

The Utilization of CAMEL Tools in the Part-of-Speech Analysis of Arabic in the Short Story *Abdullah wa al-Ushfur*

Khairina Nasution¹, Rima Mella², Rozanna Mulyani³

^{1,2} Arabic Literature Study Program, Faculty of Cultural Sciences, Universitas Sumatera Utara, Indonesia

Email: khairinanasution@usu.ac.id¹; rimamella148@gmail.com²; rozanna.mulyani@usu.ac.id³

Submission Track:

Received: 02-04-2026, Final Revision: 01-07-2026, Available Online: 15-07-2026

Copyright © 2026 Authors



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

ABSTRACT

This study aims to analyze Arabic word categories, namely *ism*, *fi'il*, and *harf*, in the short story *Abdullah wa al-Ushfur* using CAMEL Tools. The study employed a qualitative descriptive approach assisted by *Natural Language Processing* (NLP) technology. The data consisted of the text of *Abdullah wa al-Ushfur*, obtained from the *muthala'ah* materials of the Kulliyatul Mu'allimin al-Islamiyah (KMI) second grade. The analysis was conducted using the *Part-of-Speech* (POS) Tagging feature in CAMEL Tools to automatically identify word classes based on their grammatical functions. The findings revealed that CAMEL Tools successfully identified various word classes, which were subsequently classified into the three primary categories of Arabic grammar: *ism*, *fi'il*, and *harf*. The category of *ism* was frequently used to refer to characters, objects, places, and conditions within the story, while *fi'il* contributed to the development of the narrative through the characters' actions. Meanwhile, *harf* functioned as a linking element that maintained the coherence of sentence structures. The findings suggest that CAMEL Tools can support Arabic linguistic analysis in a more systematic and efficient manner. This study is expected to contribute to technology-based Arabic linguistic studies and serve as a reference for future research integrating NLP into the analysis of Arabic literary texts.

Keywords: *CAMEL Tools; Arabic word categories; Part-of-Speech Tagging; NLP; Arabic language.*

Pemanfaatan CAMEL Tools dalam Analisis Kategori Kata Bahasa Arab pada Cerpen *Abdullah wa al-Ushfur*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kategori kata bahasa Arab yang meliputi *ism*, *fi'il*, dan *harf* pada cerpen *Abdullah wa al-Ushfur* menggunakan CAMEL Tools. Penelitian menggunakan pendekatan deskriptif kualitatif berbantuan teknologi *Natural Language Processing* (NLP). Data penelitian berupa teks cerpen *Abdullah wa al-Ushfur* yang bersumber dari materi muthala'ah Kulliyatul Mu'allimin al-Islamiyah (KMI) kelas 2. Analisis dilakukan melalui fitur *Part-of-Speech* (POS) Tagging pada CAMEL Tools untuk mengidentifikasi kelas kata secara otomatis berdasarkan fungsi gramatikalnya. Hasil penelitian menunjukkan bahwa CAMEL Tools mampu mengidentifikasi berbagai kelas kata yang kemudian dikelompokkan ke dalam tiga kategori utama tata bahasa Arab, yaitu *ism*, *fi'il*, dan *harf*. Kategori *ism* banyak digunakan untuk menyebut tokoh, benda, tempat, dan keadaan dalam cerita, sedangkan kategori *fi'il* berperan dalam membangun alur melalui tindakan tokoh. Adapun kategori *harf* berfungsi sebagai unsur penghubung yang menjaga keterpaduan struktur kalimat. Temuan penelitian menunjukkan bahwa pemanfaatan CAMEL Tools dapat mendukung analisis linguistik bahasa Arab secara lebih sistematis dan efisien. Penelitian ini diharapkan dapat memperkaya kajian linguistik Arab berbasis teknologi serta menjadi referensi bagi penelitian selanjutnya yang mengintegrasikan NLP dalam analisis teks sastra Arab.

Kata kunci: *CAMEL Tools; kategori kata; bahasa Arab; Part-of-Speech Tagging; NLP.*

PENDAHULUAN

Perkembangan teknologi digital dan kecerdasan buatan telah mendorong transformasi besar dalam berbagai bidang ilmu, termasuk linguistik komputasional dan *Natural Language Processing* (NLP). NLP merupakan bidang interdisipliner yang menggabungkan linguistik, ilmu komputer, dan kecerdasan buatan untuk memungkinkan komputer memahami, memproses, serta menghasilkan bahasa manusia secara otomatis (Jurafsky & Martin, 2023). Dalam beberapa dekade terakhir, perkembangan NLP telah menghasilkan berbagai aplikasi yang digunakan secara luas, seperti *machine translation*, *sentiment analysis*, *information retrieval*, *question answering*, dan *part-of-speech tagging* (Darwish et al., 2021). Kemajuan tersebut menunjukkan bahwa teknologi bahasa tidak lagi terbatas pada penelitian akademik, tetapi telah menjadi bagian penting dalam berbagai aktivitas digital masyarakat modern. Sejalan dengan perkembangan tersebut, penelitian linguistik juga mulai memanfaatkan berbagai perangkat NLP untuk mendukung proses analisis bahasa yang

sebelumnya dilakukan secara manual sehingga menjadi lebih cepat, sistematis, dan efisien (Obeid et al., 2020).

Bahasa Arab merupakan salah satu bahasa yang memperoleh perhatian besar dalam pengembangan NLP karena memiliki jumlah penutur yang sangat luas serta posisi penting dalam bidang pendidikan, budaya, dan keagamaan. Meskipun demikian, pemrosesan bahasa Arab secara otomatis menghadapi berbagai tantangan yang tidak ditemukan pada banyak bahasa lainnya. Darwish dkk. (2021) menjelaskan bahwa bahasa Arab memiliki sistem morfologi yang sangat kaya dan kompleks, ditandai dengan penggunaan akar kata (*root*), pola (*pattern*), prefiks, sufiks, dan infiks yang dapat menghasilkan berbagai bentuk kata dari satu akar yang sama. Selain itu, banyak teks bahasa Arab ditulis tanpa harakat sehingga menimbulkan ambiguitas dalam proses identifikasi makna dan kategori gramatikal suatu kata (Alluhaibi, 2021). Kompleksitas tersebut menyebabkan analisis bahasa Arab membutuhkan pendekatan dan perangkat yang dirancang secara khusus agar mampu menghasilkan keluaran yang akurat. Oleh karena itu, pengembangan teknologi NLP bahasa Arab terus menjadi fokus penelitian dalam beberapa tahun terakhir.

Salah satu tahapan penting dalam NLP adalah *Part-of-Speech* (POS) Tagging, yaitu proses pemberian label kategori gramatikal pada setiap kata dalam suatu teks berdasarkan fungsi sintaktisnya (Alluhaibi, 2021). POS Tagging memiliki peran yang sangat penting karena menjadi dasar bagi berbagai proses analisis bahasa lainnya, seperti analisis sintaksis, analisis semantik, ekstraksi informasi, serta pembangunan korpus linguistik. Menurut Habash (2010), keberhasilan proses POS Tagging akan memengaruhi kualitas tahapan pemrosesan bahasa berikutnya karena kategori kata merupakan informasi dasar yang digunakan untuk memahami struktur suatu teks. Dalam konteks bahasa Arab, proses POS Tagging menjadi semakin penting mengingat tingginya tingkat ambiguitas morfologis yang dimiliki bahasa tersebut. Oleh sebab itu, berbagai perangkat POS Tagging bahasa Arab terus dikembangkan untuk meningkatkan akurasi analisis linguistik berbasis komputer.

Dalam tata bahasa Arab, kata secara umum dibagi menjadi tiga kategori utama, yaitu *ism*, *fi'il*, dan *harf*. *Ism* mencakup kata benda, nama diri, kata sifat, maupun bentuk nominal lainnya. *Fi'il* merupakan kategori kata yang menunjukkan tindakan, aktivitas, atau keadaan

yang berkaitan dengan waktu tertentu. Sementara itu, *harf* berfungsi sebagai unsur penghubung yang menjelaskan hubungan antarkata maupun antarklausa dalam suatu kalimat (Habash, 2010). Ketiga kategori tersebut merupakan fondasi utama dalam pembentukan struktur kalimat bahasa Arab sehingga identifikasi dan analisisnya menjadi aspek penting dalam kajian linguistik Arab. Melalui analisis kategori kata, peneliti dapat memperoleh gambaran mengenai karakteristik kebahasaan suatu teks, termasuk kecenderungan penggunaan unsur nominal, verbal, maupun partikel yang membangun makna dan struktur teks tersebut.

Penelitian mengenai pemanfaatan teknologi NLP dalam analisis bahasa Arab telah berkembang pesat dalam beberapa tahun terakhir. Salah satu perangkat yang banyak digunakan adalah CAMEL Tools yang diperkenalkan oleh Obeid dkk. (2020) sebagai *open-source toolkit* untuk NLP bahasa Arab yang menyediakan berbagai fitur, seperti tokenisasi, analisis morfologi, disambiguasi, analisis dialek, pengenalan entitas bernama, analisis sentimen, serta *part-of-speech tagging*. Kehadiran CAMEL Tools memberikan kemudahan bagi peneliti untuk melakukan berbagai analisis linguistik bahasa Arab secara lebih sistematis dan efisien. Dibandingkan dengan beberapa perangkat sebelumnya, CAMEL Tools dirancang secara khusus untuk mengakomodasi kompleksitas bahasa Arab dengan menyediakan berbagai sumber daya linguistik yang terintegrasi dalam satu platform (Obeid et al., 2020).

Sebelum hadirnya CAMEL Tools, berbagai perangkat NLP bahasa Arab telah dikembangkan untuk mendukung analisis linguistik secara otomatis. Habash dkk. (2009) memperkenalkan MADA+TOKAN sebagai perangkat yang mampu melakukan tokenisasi, diakritisasi, disambiguasi morfologis, POS tagging, stemming, dan lemmatisasi bahasa Arab. Selanjutnya, Pasha dkk. (2014) mengembangkan MADAMIRA sebagai integrasi dari MADA dan AMIRA yang menawarkan kecepatan dan akurasi yang lebih baik dalam analisis morfologi bahasa Arab. Abdelali dkk. (2016) juga memperkenalkan Farasa sebagai segmenter bahasa Arab yang efisien dan memiliki performa kompetitif. Kehadiran berbagai perangkat tersebut menunjukkan perkembangan pesat NLP bahasa Arab, namun CAMEL Tools menawarkan keunggulan berupa integrasi berbagai fitur NLP dalam satu kerangka

kerja yang bersifat terbuka, mudah digunakan, dan terus dikembangkan oleh komunitas peneliti (Obeid et al., 2020).

Penelitian ini memilih CAMEL Tools sebagai perangkat analisis utama karena ekosistemnya yang terintegrasi dan statusnya sebagai open-source toolkit yang terus dikembangkan oleh komunitas peneliti. Hal ini memberikan keunggulan aksesibilitas dan pemeliharaan fitur yang lebih baik dibandingkan dengan perangkat terdahulu seperti tiMADA+TOKAN atau MADAMIRA yang pengembangannya lebih terbatas.

Selain pengembangan perangkat NLP, sejumlah penelitian juga berfokus pada evaluasi performa *POS tagger* bahasa Arab. Alluhaibi (2021) melakukan penelitian komparatif terhadap lima *POS tagger* bahasa Arab, yaitu Stanford Arabic Tagger, CAMEL Tools, MADAMIRA, Farasa, dan Arabic Linguistic Pipeline (ALP), menggunakan sampel teks novel Arab. Hasil penelitian tersebut menunjukkan bahwa setiap perangkat memiliki karakteristik dan tingkat akurasi yang berbeda bergantung pada jenis korpus yang digunakan. Penelitian lain oleh Alfaidi dkk. (2023) membandingkan performa Farasa dan CAMEL Taggers pada data media sosial berbahasa Arab dan menemukan bahwa CAMEL Tools memiliki performa yang kompetitif dalam proses identifikasi kategori kata. Temuan tersebut menunjukkan bahwa pemanfaatan teknologi *POS tagging* terus berkembang dan diaplikasikan pada berbagai jenis data bahasa Arab.

Di sisi lain, penelitian mengenai NLP bahasa Arab juga telah diterapkan dalam berbagai bidang kajian. Elshabrawy dkk. (2023) mengembangkan parser bahasa Arab berbasis CAMEL untuk menghasilkan analisis sintaktis yang lebih akurat. Kallas dkk. (2024) mengusulkan penyelarasan berbagai *tagset* POS bahasa Arab melalui pengembangan *bridge tagset* guna mengatasi perbedaan anotasi antarperangkat. Alrayani dkk. (2024) memanfaatkan fitur linguistik bahasa Arab untuk mendukung analisis ulasan pelanggan pada aplikasi digital di Arab Saudi. Sementara itu, Sleiman dkk. (2024) menggunakan fitur POS sebagai salah satu indikator dalam identifikasi era puisi Arab klasik. Penelitian Hamed dkk. (2024) juga menunjukkan bahwa informasi lemma dan POS dapat dimanfaatkan dalam pengembangan korpus Arab-Inggris multidialek. Berbagai penelitian tersebut

memperlihatkan bahwa teknologi NLP bahasa Arab memiliki cakupan pemanfaatan yang sangat luas, baik dalam pengembangan sistem maupun analisis linguistik.

Perkembangan NLP bahasa Arab juga didukung oleh meningkatnya perhatian terhadap pengembangan sumber daya linguistik dan model bahasa. Zalmout dkk. (2021) menunjukkan bahwa adaptasi model bahasa yang sesuai dengan karakteristik bahasa Arab mampu meningkatkan performa berbagai tugas NLP. Alkuhlani dkk. (2023) menegaskan bahwa analisis morfologi dan disambiguasi masih menjadi tantangan utama dalam pemrosesan bahasa Arab akibat kompleksitas struktur katanya. Selain itu, Mubarak dkk. (2021) menyatakan bahwa keterbatasan sumber daya linguistik dan keragaman bentuk bahasa Arab merupakan tantangan yang perlu diatasi untuk memperluas penerapan NLP bahasa Arab di berbagai bidang. Temuan-temuan tersebut menunjukkan bahwa penelitian NLP bahasa Arab masih memiliki ruang pengembangan yang luas, baik dari sisi teknologi maupun penerapannya dalam kajian linguistik.

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada pengembangan perangkat, evaluasi performa sistem, atau pemanfaatan NLP pada korpus berskala besar, seperti media sosial, berita, dan korpus multidialek. Kajian yang secara khusus memanfaatkan perangkat NLP untuk menganalisis aspek kebahasaan pada teks sastra Arab masih relatif terbatas. Padahal, teks sastra memiliki karakteristik linguistik yang khas karena memadukan unsur naratif, deskriptif, dan ekspresif dalam penyampaian pesan. Analisis terhadap kategori kata dalam teks sastra dapat memberikan gambaran mengenai cara penulis membangun tokoh, peristiwa, latar, serta hubungan antarkomponen cerita melalui pilihan bahasa yang digunakan.

Cerpen *Abdullah wa al-Ushfur* merupakan salah satu teks sastra berbahasa Arab yang menarik untuk dikaji karena menggunakan bahasa Arab baku (*Modern Standard Arabic*) dengan struktur kebahasaan yang relatif jelas. Teks ini memuat berbagai bentuk kategori kata yang dapat dianalisis untuk memperoleh gambaran mengenai karakteristik linguistiknya. Namun demikian, berdasarkan penelusuran terhadap penelitian terdahulu, belum ditemukan penelitian yang secara khusus memanfaatkan CAMEL Tools untuk menganalisis kategori kata bahasa Arab pada cerpen *Abdullah wa al-Ushfur*. Dengan

demikian, masih terdapat celah penelitian yang perlu diisi melalui kajian yang mengintegrasikan pendekatan linguistik Arab dengan teknologi NLP.

Berdasarkan uraian tersebut, penelitian ini memiliki kebaruan pada pemanfaatan CAMEL Tools sebagai alat bantu dalam analisis kategori kata bahasa Arab pada teks sastra berupa cerpen. Berbeda dengan penelitian sebelumnya yang lebih banyak berorientasi pada evaluasi sistem atau pengembangan perangkat NLP, penelitian ini menitikberatkan pada penerapan teknologi tersebut dalam kajian linguistik Arab. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kategori kata bahasa Arab pada cerpen *Abdullah wa al-Ushfur* menggunakan CAMEL Tools melalui fitur *Part-of-Speech Tagging*. Hasil penelitian diharapkan dapat memperkaya kajian linguistik Arab berbasis teknologi, sekaligus menjadi referensi bagi penelitian selanjutnya yang mengkaji pemanfaatan NLP dalam analisis teks sastra Arab.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kualitatif dengan memanfaatkan teknologi *Natural Language Processing* (NLP) untuk menganalisis kategori kata bahasa Arab pada cerpen *Abdullah wa al-Ushfur*. Pendekatan deskriptif dipilih karena penelitian bertujuan untuk mendeskripsikan hasil identifikasi kategori kata yang diperoleh melalui proses *Part-of-Speech* (POS) Tagging menggunakan CAMEL Tools tanpa melakukan pengujian hipotesis statistik.

Data penelitian berupa teks cerpen berbahasa Arab berjudul *Abdullah wa al-Ushfur* yang bersumber dari materi muthala'ah Kulliyatul Mu'allimin al-Islamiyah (KMI) kelas 2. Teks diperoleh melalui situs Putra Kapuas yang memuat materi muthala'ah beserta terjemahannya. Cerpen dipilih karena menggunakan bahasa Arab baku (*Modern Standard Arabic*) serta memuat berbagai kategori kata yang memungkinkan dilakukannya analisis linguistik berbasis teknologi.

Instrumen utama dalam penelitian ini adalah CAMEL Tools, yaitu *open-source toolkit* untuk pemrosesan bahasa Arab yang menyediakan berbagai fitur NLP, termasuk *Part-of-Speech Tagging* (Obeid et al., 2020). Fitur POS Tagging pada CAMEL Tools digunakan untuk

mengidentifikasi kategori gramatikal setiap kata dalam teks berdasarkan sistem pelabelan yang dimiliki perangkat tersebut.

Teknik pengumpulan data dilakukan melalui metode dokumentasi. Peneliti terlebih dahulu mengumpulkan teks cerpen *Abdullah wa al-Ushfur*, kemudian melakukan seleksi dan penyesuaian format teks agar dapat diproses menggunakan CAMEL Tools. Selanjutnya, teks dimasukkan ke dalam sistem untuk memperoleh hasil pelabelan kategori kata secara otomatis.

Analisis data dilakukan melalui beberapa tahapan. Pertama, teks cerpen dipersiapkan dalam format yang sesuai untuk diproses menggunakan CAMEL Tools. Kedua, dilakukan proses *Part-of-Speech Tagging* untuk mengidentifikasi kategori kata pada setiap token dalam teks. Ketiga, hasil pelabelan diklasifikasikan ke dalam kategori kata bahasa Arab yang meliputi *ism*, *fi'il*, dan *harf*. Keempat, hasil identifikasi dianalisis secara deskriptif untuk menjelaskan karakteristik kategori kata yang muncul dalam cerpen. Tahap terakhir dilakukan interpretasi terhadap hasil analisis dengan mengaitkannya pada kajian linguistik bahasa Arab dan penelitian terdahulu yang relevan.

Untuk menjaga keabsahan data, peneliti melakukan pengecekan ulang terhadap hasil identifikasi yang diperoleh dari CAMEL Tools dengan membandingkannya pada kaidah dasar tata bahasa Arab yang berkaitan dengan kategori kata. Langkah ini dilakukan untuk meminimalkan kemungkinan kesalahan interpretasi dalam proses analisis.

HASIL PENELITIAN & PEMBAHASAN

1. Hasil Analisis POS Tagging Menggunakan CAMEL Tools

Analisis data pada penelitian ini dilakukan menggunakan CAMEL Tools melalui fitur Part-of-Speech (POS) Tagging. Fitur tersebut digunakan untuk mengidentifikasi kelas kata yang terdapat dalam teks cerpen *Abdullah wa al-Ushfur* secara otomatis. Teks cerpen yang telah dibersihkan dari harakat dimasukkan ke dalam sistem CAMEL Tools secara bertahap, kemudian sistem menghasilkan informasi mengenai kategori gramatikal setiap kata yang muncul dalam teks.

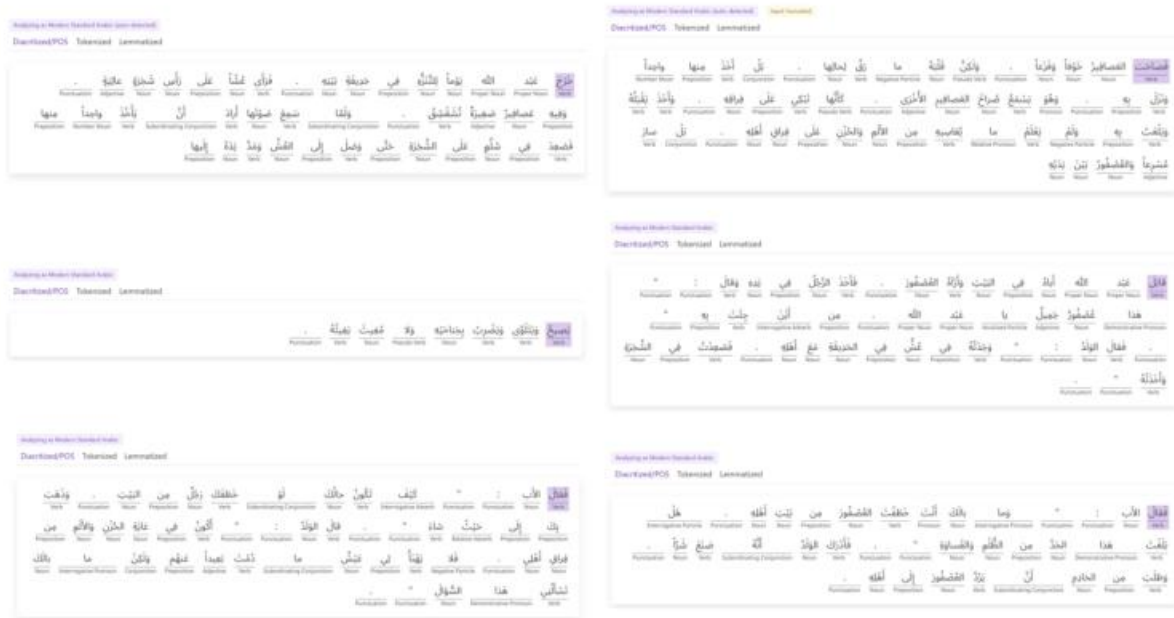
Terkait tantangan ambiguitas morfologis dalam bahasa Arab, CAMEL Tools mengatasi teks tanpa harakat dengan memanfaatkan model statistik berbasis konteks. Meskipun

sistem ini memiliki akurasi yang tinggi, peneliti tetap melakukan verifikasi manual terhadap hasil POS Tagging untuk memastikan akurasi klasifikasi, khususnya pada kata-kata yang bersifat polisemi dalam cerpen *Abdullah wa al-Ushfur*, guna meminimalkan kesalahan klasifikasi dan memastikan label sesuai dengan konteks naratif yang dimaksud.

Berdasarkan hasil analisis, CAMEL Tools berhasil mengidentifikasi berbagai kelas kata yang terdapat dalam cerpen. Kelas kata yang ditemukan meliputi noun, proper noun, adjective, number noun, verb, preposition, conjunction, subordinating conjunction, relative pronoun, demonstrative pronoun, interrogative pronoun, interrogative adverb, negative particle, pseudo verb, dan beberapa kategori lainnya. Hasil tersebut menunjukkan bahwa CAMEL Tools mampu melakukan identifikasi unsur-unsur kebahasaan dalam teks bahasa Arab dengan baik sehingga dapat dimanfaatkan sebagai alat bantu dalam penelitian linguistik komputasional.

Pada tahap ini, setiap kata dalam cerpen memperoleh label kelas kata sesuai dengan fungsi gramatikalnya. Sebagai contoh, kata *خرج* diidentifikasi sebagai verb, kata *عبد* dan *هلا* diidentifikasi sebagai proper noun, kata *العصفور* diidentifikasi sebagai noun, sedangkan kata *في* dan *على* diidentifikasi sebagai preposition. Hasil identifikasi tersebut menjadi dasar dalam proses pengelompokan kata ke dalam tiga kategori utama tata bahasa Arab, yaitu ism, fi'il, dan harf.

Selain menghasilkan informasi kelas kata, analisis menggunakan CAMEL Tools juga memberikan gambaran mengenai keragaman unsur linguistik yang terdapat dalam teks. Keragaman kategori yang muncul menunjukkan bahwa cerpen *Abdullah wa al-Ushfur* memiliki struktur kebahasaan yang cukup lengkap, baik dari segi unsur nominal, verbal, maupun unsur penghubung. Oleh karena itu, hasil POS Tagging dapat digunakan sebagai dasar untuk menganalisis kategori kata bahasa Arab serta karakteristik kebahasaan yang terdapat dalam teks.



Gambar 1. Hasil POS Tagging Cerpen Abdullah wa al-Ushfur Menggunakan CAMEL Tools

2. Analisis Kategori Kata Bahasa Arab

Setelah proses POS Tagging selesai dilakukan, langkah berikutnya adalah mengelompokkan hasil identifikasi kelas kata ke dalam tiga kategori utama tata bahasa Arab, yaitu ism, fi'il, dan harf. Pengelompokan ini dilakukan berdasarkan kesesuaian fungsi gramatikal masing-masing kelas kata dengan konsep kategori kata dalam tata bahasa Arab.

No	Kategori	Label CAMEL Tools	Keterangan
1	Ism	noun, proper noun, adjective, number noun	Nomina, nama diri, kata sifat, dan bilangan
2	Fi'il	verb, pseudo verb	Kata kerja dan bentuk verbal
3	Harf	preposition, conjunction, subordinating conjunction, relative pronoun,	Partikel dan unsur penghubung

demonstrative
 pronoun, interrogative
 pronoun, interrogative
 particle, negative
 particle

Kategori ism mencakup seluruh kata yang berfungsi sebagai nomina, nama diri, kata sifat, maupun bentuk nominal lainnya. Dalam penelitian ini, kategori noun, proper noun, adjective, dan number noun dikelompokkan ke dalam kategori ism. Sementara itu, kategori verb dimasukkan ke dalam kelompok fi'il karena menunjukkan tindakan, aktivitas, atau peristiwa yang dilakukan oleh tokoh dalam cerita. Adapun kategori harf mencakup berbagai unsur penghubung dan partikel gramatikal seperti preposition, conjunction, subordinating conjunction, relative pronoun, demonstrative pronoun, interrogative pronoun, interrogative particle, negative particle, dan kategori sejenis lainnya.

Berdasarkan hasil analisis, ditemukan bahwa ketiga kategori utama, yaitu ism, fi'il, dan harf, muncul secara konsisten dalam seluruh bagian cerpen. Kategori ism banyak digunakan untuk menyebut tokoh, benda, tempat, maupun keadaan yang berkaitan dengan jalannya cerita. Beberapa contoh kategori ism yang ditemukan antara lain *عبد الله، العصفور، الحديقة، الشجرة، الحزن، الألم، والعش*. Kehadiran unsur nominal tersebut menunjukkan bahwa cerpen memuat berbagai unsur deskriptif yang berfungsi memperjelas tokoh, latar, serta situasi yang dialami dalam cerita.

Kategori fi'il juga ditemukan dalam jumlah yang cukup banyak. Kemunculan kategori ini menunjukkan adanya aktivitas dan tindakan yang dilakukan tokoh sepanjang cerita. Kata kerja seperti *خرج، رأى، سمع، أخذ، وجد، خطف، طلب* menjadi unsur penting dalam membangun perkembangan alur cerita. Melalui penggunaan kata kerja tersebut, pembaca dapat mengikuti rangkaian peristiwa yang dialami tokoh utama dari awal hingga akhir cerita.

Sementara itu, kategori harf berfungsi sebagai unsur pendukung yang menghubungkan berbagai bagian kalimat. Beberapa bentuk harf yang ditemukan dalam teks antara lain *في، أن، على، إلى، من، و، لكن، وما*. Meskipun jumlahnya tidak sebanyak kategori ism, keberadaan

harf sangat penting karena membantu membentuk hubungan sintaktis antarkata dan antarklausa sehingga struktur kalimat menjadi lebih utuh dan mudah dipahami.

3. Pembahasan

Berdasarkan hasil analisis menggunakan CAMEL Tools, ditemukan bahwa teks cerpen *Abdullah wa al-Ushfur* mengandung berbagai kategori kata yang dapat dikelompokkan ke dalam tiga kelas utama tata bahasa Arab, yaitu ism, fi'il, dan harf. Ketiga kategori tersebut memiliki fungsi yang berbeda dalam membangun struktur dan makna teks. Hasil identifikasi menunjukkan bahwa kategori ism, fi'il, dan harf digunakan secara beragam sesuai dengan kebutuhan naratif cerita. Penggunaan kategori-kategori tersebut memperlihatkan karakteristik kebahasaan teks sastra Arab yang memadukan unsur deskriptif, tindakan, serta hubungan antarkalimat.

Lebih dari sekadar identifikasi otomatis, penelitian ini memberikan kontribusi pada model linguistik komputasional-terapan. Distribusi kategori kata ism, fi'il, dan harf dalam cerpen ini mencerminkan gaya bahasa penulis dalam membangun deskripsi tokoh (ism) dan alur aksi (fi'il), yang memperkaya kajian stilistika Arab berbasis teknologi.

Tingginya penggunaan kategori ism menunjukkan bahwa cerpen ini berfokus pada penggambaran tokoh, objek, latar, serta keadaan yang mendukung jalannya cerita. Beberapa kata yang termasuk kategori ism antara lain *الألم*, *العش*, *الشجرة*, *الحديقة*, *العصفور*, *عبد الله*, dan *الحزن*. Kata-kata tersebut berperan penting dalam membangun unsur naratif sehingga pembaca dapat memahami siapa pelaku dalam cerita, di mana peristiwa berlangsung, serta kondisi yang dialami tokoh. Dominasi kategori ism merupakan karakteristik yang lazim ditemukan dalam teks naratif karena cerita membutuhkan banyak referensi terhadap tokoh, tempat, benda, dan konsep yang menjadi bagian dari alur cerita.

Selain kategori ism, kategori fi'il juga ditemukan dalam jumlah yang cukup signifikan. Kategori ini ditunjukkan oleh kemunculan berbagai kata kerja seperti *وجد*, *أخذ*, *سمع*, *رأى*, *خرج*, *خطف*, dan *طلب*. Penggunaan kata kerja tersebut menunjukkan adanya rangkaian tindakan yang dilakukan tokoh dalam cerita. Kehadiran fi'il berfungsi untuk membangun alur naratif dan menunjukkan perkembangan peristiwa dari awal hingga akhir cerita. Melalui kata kerja tersebut, pembaca dapat mengikuti tindakan Abdullah mulai dari menemukan sarang

burung, mengambil anak burung, hingga akhirnya menyadari kesalahannya dan mengembalikan burung tersebut kepada keluarganya.

Sementara itu, kategori harf berfungsi sebagai unsur penghubung yang mengaitkan kata, frasa, maupun klausa dalam teks. Berdasarkan hasil analisis, ditemukan berbagai bentuk harf seperti في (di), على (di atas), إلى (ke), من (dari), و (dan), لكن (tetapi), وما (dan), serta أن (bahwa). Meskipun jumlahnya tidak sebanyak kategori ism, keberadaan harf sangat penting karena membantu membentuk hubungan gramatikal antarbagian kalimat. Tanpa penggunaan harf, struktur kalimat dalam teks tidak akan terbentuk secara utuh dan makna yang disampaikan penulis menjadi kurang jelas.

Apabila ditinjau dari isi cerita, dominasi kategori ism menunjukkan bahwa penulis lebih banyak memberikan deskripsi mengenai tokoh, benda, dan keadaan yang berkaitan dengan pesan moral cerita. Cerpen Abdullah wa al-Ushfur tidak hanya berisi rangkaian tindakan tokoh, tetapi juga menggambarkan hubungan emosional antara anak burung dengan keluarganya. Hal ini terlihat dari penggunaan kata-kata seperti الألم dan الحزن yang menunjukkan keadaan emosional akibat perpisahan. Dengan demikian, penggunaan kategori ism berperan dalam memperkuat aspek deskriptif sekaligus pesan moral yang ingin disampaikan kepada pembaca.

Hasil penelitian ini juga menunjukkan bahwa CAMEL Tools dapat dimanfaatkan sebagai perangkat analisis linguistik yang efektif dalam mengidentifikasi kategori kata bahasa Arab secara otomatis. Melalui fitur POS Tagging, proses identifikasi kelas kata yang sebelumnya dilakukan secara manual dapat dilakukan dengan lebih cepat dan sistematis. Temuan ini sejalan dengan penelitian Obeid dkk. (2020) yang menyatakan bahwa CAMEL Tools menyediakan berbagai fitur analisis linguistik bahasa Arab yang dapat dimanfaatkan dalam penelitian berbasis NLP. Selain itu, hasil penelitian ini juga mendukung penelitian Alluhaibi (2021) yang menunjukkan bahwa teknologi POS Tagging dapat membantu proses analisis struktur bahasa Arab secara efisien.

Secara keseluruhan, hasil penelitian menunjukkan bahwa kategori ism menjadi kategori yang paling dominan dalam cerpen Abdullah wa al-Ushfur, diikuti oleh kategori fi'il dan harf. Dominasi kategori ism menunjukkan bahwa teks naratif lebih banyak

memanfaatkan unsur nominal untuk menggambarkan tokoh, latar, objek, dan keadaan. Sementara itu, kategori fi'il berfungsi membangun alur cerita melalui tindakan tokoh, sedangkan kategori harf berperan menjaga keterkaitan antarkomponen kalimat sehingga cerita dapat dipahami secara utuh oleh pembaca.

4. Perbandingan Temuan dengan Penelitian Terdahulu

Hasil penelitian ini menunjukkan bahwa CAMEL Tools mampu mengidentifikasi kategori kata bahasa Arab secara otomatis dan sistematis. Temuan ini sejalan dengan penelitian Obeid dkk. (2020) yang menyatakan bahwa CAMEL Tools merupakan perangkat NLP bahasa Arab yang mendukung berbagai tugas linguistik, termasuk analisis morfologi dan POS tagging. Kemampuan tersebut memungkinkan proses identifikasi kategori kata dilakukan dengan lebih cepat dibandingkan analisis manual.

Selain itu, hasil penelitian ini juga mendukung penelitian Alluhaibi (2021) yang menunjukkan bahwa teknologi POS Tagging dapat digunakan untuk mengidentifikasi kelas kata bahasa Arab dengan baik. Dalam penelitian ini, hasil POS Tagging yang diperoleh kemudian dimanfaatkan untuk mengelompokkan kata ke dalam kategori ism, fi'il, dan harf sehingga karakteristik kebahasaan teks dapat diamati secara lebih sistematis.

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada objek kajian yang digunakan. Penelitian terdahulu lebih banyak berfokus pada pengembangan sistem atau evaluasi performa perangkat NLP, sedangkan penelitian ini memanfaatkan CAMEL Tools untuk memanfaatkan CAMEL Tools untuk mengkaji kategori kata bahasa Arab pada teks sastra berupa cerpen. Dengan demikian, penelitian ini memberikan kontribusi berupa penerapan teknologi NLP dalam analisis linguistik teks sastra Arab.

5. Implikasi dan Keterbatasan

Temuan penelitian ini memiliki implikasi bagi pengembangan kajian linguistik Arab berbasis teknologi, khususnya dalam pemanfaatan alat NLP untuk mempermudah analisis kategori kata secara sistematis. Penelitian ini memiliki kelebihan karena memanfaatkan perangkat CAMEL Tools yang mampu melakukan identifikasi kategori kata secara otomatis sehingga proses analisis linguistik menjadi lebih efisien dibandingkan metode manual. Hasil

penelitian ini dapat menjadi referensi bagi mahasiswa dan peneliti yang ingin mengintegrasikan teknologi NLP dalam kajian morfologi dan sintaksis bahasa Arab.

Namun, penelitian ini memiliki keterbatasan dalam hal validitas empiris, karena analisis dilakukan secara deskriptif kualitatif tanpa pengukuran *inter-rater agreement* maupun metrik statistik seperti *F1-score*. Kami menyadari bahwa penggunaan *mixed methods* (menggabungkan statistik frekuensi dan analisis kualitatif) akan memberikan hasil yang lebih komprehensif. Selain itu, karena objek kajian hanya terbatas pada satu cerpen, temuan ini bersifat sebagai studi kasus awal (*pilot study*) dan belum dapat digeneralisasi untuk genre sastra lain seperti puisi atau drama yang memiliki karakteristik linguistik berbeda. Oleh karena itu, penelitian selanjutnya disarankan untuk melibatkan validasi oleh ahli nahwu (*human annotators*) guna menghitung akurasi sistem secara kuantitatif serta memperluas korpus data agar representasi penggunaan kategori kata lebih luas dan akurat.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis kategori kata bahasa Arab yang meliputi ism, fi'il, dan harf pada cerpen *Abdullah wa al-Ushfur* menggunakan CAMEL Tools. Analisis dilakukan melalui fitur *Part-of-Speech* (POS) Tagging yang mampu mengidentifikasi kelas kata secara otomatis berdasarkan fungsi gramatikalnya dalam teks bahasa Arab.

Hasil penelitian menunjukkan bahwa CAMEL Tools berhasil mengidentifikasi berbagai kelas kata yang kemudian dapat dikelompokkan ke dalam tiga kategori utama tata bahasa Arab, yaitu ism, fi'il, dan harf. Kategori ism banyak digunakan untuk menggambarkan tokoh, benda, tempat, serta keadaan yang terdapat dalam cerita. Sementara itu, kategori fi'il berperan dalam membangun alur cerita melalui tindakan tokoh, sedangkan kategori harf berfungsi sebagai unsur penghubung yang menjaga keterpaduan struktur kalimat.

Temuan penelitian ini menunjukkan bahwa cerpen *Abdullah wa al-Ushfur* memiliki karakteristik sebagai teks naratif yang memadukan unsur deskriptif dan unsur tindakan. Dominasi kategori ism menunjukkan kuatnya aspek deskripsi dalam penyampaian cerita, sedangkan penggunaan fi'il dan harf mendukung perkembangan alur serta hubungan antarkalimat dalam teks. Selain itu, penelitian ini membuktikan bahwa CAMEL Tools dapat

dimanfaatkan sebagai perangkat bantu yang efektif dalam analisis linguistik bahasa Arab berbasis Natural Language Processing (NLP).

Penelitian selanjutnya dapat dilakukan dengan menggunakan korpus yang lebih besar, seperti kumpulan cerpen, novel, atau teks pembelajaran bahasa Arab lainnya sehingga distribusi kategori kata yang diperoleh dapat memberikan gambaran yang lebih luas mengenai karakteristik kebahasaan teks bahasa Arab.

Pernyataan Penggunaan Kecerdasan Buatan (AI)

Para penulis menyatakan bahwa alat Kecerdasan Buatan (Artificial Intelligence/AI) hanya digunakan sebagai instrumen bantu selama proses penyusunan naskah ini. Secara khusus, ChatGPT, Grammarly, Quillbot digunakan untuk mendukung kejelasan bahasa, tata bahasa, dan format penulisan. Alat AI tersebut tidak digunakan untuk menghasilkan, merekayasa, atau memanipulasi data penelitian, analisis, interpretasi, maupun referensi. Seluruh keluaran yang dihasilkan oleh AI telah ditinjau, diverifikasi, dan disunting secara cermat oleh para penulis, yang sepenuhnya bertanggung jawab atas isi naskah ini. Penggunaan AI ini telah sesuai dengan Pernyataan Etika Publikasi dan Malpraktik (Publication Ethics and Malpractice Statement) dari *Journal of Linguistics, Culture, and Communication*.

Ucapan Terima Kasih

Para peneliti ingin mengucapkan terima kasih kepada Prof. Khairina Nasution atas dukungan dan hibah yang diberikan dalam menyelesaikan penelitian ini, juga semua pihak yang telah membantu penyusunan penelitian ini.

DAFTAR PUSTAKA

- Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A fast and furious segmenter for Arabic. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 11–16).
- Alfaidi, A., Alwadei, H., Alshutayri, A., & Alahdal, S. (2023). Exploring the performance of Farasa and CAMEL taggers for Arabic dialect tweets. *International Arab Journal of Information Technology*, 20(3), 349–356.

- Alkuhlani, S., Habash, N., & Roth, R. (2023). Advances in Arabic morphological analysis and disambiguation: A review. *ACM Computing Surveys*, 56(4), 1–35.
- Alluhaibi, R. (2021). A comparative study of Arabic part of speech taggers using literary text samples from Saudi novels. *Information*, 12(12), 523.
<https://doi.org/10.3390/info12120523>
- Alrayani, B., Kalkatawi, M., Abulkhair, M., & Abukhodair, F. (2024). From customer's voice to decision-maker insights: Textual analysis framework for Arabic reviews of Saudi Arabia's super app. *Applied Sciences*, 14(16), 6952.
<https://doi.org/10.3390/app14166952>
- Darwish, K., Habash, N., Abbas, M., Al-Khalifa, H. S., Al-Natsheh, H. T., El-Beltagy, S. R., Bouamor, H., Bouzoubaa, K., Cavalli-Sforza, V., El-Hajj, W., Jarrar, M., & Mubarak, H. (2021). A panoramic survey of natural language processing in the Arab world. *Communications of the ACM*, 64(12), 72–81.
- Elshabrawy, A., AbuOdeh, M., Inoue, G., & Habash, N. (2023). CamelParser 2.0: A state-of-the-art dependency parser for Arabic. In *Proceedings of the First Arabic Natural Language Processing Conference* (pp. 170–180).
- Eryani, F., Khalifa, S., & Habash, N. (2024). Recent trends in Arabic natural language processing resources and tools. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (pp. 1–10).
- Habash, N. (2010). *Introduction to Arabic Natural Language Processing*. Morgan & Claypool Publishers.
- Habash, N., Diab, M., & Rambow, O. (2012). Conventional orthography for dialectal Arabic. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)* (pp. 711–718).
- Habash, N., Eskander, R., & Hawwari, A. (2019). A morphological analyzer for Arabic dialects. *Natural Language Engineering*, 25(2), 195–225.
- Habash, N., Rambow, O., & Roth, R. (2009). MADA+TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and

- lemmatization. In *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools* (pp. 102–109).
- Hamed, I., Eryani, F., Palfreyman, D., & Habash, N. (2024). ZAEBUC-Spoken: A multilingual multidialectal Arabic-English speech corpus. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (pp. 17770–17782).
- Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed., draft).
- Kallas, O., Inoue, G., & Habash, N. (2024). EMAD: A bridge tagset for unifying Arabic POS annotations. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (pp. 5637–5643).
- Mubarak, H., Darwish, K., & Magdy, W. (2021). Arabic language processing: Challenges and opportunities. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(5), 1–23.
- Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMEL Tools: An open source Python toolkit for Arabic natural language processing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 7022–7032).
- Pasha, A., Al-Badrashiny, M., Diab, M. T., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., & Roth, R. (2014). MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)* (pp. 1094–1101).
- Sleiman, N. M., Hussein, A. A., Kuflik, T., & Minkov, E. (2024). Automatic era identification in classical Arabic poetry. *Applied Sciences*, 14(18), 8240. <https://doi.org/10.3390/app14188240>
- Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2021). Adapting language models for Arabic NLP. In *Proceedings of the Arabic Natural Language Processing Workshop* (pp. 1–12).

Zeroual, I., Lakhouaja, A., & Belahbib, R. (2017). Towards a standard part-of-speech tagset for the Arabic language. *Journal of King Saud University – Computer and Information Sciences*, 29(1), 43–50.