

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahasa Jawa adalah bahasa daerah yang memiliki penutur terbanyak di Indonesia. Menurut Badan Pusat Statistik dan Badan Bahasa, Bahasa Jawa memiliki penutur sebanyak 80 juta jiwa pada tahun 2023¹. Penutur Bahasa Jawa tersebar di berbagai provinsi di Indonesia yakni Jawa Tengah, Jawa Timur, Daerah Istimewa Yogyakarta, Lampung dan sebagainya. **Gambar 1** adalah peta persebaran penutur Bahasa Jawa yang berada di Jawa Tengah, dimana Bahasa Jawa menjadi bahasa yang dominan. Namun perlu diketahui, Bahasa Jawa tidak hanya dituturkan di Indonesia tetapi juga di luar negeri seperti Belanda, Suriname, dan New Caledonia (Wedhawati et al., 2006).

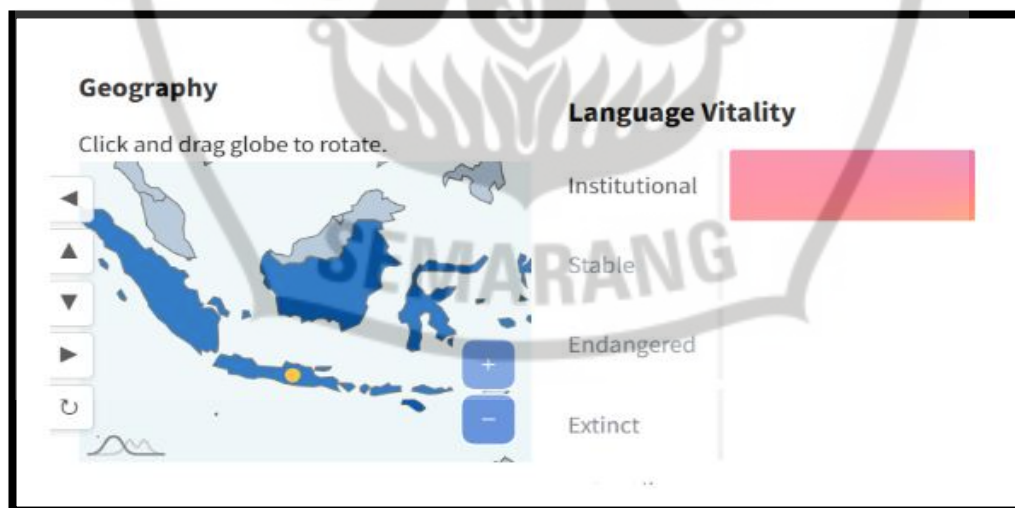


Gambar 1. Persebaran Penutur Bahasa Jawa di Jawa Tengah²

¹ <https://www.kemendikdasmen.go.id/en/siaran-pers/8438-balai-bahasa-yogyakarta-jaring-masukan-dari-pemangku-kepent>

² <https://petabahasa.kemendikdasmen.go.id/>

Selanjutnya, **Gambar 2** menunjukkan vitalitas Bahasa Jawa yang diambil dari website Ethnologue dengan adanya 4 tingkat vitalitas yaitu *institutional*, *stable*, *endangered*, dan *extinct*. Tingkat *institutional* adalah tingkatan paling tinggi karena bahasa digunakan dalam ranah institusi formal seperti pendidikan, pemerintahan, dan media serta masih diajarkan di sekolah-sekolah. Pada tingkat *stable* sebuah bahasa digunakan dalam ranah keluarga dan masih diwariskan ke anak-anak tetapi tidak digunakan dalam institusi formal. Pada tingkat *endangered* sebuah bahasa sudah tidak diwariskan ke anak-anak sehingga, sebuah bahasa berisiko mengalami kepunahan. Tingkat *extinct* adalah tingkat paling rendah karena bahasa mengalami kepunahan. Berdasarkan **Gambar 2**, Bahasa Jawa memiliki tingkat vitalitas paling tinggi yaitu tingkat institusional. Bahasa Jawa masih digunakan dalam ranah formal dan masih diajarkan di sekolah-sekolah, meskipun dalam muatan lokal.

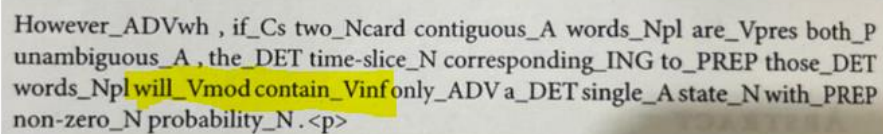


Gambar 2. Vitalitas Bahasa Jawa³

³ <https://www.ethnologue.com/language/jav/>

Verba Bahasa Jawa memiliki berbagai kompleksitas dari segi morfologi, sintaksis, dan semantis. Penelitian berfokus pada kompleksitas morfologi, khususnya afiksasi. Proses afiksasi dalam verba Bahasa Jawa berpotensi mengubah tingkat tutur dan diatesis. Oleh sebab itu, dalam penelitian ini Peneliti membuat *tagger* Bahasa Jawa, mengevaluasi performanya, dan mengidentifikasi distribusi afiksasi verba Bahasa Jawa dalam korpus sasaran.

Selain itu, afiksasi pada Bahasa Jawa adalah salah satu proses pembentukan kata yang paling produktif, seperti halnya bahasa-bahasa rumpun Austronesia, seperti, Indonesia, Sunda, Malagasi, dan sebagainya. Bahasa Jawa sebagai bahasa yang banyak dituturkan memiliki ciri yakni mempunyai afiksasi yang beragam seperti penggunaan sufiks -AKE, prefiks -M, dan sebagainya (Wulandari, 2021). Selanjutnya, Bahasa Jawa memiliki tingkat bahasa atau *speech level* yakni *ngoko* dan *krama* sebagai penanda kesantunan (Atmawati, 2021). Berdasarkan hal tersebut, Bahasa Jawa menjadi objek penelitian yang relevan untuk peneliti dalam mengkaji anotasi di ranah linguistik komputasi dan morfologi terutama untuk pembuatan perangkat lunak yang disebut *tagger*. *Tagger* dapat memberikan anotasi secara manual atau otomatis (Voutilainen, 1999). **Gambar 3** adalah contoh pembuatan *tagger* untuk kelas kata atau disebut dengan *part-of-speech tag* (POS), penanda deskriptif kelas kata untuk melakukan anotasi.



However_ADVwh , if_Cs two_Ncard contiguous_A words_Npl are_Vpres both_P
unambiguous_A , the_DET time-slice_N corresponding_ING to_PREP those_DET
words_Npl will_Vmod contain_Vinf only_ADV a_DET single_A state_N with_PREP
non-zero_N probability_N. <p>

Gambar 3. Contoh pembuatan *tagger* (Ruslan Mitkov, 2005)

Part-of-speech tagger digunakan untuk menandai berbagai kelas kata seperti nomina, adverbial, adjektiva, dan verba. Pada **Gambar 3**, kata *will* ‘akan’ dan *contain* ‘berisi’ mendapatkan anotasi ‘Vmod’ dan ‘Vinf’ karena kedua kata ini termasuk kategori kelas kata verba. Namun, kedua kata tadi memiliki perbedaan. Verba *will* ‘akan’ adalah untuk modalitas. Sedangkan, *contain* ‘berisi’ adalah verba infinitif. Penandaan kelas kata dapat dimanfaatkan untuk pembuatan kamus, dokumentasi bahasa dan tata bahasa dalam ranah linguistik.

Pembuatan *tagger* yang berkaitan erat dengan ilmu komputer ternyata dapat dikembangkan dalam ranah linguistik khususnya linguistik komputasi dan linguistik korpus. Hal ini dibuktikan dengan adanya penelitian dengan kajian serupa yang dipublikasikan pada artikel jurnal, dan disertasi yang membahas pengembangan *tagger* bukan hanya pemanfaatan *tagger*. Beberapa artikel jurnal yang membahas pengembangan *tagger* yakni Aarts et al. (1998); Bernhard & Dolińska (2025); Koleva et al. (2017); Saario et al. (2021). Bernhard & Dolińska (2025) mengembangkan POS *tagger* dalam Bahasa Dagur dan Bahasa Alsatian. Koleva et al. (2017) mengembangkan POS *tagger* dalam Bahasa Jerman dan Saario et al. (2021) mengembangkan POS *tagger* pada Bahasa Inggris. Selain itu, disertasi yang membahas pembuatan *tagger* yakni Hardie (2004) dan Prihantoro (2021). Hardie (2004) mengembangkan *part-of-speech tagger* untuk bahasa Urdu. Sedangkan, Prihantoro (2021) mengembangkan *morpheme tagger* untuk Bahasa Indonesia.

Tesis ini akan berfokus untuk membuat *tagger* kelas kata verba dalam Bahasa Jawa. Alasan peneliti memilih verba karena verba memiliki fungsi yang

penting dalam sebuah struktur klausa untuk menentukan jenis dan jumlah argumen (Carnie, 2011). Argumen biasanya berupa subjek atau objek (Carnie, 2011). Selain itu, verba dalam Universal Dependencies (UD) adalah inti dari sebuah klausa yang dapat menentukan subjek atau objek (De Marneffe et al., 2021). Verba terdiri dari tiga jenis valensi yakni transitif, intransitif, dan ditransitif (De Marneffe et al., 2021). Namun, pada penelitian ini Peneliti fokus pada notasi verba. Argumen verba tidak dibahas dalam penelitian ini.

1.2 Celah Penelitian

Pembuatan *tagger* dalam Bahasa Jawa sudah dilakukan oleh beberapa peneliti seperti Askhabi et al. (2020); Pramudita et al. (2016); dan Pratama et al. (2020), tetapi masih memiliki beberapa kelemahan seperti (1) *tagger* tersebut tidak bisa diakses untuk umum, (2) beberapa *tagger* yang memiliki akses terbuka mempunyai antarmuka tidak ramah pengguna untuk masyarakat umum, sehingga, *tagger* hanya mampu diakses untuk kalangan tertentu yang memahami bidang ilmu komputer. Kajian penelitian terdahulu, tersedia lebih detail pada **Bab 2.1**.

1.3 Kebaruan

Kebaruan penelitian tesis Peneliti yakni (1) membuat perangkat lunak yang bisa diakses oleh semua kalangan baik secara sistem dan *resource*, (2) menyediakan perangkat lunak *tagger* yang memiliki antarmuka ramah pengguna (3) pembuatan perangkat lunak *tagger* yang berfokus pada verba Bahasa Jawa ini dapat menjadi langkah awal yang dapat dikembangkan dan diperluas menjadi sebuah perangkat *tagger* yang utuh mencakup kelas kata lain seperti nomina, adjektiva, dan adverbia.

Penelitian parsial seperti memilih salah satu kelas kata untuk pembuatan *tagger* layak dilakukan. Hal tersebut dibuktikan dengan pembuatan *tagger* nomina dalam Bahasa Mandarin oleh Lin (2010).

1.4 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah pada penelitian ini yaitu bahasa-bahasa lokal atau bahasa daerah tidak mendapatkan banyak perhatian dalam hal digitalisasi, khususnya pembuatan *tagger* seperti yang sudah dilakukan dalam Bahasa Inggris dan Bahasa Indonesia. Pada kajian linguistik komputasi, sudah ada beberapa *tagger* Bahasa Jawa, namun tidak memiliki *resources* yang aksesnya terbuka dan tidak ramah pengguna (Askhabi et al., 2020; Pramudita et al., 2016; dan Pratama et al., 2020). Afiksasi dalam Bahasa Jawa sudah banyak yang dibahas, namun, belum ada yang membahas distribusinya secara kuantitatif.

1.6 Tujuan Penelitian

Tujuan penelitian tesis Peneliti adalah sebagai berikut:

- 1) Mengidentifikasi instrumen yang dibutuhkan dalam membuat *tagger* verba Bahasa Jawa
- 2) Mengungkap akurasi performa *tagger* verba Bahasa Jawa
- 3) Mengungkap produktivitas afiksasi verba Bahasa Jawa pada korpus UD Javanese

1.7 Batasan Penelitian

Limitasi penelitian ini adalah *tagger* yang dikhususkan untuk verba dalam Bahasa Jawa. Pembuatan *tagger* verba Bahasa Jawa dapat menjadi embrio untuk

memperluas penelitian selanjutnya yang berkaitan dengan kelas kata lainnya. Penelitian ini berfokus meneliti bentuk dan fungsi verba dari segi morfologi. Bentuk verba yang diteliti berupa afiksasi. Fungsi yang diteliti berupa diatesis (aktif dan pasif) dan tingkat tutur (*krama* dan *ngoko*). Pembuatan *tagger* perlu penyederhanaan *tagset*, sehingga diatesis yang digunakan aktif dan pasif serta tingkat tutur yang digunakan *krama* dan *ngoko*. Penyederhanaan ini mempunyai tujuan agar performa *tagger* dalam menganalisis verba lebih konsisten. Namun, *tagset* tetap bisa dikembangkan lebih detail secara bertahap. Penelitian afiksasi secara deskriptif sudah banyak dilakukan, tetapi secara komputasi masih sedikit peneliti yang melakukan terutama Bahasa Jawa. Oleh karena itu, penelitian ini meneliti bentuk afiksasi Bahasa Jawa khususnya verba dengan pendekatan linguistik komputasi.

1.8 Ruang Lingkup Penelitian

Secara umum, dialek Bahasa Jawa dikelompokkan menjadi dua yakni dialek geografis dan dialek sosial. Dialek geografis adalah dialek yang ditentukan berdasarkan perbedaan wilayah penutur. Ruang lingkup penelitian ini adalah dialek Bahasa Jawa standar yakni dialek yang mencakupi daerah Yogya dan Solo atau sering disebut dialek Yogya-Solo (Wedhawati et al., 2006). Dialek-dialek Bahasa Jawa lain seperti Banyumas, Osing, dan Jawa Timur tidak menjadi fokus kajian dalam penelitian ini. Menurut Wedhawati et al. (2006), dialek Bahasa Jawa standar digunakan oleh penutur di beberapa daerah seperti Yogyakarta, Solo, Purworejo, Magelang Temanggung, Klaten, Karanganyar, Sukoharjo dan Wonogiri.

1.9 Manfaat Penelitian

1.9.1 Manfaat Teoritis

- a) berkontribusi terhadap bidang ilmu linguistik khususnya linguistik komputasi dengan menyediakan *tagger* yang dapat dimanfaatkan untuk kajian dalam ranah linguistik
- b) berkontribusi terhadap dokumentasi bahasa yang dapat dimanfaatkan untuk pembuatan kamus elektronik

1.9.2 Manfaat Praktis

- a) menghasilkan perangkat lunak dengan akses yang terbuka baik secara sistem dan *resource*
- b) mempermudah analisis teks dalam skala besar