

Language Technology Evaluation and Application - LTEA

This speech/presentation is authorized by the Defense Language Institute Foreign Language Center and the Department of Defense. Contents of this presentation are not necessarily the official views of, or endorsed by, the U.S. Government, Department of Defense, Department of the Army, or the Defense Language Institute Foreign Language Center.



DLIFLC
DEFENSE LANGUAGE INSTITUTE
FOREIGN LANGUAGE CENTER



Auto-ILR

Background information:

A DOD resource to identify topic areas automatically was modified to identify foreign language text complexity as per ILR level description. (<http://govtilr.org/>)

Margin-infused relaxed algorithm (MIRA) - a machine learning algorithm for multiclass classification problems – is used to support the Auto-ILR system.

MIT-LL created the automated ILR-leveling version that is available at DLIFLC.



Auto-ILR

Automatic ILR Rating for Text Materials



Auto-ILR Instant

Determine ILR level for any text instantly





Auto-ILR Subscriptions

Subscribe to receive notification of articles by ILR level, language, and topic



Auto-ILR

Product	Description
	<p>Cut and paste text into window to be rated in seconds in the following languages: Arabic, Chinese Simplified, Chinese Traditional, Dari, English, Farsi, French, Korean, Pashto, Portuguese, Russian, Spanish, Urdu</p>
	<p>The Auto-ILR Subscriptions system allows subscribers to request documents to be searched for automatically, at levels set by users on available sites through RSS feed services, thus saving invaluable time for teachers, students and material developers.</p>
Under Development	German, Tagalog

Get estimated ILR level for any text

Supported languages: **Arabic, Chinese-Simplified, Chinese-Traditional, Dari, English, Farsi, French, Korean, Pashto, Portuguese, Russian, Spanish, and Urdu**

Instructions:

1. Paste a **whole** document avoiding partial sentences.
2. Select the document language.
3. Click **Score**.

Select language:

Select a language

Arabic

Chinese-Simplified

Chinese-Traditional

Dari

English

Farsi

French

Korean

Pashto

Portuguese

Russian

Spanish

Urdu

Enter your text:

Score



The ILR level will automatically refresh when you update the target language, using the drop-down menu.

Estimated ILR Level: 3

Language: Selected -- Arabic

Your text:

يحتفظ متحف اللوفر بلوحة الموناليزا التي تعد من أكثر اللوحات شهرة في العالم ككشف علماء آثار إيطاليون عن بقايا عظام يعتقد أنها تعود الى المرأة التي صورها الفنان الشهير ليوناردو دافنشي في لوحته الشهيرة الموناليزا. وقضى الباحثون سنوات في فحص وتدقيق بقايا الرفات التي عثر عليها في مدينة فلورانس ويرجع تاريخها إلى نحو خمسة قرون. ويقول العلماء إنها جزء من عظم ساق تعود لليزا غيرارديني، زوجة التاجر الفلورنسي فرانسيسكو ديل جيوكوندو، التي يعتقد أنها جلست أمام دافنشي ليرسمها في لوحته الشهيرة عام 1503. ويقول العلماء إن بقايا الرفات كانت في حالة متردية لا تسمح بالقيام بتحليل الحمض النووي "دي أن أيه"، كما فشل العلماء في العثور على الجمجمة. ويحتفظ متحف اللوفر بلوحة الموناليزا، التي تعرف بالإطالية باسم "الجيوكوندا"، والتي تعد من أكثر اللوحات شهرة في العالم، وتصور امرأة شابة ذات ابتسامة غامضة وقد شبكت يديها برفق في حضنها. وعلى الرغم من أن هوية هذه المرأة غير محددة بدقة، إلا أن العديد من المؤرخين يعتقدون أنها من المحتمل أن تعود إلى غيرارديني، وقد بدأ علماء الآثار حملة قبل ثلاث سنوات، للبحث عن رفاتها في دير كانت قضت فيه أيامها الأخيرة. وقام علماء الآثار بفتح قبر عائلة جيوكوندو في كنيسة في مدينة فلورانس من أجل استحصاء الحمض النووي لمطابقته. بيد أن فحص الكاربون كشف عن وجود مجموعة متفرقة من العظام يرجع تاريخها إلى مطلع القرن السادس عشر، حيث عاشت غيرارديني ورسمت لوحة الموناليزا.



Auto-ILR

Subscription Process



[Home](#) [New Subscription](#) [My Account](#) [Summary](#) [Logout](#) [Info](#)

tamas.marius@dliflc.edu

Articles: 27825

This page will automatically refresh.

[AD_AI Jazeera Top Stories L2](#)

[AD_AI-Ahram_Issues_L3](#)

[AD_AI-Jazeera_TopStories_L2](#)

[AD_AI-Mustaqbal_Health_2+](#)

[AD_AI-Qabas_Travel_1+](#)

[AD_ALQabas_Cul_2+](#)

[AD_Asharq AI-Awsat_Top_Stories_2](#)



Personalized Auto-ILR Subscriptions

AD_AI-Ahram_Issues_L3

Nickname	Language	Source	Topic	ILR	# Articles	Results	Updated	Preferences
AD_AI-Ahram_Issues_L3	Arabic	Al-Ahram	The views of issues	3	2	view	2017-07-11	change

AD_AI-Jazeera_TopStories_L2

Nickname	Language	Source	Topic	ILR	# Articles	Results	Updated	Preferences
AD_AI-Jazeera_TopStories_L2	Arabic	Al Jazeera	Top Stories	2	962	view	2017-07-11	change

AD_AI-Mustaqbal_Health_2+

Nickname	Language	Source	Topic	ILR	# Articles	Results	Updated	Preferences
AD_AI-Mustaqbal_Health_2+	Arabic	Al-Mustaqbal	Health	2+	24	view	2017-07-11	change

AD_AI-Qabas_Travel_1+

Nickname	Language	Source	Topic	ILR	# Articles	Results	Updated	Preferences
AD_AI-Qabas_Travel_1+	Arabic	Al-Qabas	Travel	1+	4	view	2017-07-11	change



Personalized Auto-ILR Subscriptions



[Home](#) [New Subscription](#) [My Account](#) [Summary](#) [Logout](#) [Info](#)

Nickname: SP_TECH_LaNacion_3

Articles: 52

RSS/Atom Source: http://contenidos.lanacion.com.ar/herramientas/rss/categoria_id=432

The following articles were found for this subscription. This page will automatically refresh. You may view the results here, or send a version of what you see to your email address (using plaintext or XML).

[Send Me Text](#)

[Send Me XML](#)

Headline: [La tecnología y los datos, vitales para el éxito en la Fórmula 1](#)

Description:

Published: Date not available

Language: Spanish

ILR: 3

Headline: [La ciudad de San Pablo, capital temporal del mundo gamer](#)

Description:

Published: Date not available

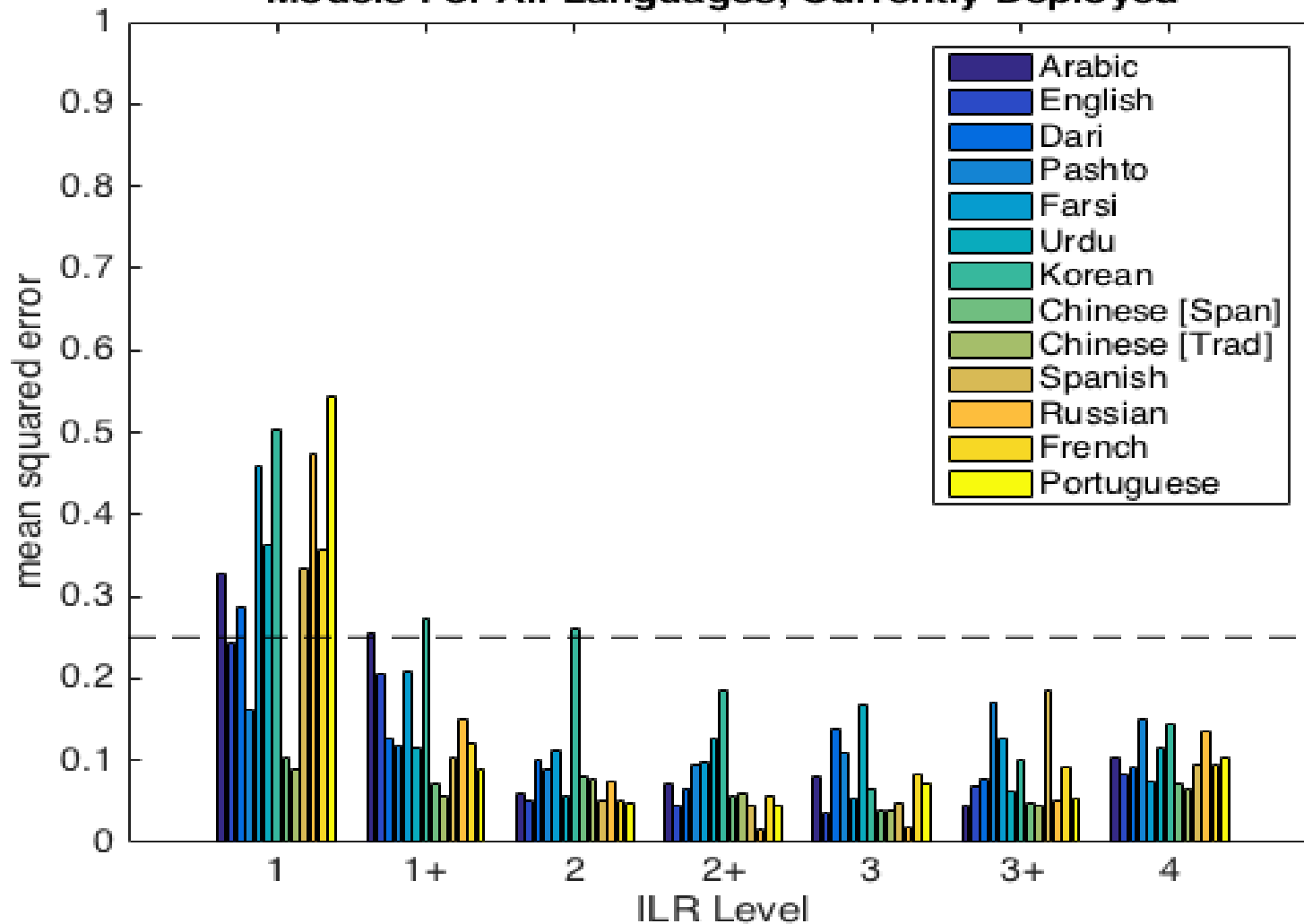
Language: Spanish

ILR: 3



Auto-ILR Data Management

Models For All Languages, Currently Deployed





Auto-ILR Data Management

FY	Lang	Code	Avg MSE	1	1+	2	2+	3	3+	4
11	Arabic	ar	0.135752	0.327253	0.255999	0.060719	0.070627	0.080276	0.044458	0.103860
14	Chinese - Simp	zh	0.062214	0.089878	0.057394	0.075900	0.059964	0.039855	0.046166	0.066935
14	Chinese - Trad	tradzh	0.076413	0.171960	0.083657	0.111841	0.052060	0.029334	0.072543	0.105162
12	Dari	da	0.126786	0.288101	0.126337	0.100020	0.067231	0.138740	0.075950	0.091809
11	English	en	0.104357	0.244079	0.205823	0.051077	0.045052	0.035449	0.067873	0.082435
13	Farsi	fa	0.159973	0.072544	0.127238	0.054948	0.096921	0.111841	0.206743	0.458625
16	French	fr	0.123911	0.358147	0.121100	0.050049	0.056863	0.082108	0.093341	0.094264
14	Korean	ko	0.217757	0.502058	0.273589	0.260457	0.184351	0.067169	0.100609	0.143984
12	Pashto	ps	0.127936	0.162397	0.119478	0.089663	0.094833	0.107613	0.171771	0.151242
16	Portuguese	po	0.136101	0.544600	0.090487	0.048341	0.045544	0.072079	0.054626	0.102579
15	Russian	ru	0.124570	0.474187	0.150333	0.075008	0.015351	0.017640	0.050049	0.135720
15	Spanish	sp	0.123190	0.333830	0.102533	0.049761	0.045272	0.046649	0.184954	0.096904
13	Urdu	ur	0.145105	0.363740	0.116122	0.057837	0.125599	0.166709	0.064017	0.113801
14	Arabic Speech	arspeech	0.072723	N/A	1.188752	0.014987	0.016428	0.081345	0.068442	0.241228



Auto-ILR Instant

<https://auto-ilr.ll.mit.edu/>

6/19/2017	6/26/2017	7/3/2017	7/10/2017	7/17/2017	7/24/2017	INCREASE past week	Languages
1,310	1,349	1,388	1,410	1,488	1,545	57	Arabic
627	635	643	666	673	689	16	Chinese-Simplified
211	212	213	213	213	215	2	Chinese-Traditional
118	118	118	138	138	138	0	Dari
1,129	1,145	1,161	1,173	1,187	1,223	36	English
1,841	1,877	1,913	2,063	2,103	2,127	24	Farsi
589	597	605	694	699	821	122	French
485	489	493	521	539	561	22	Korean
512	522	532	638	674	745	71	Pashto
50	50	50	71	71	155	84	Portuguese
775	822	869	933	957	1,015	58	Russian
416	442	468	472	476	493	17	Spanish
21	21	21	21	23	29	6	Urdu
8.084	8.279	8.474	9.013	9.241	9.756	515	YTD



Auto-ILR Subscriptions

<https://auto-ilr.ll.mit.edu/>

6/19/2017	6/26/2017	7/3/2017	7/10/2017	7/17/2017	7/24/2017	INCREASE past week	
11,837	11,864	11,868	11,868	11,963	12,097	134	Arabic
5,499	5,540	5,542	5,542	5,542	5,624	82	Chinese-Simplified
1,784	1,784	1,780	1,780	1,780	1,809	29	Chinese-Traditional
109	109	109	109	109	109	0	Dari
551	551	552	562	568	567	-1	English
2,132	2,132	2,164	2,181	2,181	2,190	9	Farsi
1,061	1,061	1,061	1,061	1,061	1,062	1	French
2,373	2,379	2,379	2,599	2,638	2,699	61	Korean
872	872	882	882	917	927	10	Pashto
29	29	29	29	29	29	0	Portuguese
3,140	3,140	3,160	3,233	3,234	3,260	26	Russian
7,290	7,290	7,326	7,464	7,464	7,505	41	Spanish
3,299	3,299	3,299	3,299	3,589	3,596	7	Urdu
39,976	40,050	40,151	40,609	41,075	41,474	399	Subscriptions Created
3,208	3,211	3,218	3,232	3,253	3,298	45	# of DLIFLC SUBSCRIBERS

<https://auto-ilr.ll.mit.edu/>



DLIFLC
DEFENSE LANGUAGE INSTITUTE
FOREIGN LANGUAGE CENTER

Abkhaz · Afrikaans · Albanian · Amharic · Arabic · Armenian · Azerbaijani · Basque · Belarusian · Bengali · Berber · Bislama · Bosnian · Bulgarian · Burmese · Catalan · Cebuano · Chinese · Croatian · Czech · Danish · Dutch · English · Esperanto · Estonian · Filipino · Finnish · French · German · Greek · Gujarati · Hebrew · Hindi · Hungarian · Indonesian · Italian · Japanese · Korean · Kyrgyz · Lao · Latvian · Lithuanian · Macedonian · Malay · Maltese · Marathi · Mongolian · Nepali · Norwegian · Persian · Polish · Portuguese · Romanian · Russian · Serbian · Sinhalese · Slovak · Slovenian · Spanish · Swahili · Swedish · Tagalog · Thai · Turkish · Ukrainian · Urdu · Vietnamese · Welsh · Xhosa · Yiddish · Zulu

Questions?

Itea@dliflc.edu



DLIFLC

DEFENSE LANGUAGE INSTITUTE
FOREIGN LANGUAGE CENTER