



# Machine Translation of Multiple-Word Expressions: Using NOOJ

**Ali Boulaalam**

 <http://orcid.org/0000-0002-0465-1914>

*Multidisciplinary Faculty of Errachidia, Morocco*

**Nisrine El Hannach**

 <http://orcid.org/0000-0001-9794-9842>

*Faculty of Sciences, Fes, Morocco*

## ABSTRACT

*This article is a presentation of a project conducted by a team of linguists that addresses the importance of the syntactic lexicon theory in the creation of methodological inputs for linguistic analyzers as being at the heart of machine translation. NooJ environment is adopted as an open-source computer platform to conduct practical operations for translating multi-word expressions from Arabic into French and English, and in the near future the team will be working on the inclusion of other languages. This linguistic phenomenon is of great importance in testing the procedural extent of machine translation systems.*

## INTRODUCTION

Machine translation (MT) constitutes a central application of artificial intelligence (AI) and natural language processing (NLP), aiming to automate the transfer of meaning between natural languages. Over the past decades, MT has evolved from early rule-based systems to statistical approaches and, more recently, to neural machine translation (NMT), which currently dominates both academic research and industrial deployment. Despite their impressive gains in fluency and global coherence, neural systems continue to encounter systematic difficulties when processing linguistically complex phenomena, most notably multiword expressions (MWEs) and idiomatic expressions.

MWEs are characterized by varying degrees of fixedness and non-compositionality, meaning that their global interpretation cannot be reliably derived from the meanings of their individual components. As Sag et al. (2002) famously observe, MWEs are “*a pain in the neck for NLP*,” precisely because they challenge models that rely on compositional semantics. Idiomatic expressions pose an even greater challenge, as they are often culturally embedded and pragmatically constrained.

DOI: 10.4018/407605

These difficulties are particularly acute in morphologically rich languages such as Arabic, where inflectional complexity interacts with idiomatic structure. Numerous studies have shown that even state-of-the-art neural systems tend to favor literal or statistically dominant translations when confronted with Arabic idioms, resulting in outputs that are fluent but semantically inadequate (Ouerhani, 2008; Savary et al., 2019).

In response to these limitations, this chapter proposes a linguistically informed approach to translating Arabic verbal idiomatic expressions using the NOOJ open-source platform. Grounded in syntactic lexicon theory, the approach models idiomatic expressions as structured lexical units endowed with explicit morphological, syntactic, and semantic constraints. The chapter presents the design and implementation of a reusable idiom database and situates this resource within contemporary discussions on hybrid AI systems that combine symbolic knowledge with neural learning.

The objectives of this contribution are threefold:

- (i) to demonstrate the relevance of syntactic lexicon theory for the computational treatment of MWEs;
- (ii) to describe a reproducible methodology for encoding idiomatic expressions in NOOJ; and
- (iii) to position linguistically explicit resources as complementary to neural MT within modern AI architectures.

## **1. OVERVIEW OF MACHINE TRANSLATION PROGRAMS**

A machine translation (MT) system is an automated computational system designed to convert text from one natural language into another. Within artificial intelligence, MT systems are understood as complex architectures that integrate linguistic knowledge, computational procedures, and, in many cases, learning mechanisms (Hutchins & Somers, 1992). Rather than functioning as a single program, an MT system typically consists of multiple interacting components that jointly support text analysis, transfer, and generation.

At a fundamental level, MT systems incorporate several core elements. First, linguistic rules and models form the backbone of rule-based and hybrid systems. These rules encode morphological, syntactic, and semantic information that guides the analysis of source-language input and supports the resolution of ambiguity during translation. As Hutchins (1993) notes, “linguistic analysis remains indispensable for languages with rich morphology and flexible word order.”

Second, specialized computational dictionaries play a central role. Unlike conventional bilingual dictionaries, these resources are structured lexicons enriched with grammatical features, argument structure, and semantic constraints. In advanced systems, dictionaries also encode multiword expressions (MWEs) and idiomatic constructions, allowing the translation engine to process them as single lexical units rather than compositional sequences. As Sag et al. (2002) emphasize, “multiword expressions are pervasive in language and must be treated as lexical units rather than syntactic accidents.”

Third, user interfaces and control modules support interaction, post-editing, and resource maintenance, enabling human–machine collaboration and iterative system improvement.

The translation process typically follows a sequence of stages: segmentation, morphological and syntactic analysis, lexical transfer, and target-language generation with local reordering. Advances in linguistic engineering have translated theoretical insights into practical tools such as morphological analyzers, syntactic parsers, annotation frameworks, and automatic recognition mech-

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/machine-translation-of-multiple-word-expressions/407605](http://www.igi-global.com/chapter/machine-translation-of-multiple-word-expressions/407605)

## Related Content

---

### Effect of Delay in EOG Signals for Eye Movement Recognition

Rajat Rakesh Jhnujhunwala and Geethanjali P. (2021). *Advancing the Investigation and Treatment of Sleep Disorders Using AI* (pp. 71-80).

[www.irma-international.org/chapter/effect-of-delay-in-eog-signals-for-eye-movement-recognition/285270](http://www.irma-international.org/chapter/effect-of-delay-in-eog-signals-for-eye-movement-recognition/285270)

### Mastering Change Management in AI-Powered Supply Chain Digital Transformation

Ehap H. Sabri (2025). *Supply Chain Transformation Through Generative AI and Machine Learning* (pp. 153-182).

[www.irma-international.org/chapter/mastering-change-management-in-ai-powered-supply-chain-digital-transformation/368669](http://www.irma-international.org/chapter/mastering-change-management-in-ai-powered-supply-chain-digital-transformation/368669)

### Identification and Prevention of Joint Gray Hole and Black Hole Attacks

Munesh C. Trivedi and Sachin Malhotra (2019). *International Journal of Ambient Computing and Intelligence* (pp. 80-90).

[www.irma-international.org/article/identification-and-prevention-of-joint-gray-hole-and-black-hole-attacks/225772](http://www.irma-international.org/article/identification-and-prevention-of-joint-gray-hole-and-black-hole-attacks/225772)

### Rethinking "Oil Nationalism": The Case of Anglo Iranian Oil Company (AIOC)

Neveen Abdelrehim (2015). *International Journal of Signs and Semiotic Systems* (pp. 33-49).

[www.irma-international.org/article/rethinking-oil-nationalism/142499](http://www.irma-international.org/article/rethinking-oil-nationalism/142499)

### Gaining New Insights Into Consumer Behavior Through Operations Management in Supply Chains

Neha Verma, Purnendu Kumar Patra, Ritesh Sharma, Nikhil Polke, Priyanka Rawal and Ganesh Waghmare (2024). *Utilization of AI Technology in Supply Chain Management* (pp. 274-288).

[www.irma-international.org/chapter/gaining-new-insights-into-consumer-behavior-through-operations-management-in-supply-chains/340897](http://www.irma-international.org/chapter/gaining-new-insights-into-consumer-behavior-through-operations-management-in-supply-chains/340897)