

OCHIQ KODLI ASR MODELLARINING QIYOSIY TAHLILI

Alayev Ruhillo Habibovich,
Texnika fanlari falsafa doktori, dotsent
alayev_r@nuu.uz
O‘zbekiston Milliy universiteti

Mullaboyeva Xurmatoy Murodilovna,
Magistrant
hurmatoy@gmail.com
ToshDO‘TAU

Annotatsiya. Mazkur maqolada o‘zbek tilidagi video kontentlar uchun avtomatik subtitrlar yaratishda ochiq kodli nutqni aniqlash (Automatic Speech Recognition – ASR) modellarining samaradorligi qiyosiy tahlil qilinadi. Tadqiqotning ilmiy aniqligini ta’minlash maqsadida o‘zbek tilining barcha uslublarini qamrab olgan 10 ta video material tanlab olindi va ularning mutaxassis tomonidan qayta tinglanib, qo‘lda yozilgan transkripsiyalari etalon sifatida shakllantirildi. Tajribalar doirasida 13 ta ochiq kodli modelning natijalari ushbu inson tomonidan tayyorlangan matnlarga nisbatan Word Error Rate (WER) va Character Error Rate (CER) metrikalari, shuningdek, ishlov berish tezligi bo‘yicha o‘zaro qiyoslandi. Natijalar shuni ko‘rsatdiki, lokal korpuslarda o‘qitilgan modellar inson nutq idrokiga yaqinroq natija ko‘rsatsa-da, tilning agglyutinativ tabiati va punktuatsiya masalalarida muayyan cheklovlar saqlanib qolmoqda. Maqola yakunida o‘zbek tili uchun eng yuqori aniqlik va tezlik muvozanatini ta’minlovchi modellar bo‘yicha amaliy tavsiyalar berilgan.

Kalit so‘zlar: *nutqni aniqlash, o‘zbek tili, ASR, ochiq kodli modellar, Whisper, avtomatik subtitr, Word Error Rate, WER, Character Error Rate, CER, etalon matn, qiyosiy tahlil.*

Abstract. This article presents a comparative analysis of the effectiveness of open-source Automatic Speech Recognition (ASR) models in generating automatic subtitles for Uzbek-language video content. To ensure the scientific validity of the study, 10 video materials covering all functional styles of the Uzbek language were

selected, and their manually reviewed and transcribed versions, prepared by a specialist, were established as reference texts. Within the scope of the experiments, the outputs of 13 open-source models were compared against these human-prepared reference transcripts using the Word Error Rate (WER) and Character Error Rate (CER) metrics, as well as processing speed. The results showed that models trained on local corpora tend to produce outputs closer to human speech perception; however, certain limitations remain, particularly with regard to the agglutinative nature of the language and punctuation-related issues. At the end of the article, practical recommendations are provided on the models that offer the best balance between accuracy and speed for the Uzbek language.

Keywords: *speech recognition, Uzbek language, ASR, open-source models, Whisper, automatic subtitles, Word Error Rate, WER, Character Error Rate, CER, reference text, comparative analysis.*

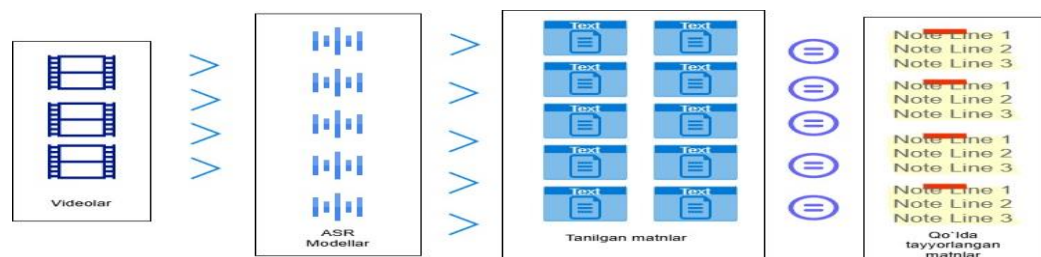
Kirish. O‘zbek tili raqamli dunyoda hali ham “kam resursli” tillar toifasiga kiradi. Sifatli va teglangan audio korpuslarning yetishmasligi o‘zbek tilidagi nutqiy signallarni matnga yuqori aniqlikda o‘girishda muayyan texnik to‘siqlarni keltirib chiqarmoqda. Videolarni subtitrlar bilan ta’minlash, nafaqat, inkluzivlikni yaratadi, balki optimallashtirish va xalqaro auditoriya uchun kontentni tushunarli qilishda muhim ahamiyatga ega. Bu holat avtomatik tizimlarga bo‘lgan ehtiyojni keskin orttirmoqda.

Mavzuning o‘rganilganlik darajasi. Nutqni avtomatik aniqlash (ASR) texnologiyalari so‘nggi o‘n yillikda chuqur o‘rganish (Deep Learning) va Transformer arxitekturalari natijasida yangi bosqichga ko‘tarildi [3:1-5]. Radford va hammualliflari (2022) 680.000 soatlik audio ma’lumotlar bazasi asosida Whisper modelini ishlab chiqishdi [8:5-10]. Bayevski va hammualliflari tomonidan taklif etilgan wav2vec 2.0 modeli nutqni avtomatik tanish sohasida self-supervised learning yondashuvining samaradorligini ko‘rsatadi [3:1-15].

O'zbek tili uchun nutq korpuslarini shakllantirish, fonetik xususiyatlarni raqamlashtirish va modellarni fine-tuning qilish bo'yicha samarali tadqiqotlar olib borilmoqda. A. Povey va K. Povey (2024) o'zbek tilidagi 60 soatlik yuqori sifatli o'qilgan nutq (read speech) korpusini shakllantirganlar [8:1-7]. B. Elov tomonidan ilk bor o'zbek tili uchun maxsus punctuation restoration (PR) modeli taklif etilib, u UzbPunct nomli korpus asosida sinovdan o'tkazilgan [4:43-52]. Ushbu tadqiqot 13 xil modelni inson tomonidan qo'lda yozilgan etalon matnlar asosida, real video kontentlar misolida qiyosiy tahlil qilishi bilan mavjud ishlardan ajralib turadi.

Metodologiya. Tadqiqot jarayoni ma'lumotlarni yig'ish, ularni qo'lda transkripsiya qilish, modellarni testdan o'tkazish va natijalarni statistik tahlil qilish bosqichlarini qamrab oldi.

3.1. *Ma'lumotlar to'plami.* Tadqiqot uchun tanlangan dataset ijtimoiy tarmoqlardagi ochiq manbalardan o'zbek tilidagi barcha uslublarni qamrab olgan videolardan shakllantirildi. Tadqiqot jarayonida ma'lumotlar quyidagi tartibda tizimlashtirildi:



1-rasm. Tadqiqot loyihasining strukturasi va ma'lumotlar arxitekturasini

1-rasmda tanlangan videolarni ochiq kodli ASR modellarida o'qitilganda olingan natijalar qo'lda tayyorlangan transkriptlar bilan qiyoslash jarayoni ko'rsatilgan.



Video	Service	Status	Runtime sec	Avg chunk ms	Success chunks	Error chunks	Total chunks	Uzbek score	Category
1.mp4	ElevenLabs	ok	7,6	2533	3	0	3	83,38	yuqori
2.mp4	ElevenLabs	ok	4,05	1349	3	0	3	74,05	o'rtacha
3.MP4	ElevenLabs	ok	4,85	1615	3	0	3	73,19	o'rtacha
4.mp4	ElevenLabs	ok	2,81	701	4	0	4	78,12	yuqori
5.mp4	ElevenLabs	ok	4,75	1582	3	0	3	89,31	yuqori
6.mp4	ElevenLabs	ok	9,04	1807	5	0	5	83,05	yuqori
7.MOV	ElevenLabs	ok	10,34	1722	6	0	6	80,53	yuqori
8.MP4	ElevenLabs	ok	8,81	1467	6	0	6	85,8	yuqori
9.mp4	ElevenLabs	ok	8,57	1712	5	0	5	85,12	yuqori
10.mp4	ElevenLabs	ok	15,63	2231	7	0	7	86,28	yuqori
12 kunuz01-rubai.mp4	ElevenLabs	ok	52,08	2169	24	0	24	82,35	yuqori
13 kunuz01.mp4	ElevenLabs	ok	53,19	2215	24	0	24	82,11	yuqori
14 TheBigBangTheory01.mp4	ElevenLabs	ok	14,66	977	15	0	15	45,29	past
15 1.mp4	Gladia	ok	17,69	5894	3	0	3	68,09	o'rtacha
16 2.mp4	Gladia	ok	10,79	3597	3	0	3	61,08	o'rtacha
17 3.MP4	Gladia	ok	10,95	3650	3	0	3	48,85	past
18 4.mp4	Gladia	ok	18,24	4559	4	0	4	52,14	past
19 5.mp4	Gladia	ok	21,79	7264	3	0	3	65,63	o'rtacha
20 6.mp4	Gladia	partial_error	45,06	7759	1	4	5	52,5	past
21 7.MOV	Gladia	partial_error	54,91		0	6	6		
22 8.MP4	Gladia	partial_error	52,95		0	6	6		
23 9.mp4	Gladia	partial_error	46,34		0	5	5		
24 10.mp4	Gladia	partial_error	65,58		0	7	7		
25 kunuz01-rubai.mp4	Gladia	partial_error	77,21		0	7	7		

2-rasm. ASR modellarining turli janrdagi video segmentlari bo'yicha qiyosiy WER ko'rsatkichlari jadvali

Jadvalda modellarning 10 ta asosiy test videosi hamda nazorat segmentlari bo'yicha aniqlik darajalari aks ettirilgan.

3.2. *Tanlangan modellar.* Sinov jarayonida HuggingFace platformasi va ochiq manbalarda taqdim etilgan o'zbek tili segmentidagi eng samarador modellar tanlab olindi.

3.3. *Baholash metrikalari.* Modellarning sifat ko'rsatkichlarini baholashda quyidagi metrikalardan foydalanildi: WER (Word Error Rate) - So'zlar darajasidagi xatolik darajasi. Formula: $WER = (S+D+I)/N$, CER (Character Error Rate) - Belgilar darajasidagi xatolik darajasi, Inference Time (ms) - Har bir segmentga ishlov berish uchun sarflangan vaqt.

Natijalar. Ushbu bo'limda 10 ta video segmenti va 13 xil ASR (Automatic Speech Recognition) modellari ustida o'tkazilgan test natijalari taqdim etiladi. Tadqiqotning metodologik asosi sifatida barcha natijalar inson tomonidan qo'lda yozilgan transkripsiyalar bilan qiyoslandi.

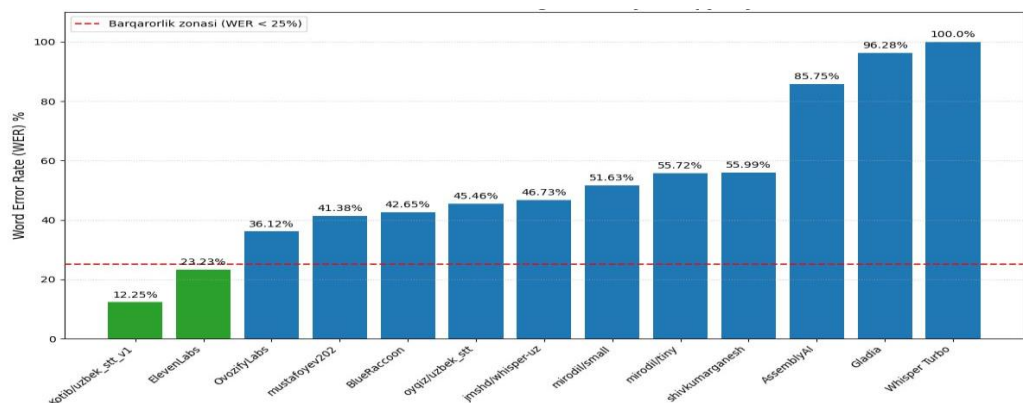
1-jadval. Modellarning qiyosiy samaradorligi

Tartib	Model nomi	Jami xatolar soni	Overall WER (%)	Eng yaxshi natija (Video)
1	Kotib/uzbek_stt_v1	135	12.25%	Video 5 (1.54%)

2	ElevenLabs	256	23.23%	Video 5 (0.0%)
3	OvozifyLabs/whisper-small-uz-v1	398	36.12%	Video 5 (1.54%)
4	mustafoyev202/whisper-uz	456	41.38%	Video 7 (18.38%)
5	BlueRaccoon/whisper-small-uz	470	42.65%	Video 5 (7.69%)
6	oyqiz/uzbek_stt	501	45.46%	Video 7 (14.71%)
7	jmskd/whisper-uz	515	46.73%	Video 5 (18.46%)
8	mirodil/whisper-small-uzbek	569	51.63%	Video 5 (27.69%)
9	mirodil/whisper-tiny-uzbek	614	55.72%	Video 5 (29.23%)
10	shivkumarganesh/whisper-small-uz-v1	617	55.99%	Video 7 (17.65%)
11	AssemblyAI	945	85.75%	Video 7 (72.06%)
12	Gladia	1061	96.28%	Video 5 (83.08%)
13	Whisper Large v3 Turbo	1102	100.00%	Video 7 (75.0%)

4.1. Grafik tahlil (Gistogrammalar)

Kotib va ElevenLabs modellari barqaror past xatolik gistogrammasini hosil qildi.



4-rasm. O‘zbek tili ASR modellarining WER bo‘yicha qiyosiy tahlili

Whisper-small turkumidagi modellar o'rtacha 35-50% oralig'ida natija ko'rsatdi. O'zbek tili korpusi bilan maxsus o'qitilmagan modellar o'zbek tili uchun 90% dan yuqori xatolik ko'rsatkichini qayd etdi.

4.2. *Sifat tahlili va xatolar tavsifi.* Modellar tomonidan yo'l qo'yilgan xatolar inson yozgan transkript bilan solishtirilganda, o'zbek tilining agglyutinativ xususiyati asosiy to'siq ekanligi ma'lum bo'ldi:

2-jadval. *Inson tomonidan yozilgan etalon matn va ASR modellari natijalari o'rtasidagi tafovutlar namunasi*

Video segmenti	Etalon matn (Inson yozgan)	Model natijasi (Hypothesis)	Xato turi
Video №1	"...zafaron tolalarini olib..."	"...zafaron to'larini olib..."	Fonetik xato
Video №1	"...yuragingizda g'ashlik..."	"...yuragingiz gashli..."	Morfologik xato

Jadvalda inson tomonidan tayyorlangan transkripsiya (Reference) va modellar generatsiya qilgan matnlar (Hypothesis) o'rtasidagi tipik xatolar aks ettirilgan.

Morfologik xatolar: Modellar so'z o'zagini to'g'ri topsa-da, qo'shimchalarni noto'g'ri shakllantirgan yoki tushirib qoldirgan.

Muhokama. O'tkazilgan qiyosiy tahlil natijalari o'zbek tili uchun avtomatik nutqni aniqlash (ASR) sohasida ham ijobiy o'sish, ham jiddiy texnik muammolar mavjudligini ko'rsatdi.

Modellarning samaradorligi. Tajribalar shuni ko'rsatdiki, global miqyosda eng kuchli hisoblangan Whisper Large v3 Turbo modeli o'zbek tili uchun deyarli yaroqsiz holatda (WER: 100%) ekanligi aniqlandi. Buning asosiy sababi sifatida ushbu modelning o'zbek tili korpusi bilan yetarli darajada fine-tune qilinmaganligi va uning asosan ingliz tiliga optimallashtirilganini ko'rsatish mumkin.



A	B	C	D	E	F	G	H	I	J	K	L	M
1	Timestamp	Qo'lda yozilgan etalon matn	hisid	Kotib/uzbek_stt_v1	dek	ovqiz/uzbek_stt	fb	BlueRaccoon/whisper-smal	wh	umarganesh/whisper-sm	wh	Whisper Large v3 Turbo
	00:00:00,000 --> 00:00:09,100	Miyangizga yomon xayollar kelaveradigan bo'lsa, doim vahima bilan yuradigan bo'lsangiz, yuragingizda g'ashlik, odamlarga qo'shilish	1117	miyangizga yomon xayollar kelaveradigan bo'lsa, doim vohima bilan yuradigan bo'lsangiz, yuragingizda g'ashlik, odamlarga qo'shilish	2559	miyezgu yomon xayolli keuradigan bo'lsa doimg vohima bilan yuradigan bosis yurigitlik g'ashli odamligi qo'shilish yoqmayotgan bo'lsi	177	Miyezgi yomon xayolli keuradigan bo'lsa, doim vohimda bilan yuradigan bo'lsa, yurigitli g'ashli odamdagi qo'shilish	1661	Miyzga yomon xayolga keuradigan bo'lsa, doim vohimda bilan yuradigan bo'lsiz, yurigitli g'ashli odamdagi qo'shni	1563	Biyezgi ama hayalli ki uradigan bossa, dahin vahime bila yuradigan boseyiz. Yurigitdi gashli, adamdagi koshul
2	1 00:00:09,100 --> 00:00:28,700	ushbu ikkita damlamani ichib yuring. Birinchisi, zafaron damlamasi. Zafaron kuchli antidepressant xususiyati bilan mashhurdir. Uch yoki to'rtta dona zafaron tolalarini olib, ikki yuz gramm qaynoq suvga damlang va ertalab och qoringa iste'mol qiling. Ikkinchisi, bu -	1404	ushbu ikkita damlamani ichib bering. birinchisi zafaron damlamasi. zafaron kuchli antidepressant xususiyati bilan mashhurdir. uch yoki to'rtta dona zafaron tolalarini olib, ikki yuz g qaynoq suvga damlang va ertalab oshqorongga iste'mol qiling. ikkinchisi bu sidr barglari.	4690	ushbu ikkita damlamani ichiburing birinchisi zafaron damlamasi zafaron kuchli antidepressant hususati bilan mashg'urdur uch yoki to'rttadona zafaron tolalarini olib. ikki ignam qaynoq suvigi damlin va erilab oshqorongga iste'mol qiling ikkinchisi bu	235	Ushbu ikkita damlamani ichibilgan. Birinchi zafaron damlamasi -- zafaron kuchli antidepressant xususati bilan mashhurdir. Uch yoki to'rtta dona zafaron tolalarini olib, 200 g qaynoq suviga damlang va ertalab oshqorongga istimal qiling. Ikkinchisi, bu sidr	3136	Ushbu ikkita damlamani ichib edi. Birinchi Zafaron damlamasi Zafaron kuchli antidepressant hususati bilan mashhurdir. Uch yoki to'rtta dona Zafaron tolalarini olib, ikki gram qaynoq suvga damlang va ertalab oshqorongga istimal	3363	Ushbu, ikide damleman echl birini. Birincisi, Zafaron damlamasi. Zafaron, kucli antidepressident xususiyati bilan, mashhurdir. Uchi yoki, torta dana Zafaron talalarini olib, ikki gram qaynoq sugi damleyni va irtalab, ajqarangani istimal
3	2 00:00:28,700 --> 00:00:45,000	ovqatdan o'ttiz daqiqa oldin ichib yuring. Bu ikkita damlamani har kuni iste'mol qilsangiz, sizga yomon o'y-xayollarning kelaverishi, sababsiz vahima va vasvasadan xalos bo'lishingizga yordam beradi va insonni tinchlantiradi.	2031	shib ovqatdan o'ttiz daqiqa oldin ichib yuring. bu ikkita damlamani har kuni iste'mol qilsangiz, sizga yomon o'y-xayollarni kiyaverishi, sababsiz vohima va vaj hosadan xalos bo'lishingizga yordam beradi va insonni tinchlantiradi. bu	4422	oshib ovqatdan o'ttiz daqiqa oldin ichib yurik bu kica damlamani har kuni isim olqisis sizga yomon o'yxayollarni keuyurishi sababsiz g'ohima va vashosadan xalos bo'lishiga yordam berid va insonni	99	Uchib ovqatdan o'ttiz daqiqa oldin ichib yurik. Bu keta damlamani har kuni istimal qisiz, sizga yomon o'y-xayolarni ko'rish, sababsiz g'ohima va vasbasadan xalos bo'lishisizga yordam beradi va insonni tinchlantiradi. Bu	2255	Uchib ovqatdan o'ttiz daqiqa oldin ichib yuring'l. Bu kuta damlamani har kuni istimal qisiz, sizga yomon o'yxayolarni ko'rish, sababsiz G'ohima va Vafosadan, qalos bo'lishigza yordam beril va AKTIVIRAZIYA Windows	2296	Bu videon daim negativ fikrilydi gen, uydan koquki chqlamydi gen, yaqllariz gli var.
4	3											
5												
6												

5-rasm. Qo'lda yozilgan etalon matn bilan tanlab olingan modellarning natijalari

Kotib/uzbek_stt_v1 (WER: 12.25%) va OvozifyLabs/whisper-small-uz-v1 (WER: 36.12%) modellarning yuqori natijalari, ochiq kodli modellarni mahalliy korpuslar asosida maqsadli o'qitish qanchalik muhimligini isbotlaydi.

Agglyutinativlik va morfologik to'siqlar. O'zbek tili morfologiyasining murakkabligi barcha modellar uchun asosiy xato manbai bo'lib qolmoqda. Bu muammo o'zbek tili uchun ASR tizimlarini yaratishda nafaqat akustik modelni, balki kuchli Language Model (LM) – til modelini ham integratsiya qilish zarurligini ko'rsatadi.

Tadqiqotning cheklovlari. Ushbu tadqiqot 10 ta video segmenti bilan cheklangan bo'lib, natijalar spikerning diksiyasi, shovqin darajasi va ishlatilgan mikrofon sifatiga bog'liq bo'lishi mumkin. Shuningdek, ma'lumotlar to'plamida sheva elementlari (dialektlar) kamroq qatnashgani sababli, modellarning hududiy nutqni tushunish darajasi alohida o'rganishni talab qiladi.

Xulosa va kelajakdagi ishlar. Mazkur tadqiqot doirasida 10 ta turli uslubdagi video materiallar asosida o'tkazilgan eksperiment natijasiga ko'ra, global miqyosda universal hisoblangan modellardan ko'ra, o'zbek tili korpusi asosida maqsadli fine-tune qilingan modellar sezilarli darajada yuqori aniqlik ko'rsatdi. Whisper Large v3 Turbo kabi zamonaviy modellarning o'zbek tilidagi past samarador. Bu ushbu modellarni lokalizatsiya qilish zaruriyatini ko'rsatadi.

Kotib/uzbek_stt_v1 modeli o'zbek tili uchun eng optimallashtirilgan tizim bo'lib, u inson yozgan matnga 87.75% (ya'ni 12.25% WER) aniqlikda muvofiq kelgan.

O'zbek tili ASR tizimlarini yanada takomillashtirish uchun kelgusida datasetni kengaytirish va eng samarali deb topilgan model asosida o'zbek tilida avtomatik subtitrlar qo'ya oladigan yangi model yaratish rejalashtirilgan.

Foydalanilgan adabiyotlar ro'yxati

1. Agarwal A., Zesch T. Robustness of End-to-End Automatic Speech Recognition Models: A Case Study Using Mozilla DeepSpeech // arXiv. – 2021. – URL: <https://arxiv.org/abs/2105.09742>
2. Baevski A., Zhou H., Mohamed A., Auli M. wav2vec 2.0: A Framework for Self-supervised Learning of Speech Representations // Advances in Neural Information Processing Systems (NeurIPS). – 2020. – Vol. 33. – URL: <https://arxiv.org/abs/2006.11477>
3. Elov B.B. Og'zaki nutqni tanib olish tizimida matndagi tinish belgilarini tiklash usullari // TATU ilmiy-texnika va axborot-tahliliy jurnali. – 2025. – № 1(73). – B. 43–52.
4. Hamdamov O'., Elov B., Alayev R. Ma'lumotlar to'plamini o'qitish, baholash va test to'plamlariga ajratish usullari // Raqamli transformatsiya va sun'iy intellekt ilmiy jurnali. – 2024. – Vol. 2. – Issue 6.
5. Hamroyeva S., Matyakubova N. Mashina tarjimasida matni moslashtirish usullari // Uzbekistan: Language and Culture. – 2(2).
6. Mukhamadiyev A. va boshq. Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language // Sensors. – 2022. – Vol. 22. – No. 10. – URL: <https://www.mdpi.com/1424-8220/22/10/3683>



7. Povey A., Povey K. FeruzaSpeech: A 60 Hour Uzbek Read Speech Corpus with Punctuation, Casing, and Context // arXiv preprint arXiv:2410.00035v1. – 2024. – URL: <https://arxiv.org/abs/2410.00035>
8. Radford A., Kim J.W. va boshq. Robust Speech Recognition via Large-Scale Weak Supervision // arXiv. – 2023. – URL: <https://arxiv.org/abs/2212.04356>