National Research University Higher School of Economics

as a manuscript

Liubov Nesterenko

FUNCTIONS OF PASSIVE WITH AGENT PHRASE: QUANTITATIVE ANALYSIS BASED ON MULTILINGUAL DATASET OF EUROPEAN LANGUAGES

Dissertation Summary
for the purpose of obtaining
academic degree Doctor of Philosophy in Philology and Linguistics

Academic Supervisor:
A.A. Bonch-Osmolovskaya, Candidate of Sciences

Moscow 2022

Overview

Passives with agent phrase were widely studied by linguists from different perspectives, there are numerous works devoted to this construction [Shibatani 1985], [Givón 1994], [Kazenin 2001], [Abraham 2006], [Keenan, Dryer 2009], [Siewierska, Bakker 2015]. On the other hand, only a few studies observe semantic and functional properties of passives with agent phrase and this topic deserves to be studied in more detail.

In general, linguists described passives with agent phrase as a syntactic phenomenon and paid their attention mostly to its formal properties. From functional point of view this construction is considered to serve as a topicalization mechanism, and to some extent it can be seen as one of the major passive traits. At the same time, little attention was paid to the issues of semantics and functional features of this construction, which require a comparative analysis of the cross-linguistic variation of marking strategies used for similar situations. In particular, we need to analyze 1) which semantically similar situations are marked with passive with agent phrase and 2) the variety of alternative strategies to encode these situations. In order to solve these issues we are going to use multilingual parallel corpus. That will allow us to not only analyze translations of situations in different languages through comparison of translational equivalents, but also involve a significantly larger amount of data than it is usually used in traditional approach. The format of multilingual parallel data gives us an opportunity to use quantitative analysis methods that allow us to reach the level of generalization that is inaccessible with manual analysis. Using this approach we are going to look at passives with agent phrase from a different perspective and get new insights about its properties that were previously unknown.

Typological analysis of grammatical phenomena demands exhaustive comparison of multiple contexts from various languages, which is a complex rigorous process, especially when it is done by hand, and examples are collected from multiple different sources. In this regard, the usage of parallel corpora has become more popular among linguists. The emergence of various natural language processing tools for morphological and syntactic parsing has contributed to an increase in the number

of annotated corpora in different languages. The availability of research material allowed linguists to prefer corpus and statistic analysis over traditional approaches based on manually collected samples.

Parallel corpora bring us to the next level of language data analysis: randomly samples sentences without much context are being replaced by an array of aligned and annotated translational units that let us consider a large amount of sentences in comparable contextual environment and analyze them more systematically. M. Cysouw and B. Wälchli were among the first who pointed out these advantages of parallel corpora in their work [Wälchli, Cysouw 2007]. The practice of the last decade shows that various language phenomena can be studied on the basis of parallel corpora. Tense and aspect variation, causatives, several types of passives, lexical units from different semantical fields have been the topics of recent parallel corpus studies. Despite its novelty, this area in the field of grammar and lexic research has already managed to attract several scholars [Cysouw & Wälchli (eds.) 2007, Dahl 2007, Aijmer 2008, Čermák & Rosen 2012, Östling 2016, Сичинава 2016, Нестеренко 2019].

Changing the data format used also entails the change of the methods applied to this data, namely, corpus data give opportunities to use quantitative research methods, that is another tendency in recent studies. For example, there are studies of causatives made by N. Levshina using subtitles corpus, and a range of works based of Bible texts, i.a. [Haspelmath 1997, de Vries 2007, Wälchli & Cysouw 2012, Östling 2016]. In these studies the authors use multidimensional scaling for automatic semantic maps construction and other statistical models for multifactorial analysis.

Thus, the **relevance** of our study is determined by the increasing popularity of multilingual parallel corpora in comparative studies and the perspectives they give for application of quantitative methods for linguistic purposes.

The object of this study is passive and relative constructions, and the subject of this study is passive with agent phrase, in other words full passive that entails the initial subject that was syntactically demoted but not fully eliminated. We adopt the term «agent phrase» from [Siewierska, Bakker 2013], and use it for any indirect

object noun phrase regardless whether it corresponds to the true agent or another participant role that can possibly act as an agent, e.g. natural force, activated instrument etc. From the formal perspective, the construction we have chosen for analysis retains all the participants of the initial diathesis, there is no omitted agent as in the basic passive, neither the agent does not correspond to the patient as in mediopassive. This means we have certain number of participants, that remain constant in translations, hence we can observe the variation in their cross-linguistic marking.

The passive with agent phrase, unlike grammatical categories like case, aspect, tense, and others, does not have a clear semantics, it is usually defined through its syntactic properties and it function. Mainly passive is associated with its discourse function of patient topicalization and agent demotion. Despite the fact that discourse function is considered as fundamental for passives, its functional scope is not limited to it. For example, A. Sansò in [Sansò 2006] describes several situation types, typical for the usage of passive and relative constructions and determining its functional scope. The results of this study indicate that passives with agent phrase can possess a variety of functions as well.

Another question about the functional scope of passive with agent phrase is related to the variety of constructions which can mark the same situations as passive with agent phrase, i.e. in contexts where some languages choose passive with agent phrase as marking strategy and other languages show several alternative marking strategies. This also does not correspond to the basic idea that passive is opposed to active, hence we expect to see it in text data. Almost every description of passives mention the passive to active relation, but as we might observe in our data, active in not the only equivalent construction to passive with agent phrase. This is an evidence that functional scope of passive with agent phrase is not that narrow as it might seem to be at first glance.

For the functional study of a particular construction it is highly important to cover as many alternative constructions the can mark the same situations as possible, because this variety of language means determine the semantical and functional scope

of the target construction. Doing comparative analysis the researcher usually face the dilemma which constructions they should or should not take into consideration. Using this approach, based on formal or informal criteria we exclude from our consideration constructions that are actually related to the construction of our interest and that can be used in similar contexts. Which means that by imaginary irrelevance one can ignore not only constructions but also situations that are related to the target construction. Multilingual parallel corpora give us opportunity not to be limited in the volume of the data and include as many contexts as possible. By including a variety of constructions related to the target one, even if they might seem irrelevant, we can fully analyze the construction in functional perspective and observe the whole range of its usage options. The novelty of our study is determined by the fact that despite the multiple studies devoted to passives very little of them are the functional studies of passives with agent phrase based on multilingual parallel corpora and perform application of quantitative methods.

The goal of this study is to analyze the «passive with agent phrase» construction in functional perspective on basis of multilingual parallel corpus using modern approaches and to describe the features of this construction. According to this we have to fulfill to following objectives:

1. Build multilingual parallel corpus and provide it with alignment, morphological and syntactic annotation.
2. Prepare the dataset for the research: extract relevant contexts that contain passives with agent phrase.
3. Perform preliminary quantitative assessment of the data and qualitative analysis, that will help to determine the relevant parameters for further annotation.
4. Perform experiments within computational approaches (factor analysis, graph models and semantic maps)
5. Interpret results and describe functional properties of passives with agent phrase.

The **theoretical significance** of the thesis is determined by the fact the it complements the existing description of passives and proposes a new interpretation of their functional properties. The results give evidence about the functional variety of passive uses hence the common point of view considering passives as discourse oriented construction can be extended.

The proposed scheme of functional analysis can be adapted to solving other linguistic tasks and applied to other constructions which makes up the **practical significance** of the thesis. Besides that, the our approach to the analysis of constructions co-occurrence with a translational unit has never been used before and can become a new instrument for linguists.

The following propositions are submitted for the defense:

1) The usage of multilingual parallel corpora the changes the format of handing the language data and extends the traditional set of analysis techniques applied to the language data. It allows to apply quantitative methods and make generalizations at a wholly different level that is hard to reach with manual analysis.

2) The passive with agent phrase is not solely discourse oriented construction, its semantical and functional sphere includes aspectual meanings, lexical semantics and relative positioning of events.

3) None of the languages uses passive with agent phrase exclusively in contexts where passive with agent phrase is opposed to active.

4) The language of our sample vary by the range of contexts in which passives with agent phrase can be used. Each language shows a particular set of situations that are typical for being marked by passive with agent phrase.

Public demonstration of results. The main finding were presented at International Conference for Computational Linguistics “Dialogue 2019” and “Parallel Corpora as Digital Resources and Their Applications” workshop at DHN 2020. There are also three Scopus publications:

- Nesterenko L.V. Multijazychnye parallel’nye korpusa: novyj istochnik dannyx dlja tipologicheskix issledovanij, perspektivy ispolzovanija i problemy (Multilingual parallel corpora: Alternative resource of language data for typological studies, usage perspectives and problems) // Voprosy jazykoznanija – 2019. – Vol. 2 C. 111-125.
- Bonch-Osmolovskaya, A. A., and L. V. Nesterenko. "Multilingual parallel corpora as a source for quantitative cross-linguistic grammar research (the case of voice constructions)." *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, 2019. P. 114-124.
- Nesterenko, Liubov. "Quantitative Analysis of Passives with Agent Phrase Based on Multilingual Parallel Data." In *DHN Post-Proceedings*, 2020. P. 5-15.

This thesis consists of an introduction, main part with five chapters, conclusion section and bibliography.

Summary of the main body of the thesis

In **Chapter 1**, we describe multilingual parallel corpora in terms of what they can bring to typological research and how they change the format of handling the language data in comparison to the traditional approach based on the analysis of manually collected samples.

For the purposes of typological and comparative research one needs representative language data that contains a maximal number of contexts that are typical for the usage of the target construction. Functional cross-linguistic analysis includes comparison of meanings of the target construction and also examination of other constructions alternative to the target one that can mark similar situations. The issue regarding alternative constructions is how the languages differ in the strategies they use for marking different situation and what relations can we find if we analyze them, what kinds of situations have similar marking, which of them differ cross-linguistically and to what extent. When we talk about the variety of marking strategies we need to specify how it manifests itself and what lexical and grammatical expressions it includes. It can not be enough to use solely traditional typological sampled to solve such problems and multilingual parallel corpora may appear to be more effective. Based on parallel corpus material we can build a dataset that will include multiple semantically different contexts, and then compare language means used for marking these situations.

The key features of parallel corpora allow us to change the way we work with language data and to process it more effectively. A **parallel corpus** is a collection of original texts in language L1 translated into one or more languages L2...Ln, an important attribute of the parallel corpus is **alignment**, which provides correspondences between different constituents of the texts. A parallel corpus usually has sentences alignment, but it can also have word alignment. Besides the alignment itself the parallel corpus may have morphological and syntactic annotation, but the alignment is considered to the key attribute of a parallel corpus, which allow us to compare situations and their realizations in different languages.

After the analysis of modern parallel corpora we can conclude that in order to use a multilingual corpus effectively one should take into account several factors that might be crucial for the study. Building a multilingual parallel corpus one should pay attention to 1) the variety of languages in the corpus; 2) the fullness of translation correspondences between different parts of the corpus; 3) the presence of word and sentence alignment; 4) morphological and syntactic annotation with universal tagset. The most common problem in modern corpora is the lack of consistency in translation correspondences, i.e. not all the parts of the corpus have a corresponding translation in all the other languages. Unfortunately, not all modern parallel corpora meet the listed criteria, and although this limits the possibilities of researchers, this is not an absolute obstacle for using them in research.

Despite the disadvantages of existing multilingual parallel corpora, there is a bunch of studies showing how various the problems that are being solved on the basis of multilingual parallel corpora can be. The papers [Sharoff 2002; Добровольский 2009; Сичинава 2015] are devoted to lexicon oriented studies, and [Wälchli, Cysouw 2012], [Östling 2016], [Levshina 2016] represent the studies of different grammatical phenomena. On hand these works one can get the idea of how broadly multilingual parallel corpora can be used, and which computational methods of analysis can be applied to parallel data.

The works we mentioned demonstrate the benefits of using multilingual corpora in comparative studies. Aligned translational texts help to optimize the technical aspects of data processing, e.g. to automatically transfer the annotation from language to another. They also provide us multiple comparable contexts, i.e. a set of situations with their encodings (=translations) in many different languages. These two aspects give linguists a plenty of new research opportunities.

The **Chapter 2** is devoted to issues of dataset building oriented to lexicogrammatical study and application of methods suitable for multilingual parallel data. Multilingual parallel corpora give us new opportunities for research and allow us to use computational approaches to analysis, and the dataset preparation is an important step. Building the dataset requires preliminary research and a formulation

of hypotheses which will be tested further, e.g. by applying computational methods. Hence, creating dataset we make the first step in the analysis of our data.

To handle multilingual parallel data we use several quantitative methods that suit our goals and the format of the data. Multi-factorial analysis helps to evaluate the contribution of the set of chosen factors in the automatic classification of contexts marked by either passive with agent phrase or active. Using network analysis we can order the set of constructions related to passive with agent phrase in a sense that are used in the same contexts. Semantic maps visualize how the situations presented in the dataset cluters together and which constructions are typical for each cluster with respect to the language. Every method solves a particular task and is oriented to a scepific types of data, e.g. constructions and their correspondences, translation units and their making, languages and their variation.

For our study we created a multilingual parallel corpus according to the criteria we proposed in the first chapter. Based on this corpus and knowledge about fundamental properties of passives, we created the dataset that was suitable for solving our problems and was further used in the study.

Constructing the corpus, we followed the idea that the text we use should be completely parallel, i.e. have corresponding translations for every part of the corpus, caption modern language and form a corpus allows us to find a sufficient number of examples containing passive constructions with agent phrase. For these purposes we used a collection of teh first seven books about Harry Potter by J.K. Rowling. There nine languages presented in corpus: English, Russian, German, Italian, Spanish, Bulgarian, Swedish and French. In total, there are about 1 million tokens for each language. The texts have sentence alignment [Gale, Church 1993], word alignment [Östling, Tiedemann 2016] and also morphological and syntactic annotation [Nivre et al. 2016].

Creating the dataset we took into account the information about the general characteristics of passives, as well as the specific features of passives with agent phrase in semantical and functional perspective. The dataset was constructed in a few steps, each step complemented and enriched the data processed at the previous stage.

After the first calculations of distributions of constructions in languages were done, our initial ideas about the nature of passives with agent phrase has changed, this allowed us to reformulate our hypotheses and determine new direction in further research.

With the help of multilingual corpus we created a dataset that includes numerous examples where passives with agent phrase are used. Additional processing of the data we obtained from the corpus, allowed us to prepare the data for the planned experiments. We were able not only to take into account the characteristics we were initially interested in, but also extend them while we were working with the data and qualitatively analyzing it.

In **Chapter 3**, we make an overview of theoretical and typological literature devoted to passives with agent phrase and related voice constructions as well as their functional properties. We use this information and our preliminary calculations to understand the nature of passives and to formulate our hypotheses.

Our choice of passive with agent phrase as a target construction is determined by the fact that it has no fully demoted constituents, which means, finding correspondences between its constituents in translations will be more clearcut. Passive with agent phrase is a result of voice transformation which does not have any eliminated participants of the initial diathesis, that is why sometimes it is called full passive. The presence of overtly expressed agent distinguishes this type of passive from other passive constructions.

Despite the fact that the formal structure of passive with agent phrase is quite clear, there is no straightforward explanation for the motivation of its presence or absence, usually it is mentioned that the passive agent phrase is optional. It can imply that several languages might have two types of passive with and without the agent phrase. There are more languages in the world that have both full passives and basic passives, that languages that have only agentless passives [Siewierska, Bakker 2013]. However the distribution of these constructions within one language would usually be such that the agentless passives are more frequent than passives with agent phrase

[Siewierska, Bakker 2013]. At the same time there are no clearcut rules that determine the presence of the agent phrase in a passive construction.

In comparison to other related constructions, passives with agent phrase have more similarities with resultatives. And the resultatives themselves have very distinct aspectual semantics, but extremely similar to passives with agent phrase and can be hard to distinguish from them if does not have special marking [Nedjalkov 1984]. This indicates that in some languages we should find resultatives in the same contexts where in the other languages passives with agent phrase are used, and they are going to be among the alternative constructions.

The central question of our study is related to semantical and functional sphere of passive with agent phrase usage. The functioning of passive with agent phrase is determined by several aspects, it is the role of the agent phrase, as a part of the situation semantics, aspectual factors and discoursive promotion and demotion mechanisms. Our task was to consider how all these aspects act together using a large amount of relevant contexts.

Before we formulated the hypothesis about the functional diversity of passives with agent phrase uses, we had made some preliminary calculations. We analyzed the distribution of passives with agent phrase across the languages in our dataset, and we calculated the number of examples marked by either passive with agent phrase, the active or the other type of construction. The distributions we obtained indicated two things: first, none of the languages uses passive with the agent phrase only for the promotion of the patient, i.e. in discourse oriented contexts, second, most of the languages use passives with agent phrase in situations of other types, namely those, where the agent phrase does not correspond to the canonical human agent and it is not a context demanding patient promotion.

Based on these results, we concluded that the use of passive with agent phrase indeed is not limited to discourse function, and its functional scope if broader. Hence, we decided to focus on those examples in our sample that were not marked either by passive with agent phrase or active, because they reflect the semantical variety of situations we look at. At the same time it became clear that we need to categorize and

order not only the language means of expression but also the situations they mark on a par with passive with agent phrase.

Chapter 4 is devoted to the discourse oriented uses of passive with agent phrase. We present the factor analysis experiment that aims to assess the features that might influence the choice between passive with agent phrase and active in a particular context. Using logistic regression models we evaluate the contribution of different factors (semantic characteristics of participants or contextual features) into classification of these two constructions. In our experiment we found out that contextual features related to mentioning one of participants in the preceding context, have big influence on the choice of construction. For example, the feature “agent was mentioned” appeared to be the trigger for using active, and the feature “patient was mentioned” was the stimulus for using passive with agent phrase. Similar result was obtained by R.Tomlin in his experiments, it has shown that priming of participants determines the use of passive and active. The other features that appeared to be significant were involvement of the patient, animacy of the agent and whether the verb denotes the negative action, it indicates that not only the contextual environment but also the semantics of the situation are relevant.

Our results are in line with the findings of R. Tomlin in his psycholinguistic experiments, which have shown that the priming of the participants motivates the type of the voice construction that will be used. In addiction to contextual features we evaluated the impact of some semantic features and moreover, we built logistic regression model and determine significant features for each language in our dataset. This information enriches the existing understanding of the mechanisms that determine the choice between active and passive with agent phrase in discourse oriented contexts.

In **Chapter 5** we show how network analysis and semantic maps can help us to illustrate the variety of the situations that are typically marked by passives with agent phrase in some languages and also the variety of alternative constructions use in these situations in th other languages. After determining the set of constructions that mark the same situations as passive with agent phrase we can group together these

situations and understand what are their semantical and functional properties. The final classification reflects the functional variety of passive with agent phrase uses and proves our hypothesis that the function of passive with agent phrase is not limited to patient promotion and agent demotion.

The first discovery for us was the fact that there are a lot of constructions alternative to the passive with agentive phrase, which act as equivalent correspondences to it in translational units, which means passive with agent phrase does not only correspond to active sentences. It is also important, that these alternative constructions are *regularly* occur in translations, and that holds for particular languages as well as for the whole sample, so it can not be interpreted as some random decisions of translators. The alternative constructions we discovered can be classified into several groups: lexical units, types voice constructions, locative constructions, connection of two clauses.

The data from the parallel corpus allow us to cover the whole variety of alternative expressions, to find explanations for this diversity and to interpret it with regard to the situation types they mark. Analyzing the groups of alternative constructions we can get some understanding of semantical and functional properties of passive with agent phrase.

Multiple examples, that we work with, can be categorized into several types, based on the semantical characteristics of the corresponding situations. Each group in our list can be describe as a particular set of participants and for these groups we use the term ‘frame’ in the sense that it was proposed by Ch. Fillmore and as it is understood in Framenet. We determined 15 frames and grouped them according to the marking strategies that are used for them:

1) The opposion of active and passive with agent phrase, discourse function

Semantics: Promotion of P, that is affected by A, mutual arrangement of A and P

Language alternatives to passive with agent phrase: transitive active clauses, coordinative and subordinative conjunctions

Frames: *Absorption*, Interruption/*collision*, *Events*, *Generic*

Examples:

(1) Interruption/collision

passive with agent phrase

EN: ***He*** was interrupted by a knock on the door.

active

DE: [Ein Klopfen an der Tür] unterbrach ***ihn***.

(2) Absorption

passive with agent phrase

EN: ***Dudley's snores*** were drowned by the low rolls of thunder that started near midnight.

active

RU: Дадли захрапел, но ***его храп*** заглушали [низкие раскаты грома]: началась гроза.

(3) Events

passive with agent phrase

FR: Mais son attention fut détournée [par les parents d'Hermione]

two clauses

RU: Но тут [увидел родителей Гермионы] и сразу же про него забыл.

2) Metaphorical uses of passive with agent phrase

Semantics: Situtations denoting states experienced by a human or an object

Language alternatives to passive with agent phrase: adjectives, verbs, denoting processes and states (*giggle*, *shake*, *panick*), reflexives, locative expressions with auxiliary verbs

Frames: *Mental triggers*, *Internal movement*, *Fill (P filled with A)*, *Facial expressions*

Examples:

(4) Mental triggers

passive with agent phrase

SE: "JAG DÖR!" tjöt Malfoy och ***hela klassen*** greps [av panik].

lexeme

EN: 'I'm dying!' Malfoy yelled, as ***the class*** panicked.

lexeme

RU: — Я умираю! – громко стонал Малфой, окруженный испуганными ***одноклассниками***.

(5) Internal movement

passive with agent phrase

ITA: ***Le sue mani*** erano scosse [da un lieve tremito].

lexeme

EN: ***Her hands*** shook slightly.

(6) Facial expressions

passive with agent phrase

ITA: ***Il suo volto pallido*** fu attraversato [da un ghigno malvagio].

active with A as indirect object

EN: ***His pale face*** split [in a malevolent grin].

active with A as indirect object

RU: ***Лицо его*** расплылось [в злобной ухмылке].

active

ES: Hizo una mueca malévola—.

(7) Fill (P filled with A)

passive with agent phrase

RU: <...> иногда ***мозг*** так переполнен [мыслями], что их необходимо куда-то выплеснуть.

adjective

EN: a ***head*** becoming so full of [thoughts]

construction with the verb of possession

ES: cuando le había dicho que tenía [tantos pensamientos] ***en la cab***

(8) Fill (P filled with A)

passive with agent phrase

SE: Dobby log strålande medan **hans ögon** på nytt **fylldes av glädjetårar**.

passive *with in-phrase*

EN: Dobby beamed very brightly, and **happy tears welled in his eyes** again.

3) Resultatives

Semantics: A state that occurred as a result of some visual change that affected the object

Language alternatives to passive with agent phrase: пассивы с косвенным дополнением, оформленным иначе, чем агентивное дополнение (преимущественно инструментальный показатель), конструкции с there is/are, глаголы обладания.

Frames: *Локативные результативы. Спрятать/скрыть*

Examples:

(9) *passive with agent phrase*

RU: Гарри заметил, что вид у нее довольно жалкий: <...>, а ***лицо и руки*** испещрены [царапинами].

(10) *passive with agent phrase*

RU: ***Каждый*** слегка извивался и был усеян [множеством блестящих припухлостей], наполненных какой-то жидкостью.

construction with the verb of possession

EN: ***Each*** was squirming slightly, and had [a number of large, shiny swellings] upon it, which appeared to be full of liquid.

construction with the verb of possession

ES: ***Todas*** estaban algo retorcidas, y tenían [una serie de bultos grandes y brillantes] que parecían llenos de líquido.

(11) Hide/conceal

passive with agent phrase

EN: He sat up quickly and looked over at Colin's bed, but ***it*** had been blocked from view [by the high curtains Harry had changed behind yesterday].

passive with agent phrase

RU: Сев на постели, Гарри увидел, что ***кровать***, где лежал Колин, отгорожена [ширмой].

passive with agent phrase

ITA: Si mise a sedere e sbirciò il letto di Colin, che però era stato escluso alla vista [dalle lunghe tende dietro cui Harry si era cambiato il giorno prima].

4) Mixed types

Frames that have clear, defined meaning but have resultative/stative and eventive uses

Language alternatives to passive with agent phrase: transitive active sentences, intransitive active sentences with indirect obeject, passive with differently marked agent phrase, locative expressions and adjectives.

Frames: *Emotions*, *Interest*, *Occupy*, *Light*

Examles:

(12) Emotions

passive with agent phrase

FR: Mais il s'était très vite rendu compte que sa plaisanterie n'amusait personne car ***tout le monde*** avait été impressionné [par la virtuosité avec laquelle Harry avait réussi à s'accrocher à son balai].

passive with *at-phrase*

EN: <…>, because ***they*** were ***all*** so impressed [at the way Harry had managed to stay on his bucking broomstick].

(13) Interest

passive with agent phrase

Никой не забеляза, че са изчезнали, ***всички*** бяха погълнати [от произведенията на Фред и Джордж].

passive with in-phrase

Nobody noticed them vanish; ***they*** were all too interested [in Fred and George's products].

(14) Occupy

passive with agent phrase

FR: ***La page*** était en grande partie occupée [par la photo noir et blanc d'un homme aux cheveux épais comme une crinière de lion et au visage marqué].

passive *with with-phrase*

EN: ***Most of this front page*** was taken up [with a large black-and-white picture <...>].

passive with agent phrase

ITA: ***Gran parte della prima pagina*** era occupata [da una grande foto <...>].

active

ES: *La mayor parte de la primera plana* ***la*** ocupaba [una gran fotografía <...>].

(15) Occupy

passive with agent phrase

FR: Harry, dont **l'attention avait été** entièrement **occupée** jusqu'alors **par Madame Maxime**, <...>

passive with upon-phrase

EN: Harry, whose **attention had been focused** completely **upon Madame Maxime**, <...>

(16) Light

passive with agent phrase

ITA: ***Le pareti di pietra*** erano illuminate [da torce fiammeggianti come quelle della Gringott], <...>

passive with agent phrase

SE: ***Stenväggarna*** lystes upp [av flammande facklor precis som hos Gringotts], <...>

пассив с with-дополнением

EN: ***The stone walls*** were lit [with flaming torches like the ones at Gringotts], <...>

active with P as indirect object

RU: ***На каменных стенах*** – точно так же, как в «Гринготтс», – горели [факелы], <...>

active

DE: Wie bei Gringotts beleuchtete [das flackernde Licht von Fackeln] ***die Steinwände***<...>

passive with con-phrase

ES: ***Las paredes de piedra*** estaban iluminadas [con resplandecientes antorchas como las de Gringotts], <...>

active

CZ: ***Kamenné stěny*** ozařovaly [planoucí pochodně stejně jako u Gringottových], <...>

passive *with c-phrase*

BG: ***Каменните стени*** бяха осветени [с факли като онези в „Гринготс“], <...>

construction «P on A»

FR: ***Des torches enflammées*** étaient fixées [aux murs de pierre, comme à Gringotts], <...>

5) Typical uses of passive with agent phrase

Situations that are typically marked by passive with agent phrase

Frames: *Resolutions*, partially Hide*/conceal* and *Light*

Examples:

(17) Resolutions

passive with agent phrase

EN: Morfin and his father attempted to fight, but ***both*** were overpowered, removed from the cottage and subsequently convicted [by the Wizengamot].

passive with agent phrase

RU: <...>, и в конце концов ***они*** были осуждены [Визенгамотом].

passive with agent phrase

DE: <...>, wurden aber ***beide*** überwältigt, von dem Haus weggebracht und anschließend [vom Zaubergamot] verurteilt.

passive with agent phrase

ITA: <...>, ma furono sopraffatti entrambi, portati via e in seguito condannati [dal Wizengamot].

active

ES: <...>, y más tarde [el Wizengamot] ***los*** condenó.

passive with před-phrase

CZ: <...>, byli ale přemoženi, odvedeni z domu a následně postaveni [před Starostolec].

passive with agent phrase

BG: <...>, но <u>били озаптени, изведени от къщурката и после осъдени</u> [от Магисъбора].

passive with agent phrase

SE: <...>, men ***båda*** <u>övermannades, fördes bort från stugan och fälldes därefter</u> [av Wizengamotrådet].

passive with agent phrase

FR: ***Ils ont*** <u>été emmenés et jugés coupables</u> [par le Magenmagot].

The information about variability in the marking strategies of the situations, makes it possible to combine into semantically and functionally motivated groups. Each language in our dataset tend to mark by passive with agent phrase some of the groups we determined. We illustrate the distribution of the groups marked by passive with agent phrase across languages with the help of a heatmap, where the cells with lighter color indicate high frequency of passive with agent phrase in particular group in a language, but the darker color denotes lower frequency.

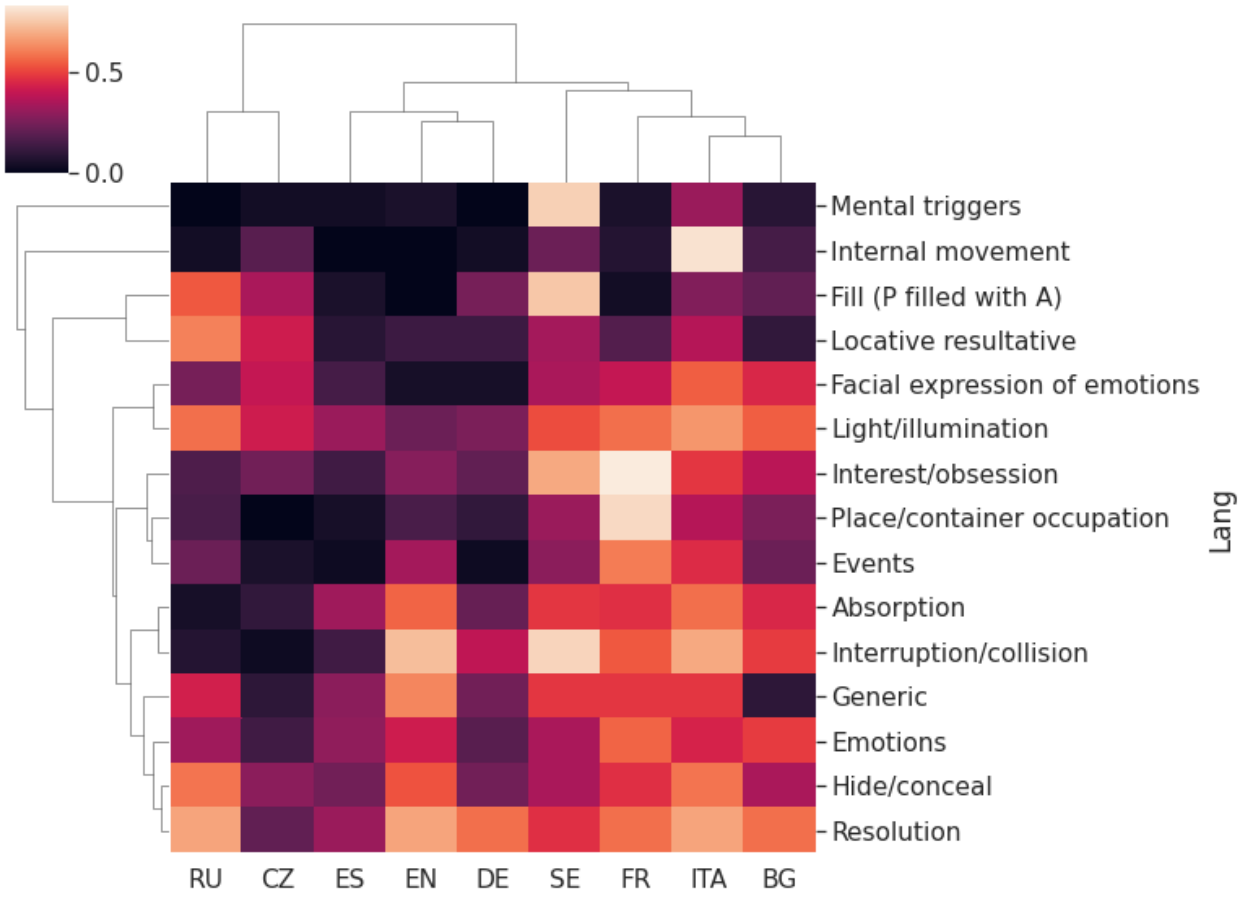


Figure. 1. The distribution of passives with agent phrase in languages according to the semantical/functional group

This figure shows that some of the languages are similar in a way they use passive with agent phrase for marking particular semantical/functional groups.

We were able to confirm the consistency of the connection between the semantics of situations and the variety of expressions by constructing a semantic map using the multidimensional scaling method, for that we use information about the types of constructions were used in each translation unit. The semantic maps show how semantical/functional groups are related to each other, which of them appear to be close, and which are opposed to each other. We can also observe various language expressions are distributed among the semantical/functional groups and detect the areas on the map that are common for a certain type of marking in a each language.

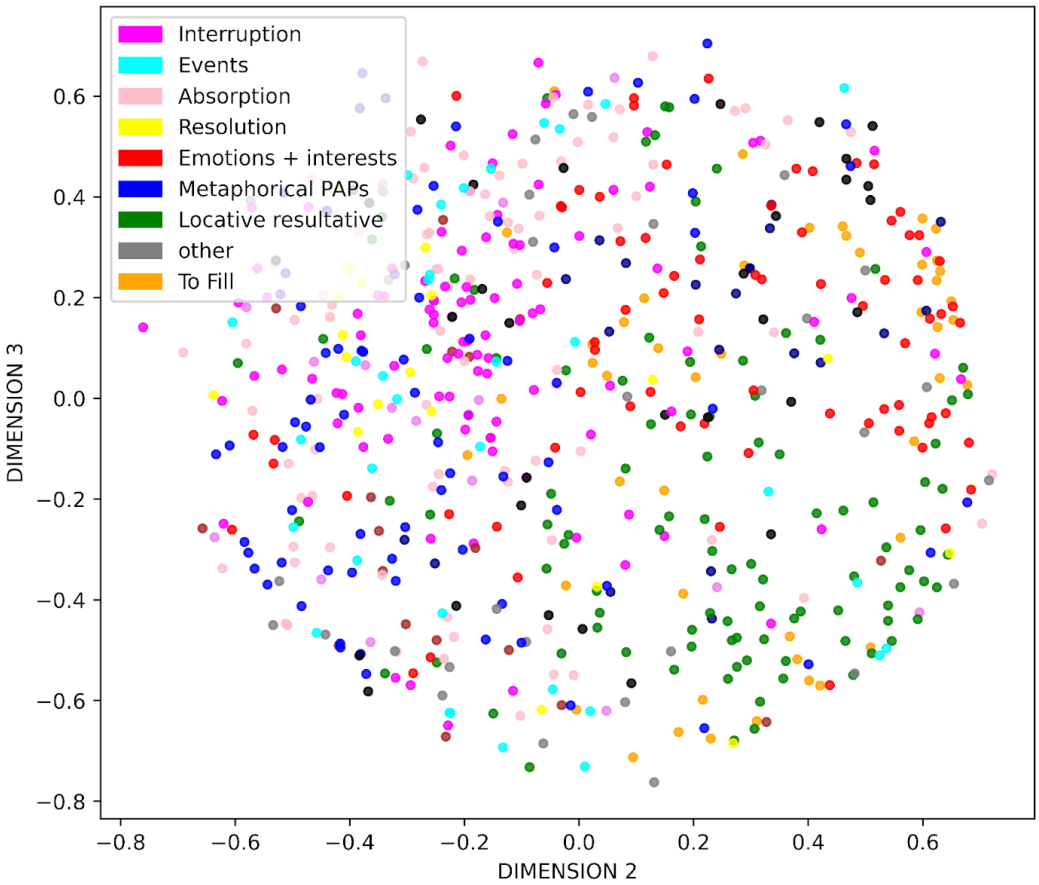
The semantic map is based on the distance matrix which is captions the information about the differences of translation units which is calculated by comparing the sets of language expressions used in translation units. Then by applying the multidimensional scaling we get the final matrix with a determined number of dimensions. We choose the optimal number of dimensions that will be enough to caption the relation in our dataset and to interpret the results. In our experiment we set the optimal number of dimensions to 3.

Each dimension is responsible for the principle that splits the translational units into groups opposed to each other, that can be linguistically interpreted and tell us whether is consistent with our ideas. The combination of dimension 1 and 2 reflects the major opposition in our data, it corresponds to the division into discourse oriented vs. semantically/functionally oriented contexts. Dimensions 2 and 3 make up a more detailed clustering, in table 2 we captioned the main clustering principles that are illustrated by figure 3.

Table 2. Interpretation of dimensions 2 and 3

	Dimension 2, negative pole	**Dimension 2, positive pole**
Dimension 3, positive pole	Interruption/collision, Absorption, Resolution, partially Events	Metaphorical usages of PAP and partially Fill (P filled with A)
Dimension 3, negative pole	Interest/obsession, emotions and Place occupation	Locative resultatives, partially Fill (P filled with A)

The figure below is the semantic map itself, there are four areas that correspond to the semantical/functional groups, that can be typical context for passives with agent phrase in some languages. In our initial classification we have more groups but the semantic map gives us a more generalized view, groups with similar marking are located close to each other on the map.



Firgure 3. Semantic map, dimensions 2 and 3 (the semantical/functional groups are differently colored)

The clustering that we see on the map mostly coincides with the classification we proposed expect for some nuances. Each of the four big clusters has the dominant category and some translational units from smaller groups that were clustered with it together. Small groups that have very specific meanings/functions have the following behavior, they either joined some of the big clusters or split into groups of eventive and stative units each of which found its position in different parts of the map.

After we found out the positions of the translation units on the map we could plot the marking strategies with regard to the language. So, for each language we make a map which captures the distribution of constructions in semantical/functional groups. It allows to see which semantical/functional groups in this language tend to be marked by passive with agent phrase and which of them are prone to be marked by alternative constructions. Having the maps analyzed, we can conclude that some of the semantical/functional groups are typical passive with agent phrase contexts for most of our languages, these are resolution and light/illumination. The other groups are more specific and are related to a few languages, e.g. internal movement for Italian, locative resultatives for Russian and Czech, place occupation for French.

Thus we analyzed multiple situations that in some languages are typically marked with passive with agent phrase and examined functional properties of this construction. We also considered language expressions alternative to passive with agent phrase and that helped us to understand how broad is the semantical/functional sphere of passive with agent phrase uses is.

We also made some conclusions about the agent phrase itself. In many contexts the agent phrase corresponds to a participant that is tightly integrated into the semantic frame of the situation, hence using passive with agent phrase in such cases have strong semantic motivation. This can be demonstrated by examples where passive with agent phrase corresponds to lexical units and also the sentences with passive with agent phrase that can not be reformutated as active or as agentless passive, e.g. locative resultatives in Russian.

We conclude the thesis by highlighting the major findings and results:

1) For the study of functional properties of passive with agent phrase we created a multilingual parallel corpus that contains translations of Harry Potter books by J.K. Rowling in nine languages. The texts have sentence and word alignment as well as morphological and syntactic annotation.

2) Using the corpus we compiled the dataset with translation unit relevant fot the study of passive with agent phrase. It contains multiple examples marked not only with passive with agent phrase and active but also with various alternative expressions. While constructing the dataset, we manually corrected the error in annotation and alignment, added missing translations. We also came up with a classification of constructions used as alternatives to passive with agent phrase in translations (lexical items, locative constructions, clausal connectors etc.)

3) For the set of discourse oriented situations we evaluated how different semantic and contextual features influence the choice between passive with agent phrase and active construction. Our results, based on corpus data, are consistent with those obtained by R.Tomplin in his psycholinguistic experiments.

4) The contexts that were not qualified as discourse oriented were categorized into semantic/functional groups. Our initial classification is supported by the semantic map. Based on the semantic maps we built we were able to to make generalizations about the functional properties of passives with agent phrase.

5) Analyzing the distributions of passive with agent phrase and other constructions we evaluated its performance in the languages in our dataset. We found out that passive with agent phrase not only has different functions, but these functions are typical for passives with agent phrase in different languages and every language has its own passive with agent phrase profile.

References

1. Abraham W. Introduction: Passivization and typology 2006.C. 1–27.
2. Aijmer K. Parallel and comparable corpora // Corpus linguistics: An international handbook. 2008. (1). C. 275–291.
3. Čermák F., Rosen A. The case of InterCorp, a multilingual parallel corpus // International Journal of Corpus Linguistics. 2012. № 3 (17). C. 411–427.
4. Cysouw M., Wälchli B. Parallel texts: using translational equivalents in linguistic typology // Language typology and universals. 2007. № 2 (60). C. 95–99.
5. Dahl Ö. From questionnaires to parallel corpora in typology // Language Typology and Universals. 2007. № 2 (60). C. 172–181.
6. Gale W. A., Church K. W. A Program for Aligning Sentences in Bilingual Corpora // Computational Linguistics. 1993. № 1 (19). C. 75–102.
7. Voice and Inversion под ред. T. Givón, Amsterdam: John Benjamins Publishing Company, 1994.
8. Kazenin K. I. Passive voice 2001.C. 899–915.
9. Keenan E. L., Dryer M. S. Passive in the world’s languages под ред. T. Shopen, Cambridge: Cambridge University Press,C. 325–361.
10. Levshina N. Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages // Folia Linguistica. 2016. № 2 (50).
11. Nedjalkov V. P., Jaxontov S. J. The Typology of Resultative Constructions 1988.C. 3.
12. Nesterenko L. V. Multilingual parallel corpora: Alternative source of language data for typological studies, applying perspectives and problems // Voprosy Jazykoznanija. 2019. № 2. C. 111–125.
13. Nivre J. [и др.]. {U}niversal {D}ependencies v1: A Multilingual Treebank Collection Portorož, Slovenia: European Language Resources Association (ELRA), 2016.C. 1659–1666.
14. Östling R. 6. Studying colexification through massively parallell corpora Berlin, Boston: De Gruyter, 2016
15. Östling R., Tiedemann J. Efficient Word Alignment with Markov Chain Monte Carlo // The Prague Bulletin of Mathematical Linguistics. 2016. № 1 (106). C. 125–146.
16. Tomlin R. S. Focal attention, voice, and word order 1995.C. 517.
17. Sansò A. «Agent defocusing» revisited 2006.C. 232–273.
18. Say S. Nominal causal constructions across Slavic: semantic contrasts in a parallel corpus perspective // Slavia-časopis pro slovanskou filologii. 2021. № 2 (90). C. 182–201

19. Sharoff S. Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics // 2002.
20. Shibatani M. Passives and Related Constructions: A Prototype Analysis // Language. 1985. № 4 (61). С. 821.
21. Siewierska A., Bakker D. Passive agents: prototypical vs. canonical passives под ред. D. Brown, M. Chumakina, G. G. Corbett, 2013.
22. Vries L. De Some remarks on the use of Bible translations as parallel texts in linguistic research // Language Typology and Universals. 2007. № 2 (60). С. 148–157.
23. Wälchli B., Cysouw M. Lexical typology through similarity semantics: Toward a semantic map of motion verbs // Linguistics. 2012. № 3 (50). С. 671–710.
24. Добровольский Д. О. Корпус параллельных текстов в исследовании культурно-специфичной лексики 2009.С. 383–400.
25. В. А. Плунгян Введение в грамматическую семантику: грамматические значения и грамматические системы языков мира, РГГУ, Москва, 2011
26. Сичинава Д. В. Европейский перфект сквозь призму параллельного корпуса // Acta Linguistica Petropolitana. Труды института лингвистических исследований. 2016. № 2 (12). С. 85–114.