

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИИ



Федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «РГГУ»)

Институт лингвистики

УНЦ компьютерной лингвистики

Рабочая программа дисциплины

«Лингвистическое аннотирование/разметка текстов»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Магистерская программа: Фундаментальная и компьютерная лингвистика

Квалификация выпускника: магистр

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

Лингвистическое аннотирование/разметка текстов

Рабочая программа дисциплины

Составители:

В. П. Селегей

Ответственный редактор:

д. филол. н., профессор В.И.Подлеская

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной
лингвистики

№ 1 от «28» августа 2019г.

Оглавление

1. Пояснительная записка

- 1.1. Предмет
- 1.2. Цель и задачи дисциплины
- 1.3. Формируемые компетенции и результаты освоения дисциплины
- 1.4. Место дисциплины в структуре образовательной программы

2. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

- 5.1. Система оценивания
- 5.2. Критерии выставления оценок
- 5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

6. Учебно-методическое и информационное обеспечение дисциплины

- 6.1. Список литературы

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

9. Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

1. Пояснительная записка

1.1 Предмет

Предметом дисциплины (модуля) является теория и практика лингвистической разметки корпусов. Вопросы применения разметки для корпусных лингвистических исследований и машинного обучения рассматриваются и в других курсах, в частности «Корпусная лингвистика».

1.2 Цель и задачи курса

Курс направлен на решение следующих задач:

- ознакомление магистрантов с методами лингвистического аннотирования;
- знакомство с международными стандартами разметки в области морфологии, синтаксиса, семантики, структуры текста; обучение технологиям работы групп аннотаторов.

1.3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина (модуль) направлена на формирование компетенций выпускника:

владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);

способностью разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты интеллектуальных и информационных электронных систем (ПК-8);

способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);

способностью разрабатывать и совершенствовать системы автоматизации и информационной поддержки лингвистических исследований (ПК-10)

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать: основные методы лингвистического аннотирования и международные стандарты разметки в области морфологии, синтаксиса, семантики, структуры текста;

Уметь: выбрать необходимый тип лингвистической разметки для решения требуемой задачи;

Владеть: технологиями оценки качества лингвистической разметки, в частности - при использовании групп аннотаторов.

1.4 Место дисциплины в структуре образовательной программы

Дисциплина (модуль) «Лингвистическое аннотирование / разметка текстов» является дисциплиной по выбору вариативной части цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 2 курса (3 семестр). Дисциплина (модуль) реализуется УНЦ компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: подготовка к семинарским занятиям, реферат, коллоквиум по контрольным вопросам с выполнением практического задания; промежуточная аттестация в форме: зачет.

Общая трудоемкость освоения дисциплины (модуля) составляет 2 зачетные единицы, 72 часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – 20 часов; самостоятельная работа студента – 52 часа.

2. Структура дисциплины

№ п/п	Раздел дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				лекции	семинары	самостоятельная работа		
1	Корпусная разметка: технологии, решаемые исследовательские задачи, методы использования в компьютерной лингвистике	3	1		2	5		ДЗ1
2	Технология аннотирования. Последовательности решаемых задач: выбор содержательной теории, постановка задачи. Оценки сложности разметки и операционности методик аннотирования. Оценка согласия аннотаторов. Множественная разметка.	3	2		2	5		ДЗ2
3	Автоматическая и ручная разметка. Преимущества и недостатки. Особенности автоматической разметки в задачах создания интернет-корпусов большого объема (миллиарды документов).	3	3		2	5		ДЗ3
4.	Размеченные корпуса как основа статистического анализа и машинного обучения: виды решаемых задач (исследовательские и	3	4		2	5		ДЗ4

	прикладные).						
5.	Словарные разметки. Интегрированные системы словарь-корпус, основанные на разметке. Примеры: проект DANTE, система LingvoPro.	3	5		2	5	Д35
6.	Современные методы разметки на разных уровнях системы языка: морфология, синтаксис, семантика, дискурс. Shallow and deep синтаксис и семантика. Элементы онтологической разметки. Разметка в системе ABBYY Compeno.	3	6		2	5	Д36
7.	Особенности мета-текстовой разметки: жанровая, социальная, региональная классификации.	3	7		2	5	Д37
8.	Современные размеченные корпуса: Penn TreeBank, Prague Dependency	3	8		2	5	Д38
9.	Treebank, проект OntoNotes, размеченные подкорпуса Русского национального корпуса.	3	9		2	5	Д39
10.	Linked Data в современной лингвистике. Стандартизация разметок (стандарты ISO). Проекты связывания лингвистических ресурсов: *Net (WordNet, FrameNet, PropBank)	3	10		2	5	Д310
	Зачет					2	Контрольные вопросы
					20	52	

3. Содержание дисциплины

- Корпусная разметка: технологии, решаемые исследовательские задачи, методы использования в компьютерной лингвистике.
- Технология аннотирования. Последовательности решаемых задач: выбор содержательной теории, постановка задачи. Оценки сложности разметки и операционности методик аннотирования. Оценка согласия аннотаторов. Множественная разметка.
- Автоматическая и ручная разметка. Преимущества и недостатки. Особенности автоматической разметки в задачах создания интернет-корпусов большого объема (миллиарды документов).
- Размеченные корпуса как основа статистического анализа и машинного обучения: виды решаемых задач (исследовательские и прикладные).
- Словарные разметки. Интегрированные системы словарь-корпус, основанные на разметке. Примеры: проект DANTE, система LingvoPro.
- Современные методы разметки на разных уровнях системы языка: морфология, синтаксис, семантика, дискурс. Shallow and deep синтаксис и семантика. Элементы онтологической разметки. Разметка в системе ABBYY Compeno.
- Особенности мета-текстовой разметки: жанровая, социальная, региональная классификации.

- Современные размеченные корпуса: Penn TreeBank, Prague Dependency
- Treebank, проект OntoNotes, размеченные подкорпуса Русского национального корпуса.
- Linked Data в современной лингвистике. Стандартизация разметок (стандарты ISO). Проекты связывания лингвистических ресурсов: *Net (WordNet, FrameNet, PropBank).

4. Образовательные технологии

В соответствии с требованиями ФГОС по направлению 45.04.03 «Фундаментальная и прикладная лингвистика» и с учетом специфики магистерской программы «Фундаментальная и компьютерная лингвистика» занятия лекционного типа составляют не более 20% аудиторных занятий, а удельный вес занятий, проводимых в интерактивных формах, составляет не менее 50% аудиторных занятий. Интерактивные формы обучения в данном курсе предполагают:

1. систематическое использование компьютерных презентаций (как преподавателем, так и студентами);
2. использование корпусов и других лингвистических ресурсов, доступных в сети Интернет.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82			C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

Распределение баллов по видам учебной деятельности таково:

- посещение семинарских занятий – до 8 баллов,
- уровень активности студента при подготовке к занятиям (конспектирование специальной литературы, готовность отвечать на вопросы по анализу кейсов, активное участие в дискуссиях, коллоквиумах и мозговом штурме и проч.) и во время проведения занятий (участие в обсуждениях и выполнении коллективных заданий) – всего до 32 баллов,
- качество выполнения контрольной работы (текущая аттестация) – до 20 баллов,
- успешность выполнения итогового творческого задания – до 40 баллов.

Оценка «зачтено» выставляется, если студент набрал в сумме не менее 50 баллов. Магистрант, не набравший в сумме 50 баллов, сдает зачет по всему курсу и предъявляет

преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

5.2. Критерии выставления оценок

При выставлении оценки преподаватель ориентируется на следующие содержательные критерии.

Количество баллов	Критерии оценки
95–100 (А)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».</p>
83–94 (В)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, почти все задания, предусмотренные рабочей программой дисциплины, выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.</p> <p>Обучающийся адекватно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Достаточно свободно ориентируется в учебной и профессиональной литературе.</p> <p>Почти все компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».</p>
68–82 (С)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p>

Количество баллов	Критерии оценки
	<p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне «хороший».</p>
56–67 (D)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
50–55 (E)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные рабочей программой дисциплины учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».</p>
21–49 (FX)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и</p>

Количество баллов	Критерии оценки
	<p>приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>
0–20 (F)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины не освоено. Необходимые практические навыки работы не сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены с грубыми ошибками. Дополнительная самостоятельная работа над материалом дисциплины не приведет к какому-либо значимому повышению качества выполнения учебных заданий.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Текущий контроль успеваемости студентов проводится в следующих формах: подготовка к семинарским занятиям (8 занятий – 5 баллов максимум каждое); реферат (20 баллов максимум); коллоквиум по контрольным вопросам с выполнением практического задания (максимально - 40 баллов). Для получения удовлетворительной оценки необходимо набрать минимум 60 баллов. Самостоятельная работа: при подготовке к семинару студент заблаговременно знакомится с источниками, указанными в плане семинарских занятий. Студенты по согласованию с преподавателем выбирают научный текст для реферирования; реферат представляется в форме подробного хэндаута и презентации. При подготовке к коллоквиуму предполагается знакомство с полным списком обязательной литературы.

Зачет ориентирован на следующие контрольные вопросы

1. Корпусная разметка.
2. Технологии аннотирования.
3. Пакет NLTK. Назначение. Применение для обработки лингвистических данных.
4. Методы работы с аннотаторами.
5. Критерии согласия аннотаторов.
6. Методы машинного обучения. Основные понятия.
7. Применение машинного обучения к задачам классификации.
8. Linked Data в современной лингвистике.
9. Современные методы разметки на разных уровнях системы языка.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список литературы

Основная литература

1. Казбекова Е.В. Разметка листа в средневековых латинских рукописях: рабочая гипотеза / Е. В. Казбекова // Средние века. - Москва: Наука, 2015. - Вып. 76 (3-4). - С. 193-229.
2. Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «ДИАЛОГ» - Москва, 2018
http://www.dialog-21.ru/media/4529/_-dialog2018scopus.pdf
3. Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «ДИАЛОГ» - Москва, 2016
<http://www.dialog-21.ru/media/3478/dialogue2016.pdf>
4. Torotoev G., Torotoeva A. LINGUISTIC ANNOTATION OF GRAMMATICAL CATEGORIES OF SAKHA LANGUAGE (ON EXAMPLE OF NOUN) // Proceedings of the international conference TURKLANG 2015. – Kazan, 2015 – 363-373 p.
<http://www.turklang.tatar/wp-content/uploads/201>

Рекомендованная литература

1. Lipsitz, S.R., N.M. Laird, and D.P Harrington. 1991. Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* 78(1): 156–160.
2. Bortz, J. 2005. *Statistik für Human- und Sozialwissenschaftler*. Springer Verlag.
3. Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, pp 37–46.
4. Hovy, E.H., M. Marcus, M. Palmer, S. Pradhan, L. Ramshaw, and R. Weischedel. 2006. OntoNotes: The 90% Solution. Short paper. Proceedings of the Human Language Technology / North American Association of Computational Linguistics conference (HLT-NAACL 2006).
5. Pradhan, S., E.H. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R.
6. Weischedel 2007. OntoNotes: A Unified Relational Semantic Representation. Proceedings of the First IEEE International Conference on Semantic Computing (ICSC-07).
7. Сайт ELRA (European Language Resources Association). www.elra.info
8. Сайт LDC (Linguistic Data Consortium) www ldc.upenn.edu/

7. Материально-техническое обеспечение дисциплины

Занятия по курсу «Лингвистическое аннотирование/разметка текстов» можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимо также наличие доски, чтобы преподаватель мог разбирать примеры по ходу объяснения и записывать задания.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

При необходимости рабочая программа дисциплины может быть адаптирована для обеспечения образовательного процесса лицам с ограниченными возможностями здоровья, в том числе для дистанционного обучения. Для этого от студента требуется представить заключение психолого-медико-педагогической комиссии (ПМПК) и личное заявление (заявление законного представителя).

В заключении ПМПК должно быть прописано:

- рекомендуемая учебная нагрузка на обучающегося (количество дней в неделю, часов в день);
- оборудование технических условий (при необходимости);
- сопровождение и (или) присутствие родителей (законных представителей) во время учебного процесса (при необходимости);
- организация психолого-педагогического сопровождения обучающегося с указанием специалистов и допустимой нагрузки (количества часов в неделю).

Для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся, при необходимости могут быть созданы фонды оценочных средств, адаптированные для лиц с ограниченными возможностями здоровья и позволяющие оценить достижение ими запланированных в основной образовательной программе результатов обучения и уровень сформированности всех компетенций, заявленных в образовательной программе.

Форма проведения текущей и итоговой аттестации для лиц с ограниченными возможностями здоровья устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно (на бумаге, на компьютере), в форме тестирования и т.п.). При необходимости студенту предоставляется дополнительное время для подготовки ответа на зачете или экзамене.

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
 - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
 - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
 - письменные задания оформляются увеличенным шрифтом;
 - экзамен и зачет проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
 - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
 - письменные задания выполняются на компьютере в письменной форме;
 - экзамен и зачет проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
- для лиц с нарушениями опорно-двигательного аппарата:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

- письменные задания выполняются на компьютере со специализированным программным обеспечением;
- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
 - в печатной форме увеличенным шрифтом;
 - в форме электронного документа;
 - в форме аудиофайла.
- для глухих и слабослышащих:
 - в печатной форме;
 - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - в печатной форме;
 - в форме электронного документа;
 - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
 - устройством для сканирования и чтения с камерой SARA CE;
 - дисплеем Брайля PAC Mate 20;
 - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
 - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
 - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - передвижными, регулируемые эргономическими партами СИ-1;
 - компьютерной техникой со специальным программным обеспечением.

9. Приложения

Приложение 1. Аннотация дисциплины

Предметом дисциплины (модуля) является теория и практика лингвистической разметки корпусов. Вопросы применения разметки для корпусных лингвистических исследований и машинного обучения рассматриваются и в других курсах, в частности «Корпусная лингвистика».

Курс направлен на решение следующих задач:

- ознакомление магистрантов с методами лингвистического аннотирования;
- знакомство с международными стандартами разметки в области морфологии, синтаксиса, семантики, структуры текста; обучение технологиям работы групп аннотаторов.

Дисциплина (модуль) направлена на формирование компетенций выпускника:

владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);

способностью разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты интеллектуальных и информационных электронных систем (ПК-8);

способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);

способностью разрабатывать и совершенствовать системы автоматизации и информационной поддержки лингвистических исследований (ПК-10)

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать: основные методы лингвистического аннотирования и международные стандарты разметки в области морфологии, синтаксиса, семантики, структуры текста;

Уметь: выбрать необходимый тип лингвистической разметки для решения требуемой задачи;

Владеть: технологиями оценки качества лингвистической разметки, в частности - при использовании групп аннотаторов.

Дисциплина (модуль) «Лингвистическое аннотирование / разметка текстов» является дисциплиной по выбору вариативной части цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 2 курса (3 семестр). Дисциплина (модуль) реализуется УНЦ компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: подготовка к семинарским занятиям,

реферат, коллоквиум по контрольным вопросам с выполнением практического задания; промежуточная аттестация в форме: зачет.

Общая трудоемкость освоения дисциплины (модуля) составляет 2 зачетные единицы, 72 часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – 20 часов; самостоятельная работа студента – 52 часа.

Приложение 2. Лист изменений

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение к листу изменений №1	31.08.2020г	1

Приложение к листу изменений №1

1. Структура дисциплины (к п. 2 РПД на 2020)

Общая трудоёмкость дисциплины составляет 2 з.е., 76 ч., в том числе контактная работа обучающихся с преподавателем 20 ч., самостоятельная работа обучающихся 56 ч.

2. Образовательные технологии (к п.4 на 2020 г.)

В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

3. Перечень БД и ИСС (к п. 6 на 2020 г.)

№п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru

4. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)

1	Microsoft Share Point 2010	Microsoft	лицензионное
2	Windows 10 Pro	Microsoft	лицензионное
3	Kaspersky Endpoint Security	Kaspersky	лицензионное
4	Microsoft Office 2016	Microsoft	лицензионное
5	Zoom	Zoom	лицензионное