

Pengurai Kalimat Bahasa Banjar Dengan Menggunakan *Parser PC-PATR*

Ali Muhammad¹, Kamariah²

¹ Pendidikan Teknologi Informasi, STKIP PGRI BANJARMASIN

² Pendidikan Bahasa dan Sastra Indonesia, STKIP PGRI BANJARMASIN

¹alimuhammad@stkipbjm.ac.id

²kamariah@stkipbjm.ac.id

Abstrak— Pengurai (Parse) bahasa Banjar merupakan pohon pola yang merepresentasikan struktur sintaksis kalimat berdasarkan kaidah tata bahasa (*grammar*) bahasa Banjar. Tujuan dari penelitian ini adalah menghasilkan aplikasi pengurai (*parsing*) kalimat untuk proses sintaksis kalimat pada dokumen berbahasa Banjar. Tulisan ini menjelaskan langkah-langkah dalam menghasilkan pengurai bahasa Banjar menggunakan *parser PC-PATR* beserta analisisnya. Permasalahan terbesar dari penelitian ini adalah belum adanya korpus kalimat bahasa banjar, korpus file leksikon, *treebank*, dan file *grammar* untuk bahasa banjar. File leksikon dan file *grammar* bahasa banjar digunakan sebagai masukan *parser PC-PATR* dalam menghasilkan pohon pola sesuai aturan bahasa Banjar yang berlaku. Hasil penelitian menunjukkan bahwa persentase kealamian kalimat uji paling rendah sebesar 38% dan persentase kesesuaian pohon pengurai sebesar rata-rata sebesar 100%. Hal ini menunjukkan bahwa pengurai kalimat dapat diterima walaupun kalimat uji tidak 100% alami.

Kata Kunci— Pohon Pengurai, Bahasa Banjar, Parser PC-PATR

1. Pendahuluan

Pengurai kalimat adalah elemen penting dalam pemrosesan bahasa alami seperti analisis sintaksis berdasarkan tata bahasa (*grammar checking*), mesin translasi, mesin penanya dan penjawab, mesin pengekstrak informasi, mesin peringkasan kalimat, mesin pemrosesan kata, aplikasi leksikografi, pengenalan ucapan (*speech recognizers*) [1], aplikasi parafrase kalimat berdasarkan konteks kalimat [2], dan aplikasi pengukur kesukaran teks [3]. Representasi struktur sintaksis dengan pohon pola memungkinkan memberikan pemahaman yang lebih baik pada suatu kalimat [4].

Referensi terkait struktur bahasa Banjar dalam kamus bahasa Banjar [5] dan ragam penggunaan bahasa Banjar dalam artikel [6] sangatlah terbatas. Sejauh ini analisis sintaksis kalimat bahasa Banjar hanya terbatas pada pendekatan rasional tradisional. Pohon pengurai direpresentasikan secara manual berdasarkan aturan baku dalam tata bahasa Banjar.

Makalah ini menyajikan percobaan awal dalam membangun pohon pengurai (*parser*) probabilistik untuk bahasa Banjar menggunakan *parser PC-PATR*. Pohon pola

akan dibangkitkan secara otomatis dan disusun berdasarkan penanda kata (*tag*) dan aturan *grammar* yang digunakan pada *parser PC-PATR*.

2. Parser Bahasa Banjar

Dalam perkembangan bahasa Banjar yang ada di Kalimantan Selatan, pengurai kalimat dalam Bahasa Banjar menggunakan pohon pola hanya disusun berdasarkan pendekatan rasional tradisional secara manual. Tidak seperti dalam bahasa Indonesia. Pemeriksaan tata bahasa baku bahasa Indonesia sudah mulai diteliti dan dikembangkan pada tahun 1995 menggunakan sistem operasi unix [7]. Penelitian ini kemudian dikembangkan oleh [8, 9] secara linguistik berdasarkan aturan baku tata bahasa Indonesia [10]. Penelitian ini kemudian diperluas dengan analisis linguistik yang lebih canggih oleh [11], sehingga tercipta aturan *grammar* bahasa Indonesia menggunakan fitur *formalism PATR-II* [12]. Penelitian ini dikembangkan lebih jauh lagi oleh [1, 4]. Keluaran *parser* bahasa alami dalam bentuk grafis dikembangkan lebih jauh oleh [13] agar mudah dipahami oleh para pelajar pemula.

Permasalahan umum yang terjadi dalam pembangunan pengurai bahasa banjar ini adalah belum adanya korpus kalimat bahasa banjar, korpus file leksikon, *treebank*, dan file *grammar* untuk bahasa banjar. Keterbatasan ini membuat eksperimen penelitian ini mengadopsi tata bahasa simbolik acuan dari karya penelitian bahasa Indonesia. Meskipun penelitian tersebut mempunyai akurasi tinggi dengan masing-masing corpus uji cobanya.

2.1. Parser PC-PATR

Parser PC-PATR adalah *parser* sintaksis berbasis unifikasi. Tata bahasa (*grammar*) PC-PATR terdiri dari seperangkat aturan dan leksikon. Setiap aturan terdiri dari aturan struktur frase bebas konteks dan serangkaian batasan fitur. Leksikon menyediakan item yang dapat menggantikan simbol terminal (kelas kata) dari aturan struktur frasa, yaitu, huruf-huruf alfabet bersama-sama dengan fitur-fiturnya yang relevan [14].

2.2. File Grammar

File tata bahasa mencakup semua aturan susunan suatu kalimat, kecuali aturan di mana sisi kanan adalah terminal (kelas kata). Sintaks dari file tata bahasa adalah sebagai berikut:

```

Rule X -> X_1 CJ X_2
  <X cat> = <X_1 cat>
  <X cat> = <X_2 cat>
  <X arg1> = <X_1 arg1>
  <X arg1> = <X_2 arg1>
    
```

Dimana Simbol X dapat digunakan untuk simbol terminal atau non-terminal. Setiap kategori dalam aturan tata bahasa dapat diganti dengan dua salinan (X_1 dan X_2) dari kategori yang dipisahkan oleh CJ.

2.3. File Leksikon

File leksikon mencakup semua aturan pembentukan $A \rightarrow \sigma$ dimana σ adalah terminal. Himpunan terminal dianggap oleh PC-PATR untuk menjadi himpunan simbol di sisi kanan aturan dalam file leksikon. Untuk setiap aturan seperti itu file leksikon berisi catatan yang dibagi menjadi bidang, masing-masing bidang dimulai dengan penanda format standar di awal baris. Susunan file leksikon dijabarkan sebagai berikut :

```

\w baisy
\c vtran
\f <subcat> = ekatransitif
    
```

Dimana, \w adalah bentuk leksikal kata, \w juga menandai bidang awal dari setiap catatan kata. \c adalah kategori kata atau simbol nonterminal. \f adalah fitur batasan aturan (yaitu, dari struktur fitur di sisi kiri). Misalnya, aturan.

3. Desain Eksperimen

Bagian ini membahas cara menghasilkan masukan file masukan pengurai bahasa banjar menggunakan PC-PATR. Metode yang diusulkan dalam penyesuaian pengurai PC-PATR untuk bahasa banjar dijabarkan pada gambar 1. Berikut :



Gambar 1. Desain yang Diusulkan

3.1. Digitalisasi Kamus Bahasa Banjar-Indonesia

Proses digitalisasi kamus bahasa Banjar-Indonesia ini diperlukan karena belum tersedia dokumen/korpus digital kamus bahasa Banjar. Proses penyalinan (*scanning*) ini dengan menggunakan aplikasi OCR (*Optical Character Recognition*), agar hasil penyalinan dapat diedit secara langsung. OCR adalah aplikasi yang berfungsi untuk menterjemahkan gambar pada kamus cetak kedalam teks

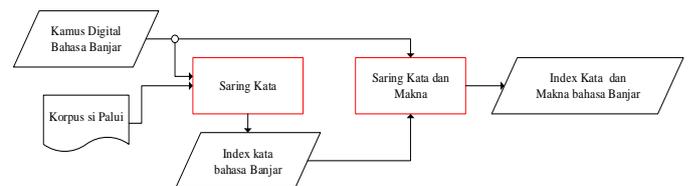
digital yang bisa diedit dan dimanipulasi. Proses Digitalisasi ini dijelaskan pada diagram alir berikut :



Gambar 2. Proses Digitalisasi Kamus Bahasa Banjar

3.2. Penyaringan Kata Bahasa Banjar

Penyaringan kata bahasa banjar digunakan untuk mendapatkan kata-kata bahasa banjar dari kamus digital bahasa Banjar-Indonesia. Penyaringan kata ini dimulai dengan membuat daftar kata bahasa banjar dari kamus bahasa Banjar-Indonesia dan dari korpus si Palui. Korpus si Palui adalah sekumpulan kata yang didapat dari artikel koran si Palui [6]. Proses penyaringan kata ini dijelaskan pada diagram alir berikut:



Gambar 3. Proses Penyaringan Kata Bahasa Banjar

Berikut adalah contoh hasil keluaran sistem :

Contoh 1. Masukan Proses Saring Kata :

Kamus Digital Bahasa Banjar : bubus – bocor atau koyak yang besar, biasanya disembelah bawah
Korpus Si Palui : Kada lawas Garbus pun datang membawa kantong plastik baisy tukaran

Hasil 1. Hasil yang Diharapkan dari Proses Saring Kata

- Indeks Kata : - bubus
 - kada
 - lawas
 - garbus
 - pun
 - datang
 - membawa
 - kantong
 - plastik
 - baisy
 - tukaran

Selanjutnya dengan indeks kata baha Banjar tersebut dapat kita gunakan untuk proses Saring Kata dan Makna Kata dari Kamus Digital Bahasa Banjar. Proses Saring Kata dan Makna dijabarkan pada Contoh 2. Dan Hasil 2. berikut :

Contoh 2. Masukan Proses Saring Kata dan Makna :

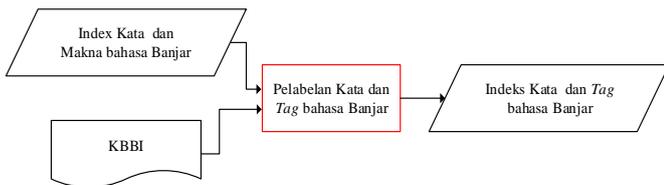
- Kamus Digital Bahasa Banjar : - bubus – bocor atau koyak yang besar, biasanya disebelah bawah
- lawas – lama
- Indeks Kata Bahasa Banjar : - bubus
- kada
- lawas
- garbus
- pun
- datang
- membawa
- kantong
- plastik
- baisi
- tukaran

Hasil 2. Hasil yang Diharapkan dari Proses Saring Kata

- Indeks Kata dan Makna Bahasa Banjar : - bubus – bocor
- lawas – lama

3.3. Pelabelan Kelas Kata

Pelabelan Kelas Kata digunakan untuk menentukan kelas setiap kata bahasa Banjar. Proses ini dilakukan secara komputerisasi dengan memadukan kelas setiap kata yang ada di kamus besar bahasa Indonesia (KBBI) [15] dengan kata-kata pada bahasa banjar. Untuk setiap kata bahasa banjar yang memiliki makna bahasa Indonesia, maka kelas kata bahasa Indonesiannya akan diambil untuk melabelin kelas kata bahasa Banjar. Proses Pelabelan kelas Kata Bahasa Banjar dijelaskan pada Gambar 4. dan Contoh 3. Berikut :



Gambar 1. Diagram Alir Proses Pelabelan Kata dan Tag (Penanda) Kata Bahasa Banjar

Contoh 3. Masukan Pelabelan Kata dan Tag (Penanda) Kata Bahasa Banjar:

- Indeks Kata dan Makna Bahasa Banjar : - bubus – bocor
- lawas – lama
- manjual - menjual
- KBBI (Kamus Besar Bahasa Indonesia) : - bocor [JJ]
- lama [JJ]
- menjual [VBT]

Hasil 3. Hasil yang Diharapkan dari Proses Pelabelan Kata dan Tag (Penanda) Kata Bahasa Banjar:

- Indeks Kata dan Tag Bahasa Banjar : - bubus [JJ]
- lawas [JJ]
- manjual [VBT]

3.4. Penyesuaian Aplikasi

Proses penyesuaian aplikasi pengurai kalimat bahasa banjar digunakan untuk menyesuaikan antara indeks kata dan tag (kelas kata) Banjar menjadi file lexicon yang diperlukan aplikasi parser PC-PATR untuk membaca simbol terminal dan non-terminal. Aturan pembentukan leksikon ini sudah baku dan mengikuti format yang sudah dijelaskan pada bab. 2.3. diatas. Proses penyesuaian aplikasi pengurai kalimat bahasa banjar tersebut dijabarkan pada Contoh 4. berikut :

Contoh 4. Masukan Proses Penyesuaian Aplikasi Pengurai Kalimat Bahasa Banjar:

- Indeks Kata dan Tag Bahasa Banjar : - bubus [JJ]
- lawas [JJ]
- manjual [VBT]

Hasil 4. Hasil yang Diharapkan dari Proses Penyesuaian Aplikasi Pengurai Kalimat Bahasa Banjar:

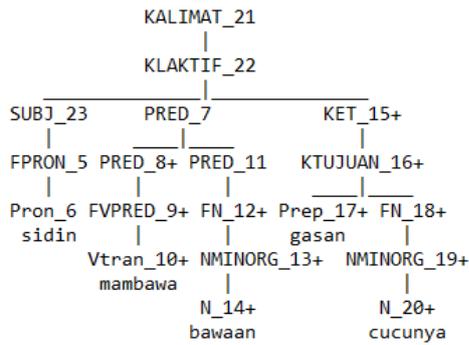
- File Lexicon Banjar : \w bubus
- \c JJ
- \w lawas
- \c JJ
- \w manjual
- \c VBT
- f <subcat> = ekatransitif

Setelah dilakukan penyesuaian file leksikon selanjut dilakukan penyesuaian file grammar. File grammar yang digunakan pada penelitian ini mangadaptasi file grammar yang sudah dibuat pada penelitian bahasa Indonesia sebelumnya [11].

4. Hasil dan Diskusi

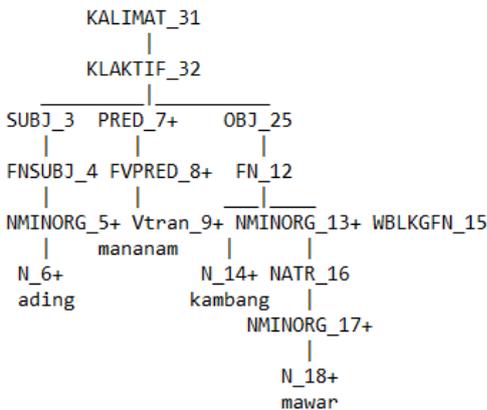
Penelitian ini menggunakan kalimat uji sebanyak 5 kalimat yang diambil secara acak dari korpus Si Palui. Kalimat uji kemudian diproses menggunakan parser PC-PATR untuk membangkitkan pohon pola. Pohon pola ini kemudian dievaluasi dan dibandingkan dengan aturan tata bahasa Banjar pada kamus bahasa Banjar. Jumlah responden penguji penelitian ini sebanyak 10 orang yang diambil dari pakar bahasa pendidikan bahasa dan sastra Indonesia STKIP PGRI Banjarmasin. Pengujian dilakukan untuk mengetahui persentase kealamian kalimat dan persentase kesesuaian pohon pola hasil keluaran parser PC-PATR. Gambar 5 berikut adalah contoh pohon pola kalimat tidak alami (ambigu) pada bahasa Banjar yang dibangkitkan menggunakan parser PC-PATR. Gambar 6 adalah contoh pohon pola alami (tidak ambigu) pada bahasa Banjar yang dibangkitkan dengan parser PC-PATR.

- Kalimat : sidin membawa bawaan gasan cucunya
- beliau membawa sesuatu untuk cucunya



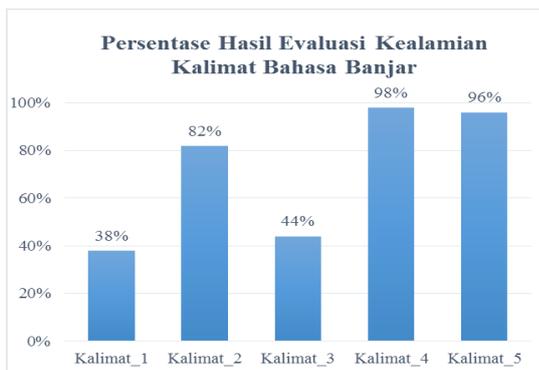
Gambar 5. Contoh pohon pola kalimat tidak alami (ambigu) hasil parser PC-PATR

Kalimat : Ading mananam kambang mawar
Adik menanam bunga mawar



Gambar 6. Contoh pohon pola kalimat alami (tidak ambigu) hasil parser PC-PATR Bahasa Banjar

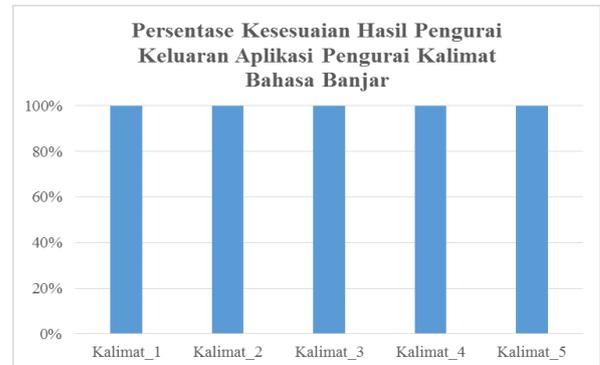
Pada gambar 5 dikatakan kalimat kurang alami karena objek dari kalimat tersebut kurang jelas. Pada kalimat formal bahasa Banjar, umumnya objek menggunakan kata benda untuk memperjelas kalimat. Kalimat tersebut dapat berubah menjadi sidin mambawa anggur gasan cucunya (beliau membawa anggur untuk cucunya). Persentase hasil pengujian kealamian kalimat uji bahasa Banjar ditampilkan pada Gambar 7. berikut :



Gambar 7. Hasil Evaluasi Kelamian Kalimat

Pada Gambar 7. diatas, persentase Kalimat_1 dan Kalimat_3 dibawah 50%. Hal ini dikarenakan kalimat tersebut tidak memenuhi kaidah tata bahasa baku bahasa Banjar. Namun, meskipun tidak memenuhi kaidah tata bahasa baku bahasa Banjar, kalimat tersebut lazim diucapkan dalam percakapan sehari-hari.

Selanjutnya, pohon pola dari masing-masing kalimat uji dievaluasi berdasarkan kaidah tata bahasa baku bahasa Banjar. Persentase hasil evaluasi pohon pola tersebut dijabarkan dalam Gambar 8.



Gambar 8. Persentase Kesesuaian Hasil Aplikasi Pengurai

Pada Gambar 8. dapat dilihat bahwa persentase kesesuaian hasil pengurai keluaran aplikasi pengurai kalimat bahasa Banjar adalah rata-rata sebesar 100%. Hal ini berbanding terbalik dengan Gambar 7. dimana kalimat_1 dan kalimat_3 dinilai oleh responden kurang alami (kurang sesuai dengan kaidah tata bahasa Banjar). Sehingga dapat dikatakan bahwa pohon pola (pengurai) kalimat bahasa Banjar sudah bekerja dengan baik dan sesuai dengan kaidah tata bahasa yang digunakan.

5. Kesimpulan

Pengujian kealamian kalimat menunjukkan hasil bahwa kalimat_1 dan kalimat_3 mempunyai persentase nilai yang sangat kecil dan dibawah 50%. Persentase tersebut terjadi karena sebagian besar responden menilai bahwa kalimat_1 dan kalimat_3 kurang sesuai dengan kaidah tata bahasa Banjar. Persentase penilaian kesesuaian hasil pengurai keluaran aplikasi pengurai kalimat bahasa Banjar rata-rata sebesar 100%. Hal ini berarti, walaupun secara kaidah tata bahasa Banjar kalimat_1 dan kalimat_3 dinilai masih kurang sesuai dan masih menyimpan ambiguitas. Namun, secara grammar (kaidah tata bahasa Banjar) kalimat_1 dan kalimat_3 telah sesuai dengan grammar (kaidah tata bahasa Banjar) yang digunakan.

6. Ucapan Terima Kasih

Terimakasih kepada STKIP PGRI Banjarmasin yang telah mendanai penelitian ini.

7. Referensi

- [1] Sukamto, R. A. & Widyantoro D. H. (2013). *Penguraian Bahasa Indonesia dengan Menggunakan Pengurai Collins*. Thesis Institut Teknologi Bandung 2009.
- [2] Barmawi, Ari Moesriami; Muhammad, Ali. *Paraphrasing Method Based on Contextual Synonym Substitution*. Journal of ICT Research and Applications, [S.l.], v. 13, n. 3, p. 257-282, dec. 2019.
- [3] Kauchak, D., Leroy, G. and Hogue, A. (2017), *Measuring text difficulty using parse-tree frequency*. Journal of the Association for Information Science and Technology, 68: 2088-2100.
- [4] Gusmita, Ria Hari & Ruli Manurung (2008). *Some initial experiments with Indonesian probabilistic parsing*. Second MALINDO Workshop. 1-5.
- [5] Abdul Jabar H. 2017. *Kamus Bahasa Banjar – Indonesia*, Cv. Rahmat Hafiz Al Mubaraq.
- [6] Sipalui. *Koran Daring Si Palui*, 25 agustus 2019 - 24 januari 2020, 16:00 p.m., <https://banjarmasin-tribunnews.com/kolom/palui>, 100 Artikel.
- [7] S. Sari. *Prototipe pemeriksa tata bahasa baku bahasa indonesia: sebuah program yang dikembangkan dengan alat bantu lex dan yacc pada sistem operasi unix*. Undergraduate Thesis, Faculty of Computer Science, University of Indonesia, 1995. Call number: SK-0287.
- [8] I. Hendrawan. *Pengurai sintaks kalimat untuuk bahasa indonesia dengan metode linguistic string analysis*. Under-graduate Thesis, Faculty of Computer Science, University of Indonesia, 1999. Call number: SK-0388.
- [9] S. Salvitri. *Pengembangan lanjut pengurai sintaks bahasa indonesia dengan metode linguistic string analysis*. Under-graduate Thesis, Faculty of Computer Science, University of Indonesia, 1999. Call number: SK-0417
- [10] H. Alwi, S. Dardjowidjojo, H. Lapoliwa, and A. Moeliono. *Tata Bahasa Baku Bahasa Indonesia*. Balai Pustaka, Jakarta, Indonesia, third edition, 1998.
- [11] Joice. *Pengembangan lanjut pengurai struktur kalimat bahasa Indonesia yang menggunakan constraint-based formalism*. Undergraduate Thesis, Faculty of Computer Science, University of Indonesia, 2002. Call number: SK-0487.
- [12] S. Shieber. *An introduction to unification-based approaches to grammar*. CSLI Lecture Notes 4, Center for the Study of Language and Information, Stanford, USA, 1986.
- [13] M. Purgina and M. Mozgovoy, "Visualizing Sentence Parse Trees with WordBricks". *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, Exeter, 2017, pp. 1-4.
- [14] S. McConnel. *PC-PATR Reference Manual*. Summer Institute for Linguistics, October 1995. <http://www.sil.org/pcpatr/manual/pcpatr.html>.
- [15] KBBI. *Kamus Besar Bahasa Indonesia*. <https://kbbi.kemdikbud.go.id>, diakses pada 12 Agustus 2019