

Enhancing Language Proficiency with Technology

Douglas A. Jones

14 May 2025

For MITHIC, Part 1



DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the Department of the Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Air Force.

© 2024 Massachusetts Institute of Technology.

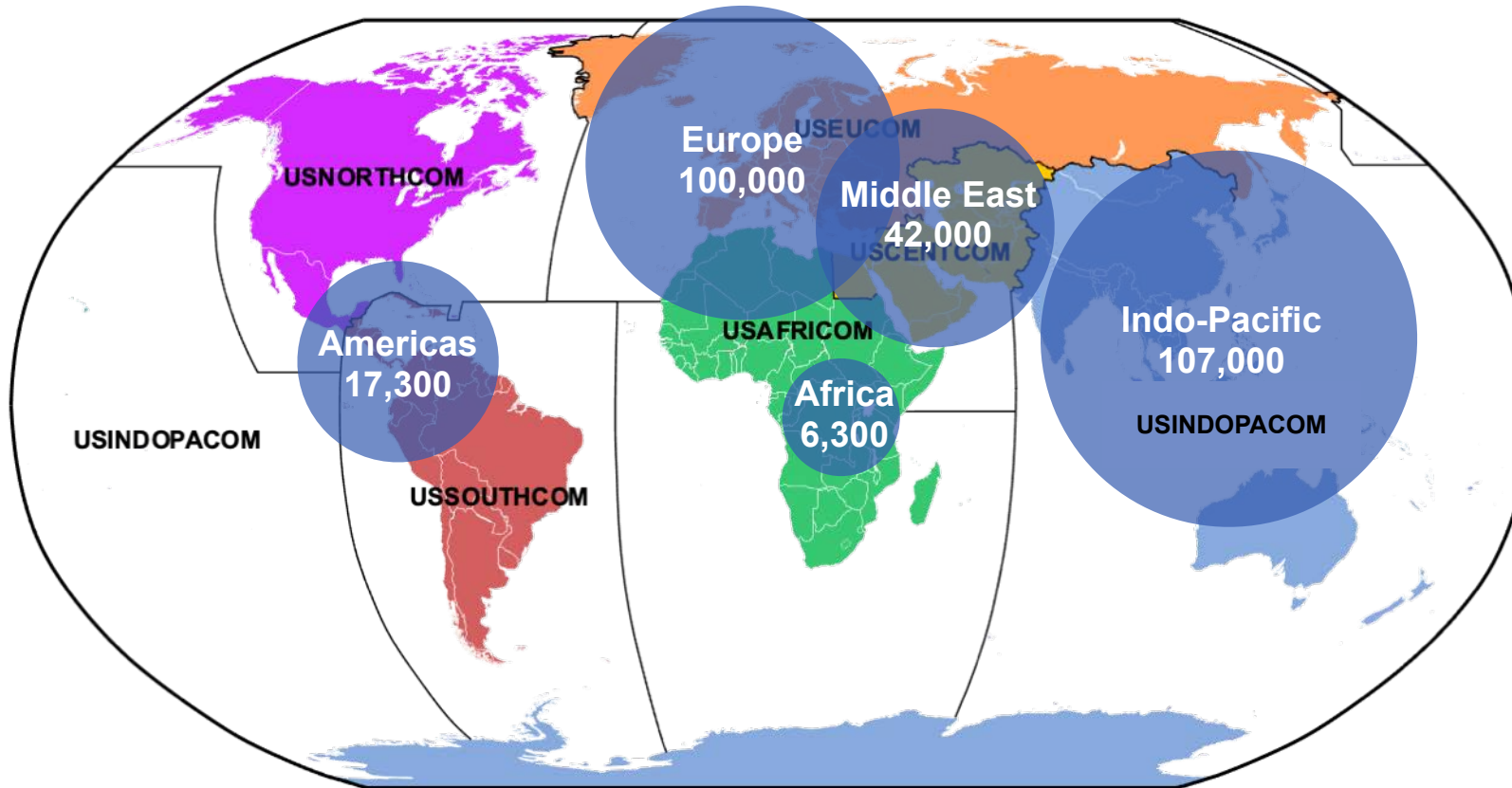
Subject to FAR52.227-11 Patent Rights - Ownership by the contractor (May 2014)

Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work. This presentation has been approved for public release by the Defense Language Institute Foreign Language Center's Public Affairs Office. For verification please e-mail: mpao@dliflc.edu. Contents of this presentation are not necessarily the official views of the Defense Language Institute Foreign Language Center, nor are they endorsed by the Department of the Army, the Department of Defense, or the U.S. Government. All third party products / materials featured in the presentation remain the intellectual property of their respective authors / originators.



US Global Presence Requires Language Expertise

Estimated number of deployed US military personnel per region



- **Countries**

- US Military and Civilian DOD Personnel in 175 countries²
- Official US Military bases in 42 countries³

- **Languages**

- 7000 languages in the world
- 3000 languages in INDOPACOM
- DOD maintains approximately 40 languages in Strategic Language List



Need for Foreign Language Learning

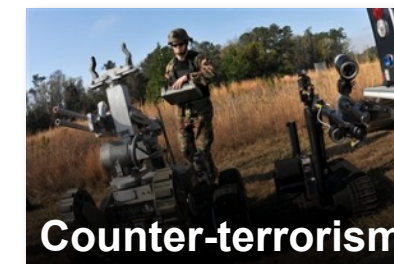
National Problem

“A critical component [is] the ability to speak languages and really engage at the right level of cultural awareness, with not only the adversaries, but also third parties.”



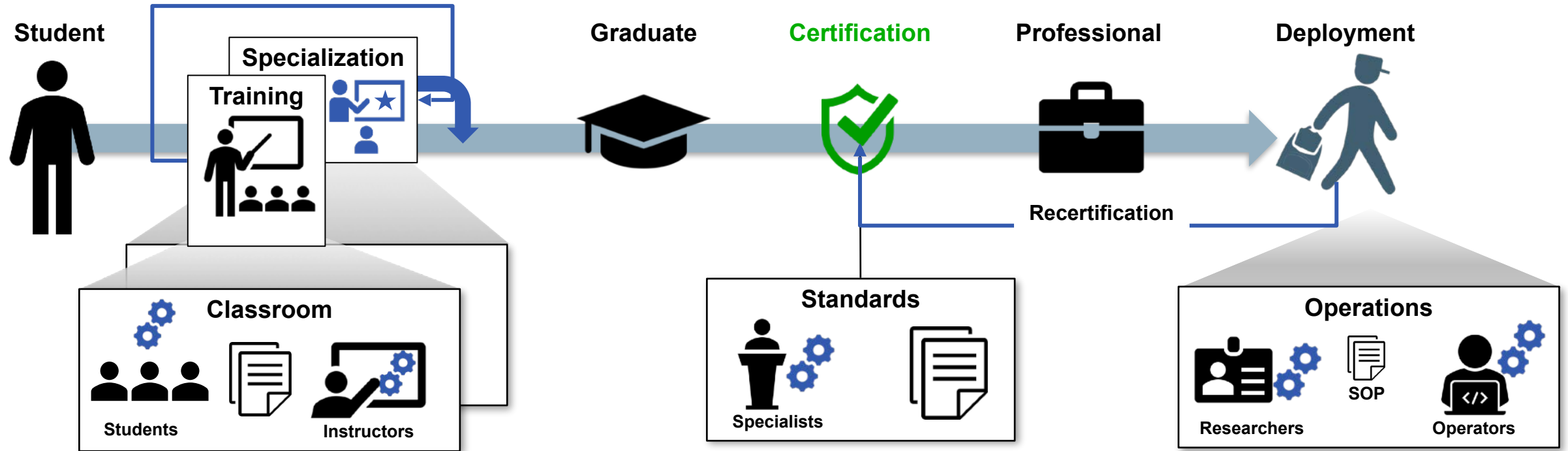
– Christopher Maier
Assistant Secretary of Defense for Special Operations and Low-Intensity Conflict¹

Sample Specializations





Generalized Training Architecture

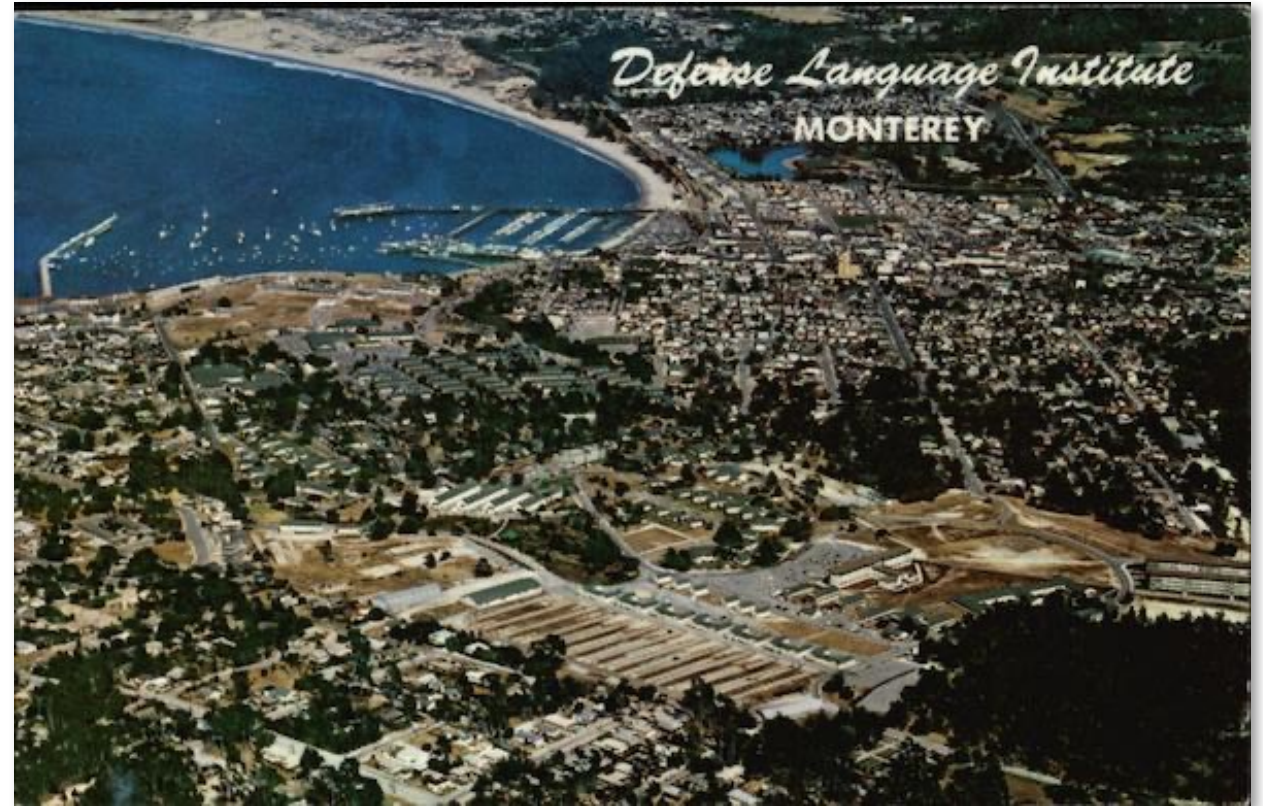


Each Human-Machine Team provides an opportunity to optimize overall process via machine network



Defense Language Institute (DLI)

- **Largest foreign language training center in the world**
 - Administered by U.S. Army TRADOC
 - Over 3,000 military students per year trained in Monterey, CA
 - Over 45,000 “distance learners”
- **Tests proficiency in 70+ languages**
- **Currently 14 languages taught on post**



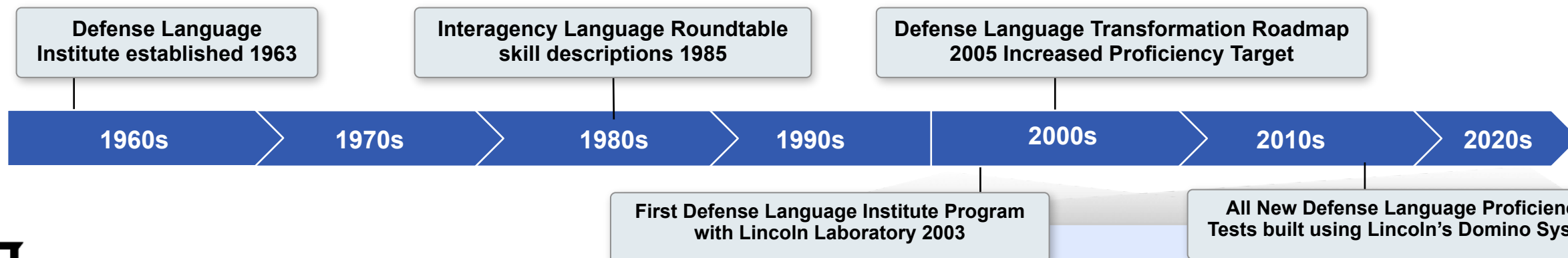
Global mission set creates demand pressure to increase national language capabilities



Historical Highlights



Foreign Language Learning



Human Language Technology



Speech vocoders



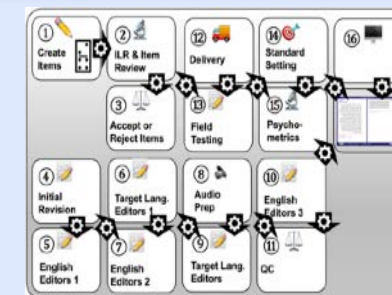
Language & speaker recognition



Machine Translation Evaluation using Foreign Language Comprehension Tests



Language Learning using Speech Recognition



Automated Workflow for Language Test Development



Automatic Pronunciation Feedback

- Use statistical models to provide phoneme and word-level pronunciation scores
- Enables focused learning of phonemes based on individual pronunciation
- Web application available to students at DLI

دقيقة minute



- Vocabulary item with reference pronunciations
- Learner pronunciation with feedback for word and for phonemes



Impact – Automatic Pronunciation Feedback



NETPROF

So if you need help pronouncing your words, especially for foreign languages, NetProF can help. The system – called Networked Pronunciation Feedback (NetProF) – has been in existence for over 8 years and includes more than 15 languages.

NetProF student use has grown steadily since the system was introduced in 2010 and reached 193,000 recordings in 2017. DLIFLC's NetProF System has proven to be very helpful for students. Studies on pronunciation have focused on various aspects of the most helpful ways to improve pronunciation. Accord-

IMPROVING Language Training One Syllable at a Time

BY DR. TAMAS MARIUS

- Available at DLI in 15 languages for students and instructors
 - Initial Prototypes in 2010
 - Fully Transitioned to DLI in 2019
 - Approximately 250,000 Netprof recordings per year
 - Automatic, individualized feedback available 24/7 over network
- Accepted and adopted for new Arabic curriculum, more anticipated



National Need for Foreign Language Proficiency Testing

Defense Language Proficiency Test¹ (DLPT)

DLPT 5 Lower Range Reading Comprehension Test for Russian Form A

Sample Passage

An excerpt from a letter to the editor of a newspaper

Я, как и многие москвичи, испытываю большие трудности в последнее время. В столице нелегко устроиться по специальности, особенно инженеру. Это всё потому, что институты и университеты каждый год выпускают тысячи специалистов, и каждый из них хочет остаться в Москве. Конкуренция очень высокая, и иногда нужны месяцы, чтобы найти работу.

End of this passage
Mark this passage for later consideration. ☐

S1. What is the writer concerned about?

- ☐ Finding good employees nowadays in Moscow.
- ☐ The cutthroat competition among university students.
- ☐ The declining quality of teaching at many universities.
- ☐ Securing employment in his field in Moscow.

End of questions for Sample Passage

Help Sample Passage 1 question << Back Next >>

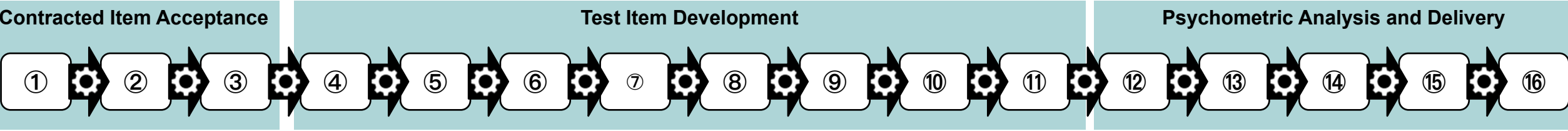
Oral Proficiency Interview (OPI)



Test development challenged by volume and variety of languages



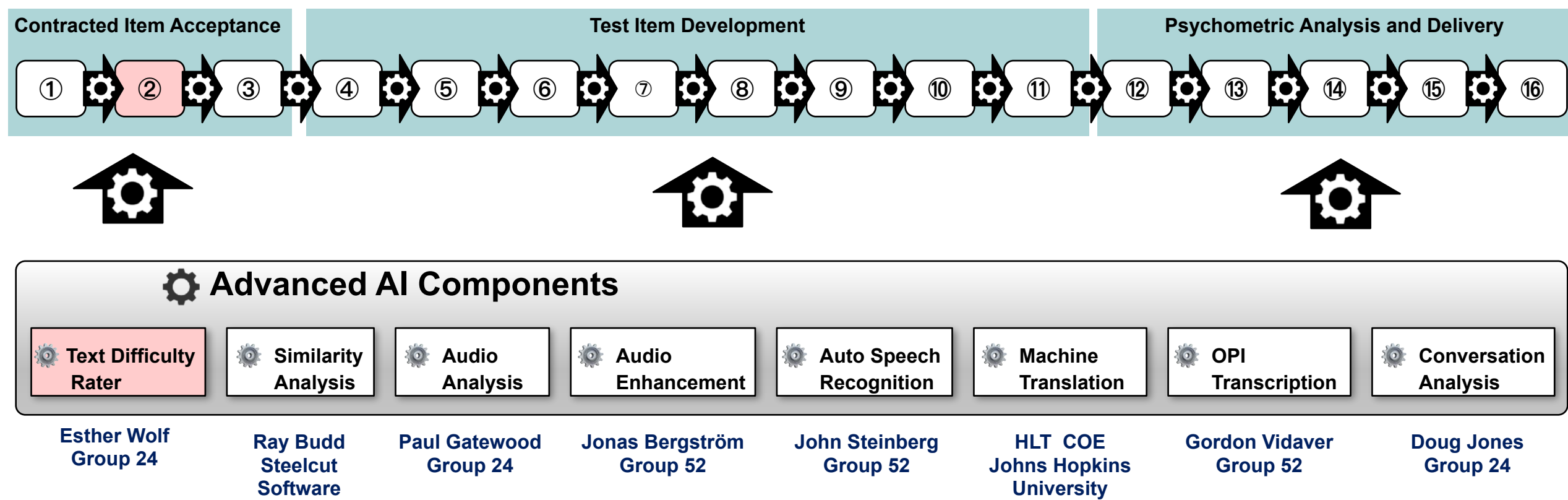
Automated Test Development Workflow Architecture



Domino system automates workflow handoffs



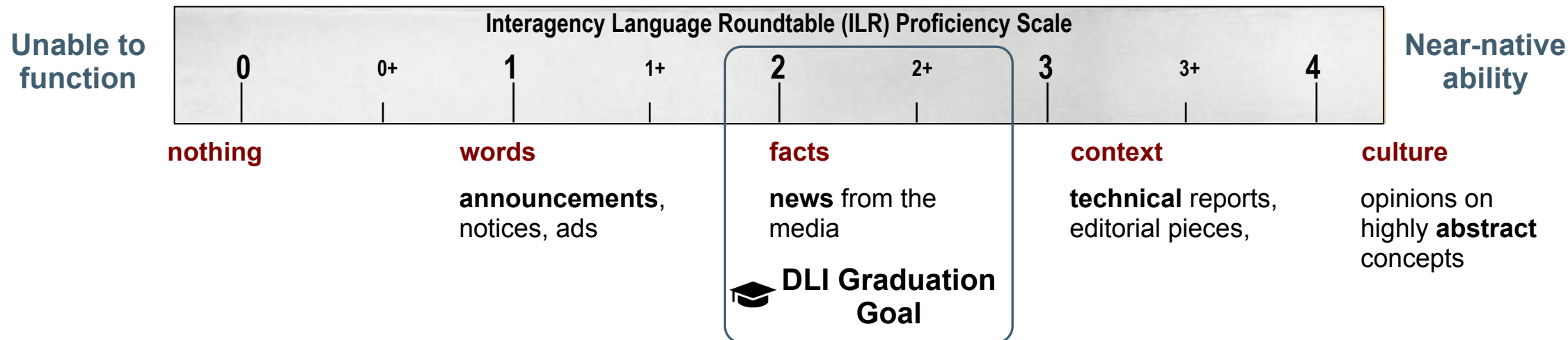
Domino Architecture for AI Component Collaboration



Step ② used to require two human reviewers for difficulty level, but now only requires one, using our automatic leveler for one of the review steps.



Military Foreign Language Skill Proficiency Levels



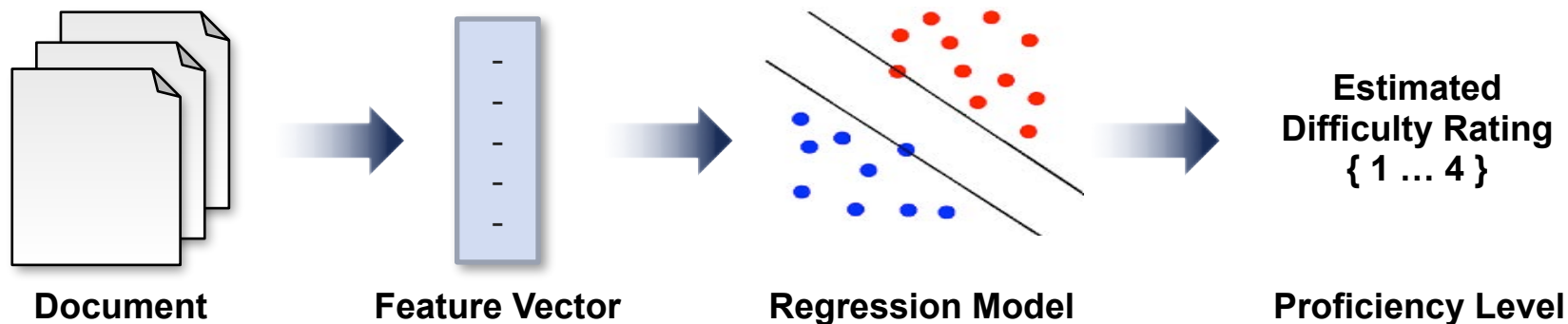
Reasons why texts must be labeled by difficulty:

- Instructors need to select docs for their students at appropriate levels
- Self-learners need to find appropriate docs quickly
- Test developers need to screen reading passages

Current process is manual and requires expensive, highly skilled professionals



Automatic Text Difficulty Rating



Sample Features
Average word length
Average sentence length
Relative entropy
Word vectors
K-Means Cluster ID

- Build a model for text levels over features extracted from documents in mission-relevant topics (Science and Technology, Geopolitical, etc.)
- Focus on language-independent features to maximize adaption to new languages

Automated process is fast and sufficiently accurate for operational use



AI Text Difficulty Rater -- Performance

Mean Squared Error

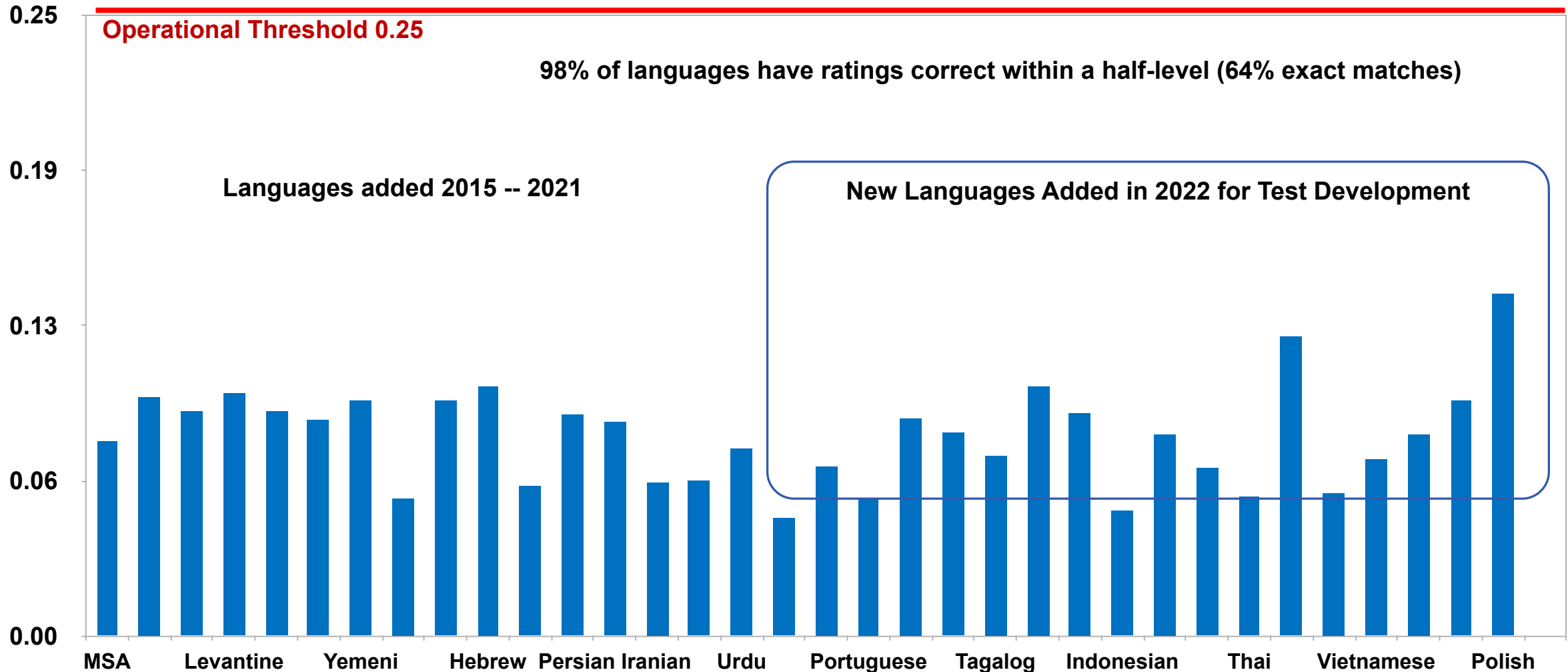
Operational Threshold 0.25

98% of languages have ratings correct within a half-level (64% exact matches)

Languages added 2015 -- 2021

New Languages Added in 2022 for Test Development

Lower Error is Best





Outline

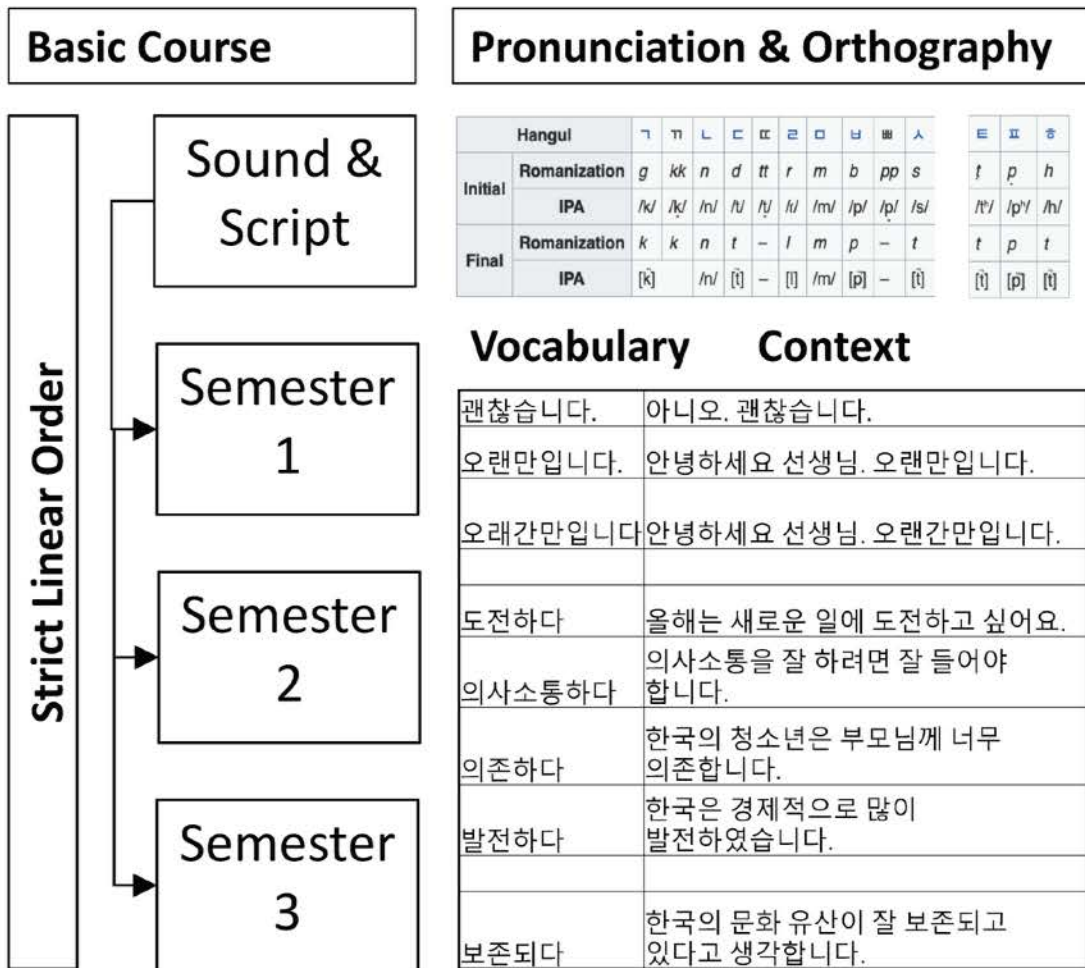
- **National Need for Foreign Language Training**
- **Foreign Language Learning & Testing**
 - **Automatic Pronunciation Feedback**
 - **Automated Test Workflow Architecture**
 - **Personalizing Language Proficiency Roadmaps**
- **Summary and Future Work**





Personalized Roadmaps for Foreign Language Proficiency

Standard Curriculum



Learner Model Framework

Vocabulary	Features
V_1	$f_{112}, f_{339}, \dots$
V_2	$f_{259}, f_{402}, \dots$
V_3	$f_{236}, f_{357}, \dots$
V_4	$f_{817}, f_{798}, \dots$
V_5	$f_{929}, f_{63}, \dots$
...	...
V_N	$f_X, f_Y, \dots$

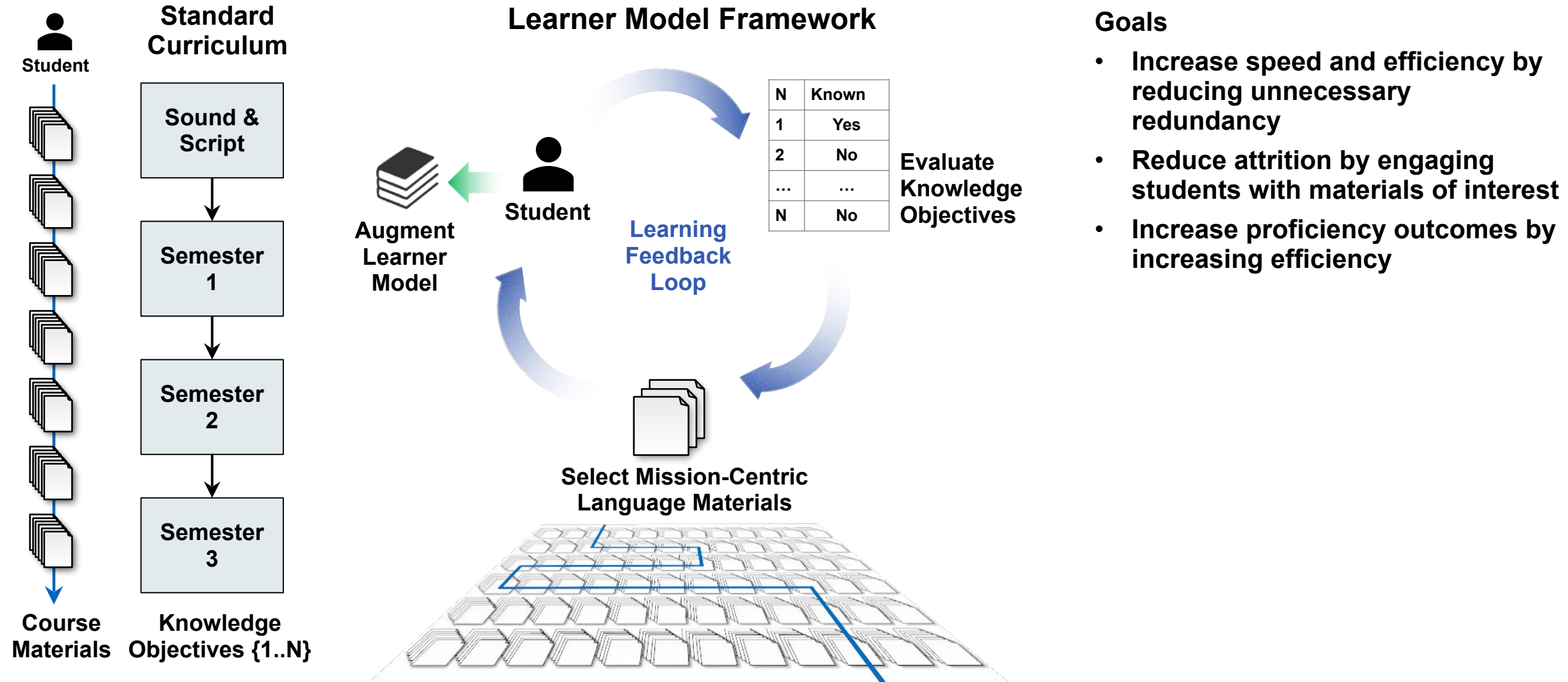
Features Include:

- Grammatical
- Pronunciation
- Context
- Cultural

Core idea is to provide learner alternative pathways through the vocabulary and features to cover “the same” learning objectives



Toward Personalized Language Roadmaps

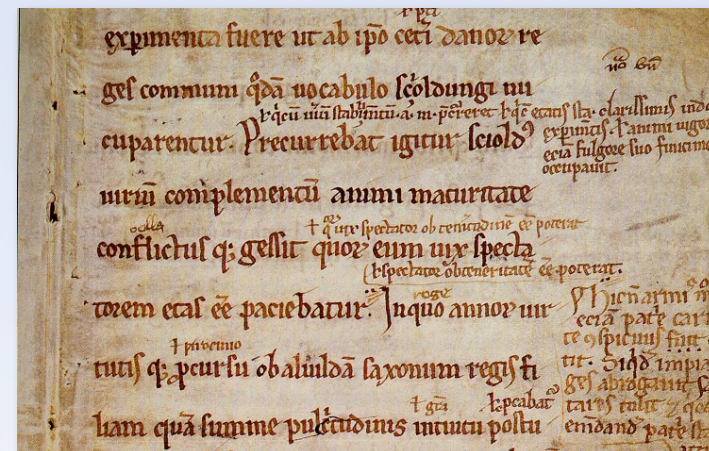




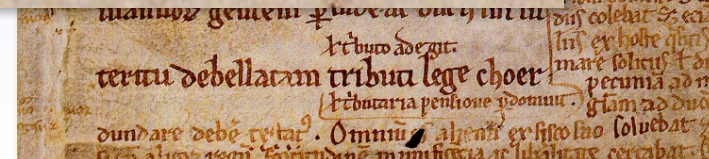
Traditional “Gloss” Concept

上 周 该 部队 发现 路边 炸弹
shàn zhōu gāi bùduì fāxiàn lù zhàdàn
g biān
last week the army found road bombs
d side

12th Century Gloss



labat cum scato allemannie

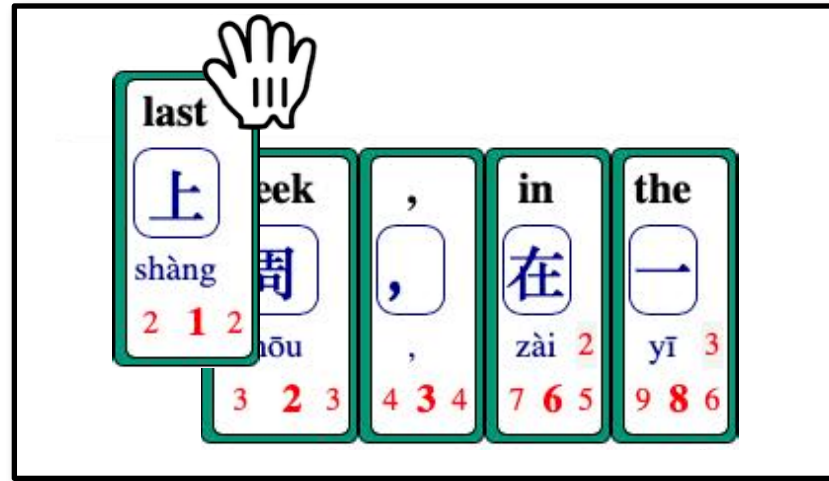


A gloss indicates the meaning and pronunciation of foreign language materials



Novel “Hypergloss” Concept for Learner Model

上 周 该 部队 发现 路边 炸弹
shàng zhōu gāi bùduì fāxiàn lù biān zhàdàn
g
last week the army found road side bombs



Manipulate Hypergloss Tiles

1. Add English “gloss” for foreign terms
2. Insert wildcard tiles as needed
3. Rearrange into valid English word order

Reader View of
Bilingual Documents

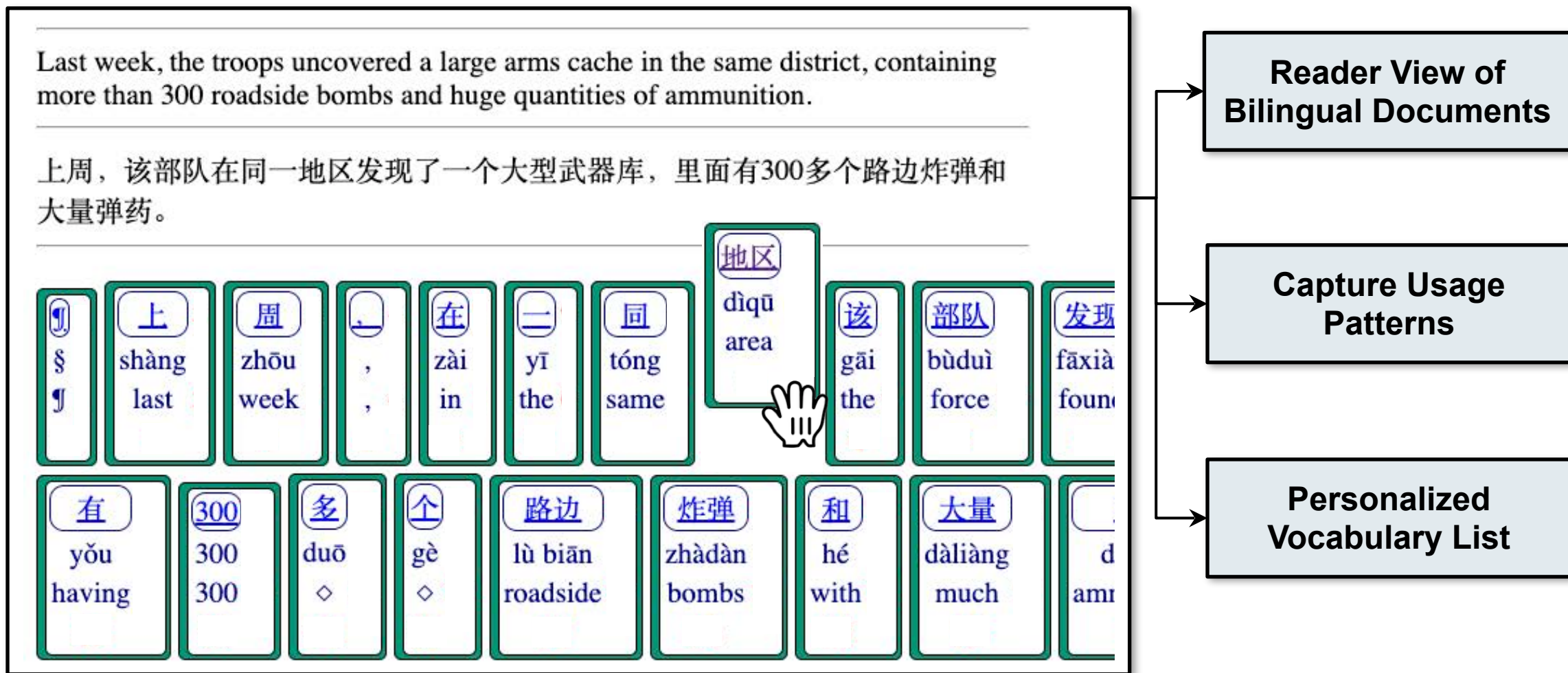
Capture Usage
Patterns

Personalized
Vocabulary List

A “hypergloss” can be manipulated in a game-like way to create personalized learning artifacts



Novel “Hypergloss” Concept for Learner Model



A “hypergloss” represents two languages at the same time



Reader View Sliders /1

Language Learning Exercise

(1) Original Word Order

(100) Frequency Percentile

斯里兰卡 军方 说 在 从前 由 已 被 击败 了 的 泰米尔 猛虎
组织 控制 的 地区 军队 又 找到 了 一 个 潜艇 型 的 船只
军方 官员 说 这 艘 能够 携带 名 机组 人员 的 潜艇 式 船只
是 在 周六 晚 被 发现 的 当时 它 淹没 在 穆莱蒂武 地区 东 北
部 距离 海岸 米 的 海水 中
斯里兰卡 数十 年 之 久 的 内战 两 个 月 前 随着 泰米尔 猛虎 组织
被 击败 组织 头目 维鲁 皮莱 普拉巴卡兰 的 死亡 而 结束 在
内战 的 最后 阶段 军队 已经 发现 一些 潜水 设备
上 周 该 部队 在 同 一 地区 发现 了 一 个 大型 武器 库 里面 有
多 个 路边 炸弹 和 大量 弹药



Reader View Sliders /2

Language Learning Exercise

(3) Original Word Order

(100) Frequency Percentile

斯里兰卡 军方 说 在 从前 由 已 被 击败 了 的 泰米尔 猛虎
the Sri Lankan military said that in an previously by the already ◇ defeated ◇ ◇ Tamil Tiger
组织 控制 的 地区 军队 又 找到 了 一 个 潜艇 型 的 船只
organization controlled ◇ area , the army also found ◇ one ◇ submarine type of vessel .
军方 官员 说 这 艘 能够 携带 名 机组 人员 的 潜艇 式 船只
military officials said that this ship can carry 3 ◇ crew members in the submarine type vessel ,
是 在 周六 晚 被 发现 的 当时 它 淹没 在 穆莱蒂武 地区 东 北
it on Saturday night being found ◇ , when it was submerged in Mullaitivu area north eastern
部 距离 海岸 米 的 海水 中
district of from coast 500 meters ◇ the seawater in the . the
斯里兰卡 数十 年 之 久 的 内战 两 个 月 前 随着 泰米尔 猛虎 组织
Sri Lanka 's ten year ◇ long ◇ civil war two ◇ months ago with the Tamil Tiger organization
被 击败 组织 头目 维鲁 皮莱 普拉巴卡兰 的 死亡 而 结束 在
being defeated , organization 's ringleader Viru pila Prabakaran of the death and the ended . in
内战 的 最后 阶段 军队 已经 发现 一些 潜水 设备
the civil_war of the final stages , the army has found some diving equipment .
上 周 该 部队 在 同 一 地区 发现 了 一 个 大型 武器 库 里面 有
last week , the force in same the area found ◇ one ◇ large arms depot , inside having 300
多 个 路边 炸弹 和 大量 弹药
◇ ◇ roadside bombs with much ammunition .



Reader View Sliders /3

Language Learning Exercise

(4) English Word Order

(100) Frequency Percentile

◇ 斯里兰卡 军方 说 ， 在 ◇ 地区 从前 的 控制 由 ◇ 已 被 击败 了
the Sri Lankan military said that in an **area** previously ◇ **controlled** by the already ◇ defeated ◇
的 泰米尔 猛虎 组织 ， ◇ 军队 又 找到 了 一 个 潜艇 型 的 船只 。
◇ Tamil Tiger organization , the army also found ◇ one ◇ submarine type of vessel .
军方 官员 说 ， 这 艘 能够 携带 3 名 机组 人员 的 ◇ 潜艇 式 船只 ，
military officials said that this ship can carry 3 ◇ crew members in the submarine type vessel ,
是 被 发现 在 周六 晚 的 ， 当时 它 ◇ 淹没 在 海水 500 米 距离 的
it **being found** on Saturday night ◇ , when it was submerged in **seawater 500 meters from** ◇
◇ 海岸 中 ◇ 东 北 部 ◇ ◇ 穆莱蒂武 地区 。
the coast in the north eastern district of **the** Mullaittivu area .
斯里兰卡 ◇ 数十 年 之 久 的 内战 结束 两 个 月 前 随着 ◇ 泰米尔 猛虎
Sri Lanka 's ten year ◇ long ◇ civil war **ended** two ◇ months ago with the Tamil Tiger
组织 被 击败 ， 而 ◇ 死亡 的 ◇ 组织 ◇ 头目 维鲁 皮莱 普拉巴卡兰
organization being defeated , **and the death of the** organization 's ringleader Viru pila Prabakaran
。 在 ◇ 最后 阶段 的 ◇ 内战 ， ◇ 军队 已经 发现 一些 潜水 设备 。
. in the **final stages** of the civil_war , the army has found some diving equipment .
上 周 ， 在 一 同 地区 该 部队 发现 了 一 个 大型 武器 库 ， 有 300 多 个
last week , **in the same area** the force found ◇ one ◇ large arms depot , **having 300** ◇ ◇
路边 炸弹 和 大量 弹药 里面 。
roadside bombs with much ammunition inside .



Reader View Sliders /4

Language Learning Exercise

(6) English Word Order

(100) Frequency Percentile

the Sri Lankan military said that in an area previously controlled by the already defeated Tamil Tiger organization , the army also found one submarine type of vessel . military officials said that this ship can carry 3 crew members in the submarine type vessel , it being found on Saturday night , when it was submerged in seawater 500 meters from the coast in the north eastern district of the Mullaittivu area . Sri Lanka 's ten year long civil war ended two months ago with the Tamil Tiger organization being defeated , and the death of the organization 's ringleader Viru pila Prabakaran . in the final stages of the civil_war , the army has found some diving equipment . last week , in the same area the force found one large arms depot , having 300 roadside bombs with much ammunition inside .



Reader View Sliders /5

Language Learning Exercise

(2) Original Word Order

(50) Frequency Percentile

斯里兰卡 军方 说 在 从前 由 已 被 了 的
the Sri Lankan military said that in an previously by the already ◇ defeated ◇ ◇ Tamil Tiger
控制 的 地区 军队 又 找到 了 一 个 的 船只
organization controlled ◇ area , the army also found ◇ one ◇ submarine type of vessel .
军方 说 这 能够 机组 人员 的 船只
military officials said that this ship can carry 3 ◇ crew members in the submarine type vessel ,
是 在 被 发现 的 当时 它 淹没 在 地区 东
it on Saturday night being found ◇ , when it was submerged in Mullaitivu area north eastern
部 的 中
district of from coast 500 meters ◇ the seawater in the . the
斯里兰卡 数十 年 之 的 内战 两 个 月 前 随着
Sri Lanka 's ten year ◇ long ◇ civil war two ◇ months ago with the Tamil Tiger organization
被 头目 的 而 在
being defeated , organization 's ringleader Viru pila Prabakaran of the death and the ended . in
内战 的 最后 军队 已经 发现 一些
the civil_war of the final stages , the army has found some diving equipment .
上 部队 在 同 一 地区 发现 了 一 个 大型 里面 有
last week , the force in same the area found ◇ one ◇ large arms depot , inside having 300
多 个 和 大量
◇ ◇ roadside bombs with much ammunition .



Hypergloss Differential

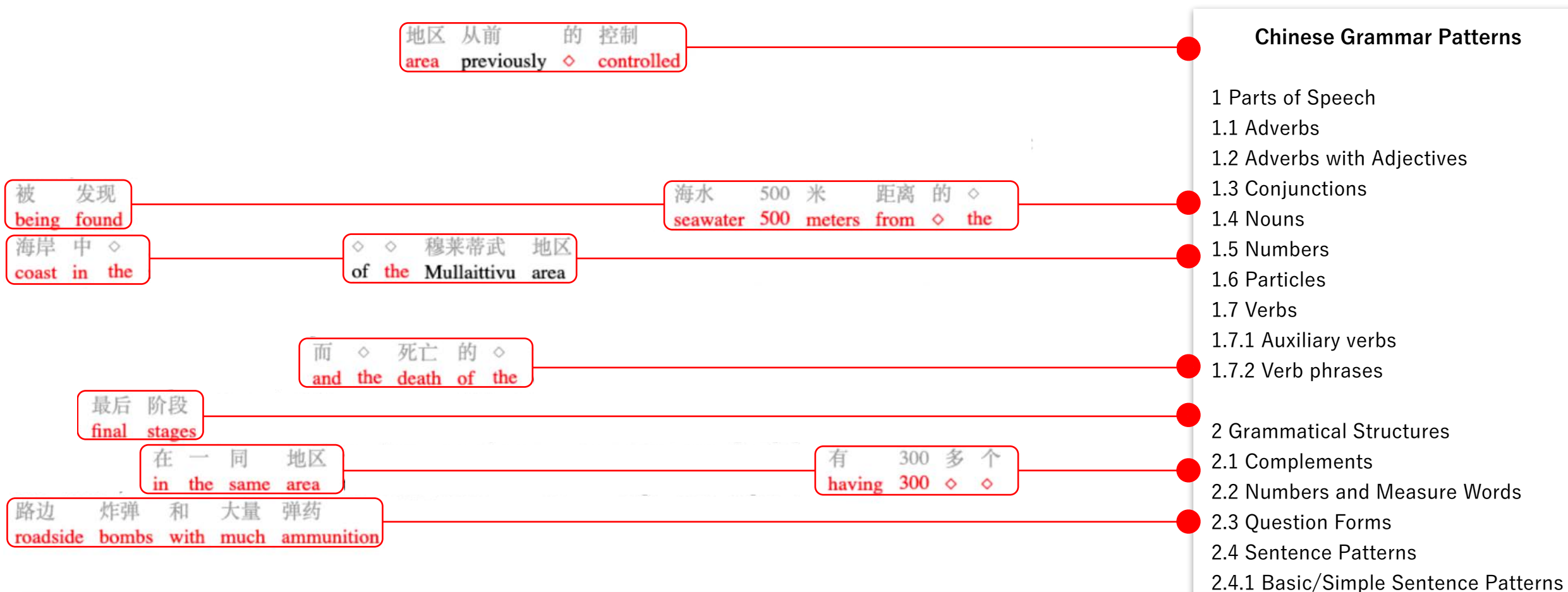
◇ 斯里兰卡 军方 说 ， 在 ◇ 地区 从前 的 控制 由 ◇ 已 被 击败 了 的
the Sri Lankan military said that in an **area** previously ◇ **controlled** by the already ◇ defeated ◇ ◇
泰米尔 猛虎 组织 ， ◇ 军队 又 找到 了 一个 潜艇 型 的 船只 。
Tamil Tiger organization , the army also found ◇ one ◇ submarine type of vessel .
军方 官员 说 ， 这 艘 能够 携带 3 名 机组 人员 的 ◇ 潜艇 式 船只 ， 是
military officials said that this ship can carry 3 ◇ crew members in the submarine type vessel , it
被 发现 在 周六 晚 的 ， 当时 它 ◇ 淹没 在 海水 500 米 距离 的 ◇
being found on Saturday night ◇ , when it was submerged in **seawater 500 meters from ◇ the**
海岸 中 ◇ 东 北 部 ◇ ◇ 穆莱蒂武 地区 。
coast in the north eastern district of **the** Mullaittivu area .
斯里兰卡 ◇ 数十 年 之 久 的 内战 结束 两 个 月 前 随着 ◇ 泰米尔 猛虎
Sri Lanka 's ten year ◇ long ◇ civil war **ended** two ◇ months ago with the Tamil Tiger
组织 被 击败 ， 而 ◇ 死亡 的 ◇ 组织 ◇ 头目 维鲁 皮莱 普拉巴卡兰 。
organization being defeated , **and the death of the** organization 's ringleader Viru pila Prabakaran .
在 ◇ 最后 阶段 的 ◇ 内战 ， ◇ 军队 已经 发现 一些 潜水 设备 。
in the **final stages** of the civil_war , the army has found some diving equipment .
上 周 ， 在 一 同 地区 该 部队 发现 了 一 个 大型 武器 库 ， 有 300 多 个
last week , **in the same area** the force found ◇ one ◇ large arms depot , **having 300 ◇ ◇**
路边 炸弹 和 大量 弹药 里面 。
roadside bombs with much ammunition inside .

- A Hypergloss represents two languages at the same time
- Difference between English and Target Language sequences are shown in red
- Notice that most of the passage is not red
- The “Differential” provides a focus for learning
- Currently manually naming the patterns in terms of Basic Course materials in Phase 1

Leverages implicit commonalities between English and Target Language to minimize learning cost



Hypergloss Differential Linked to Grammar Patterns



Leverages implicit commonalities between English and Target Language to minimize learning cost



Leverage Manual Hypergloss Differential Data for Automation

Chinese Grammar Patterns

- 1 Parts of Speech
 - 1.1 Adverbs
 - 1.2 Adverbs with Adjectives
 - 1.3 Conjunctions
 - 1.4 Nouns
 - 1.5 Numbers
 - 1.6 Particles
 - 1.7 Verbs
 - 1.7.1 Auxiliary verbs
 - 1.7.2 Verb phrases
- 2 Grammatical Structures
 - 2.1 Complements
 - 2.2 Numbers and Measure Words
 - 2.3 Question Forms
 - 2.4 Sentence Patterns
 - 2.4.1 Basic/Simple Sentence Patterns
 - 2.4.2 Comparison Patterns
 - 2.4.3 Complex Sentence Patterns

斯里兰卡 军方 说 , 在 地区 从前 的 控制 由 已 被 击败 了 的
the Sri Lankan military said that in an area previously controlled by the already defeated
泰米尔 猛虎 组织 , 军队 又 找到 了 一个 潜艇 型 的 船只 .
Tamil Tiger organization , the army also found one submarine type of vessel .
军方 官员 说 , 这 艘 能够 携带 3 名 机组 人员 的 潜艇 式 船只 , 是
military officials said that this ship can carry 3 crew members in the submarine type vessel , it
被 发现 在 周六 晚 的 , 当 时 它 淹 没 在 海 水 500 米 距 离 的
being found on Saturday night , when it was submerged in seawater 500 meters from the
海 岸 中 东 北 部 的 穆 莱 蒂 武 地 区 .
coast in the north eastern district of the Mullaitivu area .
斯里 兰 卡 数 十 年 之 久 的 内 战 结 束 两 个 月 前 随 着 泰 米 尔 猛 虎
Sri Lanka's ten year long civil war ended two months ago with the Tamil Tiger
组 织 被 击 败 , 而 死 亡 的 组 织 头 目 维 鲁 皮 莱 普 拉 巴 卡 兰 .
organization being defeated , and the death of the organization's ringleader Viru pila Prabakaran .
在 最 后 阶 段 的 内 战 , 军 队 已 经 发 现 一 些 潜 水 设 备 .
in the final stages of the civil war , the army has found some diving equipment .
上 周 , 在 一 同 地 区 该 部 队 发 现 了 一 个 大 型 武 器 库 , 有 300 多 个
last week , in the same area the force found one large arms depot , having 300
路 边 炸 弹 和 大 量 弹 药 里 面 .
roadside bombs with much ammunition inside .

VG4	b	1	0	0	0	0	0	0	VG4_b/D
NG2	a	0	0	1	0	0	0	0	NG2_a/F
NG2	b	0	1	1	0	0	0	0	NG2_b/EF
DP1	a	0	1	1	0	0	0	0	DP1_a/EF
DP1	a	0	1	1	0	0	0	0	DP1_a/EF
PD1	a	0	0	0	1	0	0	0	PD1_a/I

Learning exercises provide data for automatic identification of grammatical patterns

Introducing Hypergloss Cost Accounting for Autonomous Language Learning

Douglas A. Jones

14 May 2025

For MITHIC, Part 2



DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the Department of the Air Force under Air Force Contract No. FA8702-15-D-0001 or FA8702-25-D-B002. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Air Force.

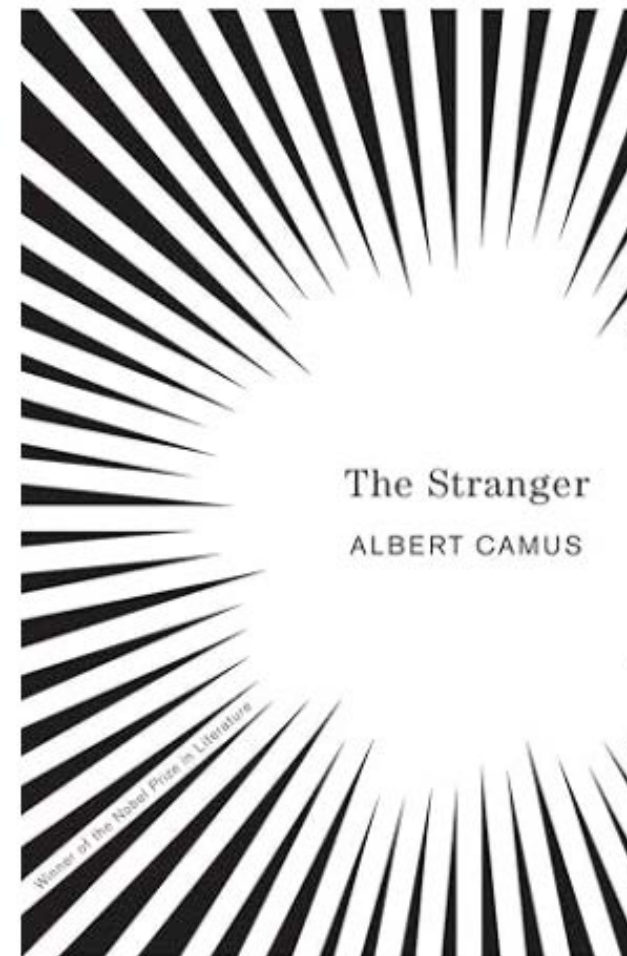
© 2025 Massachusetts Institute of Technology.

Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.



Three Translations and a Hypergloss

- f* Aujourd'hui, maman est morte. Ou peut-être hier, je ne sais pas.
*e*¹ Mother died today. Or, maybe, yesterday; I can't be sure.
*e*² Mother died today. Or maybe yesterday, I don't know.
*e*³ Mother died today. Or yesterday maybe, I don't know
e['] Today, Mom died. Or maybe yesterday, I don't know.





Three Translations and a Hypergloss and Costs

- f* Aujourd'hui, maman est morte. Ou peut-être hier, je ne sais pas. (Camus Original)
- e*¹ Mother died today. Or, maybe, yesterday; I can't be sure. (Gilbert Translation, Cost=22)
- e*² Mother died today. Or maybe yesterday, I don't know. (Laredo Translation, Cost=13)
- e*³ Mother died today. Or yesterday maybe, I don't know (Ward Translation, Cost=15)
- e*' Today, Mom died. Or maybe yesterday, I don't know (Jones Hypergloss, Cost=3)

Note: *f* and *e*' are two serialized views of the same hypergloss object, whereas *f*, *e*¹, *e*² and *e*³ are different text objects.





The Five Hypergloss Cost Types

1. **Inversion of English Words (CIe)**
2. **Extra L2 Words (CXf)**
3. **Extra L1 Words (CXe)**
4. **L2-to-L1 Word Variability (CVe)**
5. **L2-to-L1 Word Variability (CVf)**



Sample Cost Calculation

Core Hypergloss					Costs					
Row	f_order	f_token	e_token	e_order	CIe	CXf	CXe	CVe	CVf	Σ
...	...	...	...	...	...	...	...	...	...	...
6	6	parce	because	6						
7	7	qu'il	◇	7		1			11	12
8	8	a	◇	8		1			12	13
9	9	fallu	had to	11	2		1			3
10	10	que	◇	9	1	1		1	13	16
11	11	je	I	10	1					1
12	12	monte	go up	12			1			1
13	13	chez	to	13				1	2	3
14	14	Emmanuel	Emmanuel's place	14			1			1
...	...	...	...	...	...	...	...	...	...	...
					Σ 4	3	3	2	38	50

f: J'étais un peu étourdi parce qu'il a fallu que je monte chez Emmanuel pour lui emprunter une cravate noire et un brassard.

e': I was a little distracted because I had to go up to Emmanuel's place for him to lend me a black tie and an armband.



Translation versus Hypergloss

Translation

Then I saw the Aleph. I arrive now at the ineffable core of my story. And here begins my despair as a writer. All language is a set of symbols whose use among its speakers assumes a shared past. How, then, can I translate into words the limitless Aleph, which my floundering mind can scarcely encompass? ... What I want to do is impossible, for any listing of an endless series is doomed to be infinitesimal. In that single gigantic instant I saw millions of acts both delightful and awful; not one of them occupied the same point in space, without overlapping or transparency. What my eyes beheld was simultaneous, but what I shall now write down will be successive, because language is successive.

Di Giovanni Translation (e)

Hypergloss

Entonces vi el Aleph. Arribo, ahora, al inefable centro de mi relato, empieza aquí, mi desesperación de escritor. Todo lenguaje es un alfabeto de símbolos cuyo ejercicio presupone un pasado que los interlocutores comparten; ¿cómo transmitir a los otros el infinito Aleph, que mi temerosa memoria apenas abarca? ... El problema central es irresoluble: la enumeración, si siquiera parcial, de un conjunto infinito. En ese instante gigantesco, he visto millones de actos deleitables o atroces; ninguno me asombró como el hecho de que todos ocuparan el mismo punto, sin superposición y sin transparencia. Lo que vieron mis ojos fue simultáneo: lo que transcribiré sucesivo, porque el lenguaje lo es.

Borges Original (f)

Then I saw the Aleph. I arrive, now, at the ineffable center of my story – it begins here, my desperation as a writer. All language is an alphabet of symbols whose use presupposes a past that the interlocutors share; how to transmit to others the infinite Aleph, that my floundering memory barely encompasses? ... The central problem is unresolvable: enumeration, if at least partial, of an infinite collection. In that gigantic instant, I have seen millions of acts, delightful or atrocious; none amazed me like the fact that all occupied the same point, without superposition and without transparency. That which my eyes saw was simultaneous: that which I will transcribe sequentially, because language is that.

Jones Hypergloss (e')

Note: **f** and **e'** are two serialized views of the same hypergloss object, whereas **f** and **e** are different text objects.





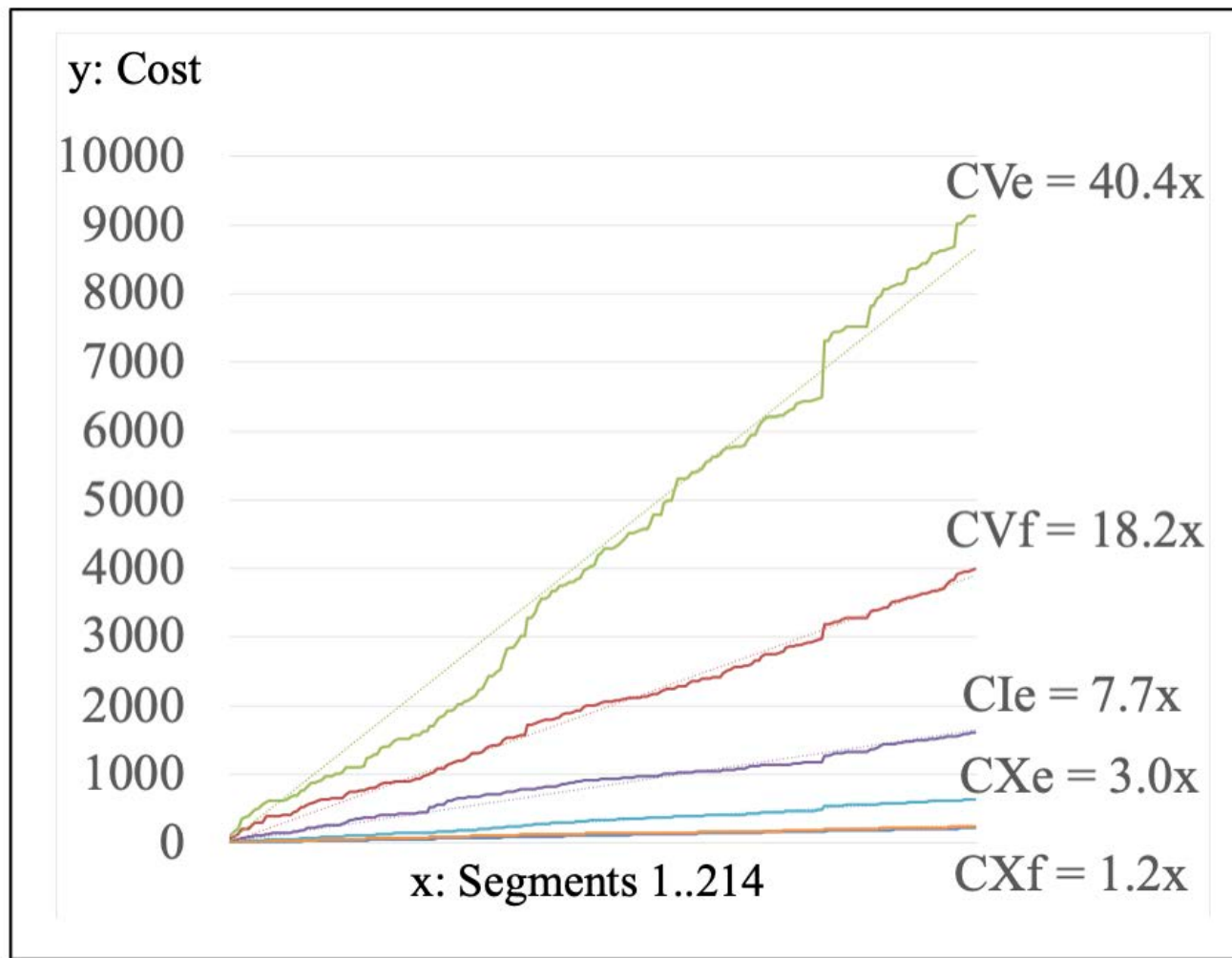
Sample Grammar Annotation

row	f_token	e_token	iΔ	grammar
1	¶	¶		
2	la	the		G3.1
3	candente	burning		G2.1
4	mañana	morning		G3.2; G2.2
5	de	of		
6	febrero	February		G4.2
7	en	on		G4.1; G5.2
8	que	which		G5.1
9	Beatriz	Beatrice		
10	Viterbo	Viterbo		
11	murió	died		
12	,	,		
13	después de	after		
14	una	an		G3.1
15	imperiosa	imperious		
16	agonía	agony		G3.2
17	que	that		
18	no	not	-1	G6.2
19	se	itself	-2	G8.1
20	◇	did	2	G6.1
21	rebajó	debase	1	G8.2

- **G1** (Noun before Adjective) Spanish typically places an adjective[1] after the noun[2] it modifies
- **G2** (Adjective before Noun) Spanish sometimes places an adjective[1] before the noun[2] it modifies.
- **G3:** (Determiner Agreement) A determiner[1] must agree with associated noun[2] in number and gender.

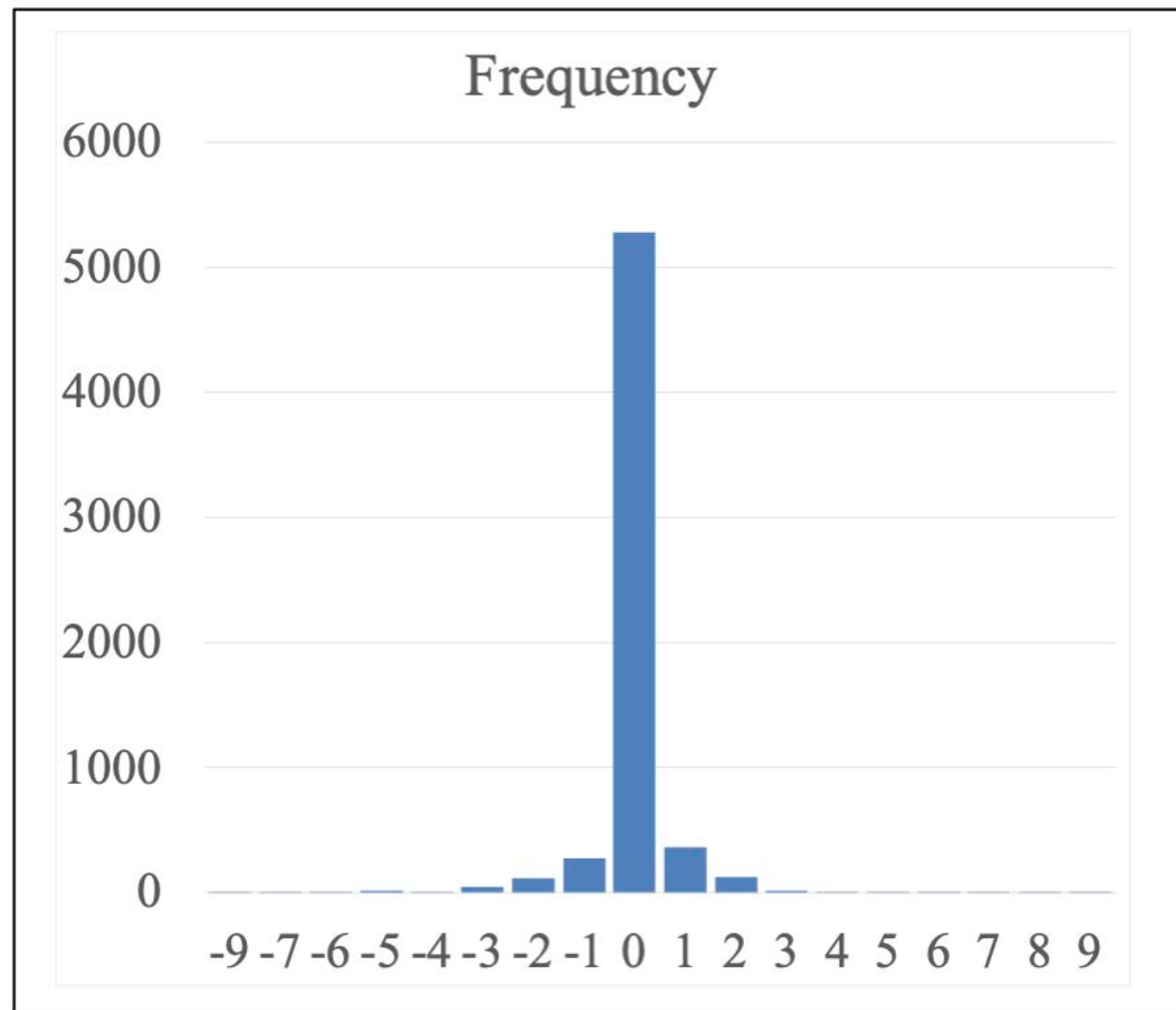


Cost Calculation over Borges *The Aleph*





Distribution of i_{delta}





Zipf Curve for Inversion Pattern Frequencies

Code	Freq	Description
G1	99	Noun before Adjective
G8	33	Preverbal Spanish Reflexive
G9	12	Complex Spanish Verb
G10	9	Prepositional Noun Compound
G11	8	Missing Preposition
G12	7	Verb Fronting
G13	6	Determiner Variance
G14	5	Negation and Verb Inversion
G15	5	Combined Pattern
G16	4	Single word as Phrase



Utility of Deeper L2 Understanding

斯里兰卡数十年之久的内战两个月前随着泰米尔猛虎组织被击败，组织头目维鲁皮莱普拉巴卡兰的死亡而结束。在内战的最后阶段，军队已经发现一些潜水设备。

Several submersible devices had been found by the military in the final phase of Sri Lanka's decades-long civil war that ended two months ago with the defeat of the Tamil Tigers and the death of the group's leader, Velupillai Prabhakaran.

§ 1 Sri Lanka 2 0 1	斯里兰卡 2 Sīlīlánkǎ 1	◇ 3 's 3 42	数十 4 shù shí ten 4 2	年 5 nián year 5 3	之 6 zhī ◇ 6 4	久 7 jiǔ long 7 5	的 8 de ◇ 8 6	内战 9 nèizhàn civil war 9 7	结束 33 jiéshù ended 10 27	两 10 liǎng two 11 8	个 11 gè ◇ 12 9	月 12 yuè months 13 10	前 13 qián ago 14 11	随着 14 suízhe with 15 12	◇ 15 the 16 43
泰米尔 16 tài mǐ'ěr Tamil 17 13	猛虎 17 měnghǔ Tiger 18 14	组织 18 zǔzhī organization 19 15	被 19 bèi being 20 16	击败 20 jībài defeated 21 17	， 21 ， 22 18	而 31 ér and 23 26	◇ 32 the 24 44	死亡 30 sǐwáng death 25 25	的 28 de of 26 24	◇ 29 the 27 45	组织 22 zǔzhī organization 28 19	◇ 23 's 29 46	头目 24 tóumù ringleader 30 20		
维鲁 25 wéi lǔ Viru 31 21	皮莱 26 pí lái pila 32 22	普拉巴卡兰 27 pǔ lābā kǎ lán Prabhakaran 33 23	。 34 。 34 28	在 35 zài in 35 29	◇ 36 the 36 47	最后 40 zuìhòu final 37 32	阶段 41 jiēduàn stages 38 33	的 38 de of 39 31	◇ 39 the 40 48	内战 37 nèizhàn civil_war 41 30	， 42 ， 42 34	◇ 43 the 43 49	军队 44 jūnduì army 44 35	已经 45 yǐjīng has 45 36	



Deep Reading Example (头目)

Several submersible devices had been found by the military in the final phase of Sri Lanka's decades-long civil war that ended two months ago with the defeat of the Tamil Tigers and the death of the group's leader, Velupillai Prabhakaran.

斯里兰卡数十年之久的内战两个月前随着泰米尔猛虎组织被击败，组织头目维鲁皮莱普拉巴卡兰的死亡而结束。在内战的最后阶段，军队已经发现一些潜水设备。



Deep Reading Example (头目)

Several submersible devices had been found by the military in the final phase of Sri Lanka's decades-long civil war that ended two months ago with the defeat of the Tamil Tigers and the death of the group's **leader** Velupillai Prabhakaran.

斯里兰卡数十年之久的内战两个月前随着泰米尔猛虎组织被击败，组织**头目**维鲁皮莱普拉巴卡兰的死亡而结束。在内战的最后阶段，军队已经发现一些潜水设备。

Does “头目” mean “leader” ?



Deep Reading Example (头目)

Several submersible devices had been found by the military in the final phase of Sri Lanka's decades-long civil war that ended two months ago with the defeat of the Tamil Tigers and the death of the group's **leader** Velupillai Prabhakaran.

斯里兰卡数十年之久的内战两个月前随着泰米尔猛虎组织被击败，组织**头目**维鲁皮莱普拉巴卡兰的死亡而结束。在内战的最后阶段，军队已经发现一些潜水设备。

Does “头目” mean “leader” ?

Noun [[edit](#)]

頭目

1. **leader** (person that leads or directs); **boss**; **ringleader**; **chieftain**

犯罪頭目 / **犯罪头目** — *fànzui tóumù* — **crime lord**

Wiktionary



Deep Reading Example (头目)

Several submersible devices had been found by the military in the final phase of Sri Lanka's decades-long civil war that ended two months ago with the defeat of the Tamil Tigers and the death of the group's **leader** Velupillai Prabhakaran.

斯里兰卡数十年之久的内战两个月前随着泰米尔猛虎组织被击败，组织**头目**维鲁皮莱普拉巴卡兰的死亡而结束。在内战的最后阶段，军队已经发现一些潜水设备。

Does “头目” mean “leader” ?

§ 1 0	斯里兰卡 2 Sīlīlānkǎ Sri Lanka 2 1	◇ 3 ◇ 's 3 42	数十 4 shù shí ten 4 2	年 5 nián year 5 3	之 6 zhī ◇ 6 4	久 7 jiǔ long 7 5	的 8 de ◇ 8 6	内战 9 nèizhàn civil war 9 7	结束 33 jiéshù ended 10 27	两 10 liǎng two 11 8	个 11 gè ◇ 12 9	月 12 yuè months 13 10	前 13 qián ago 14 11	随着 14 suízhe with 15 12	◇ 15 ◇ the 16 43	泰米尔 16 tài mǐ'ěr Tamil 17 13
猛虎 17 měnghǔ Tiger 18 14	组织 18 zǔzhī organization 19 15	被 19 bèi being 20 16	击败 20 jībài defeated 21 17	， 21 ， 22 18	而 31 ér and 23 26	◇ 32 ◇ the 24 44	死亡 30 sǐwáng death 25 25	的 28 de of 26 24	◇ 29 ◇ the 27 45	组织 22 zǔzhī organization 28 19	◇ 23 ◇ 's 29 46	头目 24 tóumù ringleader 30 20	维鲁 25 wéi lǔ Viru 31 21			
皮莱 26 pí lái pila 32	普拉巴卡兰 27 pǔ lābā kǎ lán Prabhakaran 33	。 34 。 34	在 35 zài in 35	◇ 36 ◇ the 36	最后 40 zuìhòu final 37	阶段 41 jiēduàn stages 38	的 38 de of 39	◇ 39 ◇ the 40	内战 37 nèizhàn civil_war 41	， 42 ， 42	◇ 43 ◇ the 43	军队 44 jūnduì army 44	已经 45 yǐjīng has 45	发现 46 fāxiàn found 46		



USC Guarani Corpus

Guarani, spoken in Paraguay, is a highly agglutinative or polysynthetic language. It is relatively straightforward to create hypergloss text from glossed text and the stories in the online USC Guarani Corpus are a delight to work with at <https://guaranicorpus.usc.edu/> (Maria Luisa Zubizarreta, PI)



USC Guarani Corpus

Story #1

https://guaranicorpus.usc.edu/Story1.html

Home Glossed Stories ▾ Search by Morpheme ▾

O-ho peteĩ karia'y **hína** peteĩ ñú-re.
 3Sg-go one young.man Cont one field-Prep

A young man was going on/across a field.

Iba un joven por el campo.

Jones Hypergloss Candidate

j ^f	f	e'	i ^e	gloss
1	§	§	1	
3	o·	◇	2	3Sg-
5	peteĩ	a	3	one
6	karia'y	young man	4	young.man
2	◇	was	5	
4	·ho	walk·	6	-go
7	hína	·ing	7	Cont
10	·re	across	8	-Prep
8	peteĩ	a	9	one
9	ñú·	field	10	field-
11	.	.	11	

Guarani, spoken in Paraguay, is a highly agglutinative or polysynthetic language. It is relatively straightforward to create hypergloss text from glossed text and the stories in the online USC Guarani Corpus are a delight to work with at <https://guaranicorpus.usc.edu/> (Maria Luisa Zubizarreta, PI)

The gloss has been rotated 90° to read vertically in the hypergloss shown in Figure 2. The hypergloss has been sorted along the English word order column **i^e**. Sorting it on the Guarani column **j^f** yields the original Guarani order. Recall that morpheme boundaries are indicated with a raised dot '·'.



Guarani Hypergloss Candidate

USC Gloss (e)

A young man was going on/across a field. Then he was looking at his side and his sight went [saw] towards a big snake. The snake's eyes were protruding and the young man did not know what was happening. Then he got closer and he saw that a large rock was squeezing the snake [he realized] and he could not get out from under it. When the snake saw the young man, he screamed at him: "Please take me out of this place," he said, very agitated. And [since] the snake was so big, the young man was afraid of it. "I will not be able to take you out, because you put me in danger," the guy said to him [snake].

Guarani Original (f)

Upéi omaña ohóvo ijképe ha hesa oho sapy'a peteĩ mbói guasúre. Hesa sēmba hína la mbói, ha la karia'y ndoikuaái mba'épa ojehu chupe. Upéi ojave ha ohecha nimbo peteĩ ita guasu ojopy ra'e la mbóipe ha ndoikatuvéi osẽ iguýgui. Ohechávo la mbói la karia'ype, osapukái voi ichupe: cheronohēmínako aimehá gui! — he'í, ijuku'a pepe. Ha tuicha iterei la mbói, okyhyje chugui la karia'y. Ndaikatu mo'ãi aipe'a ndéve, pórke nde cheapeligra — he'í chupe la típo.

Jones Hypergloss (e')

A young man was walking across a field. Then as he looked to his side, suddenly his eyes were drawn to a big snake. The snake's eyes were bulging, and the young man did not know what was happening to him. Then he got closer and he saw a big rock: it squeezed the snake and it could not get out from under it. And when the snake saw the young man, he screamed at him: get me out please I am here! — he said, agitated and shaking. And the snake was so big, the young man was afraid of it. I will not be able to free you now, because you may harm me — the guy said to him.

Note: **f** and **e'** are two serialized views of the same hypergloss object, whereas **f** and **e** are different text objects.





Three Translations and a Hypergloss

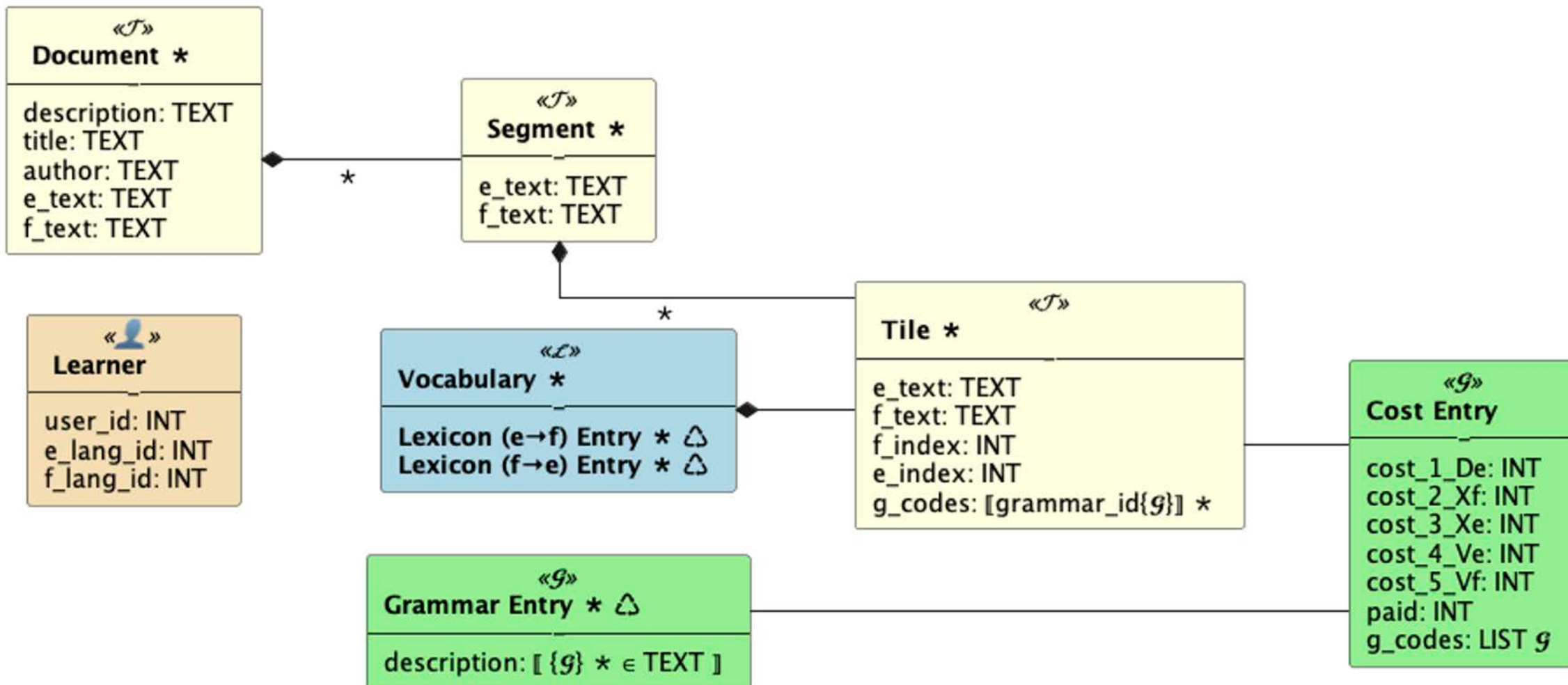
(An Epistemological Claim, not an Ontological Claim)

- f* Aujourd'hui, maman est morte. Ou peut-être hier, je ne sais pas.
- e*¹ Mother died today. Or, maybe, yesterday; I can't be sure.
- e*² Mother died today. Or maybe yesterday, I don't know.
- e*³ Mother died today. Or yesterday maybe, I don't know
- e*' Today, Mom died. Or maybe yesterday, I don't know.



Glimpse of Learner Model

(and Hint of an LLM Exit Strategy)



A Mirror for Any Language

Toward Universal Accessibility for Our Shared Cultural Heritage

Doug Jones

MIT 14 May 2025

For MITHIC, Part 0



The Way of Lao Tzu (Chinese)

Jones Hypergloss

The walkable way is the broken way.

The nameable name is the broken name.

The origin of Heaven and Earth is nameless.

Naming is the mother of all things.

So release desires, to always recognize subtlety.

keep desires, to always recognize appearance.

Then see both sides, from the same source,

of the same depth of meaning, yet different names.

From the deepest depth, it is the door of all subtleties.

Lao Tzu (Modernized)

道可道，非常道。

名可名，非常名。

無名天地之始；

有名萬物之母。

故常無，欲以觀其妙；

常有欲，以觀其徼。

此兩者，同出

而異名，同謂之玄。

玄之又玄，衆妙之門。





Death and the Dervish (Selimović)

Rakić and Dickey Translation

I begin my story for nothing, without benefit for myself or anyone else, from a need stronger than benefit or reason. I must leave a record of myself, the chronicled anguish of my inner conversations, in the vague hope that a solution will be found when all accounts have been settled (if they may ever be), when I have left my trail of ink on this paper, which lies in front of me like a challenge.

Selimovic Original

Почињем ову своју причу,
низашто, без користи за себе и за
друге, из потребе која је јача од
користи и разума, да остане
запис мој о мени, записана мука
разговора са собом, с далеком
надом да ће се наћи неко
рјешење кад буде рачун сведен,
ако буде, кад оставим траг
мастила на овој хартији што чека
као изазов.

Jones Hypergloss Rendering

I begin this story of mine, for nothing, without benefit to myself or to others, out of a need which is stronger than use or reason, to keep my record for myself, a written tormented interview with oneself, in the vague hope that some decision will be found when the account is settled, if it will be, when I have left the trace of ink on this paper that is waiting as a challenge.





Half Day (Mahfouz, Arabic)

Johnson-Davies Translation

HALF A DAY

I proceeded alongside my father, clutching his right hand,

running to keep up with the long strides he was taking.

All my clothes were new: the black shoes, the green school uniform, and the red tarboosh.

My delight in my new clothes, however, was not altogether unmarred, for this was no feast day but the day on which I was to be cast into school for the first time.

Mahfouz Original

نصف نؤوم

سرت إلى جانب أبي مته متعلقا بنؤومناه

جرئوت لألحق بخطاه الواسعة

ملابسي كلها جدنودة، الحذاء الأسود و
المرئولة الخضراء و الطربوش الأحمر

غنور أنى لم أسعد بالملابس الجدنودة
سعادة صافنودة، فنؤومي لم نؤكن نؤوم
عنود ولكنه أول نؤوم نؤلقى بى فى
المدرسة

Jones Hypergloss Rendering

Half a day

I walked to the right side of my father, clutching his hand.

I dared to follow his wide footsteps.

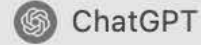
All my clothes are new: the black shoes and the green trousers and the red fez.

Except that I didn't find happiness in the new clothes, not pure happiness, because the first day was not a holiday, but rather the first day I was thrown into school.





Language and the Nation State



What was the role between language and formation of the nation states in the early modern period?



Thought about language and nation state formation for 27 seconds >

In the early modern period (roughly the 15th through the 18th centuries), language played a pivotal—and often underestimated—role in the consolidation of political power and the emergence of nation-states. While the fully modern notion of the “nation-state” is usually associated with the 19th century, many of the underlying processes that linked language and political identity began earlier. Here are some of the key ways in which language influenced the formation of nation-states in the early modern era:



Thanks to Our Hosts at MIT Global Languages

