Domain Adaptation in Neural Machine Translation using a Qualia-Enriched FrameNet

Alexandre Costa, Mateus Marim,
Ely Matos & Tiago Torrent
Federal University of Juiz de Fora

TL; DL ///

This paper presents two implementations of a methodology for domain adaptation in NMT systems using a qualia-enriched FrameNet as a semantically structured external resource.

The Problem
/// Domain Adaptation

(1) O jogador de basquete converteu a bandeja
The basketball player scored the lay-up

(2) O garçom colocou as tijelas na bandeja
The waiter put the bowls on the tray

Winning_moves

Definition

A competitor or team, the Athlete, makes a move that awards points.

Example(s)

Core Frame Elements

FE Core:

Athlete [Athlete] The individual or team who scores the point.

Point [Point] Outcome of the successful move played by the Athlete.

Non-Core Frame Elements

Relations

Lexical Units

landing.n | layout.n | layup.n | leap off.v | leaping.n | lock.n

Utensils

Definition

A Utensil is a container, a tool or something that is especially for household use. It is created for a specific Use. Several properties of the Utensil can be specified, such as its Creator, Time_of_creation, Name, Type, Origin, Constituent_parts, Description, Material of which it is composed, Quantity and the Place where it is.

Example(s)

Core Frame Elements

FE Core:

Utensil [Utensil] It indicates a utensil made for a specific Use.

Non-Core Frame Elements

Relations

Lexical Units

saucer.n | scale.n | tray.n

Frame-to-frame relations

/// Winning_moves inherits Moves

/// Moves is a subframe of Sports_event

/// Moves uses Athletes and Sports

Frame element-to-frame relation

/// The Athlete FE in Winning_moves frame is linked to the following frames

Athletes (LUs: athlete.n, competitor.n)

Athletes_by_sport (LUs: boxer.n, golfer.n)

Athletes_by_position (LUs: center.n, wing.n)

Frame-to-frame and FE-to-frame relations are not LU-specific.

The relation between Winning_moves and Athletes is not able to represent that the lay-up is a winning move performed by a basketball player.

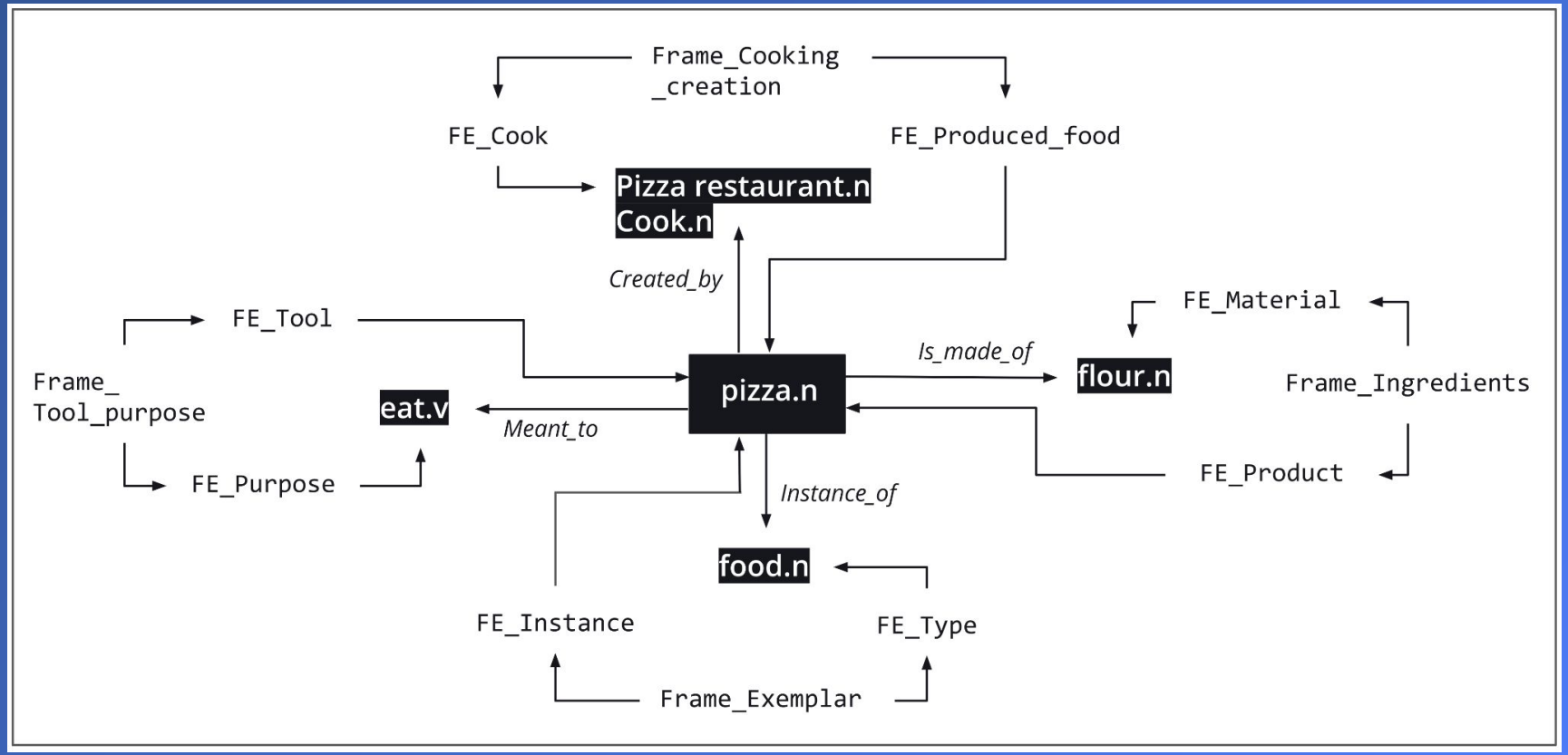
Ternary Qualia Relations
/// FrameNet meets the Generative Lexicon

Qualia

Four qualia roles originally proposed

Formal	*(type of)*
Telic	*(made for)*
Constitutive	*(made of)*
Agentive	*(created by)*

$$\begin{bmatrix} \textit{pizza.n} \\ \\ \textit{QUALIA} & \begin{bmatrix} \textit{F} = \textit{food.n} \\ \textit{T} = \textit{eat.v} \\ \textit{C} = \textit{flour.n} \\ \textit{A} = \textit{cook.n, pizza restaurant.n} \end{bmatrix} \end{bmatrix}$$


Frame_Cooking
_creation
FE_Cook
FE_Produced_food
Pizza restaurant.n
Cook.n
Created_by
FE_Material
FE_Tool
Is_made_of
flour.n
Frame_
Tool_purpose
eat.v
pizza.n
Frame_Ingredients
Meant_to
FE_Purpose
FE_Product
Instance_of
food.n
FE_Instance
FE_Type
Frame_Exemplar

Intentionally_act

Definition

This is an abstract frame for acts performed by sentient beings.

Example(s)

Core Frame Elements

FE Core:

Agent [Agent] **semantic_type:** @sentient	Someone who performs the intentional act.

FE Core-Unexpressed:

Act [Act] **semantic_type:** @state_of_affairs	It identifies the Act that the Agent performs intentionally.

basketball player.n

lay up.n

Scylla
/// domain adaptation using frames and qualia

A Two-Step Process

/// Frame Disambiguation

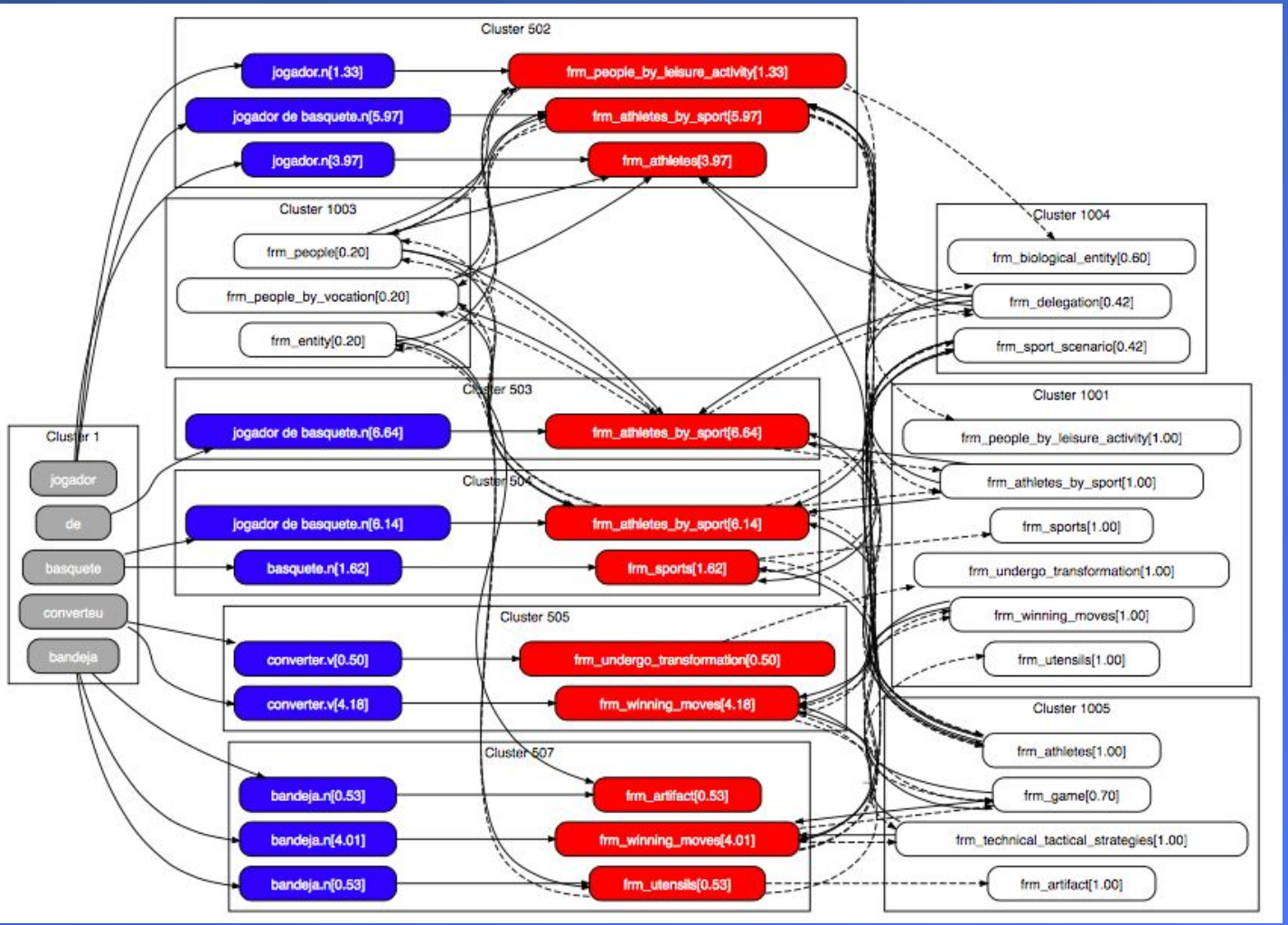
/// Terminology Injection

Pre-Processing Stage

Post-Editing Stage

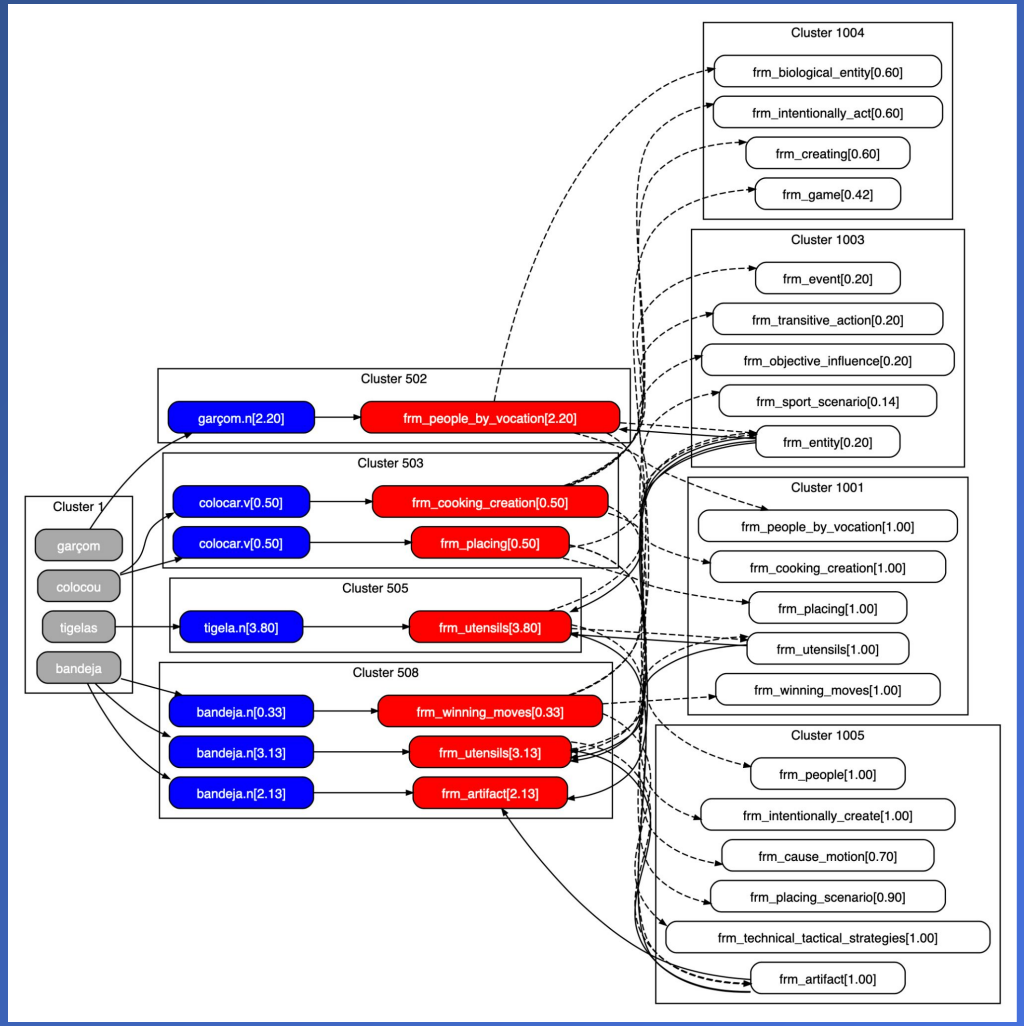
DAISY

1. Input sentence is parsed for dependencies
2. MWEs are retrieved from FN-Br
3. Lemma clusters are defined
4. LUs associated to lemmas are retrieved
5. Qualia relations between LUs are retrieved
6. Frames evoked by each LU are retrieved
7. FE-to-frame relations are retrieved



O jogador de basquete converteu a bandeja

The basketball player scored the lay-up

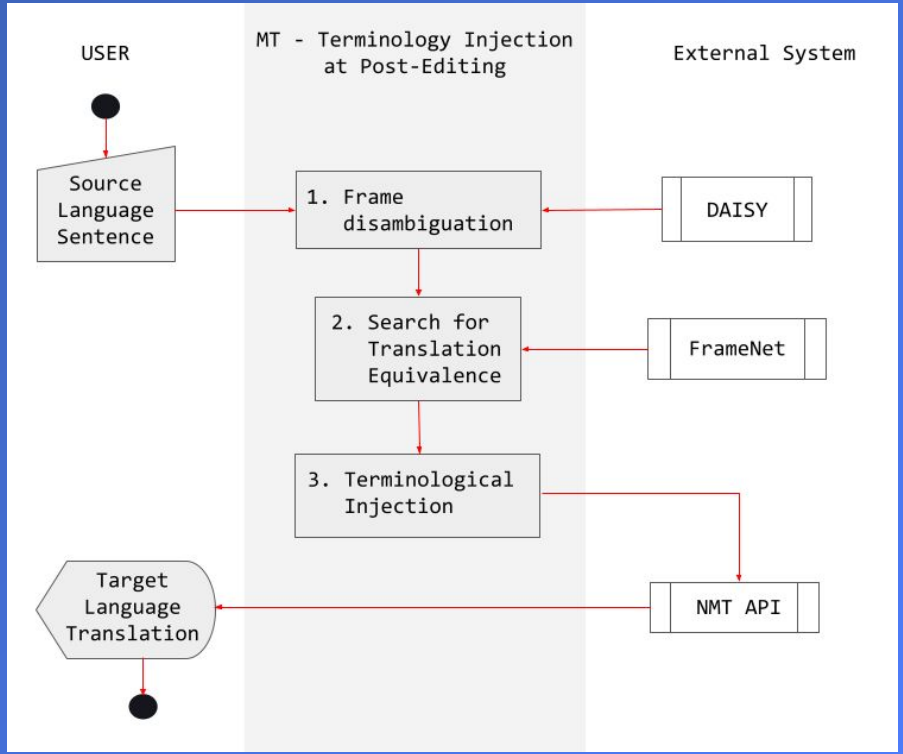


O garçom colocou as tijelas na bandeja

The waiter put the bowls on the tray

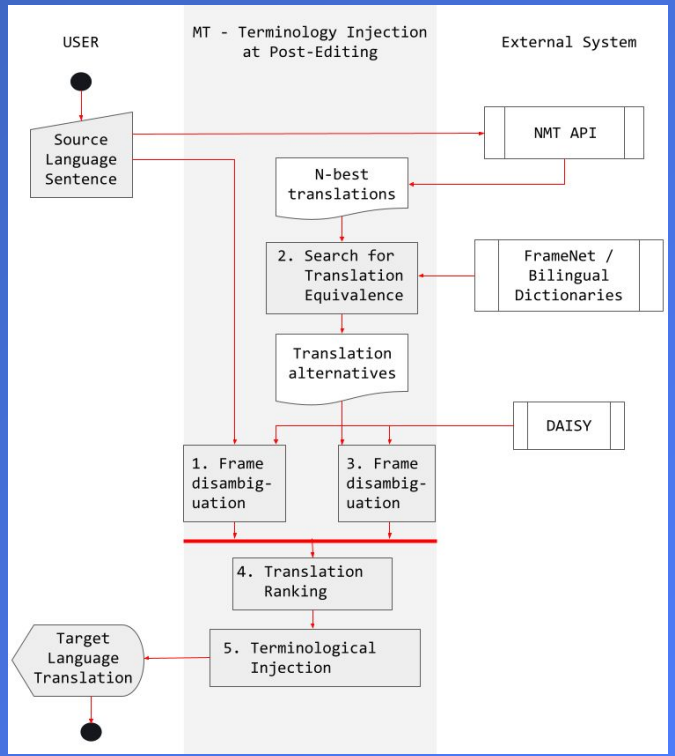
Scylla-S

/// Terminology Injection during the pre-processing stage



Scylla-T

/// Terminology Injection during the post-editing stage



Evaluation
/// domain adaptation in sports for the br-pt/en language pair

Dataset

/// 50 br-pt source sentences featuring at least one polysemous lemma

/// 50 en reference translations

/// 72.4% of in-domain frame preservation

Experiments & Metrics

/// Source sentences were submitted to a commercial NMT API (baseline) and to Scylla-S and Scylla-T

/// Machine translations were evaluated for BLEU, TER and HTER (using professional translators)

Results

	Baseline	Scylla-S	Scylla-T
BLEU	53.13	48.12	**53.66**
TER	**36.23**	42.63	36.47
HTER	13.80	10.44	**7.38**

Discussion

O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada

Source sentence

The winger is the player with less time to think about setting up a strike

Gold standard translation

Discussion

O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada

Source sentence

The forward is the player who has less time to think about setting up a move

Baseline system (TER=26.66 / HTER=0.08)

Discussion

O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada

Source sentence

The wing is the player who has less time **to think in the setup of** a play

Scylla-S (TER=53.33 / HTER=0.06)

Discussion

O ponta é o jogador que menos tempo tem para pensar na armação de uma jogada

Source sentence

The winger is the player who has less time to think about setting up a play

Scylla-T (TER=20.00 / HTER=0.00)

Conclusions and Limitations

Scylla-T improves the performance of the baseline system by 47% in HTER

No fine tuning is needed

Dataset is small and experiments represent a proof of concept

Baseline is a commercial system

Thank you!

tiagotorrent.com

@TorrentTiago