliers who accepted them.

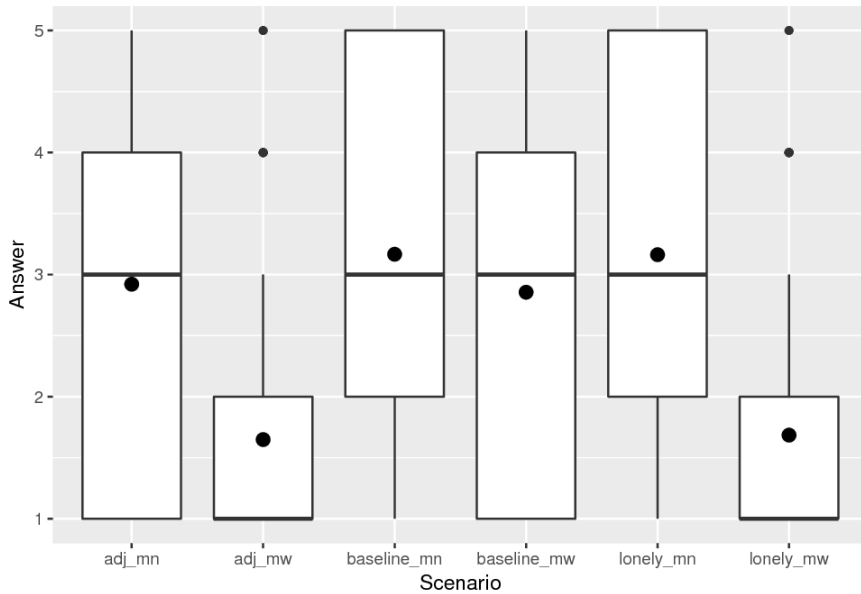


Figure 4.5: The box plot graph of Czech

For the condition `baseline_mn`, the median is 3, the mean is a bit more than 3, from figure 4.3 we know it is 3.166667. There were sub-

jects who rated sentences with this condition as *not acceptable*, but the majority gave from 2 to 5 on the scale.

The second graph in figure 4.6 says that if we run the same experiment again with other subjects, it would be different from the present this much. This graph does not show the real variation between respondents, but an expected variation between experiments. The error bars in this case represent **standard deviation**, a measure of the spread of data points. On the y axis of this graph there are numbers from 1 to 3, because standard deviation represents how far from the mean the typical data lie, and the highest means of Czech conditions are a bit more than 3.

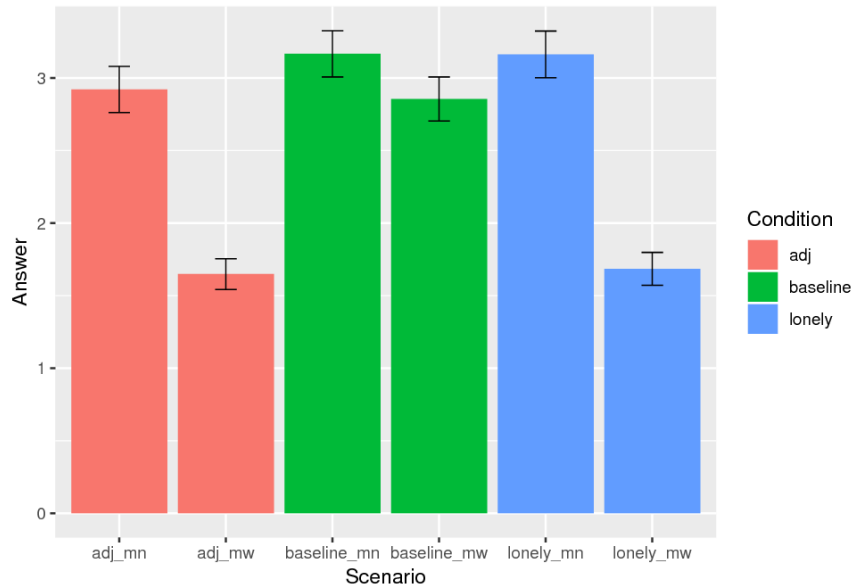


Figure 4.6: Standard deviation of the Czech version of the experiment

However, to test all data above, three mixed linear model were constructed in R. They tested if the answers can be predicted from the conditions (fixed effects) and whether the conditions are statistically significantly different or are parts of the static. Each model had one predictor, i.e. **reference level condition**: `baseline_mn`, `adj_mn` and `lonely_mn`. As far as subjects may or may not have acquiescence bias,

i.e. the tendency to respond *yes* or in this case *completely acceptable*, standard deviation of subjects' answers for each model is 0.5902. Moreover, not all tested items might have been constructed good, so standard deviation of items is 0.3212.

The reports of whether fixed effects were significant or not will follow, in brackets there is the reference level condition.

```
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      3.15659    0.18935  68.78179  16.670 < 2e-16 ***
## Scenarioadj_mn    -0.22740    0.18322  522.62920  -1.241    0.215
## Scenarioadj_mw    -1.54506    0.18172  521.91457  -8.503 < 2e-16 ***
## Scenariobaseline_mw -0.26682    0.18240  522.23670  -1.463    0.144
## Scenariolonely_mn   0.03664    0.18120  521.88301   0.202    0.840
## Scenariolonely_mw  -1.49612    0.17871  512.62911  -8.372 5.45e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.7: The report of the mixed linear model 1 (baseline_mn)

The first model presents how statistically far are fixed conditions from the referenced condition `baseline_mn`, that is why there are only five scenarios in the report. The last column $\text{Pr}(>|t|)$ is valid for us. If '.', '**', '***', or '****' is present alongside the number in this column, it means that the condition is statistically significant, and with this probability it is possible that the reference condition and the fixed condition are in statistical noise, in other words if they are random irregularities. Again, how small is the number in the column that big is statistical significance between the referenced conditional and the fixed conditional.

So, from the investigation of a correlation between `baseline_mw` and `baseline_mn`, which is 0.144 according to model 1, it follows that `baseline_mw` is worse than `baseline_mn`, but both they are in statistical noise. The probability is not too small, therefore no symbol from above is present. From it follows that there is no statistical difference between them, how it was predicted.

If we consider another correlation with '***', for example the given reference `baseline_mn` and the fixed `lonely_mw`, it simply means that the statistical difference between them is significant. With the probability $<5.45 \cdot 10^{-16}$ it is possible that these two conditions are random

4. The number is really small, it is $5.45 \cdot 10^{-16}$, or 0.000000000000000545

irregularities. This is very strong evidence, it shows how baseline sentences with *kus* being better accepted than sentences with lonely measure words *gram*, *kilogram*, *litr*, *etc.* and it is not a coincidence.

To sum up results for model 1, there are two statistically significantly different conditions against the reference one, they are `adj_mw` and `lonely_mw`. However, the conditions `adj_mn`, `baseline_mw` and `lonely_mn` are in statistical noise, how it was expected.

The model 2 has the reference condition `adj_mn` and it is significantly different from the related fixed condition `adj_mw`, the probability is $3.50e-12$. The number is bigger than the previous one, there are twelve zeros after decimal point, but still it is not big enough to be in statistical noise.

```
...
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      2.92920    0.19254  73.29035  15.214 < 2e-16 ***
## Scenariobaseline_mn  0.22740    0.18322 522.62920   1.241  0.215
## Scenarioadj_mw     -1.31767    0.18496 521.52881  -7.124 3.50e-12 ***
## Scenariobaseline_mw -0.03942    0.18288 511.88922  -0.216  0.829
## Scenariolonely_mn   0.26403    0.18464 521.85014   1.430  0.153
## Scenariolonely_mw  -1.26872    0.18454 521.68583  -6.875 1.77e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.8: The report of the mixed linear model 2 (`adj_mn`)

The model 3 has the reference condition `lonely_mn`. The difference between it and the related fixed condition `lonely_mw` is $4.83e-16$, so it is significant. However, if two previous reference conditions were in statistical noise with the fixed condition `baseline_mw`, this one is significant against it. Their difference is smaller than between predicted ones, but `'.'` is present alongside the number 0.0996 .

```
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      3.19323    0.19097  71.13572  16.721 < 2e-16 ***
## Scenarioadj_mn     -0.26403    0.18464 521.85014  -1.430  0.1533
## Scenariobaseline_mn -0.03664    0.18120 521.88301  -0.202  0.8398
## Scenarioadj_mw     -1.58170    0.18089 511.78490  -8.744 < 2e-16 ***
## Scenariobaseline_mw -0.30345    0.18393 521.48897  -1.650  0.0996 .
## Scenariolonely_mw  -1.53275    0.18284 521.30185  -8.383 4.83e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.9: The report of the mixed linear model 3 (`lonely_mn`)

To sum up, the description statistic of Czech has proved that classifier *kus* is a lexical noun because it does not behave like measure words *gram*, *kilogram*, *litr*, *etc.* which are less acceptable and it is not a coincidence according to there mixed linear modes above. They are not felicitous without any number (*lonely_mw* condition) and cannot be modified by adjectives (*adj_mw* condition) while *kus* is fine in both.

4.6.2 Russian

Unlike Czech, Russian classifier *štuka* was opposed to measure nouns *kusok*, *butylka*, *čашka*, *pačka*, *etc.* which are lexical nouns. If it was compared to the proper measure word e.g. *litr*, *kilogram*, *metr*, *etc.*, values of all conditions would be low, irrelevant and useless. As far as descriptive statistics of Russian shows that *štuka* does not behave as a measure noun, the experiment is valid.

In figure 4.10 there are means and medians⁵ of Russian conditions. The means are a bit different but basically correspond to the Czech ones. The medians of measure nouns condition seem to be higher, so Russian subjects liked sentences with them more then Czech ones with *kus*. In the end, the pattern of two versions seems to be the same: subjects accepted conditions with measure nouns and do not accepted measure words.

##	Scenario	Means	Medians
## 1	adj_mn	3.698529	5
## 2	adj_mw	1.632353	1
## 3	baseline_mn	3.835714	5
## 4	baseline_mw	2.014815	1
## 5	lonely_mn	3.482270	4
## 6	lonely_mw	1.503650	1

Figure 4.10: The means and medians of Russian

5. All basic terminology was explained in *section 4.6.1 Czech*, it will not be repeated here, so if it is necessary, see *section 4.6.1*.

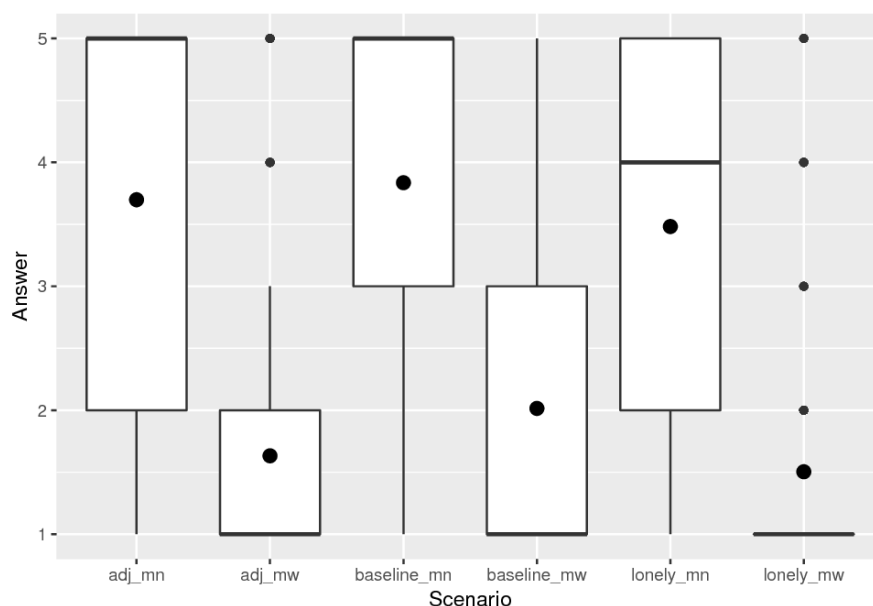


Figure 4.11: The box plot graph of Russian

The box plot graph of Russian is in figure 4.11. It was predicted that subjects would not accept sentences with *štuka*, so it did happen. Conditions `adj_mw` and `lonely_mw` were worse accepted than `baseline_mw` and it was predicted as well. It is interesting that Russian subjects did not like `lonely_mw` even more than subjects in Czech version. But according to whiskers and outliers there were subjects who more or less accepted conditions `adj_mw` and `lonely_mw`.

Even if the general patterns of the box plot graphs are not the same for Czech and Russian, e.g. there is no box in the condition `lonely_mw`, or the box of Russian condition `adj_mn` is higher than the Czech one, Russian medians are in some cases higher as well, exactly this result was expected. Subjects in both cases accepted measure nouns better than measure words and did not accepted measure words.

In figure 4.12 is standard deviation of the Russian experiment. Again, this graph does not report how the subjects reacted to conditions, it shows how the possible experiment may differ from the present one. The error bars seem to be shorter than the same Czech bars, perhaps because of the participants number, in Russian version there were more of them.

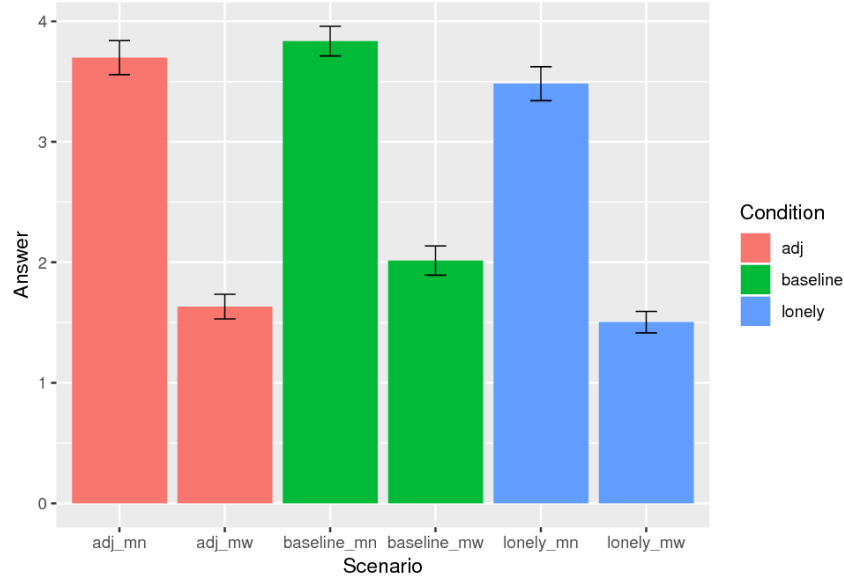


Figure 4.12: Standard deviation of the Russian version of the experiment

As well as in the Czech version, here are three mixed linear models in R with the same reference level conditions: `baseline_mn`, `adj_mn` and `lonely_mn`. Standard deviation of subjects is 0.6101, standard deviation of items is 0.1687.

In all of them conditions with measure word *štuka* are statistically significantly different and the probability of their random irregularity is the same $2e-16$. This was expected.

```
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    3.8353    0.1556  97.1100  24.649  <2e-16 ***
## Scenarioadj_mn  -0.1217    0.1546  780.8265  -0.788  0.4312
## Scenarioadj_mw  -2.2098    0.1552  787.2495 -14.242  <2e-16 ***
## Scenariobaseline_mw -1.8334    0.1555  787.2847 -11.793  <2e-16 ***
## Scenariolonely_mn  -0.3369    0.1532  779.9099  -2.199  0.0282 *
## Scenariolonely_mw  -2.3431    0.1546  784.9579 -15.160  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.13: The report of the mixed linear model 1 (`baseline_mn`)

```
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    3.7135     0.1567   99.7311  23.701   <2e-16 ***
## Scenariobaseline_mn  0.1217     0.1546  780.8265   0.788     0.431
## Scenarioadj_mw    -2.0881     0.1562  787.0257 -13.367   <2e-16 ***
## Scenariobaseline_mw -1.7117     0.1562  784.7009 -10.959   <2e-16 ***
## Scenariolonely_mn  -0.2151     0.1543  781.2929  -1.394     0.164
## Scenariolonely_mw  -2.2214     0.1559  787.0517 -14.245   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.14: The report of the mixed linear model 2 (adj_mn)

```
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    3.4984     0.1553   96.4811  22.520   <2e-16 ***
## Scenarioadj_mn   0.2151     0.1543  781.2929   1.394     0.1638
## Scenariobaseline_mn  0.3369     0.1532  779.9099   2.199     0.0282 *
## Scenarioadj_mw   -1.8730     0.1546  785.4678 -12.114   <2e-16 ***
## Scenariobaseline_mw -1.4966     0.1552  787.3488  -9.641   <2e-16 ***
## Scenariolonely_mw  -2.0063     0.1546  787.2875 -12.976   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.15: The report of the mixed linear model 3 (lonely_mn)

What is rather unexpected that in model 1 fixed condition lonely_mn is significant to reference baseline_mn. The same probability 0.0282 is in model 3 which is basically the reflection of model 1, reference condition is lonely_mn and fixed is baseline_mn. Model 2 does not report any unexpected data.

To sum up, the description statistic of Russian has proved that *štuka* does not behave like a regular lexical word. It is not acceptable alone without any number (lonely_mw condition) and it cannot stand the adjectival modification (adj_mw condition). Moreover, it is not acceptable in baseline_mw condition⁶, I will try to explain it in *section 4.6.3 Discussion*. The fact is that *kus* is fine in these conditions while *štuka* is not.

4.6.3 Discussion

The main result of the experiment is that the hypothesis in (52) was proved, *kus* and *štuka* are different terms. It would not be proved, if

6. Czech median of condition baseline_mw is 3, while Russian is 1

Table 4.1: The means and medians of *kus* and *štuka*

means		
	<i>kus</i> (mn)	<i>štuka</i> (mw)
baseline	3.166667	2.014815
lonely	3.163043	1.503650
adj	2.921348	1.632353
medians		
	<i>kus</i> (mn)	<i>štuka</i> (mw)
baseline	3	1
lonely	3	1
adj	3	1

štuka had the same acceptability as measure nouns in Russian version of the experiment, or if *kus* behaved like measure words in Czech version. In table 4.1 there are final means and medians of two investigated classifiers, and it shows how far these terms are from each other.

The only thing that was unexpected in Czech version is acceptability of baseline condition. It was predicted that the condition should have been *completely acceptable* but it has the median 3 in the end. But the main goal of the experiment was achieved: *kus* does not behave like *gram*, *metr*, *litr*, *ect.*, it can stand adjectival modification and it is fine alone.

In case of Russian the experiment has proved that *štuka* is not a lexical term because it cannot be modified by adjectives and be alone. But unlike Czech, *štuka* was not acceptable at all in baseline condition. I assume it is because of its relation with mass nouns and fake mass nouns.

Lexical terms *čашka*, *butylka*, *ložka*, *bokal* 'cup, bottle, spoon, wine glass' seem to be used as classifiers with canonical mass nouns (53). Lexical terms as *korobka*, *pačka*, *upakovka* 'box, pack, packaging' can be found alongside plural count nouns, mass nouns and fake mass nouns at the same time (54). *Kusok* 'piece' seems to accept mass nouns and singular nouns only (55a), while its plural counterpart *kuski* is felicitous with singulars, plurals, mass, and fake mass (55b). And *štuka* can be with plural count nouns and fake mass nouns only. I used all

of them without this particular distribution, that is why *štuka* was so unacceptable.

- (53) a. *čáška/butylka/ložka/?bokal moloka*
a cup/bottle/spoon/?wine glass of milk
- b. *?čáška/butylka/ložka/bokal vina*
a ?cup/bottle/spoon/wine glass of wine
- (54) a. *korobka konfet/vina/knig/odeždy*
a box of candies/wine/books/clothes
- b. *pačka moloka/salfetok/pečenja*
a pack of milk/wipes/biscuits
- (55) a. *kusok mjaso/futbolki/tkani*
a piece of meat/t-shirt/fabric
- b. *kuski mjaso/futbolki/futbolok/mebeli/tkani*
pieces of meat/t-shirt/t-shirts/furniture/fabric

The point is that if I had an opportunity to construct the second Russian experiment but with this distribution, it would be clear for sure that *štuka* is not a lexical noun. Since the comparison of Czech and Russian classifiers *kus* and *štuka* was the purpose of the present experiment, the Russian version was needed to be made in the way how it was made.

Conclusion For the experiment I constructed two questionnaires, Czech and Russian. According to the answers from 55 subjects, two descriptive statistics and six linear mixed models were created. The data have proved the hypothesis in (52) that Czech and Russian terms are different classifiers, *kus* is a measure nouns because it a lexical word, and *štuka* is a measure word analogous to *liter*.

5 Summary

The goal of the present BA thesis was the study of classifiers. In the first chapter I described mass/count distinction which is inseparable with classifiers. There were examples from different types of languages presenting how mass/count distinction reflects subject/object relation in real life. There were three properties of mass nouns and the first mention of classifiers themselves. Classifiers, however, were described in classifier languages where they play the essential role because they belong to the separate functional category, to the closed class of terms. In non-classifier languages there is term measure phrase which does almost the same work but does not belong to the closed class and can be used as a lexical noun, i.e. it is not a part of a special function category. In the end of the first chapter there was described an interesting phenomenon as fake mass nouns.

If the first chapter just mentioned necessity of classifiers whether in classifier languages or in mass noun constructions, in the second chapter there were three theories which tried to explain why they are obligatory, or in other words why mass nouns are distinct from count. The first was Godehard Link and his the lattice theory which claims that unlike count nouns, mass nouns do not possess the bottom layer and because of it classifiers and measure constructions must 'help' them to be felicitous with numbers. Gennaro Chierchia argued that through the concept of supervaluation vagueness can be integrated in singular/plural structures and it may help to catch mass/count distinction. According to Manfred Krifka, who was the author of the third theory, the difference between mass and count nouns is in the OU/KU in classifier languages and the OKU classifiers in non-classifier languages. Keren Khrizman applied Krifka's assumption to Russian and presented classifiers *štuka*, *čelovek* and *golova* which are described in the third chapter.

The main idea of the Khrizman's paper is that *štuka*, *čelovek* and *golova* behave like Mandarin classifiers and are parts of separate functional category. Hence, they are not regular classifiers or measure nouns but measure words analogous to liter. In the third chapter I retold Khrizman (2016) and evidence from Russian.

The last but not least forth chapter was about the experiment I run. As far as I can call myself a fluent speaker of Czech and the native speaker of Russian, there was an opportunity to compare two related languages and see how Czech deals with classifier *kus* which at first sight seems to be a simple translation of Russian *štuka*. I have constructed two versions of the experiment and via online platform Ibex Farm collected data from native speakers. The results were statistically analyzed and revealed that Czech *kus* is not a measure word like *štuka* but is a measure noun.

And almost at the end of the thesis I would like to share some observations from two other languages. According to Vasmer, *štuka* was borrowed from Polish *sztuka*, while Polish term comes from Middle High German *stücke* (Sussex, 1976). In German and in Polish they use these terms in the same contexts as in (47)-(48), see (56)-(57).

(56) a. Sensitive Feuchttücher Vorteilspack 56 Stück.

b. Eier aus Bodenhaltung 10 Stück.

(57) a. Chusteczki nawilżane 120 sztuk.

b. Jaja polskie L 10 sztuk.

From it follows the question, do Polish *sztuka* and German *stücke* behave like Czech *kus* or more like Russian *štuka*? Unfortunately, I do not speak Polish or German fluently, but as far as I know it is possible to see *Muzeum Sztuki Nowoczesnej w Warszawie*¹ on the streets of Warsaw. Hence, I can assume that *sztuka* can be used as the lexical word and can stand adjectival modification, exactly like *kus*.

The sentences in (58) were found via Google on German sites, so I believe they are well-formed. In (58a) *Stücke* is a lexical noun, if the translation is right. In (58b) it is adjectivally modified, so as well as Polish *sztuka*, German *stücke* behaves like *kus*.

1. Translated as 'The museum of Modern Art in Warsaw'.

- (58) a. *Brich den Stock in Stücke und schmeiß ihn ins Feuer.*
Break the stick into pieces and throw it into the fire.
- b. *Die Lebensmittel werden gleichmäßig gegart, wenn sie in
ähnlich große Stücke geschnitten werden.*
The vegetables will be evenly cooked if you cut them into
similar-sized pieces.

So, if the first question was (perhaps) answered, there are two new. First, why does Russian *štuka* have Polish and German roots but behaves like something completely different? And the second one, why does Czech *kus* behave like German and Polish terms but has a different form? Well, I cannot answer these questions because they go far beyond the thesis, but I can provide with information from the etymological dictionary of Czech. According to Rejzek (2001) *kus* was derived from a verb *kousat* 'to bite' and has connections with *kousek* 'piece'.

kus, *kousek*, *kousíček*, *kusý*, *kusový*, *kusanec*, *kouskovat*, *rozkouskovat*. Věsl. - p. *kęs*, r. *kus(ók)*, s./ch. *kûs*. Psl. **kqsъ* je odvozeno od **kqsati* (→ *kousat*), původně tedy vlastně 'ukousnutý díl'. Srov. i angl. *bit* 'kousek' od *bite* 'kousat'

Conclusion It was a great pleasure to work on the thesis. I have read relevant and reliable theories about mass/count distinction (Link (1983); Krifka (1995); Chierchia (2010)) and the practical application of one of them onto Russian (Khrizman, 2016). I have learned how to edit the text in \LaTeX environment and how to work with Ibex Farm. Now I know far more about statistics than I knew before. And after the experiment I understand a little better how my native Russian and so beloved Czech work.

Bibliography

- [1] Keith Allan. "Classifiers". In: *Language* 53.2 (1977), pp. 285–311.
- [2] Lucas Champollion. *Distributivity, collectivity and cumulativity*. 2014.
- [3] Gennaro Chierchia. "Mass nouns, vagueness and semantic variation". In: *Synthese* 174.1 (2010), pp. 99–149.
- [4] Geoff Cumming and Robert Calin-Jageman. *Introduction to the new statistics: Estimation, open science, and beyond*. Routledge, 2016.
- [5] Mojmír Dočekal. POČITATELNÉ A NEPOČITATELNÉ JMÉNO. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*. 2017. URL: <https://www.czechency.org/slovník/P0%C4%8CITATELN%C3%89%20A%20NEP0%C4%8CITATELN%C3%89%20JM%C3%89N0>.
- [6] Jenny Doetjes. "Count/Mass Distinction Across Languages". In: *Semantics: An International Handbook of Natural Language Meaning*. de Gruyter, 2012. Chap. 96, pp. 2559–2580.
- [7] Scott Grimm. "Number and Individuation". PhD thesis. Stanford University, 2012.
- [8] Keren Khrizman. "Functional Unit Classifiers in (Non)-Classifier Russian". In: *Baltic International Yearbook of Cognition, Logic and Communication* 11.1 (2016), p. 5.
- [9] Manfred Krifka. "11 COMMON NOUNS: A CONTRASTIVE ANALYSIS OF CHINESE AND ENGLISH". In: *The generic book* (1995), p. 398.
- [10] Manfred Krifka. "Nominal reference, temporal constitution and quantification in event semantics". In: *Semantics and contextual expression* 75 (1989), p. 115.
- [11] Peter Lasnik. "Mass Nouns and Plurals". In: *Semantics: An International Handbook of Natural Language Meaning*. de Gruyter, 2011. Chap. 46, pp. 1131–1153.
- [12] Godehard Link. "The Logical Analysis of Plural and Mass Terms: a Lattice-theoretical Approach". In: *Meaning, Use, and Interpretation of Language*. de Gruyter, 1983, pp. 302–332.
- [13] P.H. Portner. *What is Meaning: Fundamentals of Formal Semantics*. Fundamentals of Linguistics. Wiley, 2005. ISBN: 9781405109185. URL: <https://books.google.cz/books?id=-RdrQgAACAAJ>.

BIBLIOGRAPHY

- [14] Jiří Rejzek. *Český etymologický slovník*. nakladatelství LEDA, 2001.
- [15] Roland Sussex. "The numeral classifiers of Russian". In: *Russian linguistics* (1976), pp. 145–155.
- [16] Josef Vlček. *Russko-český slovar*. Russkij jazyk, 1974.

Appendices

A Czech items

Item 1

Kontext: Děti si hrály s Harry Potterem sestaveným z lega, upadl jim a rozbil se.

baseline_mn	HP byl sestaven ze 125 kusů lega.
baseline_mw	HP byl sestaven z 300 gramů lega.
lonely_mn	HP z lega se rozpadl na kusy.
lonely_mw	HP z lega se rozpadl na gramy.
adj_mn	HP byl složen z různých kusů lega.
adj_mw	HP byl složen z různých gramů lega.

Item 2

Kontext: K čaji po obědě babička upekla kilo cukroví, naskládala je na jeden velký a jeden malý talíř, pak nabídla rodině.

baseline_mn	Táta si dal 11 kusů, máma si dala jen 5 kusů.
baseline_mw	Táta si dal 220 gramů, máma si dala jen 100 gramů.
lonely_mn	Táta a máma si dali kusy z malého talíře.
lonely_mw	Táta a máma si dali gramy z malého talíře.
adj_mn	Vnoučata si vzala pěkné kusy z velkého talíře.
adj_mw	Vnoučata si vzala pěkné gramy z velkého talíře.

Item 3

Kontext: Paní Nováková pořídila dvě balení různých bonbónů, aby motivovala děti k lepšímu prospěchu ve škole. Míša a Zdena je dostávali jako odměnu za dobré známky.

baseline_mn	Na konci týdne měl Míša 13 kusů a Zdena měla 9 kusů.
baseline_mw	Na konci týdne měl Míša 159 gramů a Zdena měla 110 gramů.
lonely_mn	Na konci týdne Míša měl víc kusů než Zdena.
lonely_mw	Na konci týdne Míša měl víc gramů než Zdena.
adj_mn	Za každou jedničku Míša a Zdena dostali čokoládový kus, za dvojku karamelový.
adj_mw	Za každou jedničku Míša a Zdena získávali čokoládový gram, za dvojku karamelový.

Item 4

Kontext: Koláč byl šťavnatý a chutný, protože Jana nešetřila jablky, když ho pekla.

baseline_mn	Na jeho přípravu nachystala 10 kusů jablek.
baseline_mw	Na jeho přípravu nachystala 1500 gramů jablek.
lonely_mn	Jablka omyla vodou a nakrájela na kusy.
lonely_mw	Jablka omyla vodou a nakrájela na gramy.
adj_mn	Koláč obsahoval drobně nakrájené kusy jablek.
adj_mw	Koláč obsahoval drobně nakrájené gramy jablek.

Item 5

Kontext: V ateliéru si Tereza objednala šaty na promoci, potřebovala několik metrů černé a vínově červené látky.

baseline_mn	Po zhotovení šatů Tereze zbyly dva kusy červené látky a jeden kus vínově červené látky.
baseline_mw	Po zhotovení šatů Tereze zbyly dva metry červené látky a jeden metr vínově červené látky.
lonely_mn	Po zhotovení šatů Tereze zbyly kusy látky.
lonely_mw	Po zhotovení šatů Tereze zbyly metry látky.
adj_mn	Po zhotovení šatů by Tereza mohla ze zbylých kusů látky ušít další šaty.
adj_mw	Po zhotovení šatů by Tereza mohla ze zbylých metrů látky ušít další šaty.

Item 6

Kontext: Firma ABC si objednala v tiskárně několik tisíc akčních letáku o rozměrech 21x14,8 cm.

baseline_mn	Celkem objednala 2000 kusů.
baseline_mw	Celkem objednala 42x28 metrů.
lonely_mn	Kusy bude umisťovat ve městě s pomocí brigádníků.
lonely_mw	Centimetry bude umisťovat ve městě s pomocí brigádníků
adj_mn	Vytiskla slevy a kupony pro své zákazníky na recyklovaných kusech papíru.

adj_mw Vytiskla slevy a kupony pro své zákazníky na recyklovaných centimetrech papíru.

Item 7

Kontext: Lék X léčí akutní respirační onemocnění, balení obsahuje měkké tobolky, 10 kusů po 50 mg.

baseline_mn Balení tohoto léku obsahuje 10 kusů měkkých tobolek.

baseline_mw Balení tohoto léku obsahuje 50 miligramů měkkých tobolek.

lonely_mn Kusy si dávejte po jídle a zapíjejte velkým množstvím vody.

lonely_mw Miligramy si dávejte po jídle a zapíjejte velkým množstvím vody

adj_mn Prošlé kusy v žádném případě nepoužívejte.

adj_mw Prošlé miligramy v žádném případě nepoužívejte.

Item 8

Kontext: Hana se vracela z knihovny s těžkou taškou, ve které nesla knihy.

baseline_mn Hana si půjčila v knihovně sedm kusů knih.

baseline_mw Hana si půjčila v knihovně pět kilogramů knih.

lonely_mn Hana by měla vrátit kusy do knihovny během měsíce.

lonely_mw Hana by měla vrátit kilogramy do knihovny během měsíce.

adj_mn Hana si půjčila krásné kusy knih.

adj_mw Hana si půjčila krásné kilogramy knih.

Item 9

Kontext: Máma si koupila 2 kilogramy brambor.

baseline_mn Na oběd potřebovala jen 9 kusů brambor.

baseline_mw Na oběd potřebovala jen 600 gramů brambor.

lonely_mn Oškrábala brambory a nakrájela je na kusy.

lonely_mw Oškrábala brambory a nakrájela je na gramy.

adj_mn	Brambory oškrábala a nakrájela na malé a velké kusy.
adj_mw	Brambory oškrábala a nakrájela na malé a velké gramy.

Item 10

Kontext: Lucie si koupila několik obrovských a velmi těžkých váz.

baseline_mn	Lucie si koupila 5 kusů obrovských a těžkých váz.
baseline_mw	Lucie si koupila 20 kilogramů obrovských a těžkých váz.
lonely_mn	Luciin pes rozbil jednu z váz na kusy.
lonely_mw	Luciin pes rozbil jednu z váz na gramy.
adj_mn	Luciin pes rozbil dvě vázy na malé kusy.
adj_mw	Luciin pes rozbil dvě vázy na malé gramy.

Item 11

Kontext: Lev konžský, kterého můžeme vidět v brněnské zoo, se živí jenom čerstvým masem.

baseline_mn	Lev konžský jí 5 kusů čerstvého masa během jednoho dne.
baseline_mw	Lev konžský jí 5 kilogramu čerstvého masa během jednoho dne.
lonely_mn	Kusy masa pro lva konžského upravují pracovníci zoo.
lonely_mw	Kilogramy masa pro lva konžského upravují pracovníci zoo.
adj_mn	Lev konžský potřebuje opravdu velké kusy masa, aby se nasytil na celý den.
adj_mw	Lev konžský potřebuje opravdu velké kilogramy masa, aby se nasytil na celý den.

Item 12

Kontext: Dagmar je švadlena a upravuje šaty.

baseline_mn	Dagmar odřízla z modré stuhy 3 kusy a přišla k šatům.
baseline_mw	Dagmar odřízla z modré stuhy 3 centimetry a přišla k šatům.

lonely_mn	Dagmar nařízla modrou stuhu na kusy.
lonely_mw	Dagmar nařízla modrou stuhu na centimetry.
adj_mn	Dagmar odřízla z modré stuhý dlouhé kusy.
adj_mw	Dagmar odřízla z modré stuhý dlouhé centimetry.

Item 13

Kontext: Na prodej do obchodu dovezli španělské vodní melouny.

baseline_mn	Celkem dovezli 150 kusů vodních melounů.
baseline_mw	Celkem dovezli 350 kilogramu vodních melounů.
lonely_mn	Kusy najdete v oddělení ovoce a zeleniny.
lonely_mw	Kilogramy najdete v oddělení ovoce a zeleniny.
adj_mn	Anna si koupila šťavnatý kus vodního melounu.
adj_mw	Anna si koupila šťavnatý kilogram vodního melounu.

Item 14

Kontext: Pavel má rád vejce, ví, že jsou bohatým zdrojem různých živin.

baseline_mn	Pavel denně jí 6 kusů vajec.
baseline_mw	Pavel denně jí 350 gramů vajec.
lonely_mn	Na úpravu kusů Pavel používá několik různých způsobů.
lonely_mw	Na úpravu gramů Pavel používá několik různých způsobů.
adj_mn	Pavel konzumuje jenom kuřecí kusy.
adj_mw	Pavel konzumuje jenom kuřecí gramy.

Item 15

Kontext: Bára upekla 2 druhy muffinů

baseline_mn	Bára upekla 12 kusů muffinů.
baseline_mw	Bára upekla 720 gramů muffinů.
lonely_mn	Bára ozdobila kusy krémem a marcipánem.
lonely_mw	Bára ozdobila gramy krémem a marcipánem.
adj_mn	Bára upekla čokoládové a banánové kusy muffinů.
adj_mw	Bára upekla čokoládové a banánové gramy muffinů.

Item 16

Kontext: Klára pěstuje dýně.

baseline_mn Ze třech malinkých semínek Kláře vyrostlo 5 kusů dýně.

baseline_mw Ze třech malinkých semínek Kláře vyrostlo 12 kilogramů dýně.

lonely_mn Klára pořádně zalévá kusy vodou, protože dýně milují vlhkost.

lonely_mw Klára pořádně zalévá kilogramy vodou, protože dýně milují vlhkost.

adj_mn Z malých semínek Kláře vyrostly obrovské kusy dýně.

adj_mw Z malých semínek Kláře vyrostly obrovské kilogramy dýně.

Item 17

Kontext: Jana připravuje ke snídani ovesnou kaši s banánem.

baseline_mn Jana potřebuje 2 kusy banánů.

baseline_mw Jana potřebuje 200 gramů banánů.

lonely_mn Banány Jana loupá a krájí na kusy.

lonely_mw Banány Jana loupá a krájí na gramy.

adj_mn Banány Jana krájí na malé kusy.

adj_mw Banány Jana krájí na malé gramy.

Item 18

Kontext: Tomáš jedl mléčnou čokoládu.

baseline_mn Tomáš odlomil tři kusy čokolády a nabídnul své sestře.

baseline_mw Tomáš odlomil 50 gramů čokolády a nabídnul své sestře.

lonely_mn Tomáš rozdělil čokoládu na kusy.

lonely_mw Tomáš rozdělil čokoládu na gramy.

adj_mn Tomáš rozdělil čokoládu na malé a velké kusy.

adj_mw Tomáš rozdělil čokoládu na malé a velké gramy.

Item 19

Kontext: Úřad městské části Brno-Bystrc sbírá oblečení pro charitativní šatník.

baseline_mn	Za pár měsíců úřad nasbíral 143 kusů oblečení.
baseline_mw	Za pár měsíců úřad nasbíral 24 kilogramů oblečení.
lonely_mn	Úřad bude darovat kusy oblečení lidem v nouzi.
lonely_mw	Úřad bude darovat kilogramy oblečení lidem v nouzi.
adj_mn	Úřad sbírá použitelné kusy oblečení, které se občanům už nehodí.
adj_mw	Úřad sbírá znovu použitelné kilogramy oblečení, které se občanům už nehodí.

Item 20

Kontext: Rajčatová polévka je sytá a velmi rychlá na uvaření.

baseline_mn	Na přípravu šesti porcí je potřeba nachystat 7 kusů čerstvých rajčat.
baseline_mw	Na přípravu šesti porcí je potřeba nachystat 550 gramů čerstvých rajčat.
lonely_mn	Napřed kusy máme omýt a oloupat.
lonely_mw	Napřed gramy máme omýt a oloupat.
adj_mn	Rajčata máme omýt, oloupat a nakrájet na malé kusy
adj_mw	Rajčata máme omýt, oloupat a nakrájet na malé gramy

Item 21

Kontext: Markéta a Adéla si objednaly pizzu.

baseline_mn	Markéta si dala 3 kusy pizzy, Adéla si dala 2 kusy pizzy.
baseline_mw	Markéta si dala 150 gramů pizzy, Adéla si dala 100 gramů pizzy.
lonely_mn	Markéta a Adéla nakrájely pizzu na kusy.
lonely_mw	Markéta a Adéla nakrájely pizzu na gramy.
adj_mn	Markéta a Adéla snědly velké kusy pizzy.
adj_mw	Markéta a Adéla snědly velké gramy pizzy.

Item 22

Kontext: Při domovní prohlídce policie našla u Adama zakázané věci.

baseline_mn 200 kusů střeliva.	Adam nelegálně uschoval tři pušky a přes
baseline_mw 15 kilogramů střeliva.	Adam nelegálně uschoval tři pušky a přes
lonely_mn truhle.	Adam schoval tři pušky a kusy střeliva v
lonely_mw v truhle.	Adam schoval tři pušky a kilogramy střeliva
adj_mn liva.	Adam schoval tři pušky a kulové kusy střeliva.
adj_mw střeliva.	Adam schoval tři pušky a kulové kilogramy střeliva.

Item 23

Kontext: Na svatbě Novákových byl obrovský dort.

baseline_mn	Ženich pečlivě nakrajel 4 kusy dortu pro rodiče.
baseline_mw	Ženich pečlivě nakrajel 500 gramů dortu pro rodiče.
lonely_mn	Ženich nakrajel dort na kusy.
lonely_mw	Ženich nakrajel dort na gramy.
adj_mn	Ženich nakrajel dort na velké kusy.
adj_mw	Ženich nakrajel dort na velké gramy.

Item 24

Kontext: Každou neděli Lenka své rodině připraví palačinky s jahodami.

baseline_mn dobila palačinky.	Lenka potřebuje 10 až 15 kusů jahod, aby ozdobila palačinky.
baseline_mw palačinky.	Lenka potřebuje 300 gramů jahod, aby ozdobila palačinky.
lonely_mn	Lenka krájí jahody na kusy a zdobí jimi palačinky.
lonely_mw	Lenka krájí jahody na gramy a zdobí jimi palačinky.
adj_mn nější kusy jahod.	Lenka pro svou rodinu vybírá jenom nejkvalitnější kusy jahod.
adj_mw nější gramy jahod.	Lenka pro svou rodinu vybírá jenom nejkvalitnější gramy jahod.

Item 25

Kontext: Petra upekla koláč.

baseline_mn	Petra nakrájela z koláče 8 kusů.
baseline_mw	Petra nakrájela z koláče 400 gramů.
lonely_mn	Petra nakrájela koláč na kusy.
lonely_mw	Petra nakrájela koláč na gramy.
adj_mn	Petra nakrájela koláč na malé kusy.
adj_mw	Petra nakrájela koláč na malé gramy.

B Czech fillers

Filler 1 (good)

Kontext: Naplánovali jsme cestu pěšky z Brna do Vídně. Když jsme večer dorazili do Břeclavi, měli jsme už slušný kus cesty za sebou.

Šli jsme do Vídně z Brna přes Břeclav.

Filler 2 (bad)

Kontext: František si dal 5 kusů dortů.

František si dal sudý počet kusů.

Filler 3 (good)

Kontext: Jakub měří 180 centimetrů, jeho bratr Filip měří 187 centimetrů.

Filip je o 7 centimetrů vyšší než Jakub.

Filler 4 (bad)

Kontext: Jan stihl přednést na semináři jen část referátu.

Jan přednesl celý referát.

Filler 5 (good)

Kontext: V roce 2014 město Brno získalo do vily Tugendhat bufet, lavice, stolek a lehátko.

V roce 2014 město Brno získalo do vily Tugendhat čtyři originální kusy nábytku.

Filler 6 (bad)

Kontext: Kluci si koupili 1,5 litru limonády.

Kluci si koupili několik litrů limonády.

Filler 7 (good)

Kontext: V bývalých výrobních halách koncernu Škoda je dnes muzeum.

V muzeu si můžeme prohlížet a porovnávat staré a nové kusy aut.

Filler 8 (bad)

Kontext: Eva zaplatila 500 korun za pět stejných triček.

Eva zaplatila 150 korun za jeden kus.

Filler 9 (good)

Kontext: Výška hory Mont Blanc je 4810 metrů, výška hory Everest je 8848 metrů.

Everest je vyšší než Mont Blanc.

Filler 10 (bad)

Kontext: Karolína porodila syna. Syn váží 3 kilogramy.

Syn váží několik gramů.

Filler 11 (good)

Kontext: Pomeranč je těžší než jablko.

Pomeranč váží 200 gramů, jablko váží 140 gramů.

Filler 12 (bad)

Kontext: Rajče, okurka a paprika dohromady váží 700 gramů.

Paprika je o 140 gramů těžší než zelí.

Filler 13 (good)

Kontext: Gregor Mendel posunul znalosti o genetice o velký kus vpřed, když objevil své základní zákony dědičnosti.

Mendel objevil základní zákony dědičnosti.

Filler 14 (bad)

Kontext: Cukrář měl za úkol upéct 50 sušenek. Cukrář upekl jen 25 kusů.

Cukrář splnil úkol.

Filler 15 (good)

Kontext: Africký meteorit Hoba je největší kus železa přírodního původu, který váží přibližně jako 10 slonů z téhož světadílu.

Meteorit Hoba váží přibližně jako 10 afrických slonů.

Filler 16 (bad) Kontext: Mourek váží 5 kilogramů, Mikeš váží 6 kilogramů.

Dohromady Mikeš a Mourek váží velké kilogramy.

Filler 17 (good)

Kontext: Iveta si koupila 3 kilogramy jablek za 90 korun.

Iveta zaplatila 30 korun za kilogram jablek.

Filler 18 (bad)

Kontext: Půlmaraton je běžecký závod na vzdálenost 21,0975 km, což je polovina délky maratonského běhu.

Půlmaraton je delší než maraton.

Filler 19 (good)

Kontext: Délka řeky Vltava je 430 km, délka řeky Labe je 1091 km.

Labe je delší než Vltava.

Filler 20 (bad)

Kontext: Madeleine Albrightová vydala 6 knih, všechno jsou to bestsellery.

Paní Albrightová napsala 5 bestsellerů.

Filler 21 (good)

Kontext: Šárka si koupila sukni, kalhoty a tričko.

Šárka si koupila tři kusy oblečení.

Filler 22 (bad)

Kontext: Pavlína uběhla 16 kilometrů na běžeckém páse.

Pavlína uběhla liché kilometry na běžeckém páse.

Filler 23 (good)

Kontext: Iceberg se kvůli tání odděluje od ledového šelfu a svobodně plave v oceánu.

Iceberg je obrovský kus ledového šelfu.

Filler 24 (bad)

Kontext: Lukáš šlápnul do bláta.

Lukáš měl několik blát na botách.

Filler 25 (bad)

Kontext: Štěpán kupuje mrkev v obchodě. Jeden kilogram mrkve stojí 20 korun. Štěpán má 50 korun.

Štěpán si může koupit 6 kilogramů mrkve.

C Russian items

Item 1

Контекст: До того как Ивановы заехали на новую квартиру, она была совершенно пустая.

baseline_mn В первый день Ивановы завезли семь предметов мебели.

baseline_mw В первый день Ивановы завезли семь штук мебели.

lonely_mn В первый день Ивановы завезли предметы мебели только на кухню.

lonely_mw В первый день Ивановы завезли штуки мебели только на кухню.

adj_mn Кухонные предметы мебели стояли на новой квартире Ивановых с первого дня.

adj_mw Кухонные штуки мебели стояли на новой квартире Ивановых с первого дня.

Item 2

Контекст: Надежда покупает вино в магазине.

baseline_mn Она берет две бутылки вина.

baseline_mw Она берет две штуки вина.

lonely_mn Бутылка вина в этом магазине стоит 700 рублей.

lonely_mw Штука вина в этом магазине стоит 700 рублей.

adj_mn Каждую субботу Надежда покупает любимую бутылку вина.

adj_mw Каждую субботу Надежда покупает любимую штуку вина.

Item 3

Контекст: Ржаной хлеб можно легко испечь дома в хлебопечке или в духовке.

baseline_mn Чтобы приготовить три булки хлеба понадобится килограмм ржаной и килограмм пшеничной муки.

baseline_mw	Чтобы приготовить три штуки хлеба понадобится килограмм ржаной и килограмм пшеничной муки.
lonely_mn	Булки можно посыпать семечками или тмином по вкусу.
lonely_mw	Штуки можно посыпать семечками или тмином по вкусу.
adj_mn	Домашние булки хлеба в разы вкуснее, чем магазинные.
adj_mw	Домашние штуки хлеба ценятся в разы больше, чем магазинные.

Item 4

Контекст: Собака, которую Женя нашла на улице, была очень голодная. Женя привела собаку домой и накормила мясом.

baseline_mn	Женя дала собаке два куска мяса.
baseline_mw	Женя дала собаке две штуки мяса.
lonely_mn	Собака разорвала мясо на куски и молниеносно их съела.
lonely_mw	Собака разорвала мясо на штуки и молниеносно их съела.
adj_mn	Собака съела два больших куска мяса.
adj_mw	Собака съела две больших штуки мяса.

Item 5

Контекст: Дети играли с Гарри Поттером, составленным из конструктора Лего. В один момент игрушка упала и разбилась.

baseline_mn	Гарри Поттер был составлен из 125 кирпичиков Лего.
baseline_mw	Гарри Поттер был составлен из 125 штук Лего.
lonely_mn	Гарри Поттер развалился на кирпичики.
lonely_mw	Гарри Поттер развалился на штуки.
adj_mn	Гарри Поттер был составлен из разных кирпичиков Лего.
adj_mw	Гарри Поттер был составлен из разных штук Лего.

Item 6

Контекст: Илья купил подарки трём близким женщинам на 8 марта.

baseline_mn	Илья купил три коробки конфет.
baseline_mw	Илья купил три штуки конфет.
lonely_mn	Илья подарит по коробке конфет маме, жене и сестре.
lonely_mw	Илья подарит по штуке конфет маме, жене и сестре.
adj_mn	Каждая коробка конфет стоила 600 рублей.
adj_mw	Каждая штука конфет стоила 600 рублей.

Item 7

Контекст: Дети рисовали картины гуашью.

baseline_mn	Каждый ребёнок изрисовал четыре листа бумаги.
baseline_mw	Каждый ребёнок изрисовал четыре штуки бумаги.
lonely_mn	Воспитатели раздавали детям листы бумаги для рисования.
lonely_mw	Воспитатели раздавали штуки бумаги детям для рисования.
adj_mn	Дети могли рисовать на больших или маленьких листах бумаги.
adj_mw	Дети могли рисовать на больших и маленьких штуках бумаги.

Item 8

Контекст: Глеб купил на распродаже брюки, рубашку и футболку.

baseline_mn	Глеб купил на распродаже три предмета одежды.
baseline_mw	Глеб купил на распродаже три штуки одежды.
lonely_mn	Глеб купил себе предметы одежды.
lonely_mw	Глеб купил себе штуки одежды.
adj_mn	Глеб на распродаже купил себе три новых предмета одежды.
adj_mw	Глеб на распродаже купил себе три новых штуки одежды.

Item 9

Контекст: Люба пьёт очень много кофе.

baseline_mn	Люба за один день может выпить три чашки кофе.
baseline_mw	Люба за один день может выпить три штуки кофе.
lonely_mn	Люба выпивает по чашке каждые три часа.
lonely_mw	Люба выпивает по штуке каждые три часа.
adj_mn	Люба каждое утро выпивает большую чашку кофе.
adj_mw	Люба каждое утро выпивает большую штуку кофе.

Item 10

Контекст: Мама отправила Витю в магазин за соком.

baseline_mn	Витя купил три коробки сока.
baseline_mw	Витя купил три штуки сока.
lonely_mn	Витя купил коробку.
lonely_mw	Витя купил штуку.
adj_mn	Витя купил большую коробку сока.
adj_mw	Витя купил большую штуку сока.

Item 11

Контекст: Для приготовления бабушкиного фирменного супа необходима свежая капуста.

baseline_mn	Для восьми порций супа потребуется два кочана капусты.
baseline_mw	Для восьми порций супа потребуется две штуки капусты.
lonely_mn	Кочаны режем соломкой, кладём в мясной бульон с другими овощами и доводим до кипения.
lonely_mw	Штуки режем соломкой, кладём в мясной бульон с другими овощами и доводим до кипения.
adj_mn	Для восьми порций супа потребуется два средних кочана капусты.
adj_mw	Для восьми порций супа потребуется две средних штуки капусты.

Item 12

Контекст: Лена и Рита пили вино.

baseline_mn	Лена выпила два бокала вина, Рита выпила три бокала вина.
baseline_mw	Лена выпила две штуки вина, Рита выпила три штуки вина.
lonely_mn	Вина Лена и Рита выпили по бокалу.
lonely_mw	Вина Лена и Рита выпили по штуке.
adj_mn	Лена и Рита пили вино из дорогих бокалов.
adj_mw	Лена и Рита пили вино из дорогих штук.

Item 13

Контекст: Михаил много кашляет и у него низкий голос, потому что он курит очень много сигарет.

baseline_mn	Михаил может выкурить три пачки сигарет в день.
baseline_mw	Михаил может выкурить три штуки сигарет в день.
lonely_mn	Михаил выкуривает пачку в день.
lonely_mw	Михаил выкуривает штуку в час.
adj_mn	Михаил за один день выкуривает целую пачку.
adj_mw	Михаил за один час выкуривает целую штуку.

Item 14

Контекст: Ирина Фёдоровна сварила малиновое варенье.

baseline_mn	Ирина Фёдоровна сварила десять банок варенья.
baseline_mw	Ирина Фёдоровна сварила десять штук варенья.
lonely_mn	Ирина Фёдоровна разливает варенье по банкам.
lonely_mw	Ирина Фёдоровна разливает варенье по штукам.
adj_mn	Ирина Фёдоровна разливает варенье по трёхлитровым банкам.

adj_mw Ирина Фёдоровна разливает варенье по трёх-
литровым бутылкам.

Item 15

Контекст: Индийский лев, который содержится в московском зоопарке, питается только свежим мясом.

baseline_mn	За один раз лев съедает шесть кусков мяса.
baseline_mw	За один раз лев съедает шесть штук мяса.
lonely_mn	Куски мяса для льва подготавливают и нарезают работники зоопарка.
lonely_mw	Штуки мяса для льва подготавливают и нарезают работники зоопарка.
adj_mn	Чтобы насытиться на целый день, лев должен съесть действительно большой кусок мяса.
adj_mw	Чтобы насытиться на целый день, лев должен съесть действительно большую штуку мяса.

Item 16

Контекст: Оля варит кашу своему сыну, она пробует её на вкус и понимает, что в кашу нужно добавить больше сахара.

baseline_mn	Оля добавляет в кашу две ложки сахара.
baseline_mw	Оля добавляет в кашу две штуки сахара.
lonely_mn	Оля добавляет ещё ложку.
lonely_mw	Оля добавляет ещё штуку.
adj_mn	Оля добавляет две маленьких ложки сахара.
adj_mw	Оля добавляет две маленьких штуки сахара.

Item 17

Контекст: Таня на завтрак ела творог с яблоком и бананом.

baseline_mn	Таня съела одну пачку творога.
baseline_mw	Таня съела одну штуку творога.
lonely_mn	Таня для завтрака купила творог в пачке.
lonely_mw	Таня для завтрака купила творог в штуке.
adj_mn	Таня съела целую пачку творога.
adj_mw	Таня съела целую штуку творога.

Item 18

Контекст: Папа ест борщ только со сметаной.

baseline_mn	Обычно папа кладёт в борщ две ложки сметаны.
baseline_mw	Обычно папа кладёт в борщ две штуки сметаны.
lonely_mn	Папа кладёт сметану в борщ ложкой.
lonely_mw	Папа кладёт сметану в борщ штукой.
adj_mn	Папа кладёт в борщ две большие ложки сметаны.
adj_mw	Папа кладёт в борщ две большие штуки сметаны.

Item 19

Контекст: Два внука ели у бабушки суп.

baseline_mn	Старший Вася съел две тарелки супа, младший Коля только одну тарелку.
baseline_mw	Старший Вася съел две штуки супа, младший Коля только одну штуку.
lonely_mn	Бабушка разливала суп внукам по тарелкам.
lonely_mw	Бабушка разливала суп внукам по штукам.
adj_mn	Старший внук съел большую тарелку супа, младший маленькую тарелку супа.
adj_mw	Старший внук съел большую штуку супа, младший маленькую штуку супа.

Item 20

Контекст: Вера выращивает цветы дома и сама следит за ними.

baseline_mn	Вера ухаживает за пятнадцатью горшками цветов.
baseline_mw	Вера ухаживает за пятнадцатью штуками цветов.
lonely_mn	Вера пересаживает цветы по горшкам.
lonely_mw	Вера пересаживает цветы по штукам.
adj_mn	Большие цветы Вера выращивает в больших горшках.
adj_mw	Большие цветы Вера выращивает в больших штуках.

Item 21

Контекст: Игорь с самого детства очень любит молоко. Каждый день он выпивает около одного литра.

baseline_mn Игорь пьёт два стакана молока утром и два стакана молока вечером.

baseline_mw Игорь пьёт две штуки молока утром и две штуки молока вечером.

lonely_mn Игорь обычно пьёт молоко из стаканов.

lonely_mw Игорь обычно пьёт молоко из штук.

adj_mn Игорь пьёт молоко из гранёных стаканов.

adj_mw Игорь пьёт молоко из гранёных штук.

Item 22

Контекст: Городская библиотека переехала со старого места работы на новое. Работники библиотеки сами занимались транспортировкой книг.

baseline_mn Работники перевезли на новое место 34 коробки книг.

baseline_mw Работники перевезли на новое место 34 штуки книг.

lonely_mn Работники самостоятельно разложили все книги по коробкам.

lonely_mw Работники самостоятельно разложили все книги по штукам.

adj_mn Работники раскладывали книги по картонным коробкам.

adj_mw Работники раскладывали книги по картонным штукам.

Item 23

Контекст: Лиза из-за границы привезла своей дочери очень дорогое и вкусное печенье.

baseline_mn Лиза привезла три упаковки печенья.

baseline_mw Лиза привезла три штуки печенья.

lonely_mn Печенье, которое привезла Лиза, было в упаковке.

lonely_mw Печенье, которое привезла Лиза, было в штуке.

adj_mn Печенье, которое привезла Лиза, было в индивидуальной упаковке.

adj_mw Печенье, которое привезла Лиза, было в индивидуальной штуке.

Item 24

Контекст: Даниил три раза в день пьёт зелёный чай.

baseline_mn За один день Даниил выпивает три чашки чая.

baseline_mw За один день Даниил выпивает три штуки чая.

lonely_mn Даниил готовит чай в специальном чайнике и пьёт его из чашек.

lonely_mw Даниил готовит чай в специальном чайнике и пьёт его из штук.

adj_mn Даниил готовит чай в керамическом чайнике и пьёт его из специальных чашек.

adj_mw Даниил готовит чай в керамическом чайнике и пьёт его из специальных штук.

Item 25

Контекст: Мама принесла из магазина сыр разных сортов.

baseline_mn Мама принесла три кусочка сыра.

baseline_mw Мама принесла три штуки сыра.

lonely_mn Мама нарезала сыр, который принесла, на кусочки.

lonely_mw Мама нарезала сыр, который принесла, на штуки.

adj_mn Мама принесла три небольших кусочка сыра.

adj_mw Мама принесла три небольших штуки сыра.

D Russian fillers

Filler 1 (good)

Контекст: Бутылка вина стоит 300 рублей.

Три бутылки вина стоят 900 рублей.

Filler 2 (bad)

Контекст: Мама в магазине купила баранину, свинину и говядину.

Мама в магазине купила три мяса.

Filler 3 (good)

Контекст: За брюки и футболку Яна заплатила 2000 рублей.

Яна купила два предмета одежды.

Filler 4 (bad)

Контекст: Чтобы покрасить комнату в жёлтый цвет, понадобится две банки краски.

Две банки краски понадобятся, чтобы покрасить комнату в зелёный цвет.

Filler 5 (good)

Контекст: Марина взяла в библиотеке две книги.

Марина взяла книги в библиотеке в количестве двух штук.

Filler 6 (bad)

Контекст: Солнце – единственная звезда в Солнечной системе.

Кроме Солнца есть ещё одна звезда.

Filler 7 (good)

Контекст: Стас мелко нарезал лист цветной бумаги ножницами.

Стас нарезал лист бумаги на кусочки.

Filler 8 (bad)

Контекст: Чтобы заказать одежду в онлайн-магазине X, нужно иметь на карте достаточное количество средств. Фёдор выбрал

себе футболку за 500 рублей, но на карте у него только 320 рублей.

Фёдор может заплатить за футболку картой в онлайн-магазине Х.

Filler 9 (good)

Контекст: Песец и полярная лисица — это одно и то же животное.

Песца иногда называют полярной лисицей.

Filler 10 (bad)

Контекст: Длина реки Волга составляет 3530 километров. Длина реки Дунай 2860 километров.

Дунай ненамного длиннее Волги.

Filler 11 (good)

Контекст: На свадьбе семьи Ивановых был роскошный торт.

Жених и невеста нарезали его на кусочки и предложили всем гостям.

Filler 12 (bad)

Контекст: В школе на уроке физкультуры 7А класс бежал кросс на три километра.

Каждые ребята уложились в 15 минут.

Filler 13 (good)

Контекст: Ювелир тщательно проверяет каждое украшение, которое попадает ему в руки.

Любое украшение, которое попадает к ювелиру, подвергается тщательной проверке.

Filler 14 (bad)

Контекст: Спортсмены-боксёры пьют много воды во время тренировки.

За одну тренировку боксёр может выпить четыре бутылков воды.

Filler 15 (good)

Контекст: Даша руководитель небольшого коллектива и недавно

у неё был день рождения. Она получила от своих коллег коллективный подарок.

Каждый из коллег сдал 100 рублей на подарок Даше.

Filler 16 (bad)

Контекст: Готовую смесь для торта выложите на бумажную форму и запекайте при температуре 200С 20-30 минут.

По истечении времён достаньте форму из духовки.

Filler 17 (good)

Контекст: Стоимость одного килограмма красных яблок 50 рублей. Илона купила 5 килограммов красных яблок.

Илона заплатила 250 рублей.

Filler 18 (bad)

Контекст: Швея купила несколько пар ножниц.

Швея купила три ножницы.

Filler 19 (good)

Контекст: Игорь залил 10 литров бензина в 20-литровую канистру.

Игорь заполнил бензином только полканистры.

Filler 20 (bad)

Контекст: Тамара родила сына. Сын весит 3,7 кг.

Сын Тамары весит несколько маленьких килограммов.

Filler 21 (good)

Контекст: Арбуз дешевле дыни в два раза.

Килограмм дыни стоит 50 рублей, следовательно, килограмм арбуза 25 рублей.

Filler 22 (bad)

Контекст: Полумарафон (21 км 97,5 метров) представляет собой забег на дистанцию, вдвое меньшую марафонской (42 км 195 метров).

Полумарафон в половину метров меньше, чем марафон.

Filler 23 (good)

Контекст: Андрей пьёт чай только с сахаром.

Андрей кладёт две ложки сахара в одну чашку чая.

Filler 24 (bad)

Контекст: Швея получила заказ на отделку платья лентами разных размеров и цветов.

Швея нарезала средние сантиметры от красной ленты и пришила к платью.

Filler 25 (bad)

Контекст: Высота горы Эверест составляет 8848 метров. Высота горы Эльбрус 5642 метра.

Гора Эверест на высокие метры больше, чем гора Эльбрус.