

МРНТИ 16.01.29

**А.В.Зубов**

Международная Академия Информатизации, заведующий кафедрой информатики и прикладной лингвистики, Минского государственного лингвистического университета, доктор филологических наук, профессор, академик

**НАУЧНЫЙ ПУТЬ ДОКТОРА ФИЛОЛОГИЧЕСКИХ НАУК,  
ПРОФЕССОРА ЖУБАНОВА АСКАРА КУДАЙБЕРГЕНОВИЧА**

**Аннотация.** А.К. Жубанов занимался и занимается проблемой создания национального корпуса текстов казахского языка. Этой теме он посвятил много статей, а также выступил на многих научных конференциях, как в Казахстане, так и за рубежом. Он успешно передает свои знания аспирантам и магистрантам. В этой статье проведен анализ исследований профессора А. К. Жубанова. Профессор казахского языкознания А. К. Жубанов выступает достойным продолжателем работы своего отца, крупнейшего казахского ученого, первого профессора казахской филологии Кудайбергенова Куановича Жубанова.

**Ключевые слова:** язык, текст, термин, анализ, исследование.

**А.В.Зубов**

Халықаралық ақпараттандыру академиясы, Минск мемлекеттік лингвистикалық университетінің информатика және қолданбалы лингвистика кафедрасының меңгерушісі, филология ғылымдарының докторы, профессор, академик

**ФИЛОЛОГИЯ ҒЫЛЫМДАРЫНЫҢ ДОКТОРЫ, ПРОФЕССОР  
ЖУБАНОВ АСҚАР ҚУДАЙБЕРГЕНҰЛЫНЫҢ ҒЫЛЫМДАҒЫ ЖОЛЫ**

**Аннотация.** А. Қ. Жұбанов қазақ тілі мәтіндерінің ұлттық корпусын құру мәселесімен айналысып келеді. Бұл тақырып аясында ғалым көптеген мақалалар жазып, сонымен қатар Қазақстан мен шетелдерде өткен көптеген ғылыми конференцияларда сөз сөйледі. Ол өз білімін аспиранттар мен магистранттарға сәтті жеткізеді. Бұл мақалада профессор А. Қ. Жұбановтың зерттеулеріне талдау жасалды. Қазақ тіл білімінің профессоры А.Қ. Жұбанов өзінің әкесі, қазақтың аса ірі ғалымы, қазақ филологиясының тұңғыш профессоры Құдайберген Қуанұлы Жұбановтың еңбегін лайықты түрде жалғастырып келеді.

**Тірек сөздер:** тіл, мәтін, термин, талдау, зерттеу.

**A.V.Zubov**

International Academy of Informatization, Head of the Department of Informatics and Applied Linguistics, Minsk State Linguistic University, Doctor of philological sciences, Professor, Academician

**PATH TO SCIENCE OF DOCTOR OF PHILOLOGY, PROFESSOR  
ZHUBANOV ASKAR KUDAIBERGENOVICH**

**Annotation.** A. K. Zhubanov was engaged and is engaged in the problem of creating a national corpus of texts of the Kazakh language. He devoted many articles to this topic, and also spoke at many scientific conferences, both in Kazakhstan and abroad. He successfully passes on his knowledge to postgraduates and undergraduates. This article analyzes the research of Professor A. K. Zhubanov. Professor of Kazakh Linguistics A. K. Zhubanov is a worthy successor of the work of his father, the largest Kazakh scientist, the first professor of Kazakh philology Kudaibergen Kuanovich Zhubanov.

**Keywords:** language, text, term, analysis, research

А.К. Жубанов поступил в аспирантуру Института языкознания АН КазССР в 1967 году. Несколько позже он был командирован для повышения квалификации в Минский государственный педагогический институт иностранных языков (МГПИИЯ). В этом институте при кафедре общего языкознания функционировала секция «Статистика речи». Ее члены работали над созданием компьютерных программ получения частотных словарей по текстам. В то время МГПИИЯ не имел собственного компьютера, и члены секции работали по ночам на компьютерах других не учебных организаций. А.К. Жубанов вместе со всеми также по ночам ходил работать на ЭВМ «Минск-22». Руководитель секции МГПИИЯ А.В. Зубов научил А.К. Жубанова программировать и работать на ЭВМ «Минск-22».

В итоге, А.К. Жубанов успешно выполнил задания своего научного руководителя К.Б. Бектаева по получению с помощью ЭВМ словарей по текстам романа М. Ауэзова «Абай жолы». А.К. Жубанов стал активно выступать на научных конференциях по статистической обработке текстов в Минске, Кишиневе, Киеве, Алматы, Чимкенте и других городах Советского Союза. Здесь он проявил себя не просто как математик-программист, но и как лингвист, хорошо знающий казахский язык.

В июне 1973 года А.К. Жубанов успешно защитил в Институте языкознания Академии наук Казахской ССР кандидатскую диссертацию на тему «Статистическое исследование казахского текста с применением ЭВМ (на материале романа М. Ауэзова «Абай жолы)». В этой диссертации детально раскрылась не только процедура автоматического построения алфавитно-частотных словарей, но также частотно-алфавитных и обратных словарей. При решении этих задач на материале текстов казахского языка А.К. Жубанову, как лингвисту, пришлось разработать единую для специфических букв тюркских языков транскрипцию, помогающую программисту при создании компьютерных программ обработки текстов на тюркских языках.

В этой кандидатской диссертации А.К. Жубанов впервые попытался выявить статистические формулы, оценивающие законы распределения в казахских текстах существительных, глаголов, прилагательных и наречий. Эти формулы были использованы А.К. Жубановым при автоматической оценке статистической структуры текста романа М. Ауэзова «Абай жолы».

Развивая идеи своего учителя К.Б. Бектаева, А.К. Жубанов, обладая хорошими научными знаниями казахского языка, вновь приехал в Минск. На этот раз он совершенствовал свои знания по формализации содержания текстов на недавно образованной кафедре информатики и прикладной лингвистики МГПИИЯ. К тому времени на кафедре был установлен ЭВМ «Минск-22», где в любое время мог работать А.К. Жубанов. И это очень помогло ему добиться хороших результатов!

Он впервые перешел от статистического описания казахских текстов к формализации их семантического содержания. Им была разработана семантическая классификация знаменательных классов слов казахского языка по отношению к научно-популярным, публицистическим и художественным текстам.

Изучая на семантико-синтаксическом уровне организацию содержания трех отмеченных выше типов текстов, А.К. Жубанов выявил предметно-логическое содержание абзацев этих текстов и в дальнейшем, с опорой на специальный семантико-синтаксический формальный язык СЕМСИНТ, смог создать семантико-синтаксические формулы текстов, понятные компьютеру.

Все вышесказанное А.К. Жубанов детально описал в своей докторской диссертации «Основные принципы формализации содержания казахского текста». Она была успешно защищена в ноябре 2002 года в Институте языкознания имени А. Байгұрсынова Министерства образования и науки Республики Казахстан.

Далее остановимся на некоторых теоретически моментах исследования профессора А.К. Жубанова.

Во многих научных работах доктора филологической наук, профессора Аскара Кудайбергеновича Жубанова рассматриваются возможности использования статистических методов в изучении текстов. Так, в одной из первых его работ показаны способы использования ЭВМ (компьютеров) для изучения статистических характеристик различных авторов [1].

В дальнейшем, в своей докторской диссертации А.К. Жубанов с опорой на статистические методы построил автоматический словарь казахского языка, детально описал его структуру и принципы его использования для синтеза различных словоформ казахского языка [2, 9-16].

Высказанные А.К. Жубановым идеи о возможности использования статистического метода для выявления основного содержания текстов [3, 49], также нашли дальнейшее развитие в его докторской диссертации [2, 35-40].

Полученные в этой диссертации результаты целиком вписываются в современный вероятностно-алгоритмически подход моделирования статической составляющей содержания текста.

Многоаспектные исследования А.К. Жубанова различных казахских текстов показывают, что каждый текст по своей структуре является многоплановым и противоречивым образованием. В нем можно обнаружить социальное и индивидуальное, детерминированное и случайное, обязательное и факультативное, содержательное и формальное, глубинное и поверхностное, языковое и речевое. К тем же противоречивым особенностям текстов относятся и наличие в тексте статического и динамического аспектов.

Термины «статический» и «динамический» по отношению к тексту употребляются обычно в смысле «покой» и «движение» [4]. Как известно, текст как последовательность лексических единиц, как некоторый результат, продукт речемыслительной деятельности находится в статическом состоянии или в состоянии покоя. Текст же в процессе его порождения, восприятия и понимания считается находящимся в движении. Как отмечают многие исследователи, при воспроизведении предложений, запоминаются две группы данных о предложении: информация о его семантическом содержании и информация о его синтаксической структуре. При этом семантический аспект запоминается в качестве первого шага, а синтаксический – в качестве второго. В дальнейшем это положение было развито по отношению к тексту и свелось к констатации того, что в каждом тексте есть свой словарь и свой синтаксис. По-разному понимаются при этом такие составляющие как «словарь» и «синтаксис». Ближе всего нам та точка зрения, когда под словарем понимаются так называемые слова «содержания» или «несомые слова», а под синтаксисом имеется в виду грамматический строй вместе с «несущими» или служебными словами [5]. Таким образом, можно представить, что слова «содержания» представляют статику текста и являются отражением в тексте

некоторого множества предметов, явлений, фактов реальной действительности, а синтаксис – динамику текста, отражающую те отношения между этими предметами, фактами, явлениями, которые устанавливает автор текста в зависимости от целевой установки, типа текста, речевого опыта и целого ряда других факторов. Наличие этих двух составляющих в тексте подтверждают и эксперименты по анализу процесса понимания текста.

Как отмечает А.К. Жубанов любая ситуация, отраженная в тексте, может быть рассмотрена с той или иной позиции. Помимо этого, различный отбор элементов ситуации связан с тем, что человек воспринимает действительность в условиях некоторой вероятности, зависящей от его жизненного и языкового опыта. Все это приводит к тому, что в процессе развертывания содержания отбираются узловые элементы системы референтов, являющиеся наиболее существенными для построения данного текста. Каждый из этих элементов позволяет задавать порядок дальнейшего отбора необходимых элементов в подмножестве связанных с ним референтов [6]. А так как референтам в тексте соответствуют различные имена, то формальная структура содержания текста сводится к формальной структуре линейной последовательности материальных знаков – слов.

Анализ большого числа работ по смысловой организации текста показывает, что смысловое единство текста связано с наличием в нем некоторых «опорных слов» («топиков»), «смысловых вех», «дескрипторов», «ключевых слов», «фо-кальных точек», «ядер» и т.д.) [7], которые пронизывают весь текст от его начала и до конца.

Существуют достаточно интересные попытки формального представления смысла текста через такие опорные слова и словосочетания, связанные с построением различного рода графов. Однако, как утверждает проф. А.К. Жубанов, такая формальная связь не позволяет полностью решить задачу построения графа денотативной структуры, соответствующего всему тексту, так как его окончательная связь всех вершин, представляющих текст, осуществляется на содержательном уровне с привлечением не содержащейся в тексте информации.

Известно, что в структуре «опорных слов» или «смысловых вех» существует определенная иерархия. Часть из них является «главными», «важными» словами, «номенклатурными дескрипторами», «ключевыми» словами, «выступающими точками» всего содержания. Они определяют главный предмет сообщения и по существу являются свернутым замыслом текста. При этом, как отмечает проф. А.К. Жубанов, один и тот же референт может выражаться различными словами – контекстуальными синонимами в тех пределах, при которых возможно их взаимное приравнивание. Ввиду линейности означающего текста такие слова оказываются разнесенными по разным абзацам текста.

Анализ работ по выделению опорных слов с учетом вышесказанного позволяет предложить два формальных критерия определения главных опорных слов (ГОС): наибольшая частота употребления слова (включая все его контекстуальные синонимы и местоименные замены) и максимальное число абзацев, в которых встречается слово. Такие слова в дальнейшем становятся своеобразным центром, вокруг которого формируются другие элементы, отражающие составляющие микроситуаций (глаголы, прилагательные, наречия и т.п.).

В исследованиях проф. А.К. Жубанова говорится, что главные референты, которым в тексте соответствуют главные опорные слова, в рамках той или иной ситуации могут быть связаны с другими референтами, также являющимися относительно важными для той же ситуации, но которые являются центром не всей ситуации, а некоторых микроситуаций общей ситуации. Соответствующие им слова текста проф. А.К. Жубанов

называет второстепенными опорными словами (ВОС), и также о том, что эти слова встречаются с меньшей частотой, чем ГОС, и в меньшем числе абзацев всего текста.

Существует достаточно большое число способов определения «степени важности» для содержания текста того или иного слова или словосочетания. Их можно разделить на: 1) анкетные, 2) структурные, 3) словарные, 4) частотные, 5) синтаксические, 6) диалоговые.

Все эти методы имеют один существенный недостаток: они не исходят из текста как связной единицы. Частотные методы исходят только из частоты встречаемости слов в тексте. При таком подходе по существу учитывается не смысл текста, а свойства плана выражения текста [8].

Для выделения из текста ГОС и ВОС проф. А.К. Жубанов использует метод, где для каждого текста строится частотно-алфавитный словарь и на его основе для каждого слова вычисляется коэффициент важности по формуле:

$$K_{важ} = \frac{F \cdot m}{N \cdot n}$$

В этой формуле: F – абсолютная частота слова в тексте (в неё входит суммарная частота всех типов синонимов этого слова его ассоциативных и местоименных замен); m – число абзацев, в которых встретилось слово; N – общее число слов в тексте; n – общее число абзацев в тексте. Слова текста, имеющие наибольшие значения K<sub>важ</sub>, относят к числу ГОС и ВОС.

В данное время А.К. Жубанов занимается и занимается проблемой создания национального корпуса текстов казахского языка. На эту тему он посвятил много статей, а также выступил на многих научных конференциях, как в Казахстане, так и за рубежом. Он успешно передает свои знания аспирантам и магистрантам. Под его руководством защитили свои диссертации 5 аспирантов. Результаты своих исследований А.К. Жубанов опубликовал в 4 монографиях и 4 учебных пособиях.

Все это свидетельствует о том, что А.К. Жубанов достойно продолжает дело своего отца, крупного казахского ученого-филолога, первого профессора по казахской филологии Кудайбергена Куановича Жубанова.

## ЛИТЕРАТУРЫ:

- [1] Бектаев К.Б., Джубанов А., Зубов А.В. Автоматическое построение частотных словарей // Вестник АН КазССР. – Алма-Ата, 1970. – С. 48-53.
- [2] Жубанов А.К. Основные принципы формализации содержания казахского текста: автореф. дисс. к.ф.н. 10.02.02. – Алматы, 2002. – 57 с.
- [3] Джубанов А.К. Использование ЭВМ для выделения опорных слов текста // Актуальные проблемы компьютерной лингвистики. Тезисы докладов Всесоюзной конференции в г. Тарту, 29 – 31 мая 1990 г. – Тарту: ТУ, 1990. – 45 с.
- [4] Гальперин И.О. Текст как объект лингвистического исследования. – М.: Наука, 1981. – 31 с.
- [5] Проблемы текстуальной лингвистики. – Киев: КГУ, 1983. – 56 с.
- [6] Новиков А.И. Семантика текста и ее формализация. – М.: Наука, 1983. – 23 с.
- [7] Вейнрейх У. Опыт семантической теории // Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. – М.: Прогресс, 1981. – С. 50-176.
- [8] Гиндин С.И. Семантика текста и различные теории информации // НТИ. Сер.2, 1977. – №10. – С. 10-15.