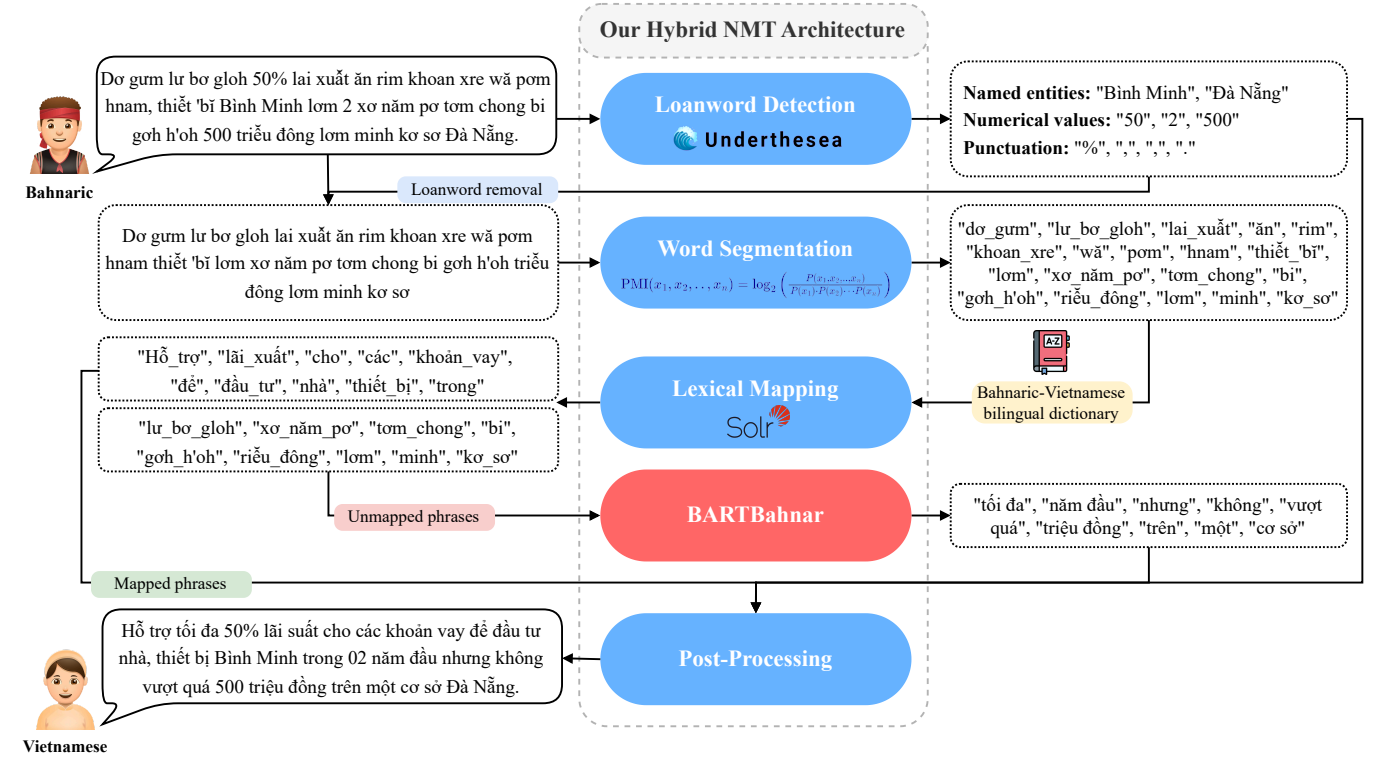
Our Hybrid NMT Architecture
Bahnaric
Dơ gưm lư bơ gloh 50% lai xuất ăn rim khoan xre wă pơm hnam, thiết 'bĩ Bình Minh lơm 2 xơ năm pơ tơm chong bi gơh h'oh 500 triệu đông lơm minh kơ sơ Đà Nẵng.
Loanword Detection
Underthesea
Named entities: "Bình Minh", "Đà Nẵng"
Numerical values: "50", "2", "500"
Punctuation: "%", ",", ",", "."
Loanword removal
Dơ gưm lư bơ gloh lai xuất ăn rim khoan xre wă pơm hnam thiết 'bĩ lơm xơ năm pơ tơm chong bi gơh h'oh triệu đông lơm minh kơ sơ
Word Segmentation
$\mathrm{PMI}(x_1, x_2, \ldots, x_n) = \log_2\left(\frac{P(x_1, x_2, \ldots, x_n)}{P(x_1) \cdot P(x_2) \cdots P(x_n)}\right)$
"dơ_gưm", "lư_bơ_gloh", "lai_xuất", "ăn", "rim", "khoan_xre", "wă", "pơm", "hnam", "thiết_'bĩ", "lơm", "xơ_năm_pơ", "tơm_chong", "bi", "gơh_h'oh", "riều_đông", "lơm", "minh", "kơ_sơ"
"Hỗ_trợ", "lãi_xuất", "cho", "các", "khoản_vay", "để", "đầu_tư", "nhà", "thiết_bị", "trong"
Lexical Mapping
Solr
Bahnaric-Vietnamese bilingual dictionary
"lư_bơ_gloh", "xơ_năm_pơ", "tơm_chong", "bi", "gơh_h'oh", "riều_đông", "lơm", "minh", "kơ_sơ"
Unmapped phrases
BARTBahnar
"tối đa", "năm đầu", "nhưng", "không", "vượt quá", "triệu đồng", "trên", "một", "cơ sở"
Mapped phrases
Post-Processing
Hỗ trợ tối đa 50% lãi suất cho các khoản vay để đầu tư nhà, thiết bị Bình Minh trong 02 năm đầu nhưng không vượt quá 500 triệu đồng trên một cơ sở Đà Nẵng.
Vietnamese