

МОДЕЛЬ ОГРАНИЧЕННОГО ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ПРОЦЕССЕ УПРАВЛЕНИЯ

Б.Н. Судаков, А.В. Першин
(Харьковский университет Воздушных Сил им. И. Кожедуба)

Рассматривается структура системы поддержки принятия решений реального времени, указывается на необходимость применения естественного языка для взаимодействия различных групп пользователей с системой. Предлагается использование нормализованного естественного языка для построения естественно-языкового интерфейса системы поддержки принятия решений, базирующейся на математическом аппарате теории категорий и многозначной логике присутствия, позволяющих учитывать неопределенность и противоречивость процессов, протекающих во внешней среде.

естественный язык, системы поддержки принятия решений, интерфейс

Постановка проблемы. Современный уровень развития АСУ требует от руководителя и персонала новых подходов и методов организации и управления, взаимодействия с техническими средствами в виде систем поддержки принятия решений (СППР). Для эффективного внедрения которых действительно необходима разработка новых научно обоснованных методов и моделей создания естественно-языкового интерфейса (ЕЯИ) общения пользователей с СППР.

Анализ литературы [1 – 3] позволяет сделать выводы, что к настоящему времени существует ряд вопросов, связанных с организацией взаимодействия пользователей и интеллектуальных систем, которые проработаны весьма недостаточно. Во-первых, основная часть специалистов в области разработки СППР не уделяют должного внимания системе взаимодействия, что входит в противоречие с самой идеей совместного с пользователем решения задач принятия решений. Во-вторых, в настоящее время отсутствуют единые теоретические подходы к разработке всех компонентов системы взаимодействия. Наконец, разнообразие лингвистических средств и отсутствие методик их выбора с целью представления различным группам пользователей, узкая направленность языков взаимодействия не позволяет учитывать такие особенности физической реальности как неполнота, противоречивость знаний, динамика процессов, неопределенность поведения объектов. Все это приводит к необходимости разработки новых методов и моделей языковых средств взаимодействия пользователей с системой, а также средств их обработки.

В соответствии с этим **цель настоящей статьи** заключается в том, чтобы предложить модель ограниченного естественного языка для СППР.

Постановка задачи разработки ограниченного естественного языка. Исходя из подхода к разработке языка взаимодействия, изложенного в [4], можно говорить о двух уровнях языка взаимодействия. Первый уровень представлен выражениями ограниченного естественного языка (ОЕЯ) (L-язык), которые будем рассматривать как поверхностную структуру языка взаимодействия. Поверхностная структура языка взаимодействия (ЯВ) отражает представление пользователей о предметной области, решаемых задачах, целях, преследуемых ими при подготовке боевых действий. В связи с тем, что пользователи общаются между собой на одном профессиональном языке, хорошо представляют моделируемую предметную область, то им не представляет труда понимать поверхностный уровень языка взаимодействия. В тоже время, уровень L-языка не может быть понятен интеллектуальной системе. Интеллектуальная система "понимает" только уровень внутреннего языка системы (С-языка). Поэтому можно говорить, что С-язык является уровнем смысла L-языка. При общении пользователей на ОЕЯ необходимо осуществить трансляцию текстов L-языка в выражения С-языка. Опишем процесс трансляции в общем виде более подробно.

Пусть задано некоторое множество фрагментов L-языка и внутренний язык интеллектуальной системы. При этом задача трансляции заключается в преобразовании выражений $l \in L$ (где l – множество всех выражений L-языка) в некоторый текст $c \in C$ (где c – множество всех выражений С-языка), или на множестве L определяется отображение $\mu: L \rightarrow C$, позволяющее по заданному представлению l построить адекватный по смыслу текст c .

Необходимо заметить, что естественному языку присущи некоторые особенности, связанные с наличием синонимии, анафорических ссылок, эллипсисов [5]. В силу чего отображение множества текстов ОЕЯ на множество смыслов не всегда является взаимнооднозначным. Поэтому переход от текста к смыслу в общем случае крайне затруднителен. Исходя из этого, как показал проведенный анализ [5 – 7] в лингвистике используется многоуровневое описание языка. Причем единой точки зрения на количество уровней, необходимых для адекватного представления ОЕЯ, не существует. Многие пользователи решают эту проблему исходя из решаемых задач, ограничений, накладываемых на профессиональный язык пользователей, особенностей построения конструкций внутреннего языка системы, семантики предметной области. Необходимо заметить, что условия военной деятельности наложили свой отпечаток на естественный язык, который используется в военной сфере при решении различных задач. Он достаточно лаконичен, имеет жесткие грамматические средства выражения, освобожден от малоупотребительных конструкций, слов, обогащен дополнительной терминологией. Можно заметить, что военный язык имеет четко выраженные

функции каждого сообщения. При передаче приказа, указания или запроса командир всегда должен точно знать чего хочет он или адресат. В то же время этот язык многозначен и недостаточно точен. Например, запись "4 зрдн" может трактоваться и как "четвертый зрдн" и как "четыре зрдн". Поэтому в процессе создания модели ОЕЯ и процедуры его обработки естественный язык пользователей целесообразно подвергнуть нормализации.

Под нормализованным языком будем понимать подмножество естественного языка, ограниченного с точки зрения семантики и грамматики. В нем могут присутствовать элементы, которые не употребляются в естественном языке (могут выделяться числительные, имена, конец директивного предложения, аббревиатуры и т.д.).

В то же время нормализованный язык сохраняет свойства естественного языка, является более гибким и практически неограниченным с точки зрения лексики. Степень нормализации определяется с одной стороны потребностями пользователей, с другой – возможностями формализации языка и реализации процедур обработки на ЭВМ. Поэтому все источники нормализации можно разделить на лингвистические и экстралингвистические. К лингвистическим будем относить источники, коренящиеся во внутренних особенностях русского языка. Они связаны с такими свойствам языка как омонимия, синонимия, избыточность, наличие скрытого подтекста и др. Экстралингвистические ограничения обусловлены тем, что в процессе общения используют определенные термины, типы предложений, специфические обороты, а в процессе восприятия необходимо восполнять отсутствующие в предложении элементы исходя из ситуации и знаний о предметной области. Так, для построения многоуровневых иерархических теорий требуется уметь извлекать знания из связанных текстов (боевых документов, инструкций, руководств по боевой работе и т.д.). Решение задачи автоматического извлечения таких знаний, даже на теоретическом уровне, пока далеко от своего разрешения.

С учетом изложенного был обработан языковой материал, используемый в процессе решения одной из задач в штабе объединения Воздушных Сил. При этом основное внимание было уделено тому, чтобы при минимальной избыточности языка максимально удовлетворить предъявляемым требованиям. Это связано с тем, что даже при незначительном увеличении выразительных возможностей языка значительно усложняются алгоритмы обработки текстов. Поэтому основные ограничения свелись к следующим:

1. Формы предложений, используемые пользователями, сводятся к следующим трем: вопросительное – для выяснения некоторого факта, уточнения необходимой информации; повествовательное – для изменения и пополнения необходимой информации в базе знаний системы; директивное – для выполнения указания по решению задачи, а также совершение каких-либо действий.

2. Ограничения на структуру предложений: в простом предложении может быть не более одного глагола; в предложениях не могут употребляться эллипсисы (пропущенные конструкции) и анафорические ссылки.

3. Особых ограничений на лексический состав предложений не накладывается, за исключением того, что запрещается употреблять вводные слова, малоупотребительные конструкции. Объем словаря, используемый пользователями, не превышает 2500 слов.

Ограничения, накладываемые на L-язык, сужают и границы использования С-языка. Анализ возможностей по формализации естественного языка [5 – 7] показывает, что на уровне отдельного предложения можно описать упрощенную структуру С-языка следующего вида:

*PF_i<роль объекта: имя объекта, служебная информация (СИ);
(роль подобъекта: имя подобъекта, СИ);
роль подобъекта: имя подобъекта, СИ)>*.

Данная структура позволяет: формировать категории предметной области, ставить задачи системе в обобщенном виде, описывать частные примеры концептов, осуществлять поиск и корректировку необходимых знаний.

Модель ограниченного естественного языка. Для перехода от текста естественного языка к его смыслу, представленному выражением С языка, обратимся к следующему выражению: "41 иап уничтожает маловысотные цели до рубежа R1". Данное повествовательное предложение говорит о том, что в базу знаний необходимо записать частный пример объекта-процесса "уничтожение". Его запись на С-языке будет иметь следующий вид:

*ЗАП <Уничтожение, объект – процесс,
(СрУН: 41 иап;
ОбУн: маловысотные цели;
МУн: до рубежа R1)>*,

где СрУн, ОбУн, МУн – соответственно роли, выполняемые подобъектами: "являться средством уничтожения", "являться объектом уничтожения", "являться местом в пространстве".

Сопоставляя поверхностный и глубинный уровни представления текста можно заметить, что процесс перехода ТЕКСТ-СМЫСЛ должен состоять в поиске входящих в объект морфизмов (ролей), имен подобъектов (признаков), через которые описывается этот объект. С учетом того, что имена объектов, а также входящие в объект и исходящие из него морфизмы (роли) описываются целыми группами слов (словосочетаниями), то необходимо учитывать направленные связи как между словами, так и между словосочетаниями различных уровней, а это позволяет осуществить аппарат системы синтаксических групп (ССГ). Аппарат ССГ, являясь обобщением деревьев синтаксического подчинения и сис-

темы составляющих [8], представляет собой материал для формирования глубинной структуры – выражений С-языка.

Сущность метода представления текста с использованием аппарата ССГ заключается в следующем. Пусть x – произвольная непустая цепочка над словарем, а X – некоторое множество непустых подмножеств x . Элементы X будем называть СГ. Граф $\{X, \rightarrow\}$ называется системой синтаксических групп на x , если он удовлетворяет следующим аксиомам [8]:

- A1. X содержит x и все одноэлементные подмножества x ;
- A2. Если $E1, E2 \in X$, то либо $E1 \cap E2 = 0$, либо $E1 \subseteq E2$, либо $E2 \subseteq E1$;
- A3. Если $E1, E2 \in X$ и $E1 \rightarrow E2$, то $E1$ и $E2$ вложены в одну и ту же СГ;
- A4. Если $E1 \rightarrow E2$ и E – произвольная СГ, то E и $E1 \cup E2$ не зацепляются;
- A5. Если $E1 \rightarrow E2, E3 \rightarrow E4$, то множества $E1 \cup E2$ и $E3 \cup E4$ не зацепляются.

Аксиомы A1, A2 описывают множество x , A3 – A5 – отношение " \rightarrow ". Система синтаксических групп представляет собой поверхностно-синтаксический уровень языка взаимодействия.

По аналогии с размеченными деревьями подчинения и размеченными системами составляющих введем размеченные ССГ (РССГ), приписывая метки как узлам, так и дугам графа (рис. 1). Формально размеченную ССГ можно определить как шестерку

$$\langle X, \rightarrow, W, Z, \varphi, \phi \rangle, \tag{1}$$

где $\{X, \rightarrow\}$ – ССГ; W и Z – конечные множества; φ – отображение X во множество всех подмножеств W ; ϕ – отображение множества дуг графа $\{X, \rightarrow\}$ в Z . Элементы W – метки при СГ, элементы Z – при стрелках.

Для наглядности построим размеченную ССГ для предыдущего примера. В отличие от принятого в литературе изображения стрелок от "хозяина" к "слуге" будем использовать обратное изображение от "слуги" к "хозяину". Это позволит получить, как будет показано в дальнейшем, структуру изоморфную направленной сети для категорной модели знаний.

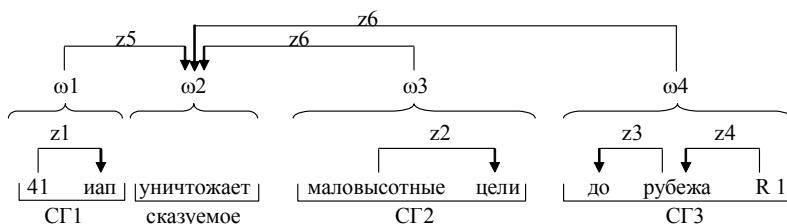


Рис. 1. Пример размеченной ССГ

В классической трактовке элементы z являются отношениями синтаксического подчинения, определяемые функцией тех или иных частей речи в предложении. Для русского языка количество таких отношений составляет от 10 до 50 и зависит как от предметной области, так и от того, где прове-

дена граница между семантикой и синтаксисом при анализе языка. В прикладной лингвистике размеченная система синтаксических групп представляет собой глубинно-синтаксический уровень естественного языка.

Представление синтаксических связей с помощью РССГ в основном совпадает с принятыми представлениями о грамматической структуре предложения, Дополнения зависят от сказуемого, определения – от определяемого и т.д. Поэтому любые предложения могут быть представлены РССГ.

Рассмотрим более подробно СГ и те отношения, которые существуют внутри СГ.

Как показал ряд исследований, наиболее полная СГ может быть представлена следующей цепочкой:

Сс, От, Пр, КС, КЧ, ПЧ, П, С, И,

где Сс – сочинительный союз; От – отрицание; Пр – предлог; КС – кванторное слово; КЧ – количественное числительное; ПЧ – порядковое числительное; П – прилагательное; С – существительное; И – имя.

Многие исследователи в области лингвистики рассматривают отношения внутри СГ, используя свою терминологию и свои взгляды [9, 10] на образование этих отношений. Опираясь на эти представления и учитывая семантику связей между словоформами, ряд названий глубинно-синтаксических отношений, которые приняты в традиционной лингвистике, можно заменить на семантические, используя особенности как предметной области, так и внутреннего языка системы. Исходя из этого можно выделить следующие типы отношений:

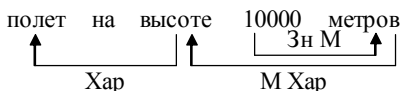
– *являться именем (Им)* – в качестве имени могут выступать ПЧ, строки, заключенные в кавычки, например,

Авианосец ←^{Им} "Нимитц", 41 →^{Им} *иап.* ;

– *являться свойством (Св)* – выражает, как правило, связь между существительным и прилагательным в СГ, например:

боеготовый ←^{Св} *дивизион* ;

– *являться характеристикой (Хар)*, *являться значением меры характеристики (ЗнМ)* и *являться мерой характеристики (М Хар)* – данные отношения возникают в группе слов, характеризующих какой-то объект, например, для группы слов "полет на высоте 10000 метров":



Если значение характеристики дано качественно, например, малая высота, то устанавливается отношение *являться качественным значением (Кач)*:

малая —^{Кач}→ высота .

В соответствии с синтаксисом русского языка с существительными могут находиться количественные числительные, кванторные слова. В этом случае могут быть установлены отношения "являться количественным значением" (Кол) и "быть квантором" (Кв). Кроме того, многие существительные могут выступать в функции определения относительно существительных других групп. Такие отношения будем называть "Атрибутивными" (Атр). В некоторых СГ – типа "зона поражения", "зона целеуказания", существительные образуют несвободное сочетание, имеющее самостоятельное толкование. Такие структуры будем рассматривать как отдельные понятия. Заметим, что в русском языке некоторые элементы, в частности, сказуемое и обстоятельства могут выражаться различными частями речи. Обстоятельства, как правило, выражаются существительными в косвенных падежах, наречиями и деепричастиями или сочетаниями слов, включающих эти части речи. Роль обстоятельств заключается в уточнении (Ут) ситуации, описываемой управляющим словом, например:

наводится ←^{Ут} — одновременно,

что приведет к образованию новой роли, уточняющей ситуацию наведение. Этот тип отношения используется и при определении связи между существительными СГ и вопросительными словами "Какой", "Сколько".

Если управляющее слово выражено сочетанием глагола и именной группы, главным словом которой является отглагольное существительное, то последнее в этом случае несет основную смысловую нагрузку, глагол же выполняет вспомогательную роль. Между ними устанавливается отношение "вспомогательное" (Всп), например:

обеспечивает ←^{Всп} — сопровождение .

Соотнесение высказывания с некоторым моментом времени может быть достигнуто использованием глагола в соответствующем времени, а модальность получена за счет использования слов, выражающих алетические, деонтические и эпистемические модальности. В этих случаях будем использовать отношение модальности (Мод), например:

необходимо —^{Мод} → распределить .

Семантика отношения определяется, как правило, семантикой "слуги".

При употреблении отрицательных частиц с управляемым словом, а также слов типа, какой-то, кто-либо с существительными СГ используется операторное отношение (Оп). Для выделения СГ необходимо учитывать следующие знания о языке: gr_1 – грамматическую информацию о словоформах (часть речи, род, число, падеж и т.д.); sin_1 – синтаксическую информацию о словоформах (с какими частями речи и как сочетаются); sem_1 – семантическую информацию о словоформах (обозначаемые понятия, смысловая сочетаемость с другими словоформами и др.).

Исходя из изложенного, можно выделить СГ и установить связи между словоформами (СГ) предложения, используя правила следующего вида:

$$H: S_i(\text{gr}_i, \text{ sint}_i, \text{ Sem}_i); S_j(\text{gr}_j, \text{ sint}_j, \text{ sem}_j): D \rightarrow R_{ij}, \quad (2)$$

где H – условие применимости правила (например, согласованность gr , sint и sem информации); D – информация о расположении анализируемых словоформ во фразе; R_{ij} – установленная связь между словоформами i и j .

Для того, чтобы установить связи СГ с другими единицами предложения заметим, что вклад некоторых словоформ в семантику ситуации является определяющим. Это, прежде всего, глаголы и глагольные формы (отглагольные существительные, прилагательные и др.). В дальнейшем группы слов, которые в предложении несут основную смысловую нагрузку, будем называть группой управляющего слова (ГУС). В естественных языках объекты предметной области, обозначаемые отдельными лексическими единицами, имеют до 6 смысловых компонентов [5, 10]. Совокупность семантической и грамматической информации о ГУС носит название модели управления (МУ). В общем виде МУ может, быть представлена в виде

$$r_{k1}: \text{СГ}(\text{gr}_{k1}, \text{ sem}_{k1}), \dots, r_{kn}: \text{СГ}(\text{gr}_{kn}, \text{ sem}_{kn}), \quad (3)$$

где r_{ki} ($i = 1, \dots, n$) – i -ая роль, выполняемая СГ в k -й модели управления; gr_{ki} , sem_{ki} – совокупность грамматической и семантической информации о синтаксической группе. Например, концепт "уничтожение" может описываться через следующие семантические признаки (входящие в него роли или морфизмы): являться средством уничтожения (СрУн), объектом уничтожения (ОбУн), местом в пространстве (МУн), моментом времени (ВрУн), целью (ЦУн). В качестве частных примеров концепта "уничтожения" выступают соответствующие СГ – 41 иап, маловысотные цели, до рубежа R I. В случае совпадения грамматической и семантической информации о синтаксических группах с соответствующей данными модели управления структура предложения, приведенная на рис. 2, примет вид:

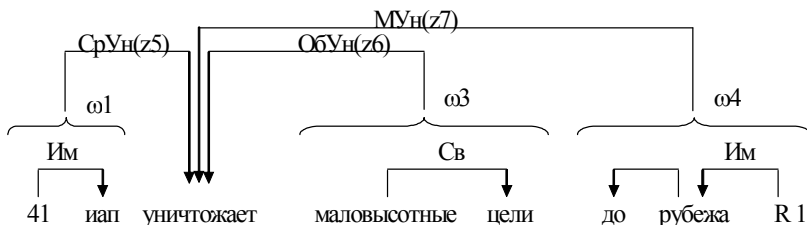


Рис. 2. Пример структуры семантико-синтаксической группы

Заметим, что $z5$, $z6$ и $z7$ (рис. 1) представляют собой имена ролей, которые выполняют подобъекты в объекте "уничтожение", $\omega1$, $\omega3$, $\omega4$ – имена подобъектов (признаков).

Как видно из рис. 2, подобная структура полностью отражает как синтаксическую, так и семантическую структуры фразы (глубинный уровень языка взаимодействия) и является промежуточным звеном между ССГ и уровнем смыслов. В дальнейшем это представление будем именовать структурой семантико-синтаксических групп (СССГ).

Исходя из изложенного, процесс перехода от поверхностного уровня языка, взаимодействия к уровню смыслов можно представить как последовательное преобразования вида

$$l \xrightarrow{\mu_1} ССГ \xrightarrow{\mu_2} СССР \xrightarrow{\mu_3} c, \quad (4)$$

где l , c – соответственно выражения ОЕЯ и внутреннего языка системы; μ_1 , μ_2 , μ_3 – процедуры, позволяющие переходить от выражений ОЕЯ к ССГ, СССР и внутреннему представлению системы.

После осуществления указанных операций предложение примет следующий вид:

<Уничтожает, объект процесс, (СрУн: 41 иап; МУн: до рубежа R1; ОбУн: маловысотные цели) >.

Можно заметить, что место подобъектов в выражении С-языка выбирается произвольно, а конкретная информация определяется по роли, выполняемой соответствующим подобъектом.

Все изложенное ранее относилось к простому предложению русского языка. Однако, описание реальной действительности происходит с использованием более сложных естественно-языковых структур, как правило, определенным образом связанных простых предложений, т.е. сложных предложений. При анализе сложных предложений можно пойти двумя путями:

1. Создать модель сложного предложения.
2. Осуществить формирование сложных высказываний как совокупность более простых сегментов, связанных операциями топологической булевой алгебры (\cap , \cup , \rightarrow , \neg).

Первый путь возможен. Уже сейчас существует достаточно много хороших моделей сложных предложений [7]. Но, в то же время, данный подход имеет ряд недостатков, которые вынуждают отказаться от него. И прежде всего это большая сложность алгоритмов обработки текстов, что требует значительных затрат памяти и увеличивает время реакции системы на сообщение. Естественно, что системы с такими характеристиками неприемлемы при работе пользователей с СППР реального времени. Исходя из этого при разработке языка взаимодействия был выбран второй подход. Тем более, что структура С-языка позволяет формировать из простых объектов более сложные.

Выводы. 1. Для перехода от текста ограниченного естественного языка к внутреннему представлению необходимо использовать модель языка,

основанную на многоуровневом описании. Тексты ограниченного – естественного языка представляют поверхностный уровень языка. Поверхностно-синтаксический уровень языка задается структурой синтаксических групп. Глубинный – структурой семантико-синтаксических групп, что соответствует размеченной системе синтаксических групп. При этом связи между словоформами (синтаксическими группами) устанавливаются не только с использованием знаний о синтаксисе языка, но и знаний о проблемной области и семантике внутреннего языка системы. Уровень смыслов, представлен выражениями С-языка.

2. Для реализации языка в соответствии с моделью необходимо разработать следующие алгоритмы анализа естественно-языковых сообщений: построения системы синтаксических групп; построения структуры семантико-синтаксических групп; преобразования данной структуры в выражения С-языка.

ЛИТЕРАТУРА

1. *Системы поддержки принятия решений: проектирование, применение, оценка эффективности* / Под ред. Б.М. Герасимова. – Севастополь: МОУ НИЦ "Государственный океанариум", 2004. – 320 с.
2. *Моделювання бойових дій військ (сил) протиповітряної оборони та інформаційне забезпечення процесів управління ними* / В.П. Городнов, Г.А. Дробах, М.О. Єрмошин, Є.Б. Смірнов, В.І. Ткаченко. – Х.: ХВУ, 2004. – 409 с.
3. *Попов Э.В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ.* – М.: Наука, 1987. – 288 с.
4. *Подход к разработке лингвистического обеспечения экспертной системы* / Б.Н. Судаков и др. // *Материалы XIII между. НПК «Информ. технологии: наука, техника, технология, образование, здоровье».* – Х.: НТУ „ХПИ”, 2005. – С. 53.
5. *Попов Э.В. Общение с ЭВМ на естественном языке.* – М.: Наука, 1982. – 360 с.
6. *Жолковский А.К. Модель "СМЫСЛ – ТЕКСТ"/ В кн. Энциклопедия кибернетики. Т. 2.* – Киев, 1974.
7. *Искусственный интеллект: В 3-х кн. Кн. 1. Системы общения и экспертные системы: Справочник* / Под ред. Э.В. Попова. – М.: Радио и связь, 1990. – 464 с.
8. *Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения.* – М.: Наука, 1985. – 234 с.
9. *Мельчук И.А. Опыт лингвистических моделей "смысл- текст". Семантика, синтаксис.* – М.: Наука, 1984. – 165 с.
10. *Апресян Ю.Д. Лексическая семантика.* – М.: Наука, 1974. – 190 с.

Поступила 2.02.2006

Рецензент: доктор физико-математических наук, профессор С.В. Смеляков, Харьковский университет Воздушных Сил им. И. Кожедуба.