

УДК 811.512

DOI 10.25587/2310-5453-2025-3-6-19

Original article

Embedded sentences as a defining feature of Altaic languages: evidence from Korean, Sakha, Even and Evenki

Ducksoo Kang

Hankuk University of Foreign Studies, Seoul, South Korea

✉ kangds@hufs.ac.kr

Abstract

The article substantiates the proposition that the syntactic construction of “embedded sentence” represents one of the foundational features of the Altaic languages, functioning as an attribute to a noun or as a complementary subordinate clause within a single clause. Despite this fact, the feature has not yet been recognized as a general linguistic characteristic distinguishing the Altaic family from other language families, particularly the Indo-European, and its diagnostic potential for establishing Altaic kinship remains understudied. Using data from Korean, Sakha (Yakut), Evenki and Even languages, the study demonstrates that embedded sentences consistently function as noun modifiers or complementary clauses, elucidating the meaning of the main clause. For instance, the constructions [Min oŋor-dox-pun] in Sakha and [Næ-ga mand-in gəs-il] in Korean (meaning “look at what I have made”) are embedded sentences serving as complements to verbs. Analysis of Tungusic languages reveals similar double-subject structures confirming the universal nature of this phenomenon in Altaic linguistics. The study shows that embedded sentences often marked by nominalization and fulfilling the roles of subject, object or predicate, constitute a central grammatical principle rather than a peripheral phenomenon. In conclusion, it is argued that embedding possesses diagnostic potential comparable to other established features of Altaic languages, such as agglutination or vowel harmony and should be considered as a key criterion for their genealogical and typological classification within the Altaic hypothesis.

Keywords: Altaic languages, embedded sentences, syntactic typology, language universals, Sakha (Yakut) language, Korean language, Tungusic languages, comparative linguistics

For citation: Kang D. Embedded sentences as a defining feature of Altaic languages: evidence from Korean, Sakha, Even and Evenki. *Arctic XXI Century*. 2025, No 3. P. 6-19. DOI: 10.25587/2310-5453-2025-3-6-19

Оригинальная научная статья

Вложенные предложения как характерная черта алтайских языков (на материале корейского, якутского, эвенского и эвенкийского языков)

Д. Канг

Университет иностранных языков Хангук, Сеул, Южная Корея

✉ kangds@hufs.ac.kr

Аннотация

В статье обосновывается положение, согласно которому синтаксическая конструкция «вложенное предложение» (embedded sentence) представляет собой одну из системообразующих черт алтайских языков, функционируя как определение к имени существительному или как дополнительное придаточное предложение в составе единой клаузы. Несмотря на это, данный признак до сих пор не получил признания в качестве общей языковой характеристики, отличающей языковую алтайскую семью от других, в особенности индоевропейской, а его диагностический потенциал для установления алтайского родства остается недостаточно изученным. На материале якутского, корейского, эвенкийского и эвенского языков демонстрируется, что вложенные предложения последовательно выступают в роли определения при имени или дополнительного придаточного, раскрывая смысл главной клаузы. Так, например, конструкции [Min ojor-bur-run] в якутском и [Næ-ga mand-in gəs-il] в корейском (со значением «посмотри то, что я сделал») являются обособленными предложениями, выполняющими функцию дополнения при глаголах *kör* и *bo-ara*. Анализ тунгусо-маньчжурских языков выявляет аналогичные структуры с двойным подлежащим, подтверждающие универсальный характер данного явления в алтаистике. Показано, что вложенные предложения, часто маркированные номинализацией и выполняющие роли подлежащего, дополнения или сказуемого, представляют собой центральный грамматический принцип, а не периферийное явление. В заключении утверждается, что вложенные предложения обладают диагностическим потенциалом, сопоставимым с такими установленными чертами алтайских языков, как агглютинация или гармония гласных, и должна рассматриваться как ключевой критерий их генеалогической и типологической классификации.

Ключевые слова: алтайские языки, вложенные предложения, синтаксическая типология, языковые универсалии, саха (якутский язык), корейский язык, тунгусо-маньчжурские языки, сравнительное языкознание

Для цитирования: Канг Д. Вложенные предложения как характерная черта алтайских языков (на материале корейского, якутского, эвенского и эвенкийского языков). *Арктика XXI век.* 2025, № 3. С. 6-19 (на англ). DOI: 10.25587/2310-5453-2025-3-6-19

Introduction

What is an embedded sentence? The concept of embeddedness in Altaic languages should be distinguished from that in Indo-European languages.

In Indo-European languages, an *embedded clause* is typically defined as a group of words that express a single idea using a subject and a verb. A dependent clause is inserted into an independent (main) clause to provide additional information.

Example 1

(a) *Jane, as soon as she heard about her mom, rushed to the hospital.*

(b) *The house, which has a beautiful blue roof, stands on a hill.*

In the sentence (a), an adverbial clause is embedded and in sentence (b), an adjectival (relative) clause is embedded – both set off by commas. If necessary, embedded clauses in Indo-European languages are enclosed in parentheses, commas, or dashes. According to Steffani, “Embedded complex sentences contain an independent clause and a dependent clause or phrase. As we know, a dependent clause and phrase must be attached to an independent one in order to have a complete meaning. Embedded phrases or clauses can be found at the beginning or end of a sentence” [1]. In the sentence “*The toy that I want is on sale,*” Steffani identifies “*that I want*” as an embedded clause. However, in this instance this clause is subordinate rather than embedded in the Altaic sense.

In Altaic languages an *embedded clause* refers to a clause that functions as a modifier of a noun or a nominal clause within a larger sentence. These clauses provide additional information about the noun or the main clause they are part of. Unlike in Indo-European languages, parentheses (such as brackets, commas or dashes) are not used to set them apart.

Example 2

Sakha: [*Min oŋor-bup-pun*] *kör.*

Korean: [*Næ-ga mand-in gəs-il*] *bo-ara.*

→ See what I made.

This paper aims to identify common characteristics of embedded sentence in Altaic languages by comparing examples from Korean, Sakha and Tungus (Evenki and Even).

Embedded sentences in Korean

The structure of incorporating sentence into another is known as embedding and it is widely used in Korean. A sentence that functions as an attribute is called “an embedded sentence” or “a lower sentence”, while the sentence that contains the embedded sentence is referred to as a “matrix sentence” or a “higher sentence” [2, p. 14].

In Indo-European languages, such structures are explained through the concept of subordination. According to Asher, “a subordinate clause is, thus, one contained within a larger and superordinate clause. Nevertheless, it is because subordination is characteristically marked internally that subordinate clauses constitute a significant syntactic class” [3, p. 3853].

Subordinate clauses are traditionally categorized based on their functional similarity to three major parts of speech:

Example 3-a

(a) *I remember [that she slapped him]* → Nominal

(b) *They arrested the man [who attacked us]* → Adjectival

(c) *She left [before it was over]* → Adverbial [3, p. 3854-3855].

However, embedded sentences in Korean are syntactically and functionally different from subordinate clauses in Indo-European languages.

Example 3-b

(a-1) *Na-nin [ginjo-ga gi-ril ttær-jæt-ta-nin-gæs-il] giækha-n-da.*

I-NOM she-NOM him-ACC slap-PST-ADNOM-NOMLZ-ACC
remember-PRS

(b-1) *Gi-dil-in [uri-ril gongjak-han] saram-il čepohæ-t-ta.*

They-PL-NOM us-ACC attack-ADNOM man-ACC arrest-PST

(c-1) *Ginjæ-nin [kkitna-gi-zæne] ttæna-t-ta.*

She-NOM be over-NOMLZ-before leave-PST

These examples demonstrate that Korean sentence structure differs fundamentally from English. In Korean, embedded sentences can function as a subject, object or predicate.

Example 4

(a) *Hyæŋ-in [næ-ga čæg-il zal il-nin-dago] malhæ-t-ta.*

Brother-NOM I-NOM book-ACC well read-PTCP[PRS]-COMPL say-PST
→ Brother said that I read a book well.

(b) *Gi-nin [Næ-ga čæg-il ilg-in gæs-il] al-go it-ta.*

he-NOM I-NOM book-ACCc read-PTCP[PST] AUX.N-ACC know-
CVB AUX.V-PR

→ He knows that I read a book.

(c) [Næ-ga ilg-in gəs-in] zo-in čæg-ida.

I-NOM read-PTCP[PST] AUX.N-NOM good-PTCP book-COPL

→ It is a good book that I read.

In Korean, embedded sentences often appear in nominalized form using the nominalizers {-im}, {-gi} or a dependent noun like *gəs* [2, p.17].

Example 5

(a) Gi-nin [čæg-il il-gi-ril] zoaha-n-da.

He-NOM [book-ACC read-NOMLZ-ACC like-PRS

→ He likes to read a book.

(b) Na-nin [gi-ga čæg-il il-gəss-im-il] al-at-ta.

I-NOMm [he-NOM book-ACC read-PST-NOMLZ-ACC] know-PST

→ I knew that he read a book.

In Korean embedded sentences are also used as adnominal components, functioning similarly to adjectives by modifying a following noun or noun phrase. This process is known as adnominalization, which is distinct from both relativization and complementation [2, p.15].

Example 6

(a) [Næ-ga žab-in gəs-in] tokki-jætta.

[I-NOM catch-ADNOM thing-NOM] rabbit-COPL[PST]

→ It was a rabbit that I caught.

(b) [Næ-ga tokki-ril žab-atta-in gəs-i] aljəž-ætta.

[I-NOM rabbit-ACC catch-PST-ADNOM thing-NOM] be known-PST

→ It was known that I caught a rabbit.

As embedding is a highly typical syntactic structure in Korean, similar phenomena are also regularly observed in the Sakha language.

Embedded sentences in Sakha

In the Sakha language, embedded sentences perform various syntactic functions including those of subject, object, predicate and adverbial clause.

Example 7-a (Sakha)

- [kini utuj-a sit-ar-in] kör-d-iim.

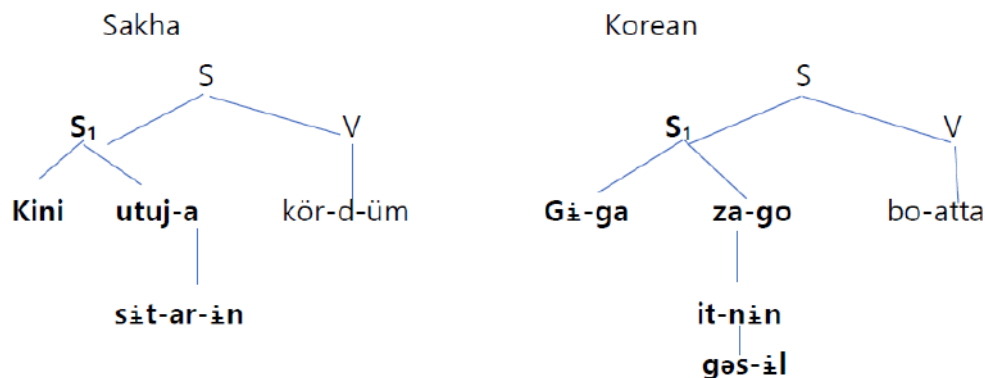
He sleep-CVB[sim] AUX.V-PTCP[PROG] -ACC see-PRET-1SG

→ I saw (that) he was sleeping [4, p. 229; 5, p. 58].

In the above example, the clause [kini utuj-a sit-ar-in] functions as an embedded object. The subject of the embedded clause is *kini* ('he'), while the main clause has *Min* ('I') as its subject. The tense of this sentence is preterite, marked by the suffix {-d-} in the main verb *kör-d-iim* in sentence-final position, which determines the tense of the entire sentence. As the embedded clause cannot carry a tense marker. The clause [kini utuj-a sit-ar-in] cannot take a tense marker. The suffix {-a} in *utuj-a* is a converb marker

indicating simultaneity, while the verbal phrase *sit-ar-in* acts as an auxiliary with $\{-ar-\}$ marking progressive aspect and $\{-in\}$ marking the accusative case. A corresponding structure in Korean as follows:

Example 7-b (Korean)



- [*Gi-ga za-go it-nin gæs-il*] *bo-atta*.

He-NOM sleep-CVB AUX.V-PROG AUX.N-ACC see-PST

→ I saw (that) he was sleeping.

Structural comparison	
Sakha	[<i>kini utuj-a sit-ar-in</i>] <i>kör-d-üm</i>
Korean	[<i>Gi-ga za-go it-nin gæs-il</i>] <i>bo-atta</i>

In both Sakha and Korean, S_1 functions as the embedded clause. In the traditional Sakha grammar, this type of structure is explained as a subordinate clause [4, p. 231; 5, p. 69]. However, the concept of subordination is insufficient to describe the syntactic structure found in Sakha.

Example 8-a (Sakha)

- [*Ujbaan silzi-bit-in*] *kini kepsee-betex*.

Ivan be-PTCP-ACC he tell-NEG.PTCP3SG

→ He didn't tell (that) Ivan was (there).

This structure is better understood through the lens of embedding. Its Korean equivalent is:

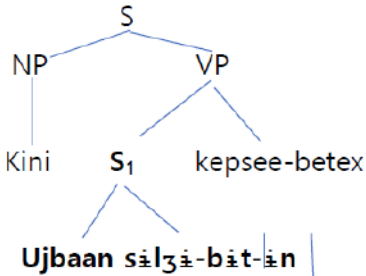
Example 8-b (Korean)

- [*Iban-i it-tanin gæ-l*] *gi-ga malha-ži an-atta*.

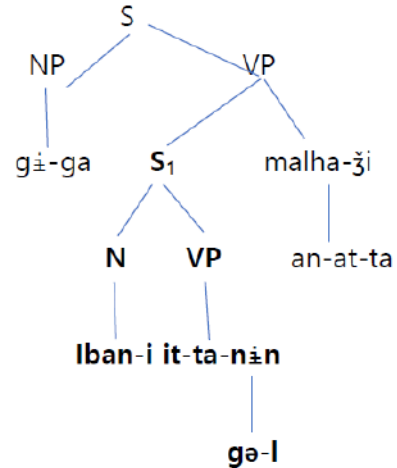
Ivan-NOM be-PTCP AUX.N-ACC he-NOM tell-CVB AUX.NEG-PST

Syntactic diagram:

Sakha



Korean



In both examples, the phrases [*Ujbaan silzi-bit-in*] in Sakha and [*Iban-i it-ta-nin ga-l*] in Korean are embedded clauses with sentential structure.

Korkina, Ubrjatova et al. do not account why participial constructions with {-bit/-batax} function as subject, modifier or complement in processes of substantivization, adjectivization or substitution. For instance, the sentence *Kiis-pin kitta körsü-betex-pitten olus xomoj-o-bun* is analyzed by the concept of subordination in their research [4, p. 232; 5, p. 69].

Example 9

- [*Kiis-pin kitta körsü-betex-pitten*] olus xomoj-o-bun

Daughter-1SG.ACC with meet-NEG.PTCP-1SG.ABL very suffer-PRS-1SG

→ I am deeply pained because I have not met my daughter.

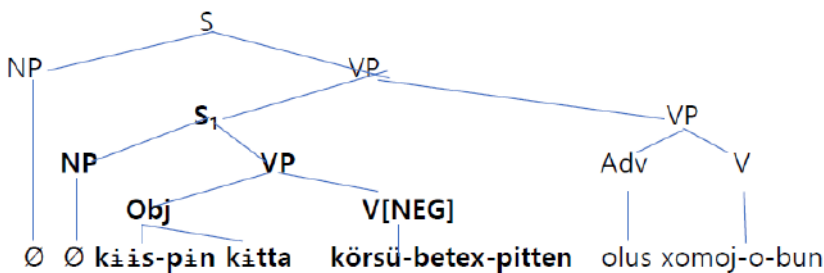
Its equivalent in Korean is as follows:

Korean: *tal-gwa manna-ji mot-hae (na-nin) mæu görob-ta.*

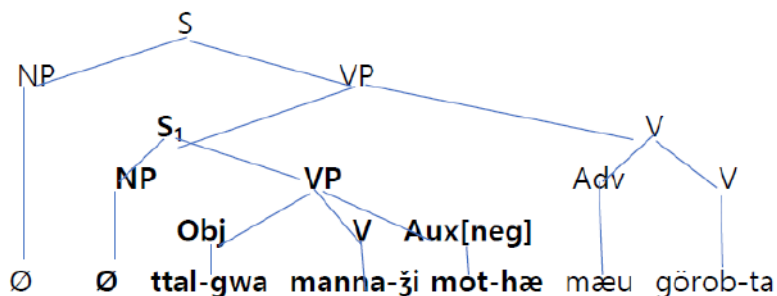
daughter-[with] meet-CVB NEG-AUX (I-NOM) very suffer-PRS

The syntactic trees for these sentences are as follows:

Sakha



Korean



Again, in both cases S₁ is more appropriately analyzed as an embedded clause rather than a subordinate clause.

A point of contention arises in cases involving modality, such as *buollara* ‘be certain’ and *dili* ‘seem’.

Example 10

Sakha: [Min da bar-ia-m] *buollara*.

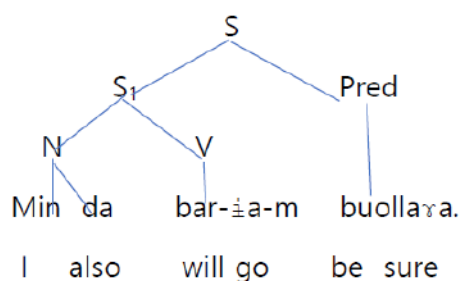
Korean: [Na-do ga-l-ge] *hwaksilhada*.

→ It is certain that I will also go.

Here, [Min da bar-ia-m] and [Na-do ga-l-ge] are embedded clauses expressing the proposition ‘I will also go.’ In traditional Sakha grammar, *buollara* is treated as a modal. If modals can serve as predicates in complex sentences, then these clauses should be understood as embedded sentences.

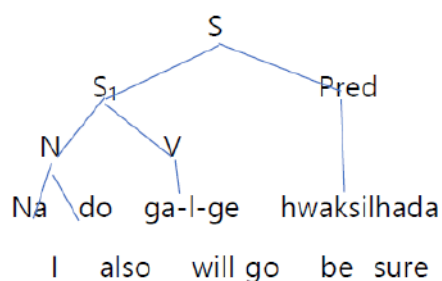
Both are formed as:

Sakha



I also will go be sure

Korean



I also will go be sure

Example 11

[Bu surax-i kini isti-betex-xe] *dili*.

This news-ACC he hear-PTCP[PST.NEG]-DAT PTCL[seem]

→ It seemed that he did not hear this news.

Here again, the modals *buollara* and *dili* function as predicates, dominating embedded sentences.

Embedded sentences in Tungusic languages (Evenki and Even)

Embedding is one of the most natural syntactic phenomena in Tungusic languages. A key structural feature to understand within embedding constructions is the double-subject sentence – a sentence that includes two distinct subjects. This construction is commonly found in both Even and Evenki [6, p. 179; 7, p. 71].

Example 12 (Even)

(a) *Oran bodel-en ŋonam.*

reindeer leg-POSS3SG long

→ Reindeer's leg is long.

(b) *Noŋan min dī-β.*

he I size-POSS1SG

→ He is my size.

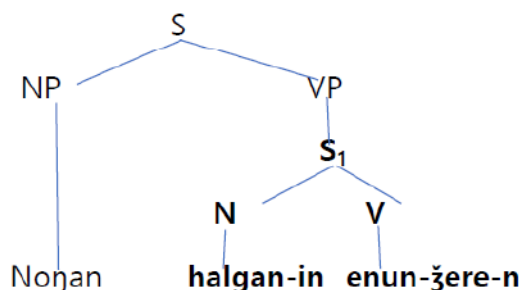
Example 13 (Evenki)

- *Noŋan halgan-in enun-žere-n.*

he foot-POSS3SG ache-PRS-3SG

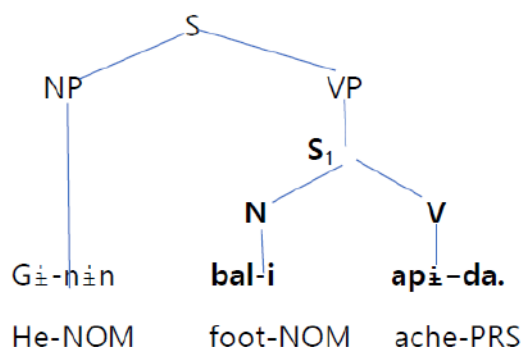
→ His foot hurts.

In Example 13 the phrase *Noŋan halgan-in* forms a possessive construction meaning 'his foot'. However, at a deeper syntactic level, the meaning is closer to "he aches in the foot," indicating a double-subject structure. The embedded clause *halgan-in enun-žere-n* is dependent on the subject *Noŋan*. Thus, this sentence is analyzed as follows:



This structure closely corresponds to Korean as follows: *Gŭ-nŭn bal-i apida* ("His foot hurts").

Korean



Example 14 (Even)

(a) *Noŋan [min ɲeɭel-ri-β] ha-n.*

He I-NOM be scared-PTCP[PROG]-ACC be aware-PRS.PF3SG
→ He was aware that I was scared.

Korean equivalent:

Gi-nin [næ-ga gəbmæg-in gəs-il] al-at-ta.

He-NOM I-NOM be scare-PTCP AUX.N-ACC be aware-PST

(b) *Erek [min ma-ri-β] huličay-u.*

this I-NOM kill-PTCP[PRS]-POSS1SG fox-POSS1SG.

Korean equivalent:

igəs-in [næ-ga žug-in] jəu-da

This-NOM I-NOM kill-PTCP fox-COP

(c) *[Gorla hər-žijə-du-β] kʲeč-čot-te.*

far dapart-PTCP[FUT]-DAT-POSS1SG see-ASP[COMM]-PRS.PF3PL

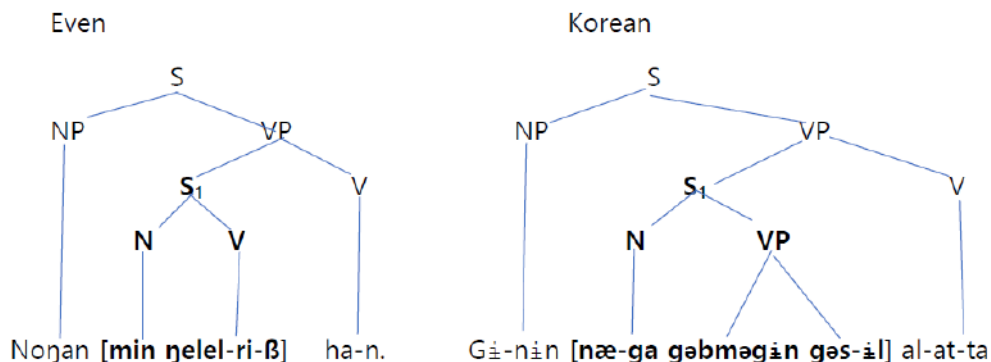
Korean equivalent:

Gi-dil-in [(næ-ga) məli galkka-bwa] bo-gon hæ-ta

They-NOM I-NOM far go-FUT.CVB see-ASP[REP] do-PST [8, p. 210-211; 9, p. 68].

The sentences in Example 14 demonstrate clear parallels in embedded sentence constructions between Even and Korean. The bracketed segments represent embedded clauses and that their syntactic structures are highly comparable.

Syntactic diagram of Example 14-(a):



Conclusion

It is widely acknowledged that the basic order in Altaic languages is SOV (Subject-Object-Verb). While this word order is significant from a typological perspective, it does not serve as sufficient evidence for genealogical classification. Therefore, SOV word order is generally not considered a defining feature of the Altaic language family.

More generally, the main typological features of Altaic languages are as follows:

- (a) absence of consonant clusters in initial position;
- (b) absence of initial *r* and *l*;
- (c) existence of vowel harmony;
- (d) absence of articles;
- (e) absence of gender;
- (f) existence of agglutination instead of inflection;
- (g) use of postpositions instead of preposition;
- (h) absence of the verb 'to have';
- (i) formation of comparative forms of adjectives with the ablative;
- (j) occurrence of modifiers before modified words and of the object before the verb [3, p. 82].

To this list, **embedding** should be added as a core syntactic and typological feature of Altaic languages. According to Choi Tong-gwon, Manchu and Mongolian also exhibit characteristic embedded sentence structures [2].

This paper has examined embedded construction in Korean, Sakha and Tungusic languages (Evenki and Even). If this analysis is extended to other languages within the Altaic family, it is highly likely that the embedding phenomenon will emerge as a consistent and recurrent trait across the group.

Given its systematic presence across Korean, Sakha and Tungusic embedding should not be regarded as a peripheral syntactic feature. Rather

it functions as a **structural principle** – comparable in its consistency and explanatory power to agglutination or vowel harmony – that underpins the grammar of Altaic languages.

Abbreviations

ABL – Ablative
ACC – Accusative
ADJ.PTCP – Adjectival participle
ADNOM – Adnominal
ADV – Adverbial
ASP – Aspect
Aux.n – Auxiliary noun
Aux.v – Auxiliary verb
COMM – Common
COMPL – Complement
COPL – Copula
CVB – Converb
NEG – Negative
NOM – Nominative
NOMLZ – Nominalizer
PF – Perfect
POSS – Possessive
POSTP – Postposition
PRET – Preterite
PROG – Progressive
PRS – Present
PST – Past
REP – Repetative
SUB – Substantive

References

1. Steffani SA. Identifying embedded and conjoined complex Sentences: making it simple. *Contemporary Issues in Communication Science and Disorders*, Spring: NSSLHA. 2007;34:44-54. DOI: 10.1044/cicsd_34_S_44.
2. Choi TG. *Comparative study of embedded sentences in Korean, Manchu and Mongolian languages*. Seoul: Korean Academic Information. 2008 (in Korean).
3. Asher RE. *The encyclopedia of language and linguistics*. Oxford: Pergamon Press Ltd. 1994.

4. Korkina EI, Ubrjatova EI, Xaritonov LN. & Petrov NE. *Grammar of the Contemporary Literary Yakut Language*. Moscow: Nauka. 1982 (in Russian).
5. Kang D. & Torotoev GG. *The Sakha (Yakut) language: An analytic grammar*. Seoul: H. Press. 2023.
6. Kang D. & Park K. *Tungus languages (Evenki and Even) and Korean*. Seoul: H. Press (forthcoming publication).
7. Lebedeva EP, Konstantinova OA. & Monakhova IV. *The Evenki language*. Leningrad: Prosveshchenie. 1985 (in Russian).
8. Cincius VI. *Introduction to the grammar of Even (Lamut)*. Leningrad: Prosveshchenie. 1947 (in Russian).
9. Kang D. & Beloljubskaja VG. *The Even language: An Analytic Grammar* (unpublished manuscript).

Литература

1. Стеффани С.А. Идентификация встроенных и сочиненных сложных предложений: упрощенный подход. *Современные проблемы науки о коммуникации и расстройствах*. Spring: Изд-во NSSLHA. 2007;34:44-54. DOI: 10.1044/cicsd_34_S_44 (на англ.).
2. Чхве Т.Г. *Hankuk-o, Manchu-o, Monggol-o naeromun bigyo yungu*. Сеул: Korean Academic Information. 2008 (на кор.).
3. Ашер Р.Э. *Энциклопедия языка и лингвистики*. Оксфорд: Pergamon Press Ltd. 1994 (на англ.).
4. Коркина Е.И., Убрятова Е.И., Харитонов Л.Н., Петров Н.Е. *Грамматика современного якутского литературного языка*. Москва: Наука. 1982.
5. Канг Д., Торотоев Г.Г. *Саха (якутский) язык: аналитическая грамматика*. Сеул: H. Press. 2023 (на англ.).
6. Канг Д., Пак К. *Тунгусские языки (эвенкийский и эвенский) и корейский*. Сеул: H. Press (готовящаяся к публикации) (на англ.).
7. Лебедева Е.П., Константинова О.А., Монахова И.В. *Эвенкийский язык*. Ленинград: Просвещение. 1985.
8. Цинциус В.И. *Очерк грамматики эвенского (ламутского) языка*. Ленинград: Просвещение. 1947.
9. Канг Д., Белолобская В.Г. *Эвенский язык: аналитическая грамматика* (неопубликованная рукопись) (на англ.).

About the author

Ducksoo KANG – Ph.D. (Linguistics), Professor Emeritus, Hankuk University of Foreign Studies, Seoul, South Korea, ORCID: 0000-0002-9075-3983, ResearcherID: S-2511-2016, Scopus Author ID: 56733981400, e-mail: kangds@hufs.ac.kr

Об авторе

КАНГ Дуксу – Ph.D. (лингвистика), почетный профессор, Университет иностранных языков Хангук, Сеул, Южная Корея, ORCID: 0000-0002-9075-3983, ResearcherID: S-2511-2016, Scopus Author ID: 56733981400, e-mail: kangds@hufs.ac.kr

Conflict of interests

The author declares no relevant conflict of interests

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов

Поступила в редакцию / Submitted 21.07.25

Принята к публикации / Accepted 27.08.25