

# Prediksi Jeda pada Ucapan Bahasa Melayu Pontianak dengan Menggunakan Metode *Shallow Parsing*

M Iqbal Kamiludin<sup>#1</sup>, Arif Bijaksana Putra Negara<sup>#2</sup>, Novi Safriadi<sup>#3</sup>

<sup>#</sup>Program Studi Teknik Informatika Universitas Tanjungpura

Jl. Prof Dr H. Hadari Nawawi, Kota Pontianak, 78115

<sup>1</sup>ibelkamiludin@gmail.com, <sup>2</sup>arifbpn@gmail.com, <sup>3</sup>bangnops@gmail.com

**Abstrak** - Penelitian ini menjelaskan bagaimana menemukan frasa jeda ucapan kalimat Bahasa Melayu Pontianak menggunakan metode *shallow parsing* yang merupakan salah satu fungsi dari NLTK (*Natural Language Toolkit*). Metode *shallow parsing* ini berfungsi untuk dapat mengetahui kejadian jeda yang terjadi pada kalimat Bahasa Melayu Pontianak melalui potongan frasa. Potongan-potongan frasa yang terbentuk berhubungan dengan frasa jeda dari penutur agar dapat digunakan untuk keperluan generasi sistem *text to speech* (TTS). Metode *shallow parsing* pada NLTK (*Natural Language Toolkit*) membutuhkan aturan *grammars* dan *PoS tagger*. Kumpulan aturan *grammars* dibuat dalam bentuk *regular expressions* untuk mengklasifikasikan kata-kata yang dapat membentuk frasa dengan menggunakan kategori sintaksis kalimat berdasarkan informasi *Part Of Speech* (PoS). Hasil dari penelitian ini adalah kalimat yang telah terpotong-potong membentuk frasa-frasa jeda. Pengujian dilakukan dengan menggunakan *precision*, *recall* serta *F-measure* terhadap kalimat tunggal dan kalimat majemuk. Kalimat tunggal terdiri dari 47 kalimat dan kalimat majemuk terdiri dari 121 kalimat. Berdasarkan pengujian yang telah dilakukan, didapatkan hasil bahwa metode *shallow parsing* yang dibentuk berdasarkan aturan *grammars* sudah dapat membentuk frasa jeda pada kalimat Bahasa Melayu Pontianak. Hasil dari pengujian diperoleh nilai *precision* dan nilai *recall* untuk kalimat tunggal adalah 0.74 dan 0.78 dengan nilai *f-measure* 0.75. Adapun pada kalimat majemuk memperoleh nilai *precision* dan *recall* sebesar 0.57 dan 0.67 dengan nilai *F-measure* 0.61.

**Kata Kunci** : frasa jeda, *Shallow Parsing*, NLTK, *Regular Expressions*, *Part Of Speech*.

## I. PENDAHULUAN

Bahasa memiliki peranan yang sangat penting bagi manusia sebagai media berkomunikasi secara lisan maupun tulisan. Melalui bahasa manusia dapat menyampaikan informasi ataupun menerima informasi. Selain bahasa Indonesia sebagai bahasa nasional, menurut UNESCO Indonesia memiliki lebih dari 40 bahasa daerah, dengan banyaknya bahasa daerah yang ada di Indonesia perlu terus dilakukan upaya pelestarian bahasa agar bahasa-bahasa ini bisa terus dikembangkan. Hasil Sensus Penduduk pada tahun 2010, dari berbagai bahasa daerah yang terdapat di Indonesia, Persentase penggunaan bahasa Melayu yang digunakan oleh masyarakat Kalimantan barat mencapai 20,45% (1.615.978 juta jiwa) dari total penduduk Kalimantan Barat [1].

Bahasa Melayu Pontianak merupakan satu di antara bahasa yang terdapat di Provinsi Kalimantan Barat. Bahasa ini dituturkan oleh orang melayu yang ada di Kota Pontianak. Untuk mengetahui jumlah pasti penutur bahasa ini memang sulit didapat, karena tidak ada data pasti mengenai jumlah penutur bahasa Melayu Pontianak [2]. Penduduk asli kota Pontianak berdomisili di beberapa kecamatan yang letaknya berdekatan dengan Keraton Kadariah yang merupakan pusat kerajaan Melayu di Pontianak. Namun, bahasa Melayu Pontianak tidak hanya ada dan digunakan di daerah-daerah yang dekat dengan bermukimnya penduduk asli kota Pontianak. Bahasa Melayu Pontianak saat ini memang jauh dari kepunahan, hal ini dikarenakan jumlah penutur bahasa ini relatif masih banyak, Bahasa Melayu Pontianak masih menjadi pilihan bahasa disituasi

nonformal. Adanya perkembangan teknologi juga berpengaruh terhadap perkembangan bahasa Melayu Pontianak, Misalnya banyak kosakata yang bukan berasal dari bentuk asli bahasa Melayu Pontianak yang terserap. Bahkan, kosakata ini banyak digunakan dari pada bentuk bahasa asli Melayu Pontianak. Ditengah ancaman bahasa lain, perlunya dilakukan upaya pemertahanan Bahasa agar tidak punah dan ditinggalkan. Salah satu caranya adalah dengan penerapan teknologi terbaru, seiring dengan perkembangan teknologi, setiap orang dituntut untuk dapat memanfaatkan perkembangan tersebut diberbagai aspek kehidupan salah satunya adalah dalam bidang bahasa yang kemudian bisa diimplementasikan untuk Bahasa daerah khususnya Bahasa Melayu Pontianak. Salah satu teknologi yang mencakup di bidang bahasa yaitu Teknologi *Text To Speech* (TTS).

Teknologi *Text To Speech* (TTS) adalah suatu sistem yang dapat melakukan konversi dari teks menjadi ucapan. Didalam sistem TTS membutuhkan informasi berupa nada, tekanan (keras-lemah), durasi, jeda (kesenyapan), dan intonasi untuk mendapatkan ucapan yang lebih alami. Penentuan jeda dalam kalimat sangatlah penting karena jeda yang terjadi dapat membentuk frase prosodi dan selain itu jeda juga dapat memperjelas informasi dari makna atau maksud pada suatu teks kalimat dapat tersampaikan dengan benar. Oleh beberapa penelitian dalam menentukan sintaksis jeda ucapan memerlukan proses yang sederhana dapat menggunakan metode *Shallow parsing*. *Shallow parsing* telah menjadi alternatif yang menarik untuk full parsing karena banyak aplikasi NLP yang tidak memerlukan banyak informasi yang diberikan *full parsing*. Sebagai tanggapan atas masalah ini dalam *full parsing*, Abney [3] memperkenalkan konsep *parsing* dengan potongan/*chunk* yang memberikan analisis sintaksis dangkal dari sebuah kalimat yang relatif tidak ambigu. *Shallow parsing* juga disebut sebagai *Chunking* yang merupakan metode pemenggalan/pemotongan kalimat menjadi frasa teks dengan aturan-aturan tertentu berdasarkan kelas kata *PoS tagger*.

Penelitian tentang *Shallow parsing* telah banyak dilakukan untuk Bahasa asing seperti Bahasa China, Arabic, Malayalam, Turki, Inggris dan Bahasa lainnya. Namun untuk penelitian *Shallow parsing* Bahasa Indonesia masih jarang dikembangkan adapun penelitian yang telah dilakukan yaitu *Syntactic Phrase Chunking* untuk Bahasa Indonesia oleh Arry Akhmad, dkk [4], dimana dipenelitian tersebut membangun sintaksis frasa Bahasa Indonesia untuk keperluan generasi informasi prosodi dalam sistem TTS. Kategori sintaksis terdiri dari lima frase karakteristik pemenggalan POS (*Part of Speech*), yaitu *noun phrases*, *verb phrases*, *connection phrases*, *numeric phrases* dan *questioning phrases*. Dengan melakukan pengujian pada 80 kalimat dan diperoleh dinilai presisi dan recall adalah 80,57 % dan 84,72 % dengan nilai F adalah 82.60 %.

Berdasarkan penjelasan yang telah dipaparkan, maka dilakukan penelitian untuk menghasilkan *rule grammar* untuk menentukan jeda ucapan pada kalimat dengan menggunakan metode *shallow parsing*, sebagai fungsi penentu jeda pada ucapan Bahasa Melayu Pontianak dan sebagai informasi yang dapat digunakan oleh sistem TTS (*text to speech*) untuk menghasilkan ucapan yang alami.

II URAIAN PENELITIAN

A. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah sebuah bagian dari penelitian dan aplikasi yang mengkaji bagaimana komputer dapat digunakan untuk mengerti dan memanipulasi bahasa alami yang berupa teks atau ucapan untuk hal-hal yang berguna. Manipulasi teks telah dikenal sebagai sebuah bidang penelitian yang penting dalam NLP. Sebuah sistem NLP yang mengolah teks dimulai dengan analisis morfologi. Teks dikonversi, dalam kueri atau dokumen, untuk mendapatkan varian morfologi kata-kata yang terlibat. Pengolahan leksikal dan sintaktis melibatkan pemanfaatan kamus untuk menentukan karakteristik dari kata-kata, pengenalan part-of-speech, menentukan kata-kata dan frasa, serta untuk penguraian kalimat [5].

B. Part of Speech Tagging

Part of Speech Tagging (PoS Tagging) adalah suatu proses yang memberikan label kelas kata secara otomatis pada suatu kata dalam kalimat [6]. Kelas kata terdiri dari kata sifat, kata benda, kata kerja, kata keterangan, kata depan, kata kepunyaan, kata penghubung, dll. Suatu kata bisa diklasifikasikan ke dalam kelas kata yang berbeda. Sebagai contoh: kata “dogs” dalam Bahasa Inggris biasanya diklasifikasikan menjadi plural noun (kata benda jamak). Akan tetapi dalam konteks tertentu, kata ini bisa diklasifikasikan menjadi verb (kata kerja) misalnya pada kalimat “Sailor dogs the hatch”. Solusi untuk mengatasi masalah ambiguitas ini adalah dengan melihat kelas kata dari kata sebelumnya. [6].

C. Shallow parsing

Shallow parsing merupakan metode yang mudah, cepat dan andal tidak memberikan analisis sintaksis penuh namun terbatas pada parsing konstituen sintaksis yang lebih kecil. ketika kita membaca sebuah kalimat, kita membacanya sepotong demi sepotong yang membentuk kesatuan kelompok kata yang disebut frase kata [3].

D. Regular Expression

Regular Expression adalah konstruksi bahasa untuk mencocokkan teks berdasarkan pola tertentu, terutama untuk kasus-kasus kompleks. Regular Expression adalah bahasa formal untuk menentukan string. Contoh misalkan mencari teks berawalan karakter tertentu, memiliki jumlah perulangan dari suatu teks, dan lain sebagainya. Regular Expression (regex) merupakan salah satu alat yang paling berguna dalam ilmu komputer. NLP sebagai bidang ilmu komputer, sangat diuntungkan dari regex: digunakan dalam fonologi, morfologi, text analysis, information extraction, dan speech recognition. Selain itu, regex juga sangat manjur (powerful) terutama untuk proses penguraian kata (text parsing). Berikut karakter Regular Expressions yang digunakan dalam shallow parsing pada tabel 1.

Tabel 1 Makna karakter Regular Expressions

karakter	Makna karakter Regular Expressions
<>	Penentuan tag part-of-speech
?	Nol atau salah satu item sebelumnya
*	Nol atau lebih dari item sebelumnya.
+	Satu atau lebih dari item sebelumnya
	Mencocokkan satu item dengan yang lainnya,

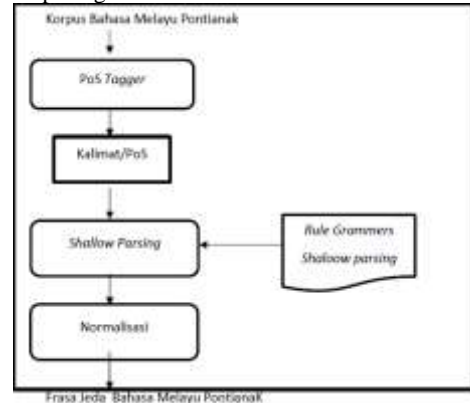
III. HASIL DAN DISKUSI

A. Data Penelitian

Data penelitian berupa korpus kalimat Bahasa Melayu Pontianak yang didapat dari data penelitian [7] sebanyak 3000 korpus kalimat dan korpus yang digunakan sebanyak 168 kalimat untuk proses shallow parsing dan 168 kalimat yang sudah ditandai jedanya oleh penutur Bahasa Melayu Pontianak.

B. Arsitektur Proses Shallow parsing Frasa Jeda.

arsitektur sistem terdiri dari beberapa proses yaitu tahap awal proses persiapan data berupa kalimat Bahasa Melayu Pontianak, PoS Tagger, Proses Shallow parsing dan normalisasi hasil parsing. Berikut adalah gambaran umum proses dari pencarian frasa jeda dapat dilihat pada gambar 1



Gambar 1 Arsitektur Proses Shallow parsing Bahasa Melayu Pontianak

C. Proses PoS Tagger

Pemberian kelas kata atau PoS Tagger menggunakan 35 tag PoS dari [8] dan dilakukan secara manual sebanyak 168 kalimat Bahasa Melayu Pontianak. Pengelompokan kelas kata dalam Bahasa Melayu Pontianak dengan 35 tag PoS disesuaikan dengan kata yang terdapat dikategori Tag PoS 35 tag Bahasa Indonesia oleh Alfan Farizki Wicaksono [8]. Tabel tag PoS Bahasa Melayu Pontianak seperti pada Tabel 2 berikut:

Tabel 2 PoS Tag Bahasa Melayu Pontianak

No	PoS	Deskripsi	Contoh Kata
1	OP	Parenthesis	(( {
2	CP	Close Parenthesis	)}])
3	GM	Slash	/
4	;	Semicolon	;
5	:	Colon	:
6	“	Quotation	“ ’
7	.	Sentence Terminator	. ! ?
8	,	Comma	,
9	-	Dash	-
10	...	Ellipsis	...
11	JJ	Adjective	Cantek, sepok, bagos, abes, payah,
12	RB	Adverb	Teros, agek, Sekarang, jak, sampat,
13	NN	Common Noun	Kawan, orang, kampung, bangku,
14	NNP	Proper Noun	Bekasi, Indonesia, Pontianak
15	NNG	Genitive Noun	Name-e, buku-e
16	VBI	Intransitive Verb	Pegi, bejalan, naek, kerje
17	VBT	Transitive Verb	Mbayar, betageh mbeli,
18	IN	Preposition	Tang, untok, dari, dalam, dengan
19	MD	Modal	Udah, bise, boleh, akan
20	CC	Coor-Conjunction	Karne, lalu, Tapi, ngan, dan
21	SC	Subor-Conjunction	Yang, mang, kalok, sambel, bahwe
22	DT	Determiner	Tu, ni, itu,
23	UH	Interjection	Baleee, Ooo, bah, eh, oi, Haa
24	CDO	Ordinal Numerals	pertame, sutik, kedua,
25	CDC	Collective Numerals	Beduak, berempat, berator-ratos
26	CDP	Primary Numerals	Satu, duak, 1, 2, 3, puloh, ratus,
27	CDI	Irregular Numerals	Semue-e, seluroh, beberape,
28	PRP	Personal Pronouns	Die, kau, Kitak, Mereke, kamek
29	WP	WH-Pronouns	Ape, apelah, Ngape, cemane
30	PRN	Number Pronouns	Satu-satunye, pertame-tame,
31	PRL	Locative Pronouns	Sinek, situk, sineklah, sanak
32	NEG	Negation	Ndak, tak, tadak, bukan
33	SYM	Symbols	@#\$%^&
34	RP	Particles	Pon, lah, kok
35	FW	Foreign	Words

D. Proses *Shallow parsing*

Setelah kata setiap kalimat berhasil di-tag dengan bentuk "kata/PoS. Selanjutnya file PoS *tagger* ini akan diolah sebagai inputan dalam proses *Shallow parsing* menggunakan *tools NLTK* pada *python*. *Output* dari proses *chunking* ini berbentuk "*ChunkTag* kata/PoS". Proses *Shallow parsing* untuk mengelompokkan kalimat input berupa teks untuk membentuk frasa-frasa sesuai dengan aturan yang telah di tentukan. Proses *Shallow parsing* pada Penelitian ini menggunakan fungsi *NLTK RegexpParser* untuk *Shallow parsing* kalimat teks, dengan menggunakan aturan tata-bahasa (*grammar*) *NLTK Tag Patterns* atau yang dikenal dengan *Regular Expressions*. Dalam proses penentuan frasa jeda *shallow parsing* membutuhkan *rule grammars* sebagai pembentuk potongan frasa dalam kalimat yang diolah. *Rule* yang digunakan dalam penelitian ini adalah *rule* dari Dr.Arif bijaksana Putra Negara, S.T, M.T.[4][9] yang berjumlah 9 *rule* yang sebelumnya digunakan untuk proses penentuan frasa *chunking* Bahasa Indonesia. *Rule* tersebut kemudian dimodifikasi oleh peneliti yang sudah disesuaikan dengan kebutuhan penelitian frasa jeda *shallow parsing* Bahasa Melayu Pontianak. *Rule* awal yang dimodifikasi tetap berjumlah 9 *rule grammars* namun ada beberapa *rule* yang diubah untuk mendapatkan potongan frasa melayu Pontianak yang sesuai. Adapun penjabaran *rule grammars* sebagai berikut:

```
grammar = r"""
    TP : { <WP>+ <NN|JJ|NNP|NNG|PRP|DT|RB|MD|NEG|RP>* }

    NP : { <NNP>+ <CC>? | <FW>+ | <NN|JJ>+ <CC|RB>? }

    KP : { <SC|CC|NEG>+ <IN>? <MD|RB>* | <MD|RB>+
    <SC|NEG|CC>* <IN>* | <IN>+ <SC|CC|NEG>* <MD|RB>* }

    BP : { <CDO|CDP|CDI|CDC>+ | <SYM> }

    VP : { <VBI|VBT>+ <PRP|DT>? | <PRP>? <VBI|VBT>+ }

    NP : { <DT>? <NP|BP|NN|PRN|PRP|PRL|NNG|UH|RP>+ <JJ>*
    <NP|BP|NN|PRN|PRP|PRL|NNG|FW>* <DT|UH>* <RP>? | <DT>?
    <JJ>+ <NP|BP|NN|PRN|PRP|PRL|NNG|FW>* <JJ>? <DT|UH>? <RP>?
    | <DT> }

    NP : { <KP> <NP> }
    VP : { <KP> <VP> }
    VP : { <VP> <KP> }
    """
```

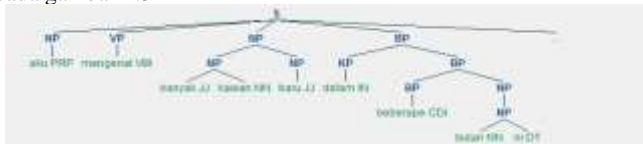
Gambar 2 Aturan Rule Grammer shallow Parsing.

Berikut contoh teks kalimat inputan : Aku mengenal banyak kawan baru dalam beberapa bulan ni .

Kalimat inputan tersebut di tag setiap kata, kemudian fungsi *RegexpParser* pada *NLTK Shallow parsing* akan memproses kalimat dengan tag PoS ke dalam bentuk potongan frasa.

- 0 (NP aku/PRP)
- 1 (VP mengenal/VBI)
- 2 (NP (NP banyak/JJ kawan/NN) (NP baru/JJ))
- 3 (BP (KP dalam/IN) (BP (BP beberapa/CDI) (NP (NP bulan/NN ni/DT))))

Pohon parsing hasil *shallow parsing* tersebut dapat dilihat seperti pada gambar 4.3



Gambar 3 Contoh *shallow parsing* kalimat

*Output* frasa jeda *shallow parsing* yang dihasilkan pada tabel 4.3 terdiri dari empat potongan chunk. **Aku** sebagai subjek kalimat kemungkinan untuk terjadinya jeda bimbang, kemungkinan juga untuk tidak terjadinya jeda menjadi **Aku Mengenal**, yang selanjutnya terjadi jeda yang dilanjutkan dengan frasa **Banyak Kawan Baru**, dan terjadi jeda lagi dilanjutkan dengan frasa **Dalam Beberapa Bulan ni**.

E. Pengujian Hasil Frasa Jeda

Pengujian dilakukan terhadap kalimat tunggal dan kalimat majemuk dengan jumlah 168 kalimat dengan frasa jeda yang dibandingkan dengan frasa jeda dari penutur. Pengujian akurasi frasa jeda pada struktur kalimat tunggal dilakukan agar didapatkan nilai acuan pada *rule grammars* dalam menghasilkan potongan frasa jeda. Pengujian pada struktur kalimat dilihat perbandingan terhadap pola kalimat subjek, predikat, objek, keterangan dan pelengkap pada kalimat uji dengan potongan frasa dari hasil *shallow parsing* yang dibangun dari *rule grammars* yang berdasarkan dari struktur kalimat. Perngujian struktur kalimat hasil *shallow parsing* terdiri dari 47 kalimat tunggal dengan lima jenis pola kalimat yaitu S-P, S-P-O, S-P-O-K, S-P-K, dan S-P-O-Pel. Setelah itu dilakukan pengujian untuk prediksi jeda dengan cara membandingkan frasa jeda dari hasil *shallow parsing* dengan frasa jeda hasil dari penandaan penutur. Pengujian dilakukan terhadap dua jenis kalimat yaitu kalimat tunggal sebanyak 47 kalimat dan kalimat majemuk sebanyak 121 kalimat. Dari hasil perbandinagn tersebut dapat dilakukan perhitungan nilai akurasi *Recall* dan *Precision* serta *f-Measure*.

1. Pengujian Akurasi kecocokan frasa jeda dalam satu kalimat penuh.

Pengujian akurasi dari frasa jeda dihitung dengan membandingkan dengan frasa jeda dari penutur yang sesuai. Pengujian akurasi ini adalah rasio antara frasa jeda shallow parsing yang sesuai dengan frasa jeda dari penutur dan total jumlah frasa jeda yang benar.

rumus untuk uji akurasi hasil frasa jeda dalam kesatuan kalimat adalah sebagai berikut:

$$\text{Akurasi frasa jeda} = \frac{\text{jumlah frasa jeda yang sesuai}}{\text{Total frasa jeda yang diuji}} \times 100\%$$

Adapun hasil dari pengujian akurasi frasa jeda gabungan antara kalimat tunggal dan kalimat majemuk pada satu kalimat penuh dapat dilihat pada tabel 3 berikut:

Tabel 3 Hasil Nilai Pengujian Akurasi Kalimat Tunggal dan kalimat Majemuk

Jenis kalimat	Jumlah	Akurasi (%)
Kalimat Tunggal	47	63.8
Kalimat Majemuk	121	31.4
Total	168	
Nilai Akurasi		40.4

Pada tabel 3 merupakan hasil akurasi frasa jeda dari kalimat tunggal dan kalimat majemuk untuk kecocokan dalam potongan frasa satu kalimat penuh. Hasil akurasi kalimat tunggal lebih tinggi dengan nilai 63.8% dari hasil akurasi kalimat majemuk yang memperoleh nilai 31.4%. Hal tersebut sangat dipengaruhi oleh batas jeda yanag ditandai oleh penutur. Kalimat tunggal lebih sederhana dan terstruktur kalimatnya sehingga potongan jeda yang ditandai oleh penutur lebih mirip dengan struktur kalimat tunggal. Sedangkan pada kalimat majemuk dalam satu kalimat penuh kebanyakan frasa jeda dari penutur tidak sepenuhnya sama dengan potongan *shallow parsing*. Adapun nilai akurasi kecocokan dari total kalimat tunggal dan kalimat majemuk yang berjumlah 168 kalimat menghasilkan nilai akurasi sebesar 40.4%.

2. Pengujian *Recall*, *Precision* dan *F-measure*

*precision* adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan oleh sistem. *Recall* adalah rasio jumlah dokumen relevan yang ditemukan kembali dengan total jumlah dokumen dalam kumpulan dokumen yang dianggap relevan. Persamaan untuk *precision* dan *recall* ditunjukkan pada berikut.

Tabel 4 Variabel Untuk Perhitungan *Precision* dan *Recall*

	Relevan	Tak Relevan
Ditemukan	True positives (tp)	False positives (fp)
Tidak ditemukan	False negatives (fn)	True negatives (tn)

$$\text{Precision} = \frac{\text{item yang relevan yang ditemukan}}{\text{item ditemukan}} = P \frac{TP}{TP+FP}$$

$$Recall = \frac{\text{item yang relevan yang ditemukan}}{\text{semua item yang relevan}} = P \frac{TP}{TP + FN}$$

*F-measure* (nilai F atau *F1-Score* atau *score*) merupakan sebuah nilai dari keakuratan sebuah tes. Nilai F menggunakan *precision* & *recall* dari tes untuk menghitung nilainya, dengan *precision* yang menyatakan jumlah hasil benar dibandingkan dengan jumlah hasil yang ditembakkan dan *recall* yang menyatakan jumlah hasil benar dibandingkan dengan jumlah hasil yang harus ditembakkan. Nilai F dapat diartikan sebagai sebuah hasil rata-rata dari *precision* & *recall*, dimana sebuah nilai F mendapat hasil 1 sebagai hasil terbaik dan 0 sebagai nilai terburuk. Rumus nilai F adalah sebagai berikut:

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

a. Pengujian *Recall* dan *Precision* Pada Frasa Jeda Kalimat Tunggal Berdasarkan Struktur Kalimat.

Pada pengujian *Recall* dan *Precision* frasa jeda kalimat tunggal berdasarkan struktur kalimat ini untuk mengetahui nilai akurasi keberhasilan *rule grammars* yang dibuat dalam menghasilkan potongan frasa jeda. Dan sebagai nilai acuan untuk pengujian frasa jeda terhadap frasa jeda yang didapat dari penutur apakah bisa menghasilkan frasa jeda yang baik sesuai penutur Bahasa Melayu Pontianak. Hasil perhitungan nilai akurasi *Recall* dan *Precision* frasa jeda berdasarkan struktur kalimat S-P-O-K-Pel pada tabel 5 berikut.

Tabel 5 Hasil Pengujian *Recall* dan *Precision* pada Struktur Kalimat

Tipe kalimat tunggal	Jumlah Kalimat	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
S-P	10	0.88	0.80	0.83
S-P-O	8	0.90	0.95	0.92
S-P-O-K	11	0.94	0.94	0.94
S-P-k	11	0.81	0.84	0.82
S-P-O-Pel	7	0.91	0.88	0.89
Jumlah	47			
Nilai Akurasi		0.89	0.88	0.88

Dari hasil pengujian *Recall* dan *Precision* pada tabel 5 didapat bahwa hasil potongan frasa dari *rule grammars* yang dibuat dapat menghasilkan frasa yang baik dan sesuai dengan struktur kalimat. Hal itu dapat dilihat dari nilai akurasi *Recall* dan *Precision* serta *F-Measure*, dimana nilai yang dihasilkan sebesar 0.89 untuk *Precision* sehingga sebagian besar frasa yang relevan dari struktur kalimat dapat dibentuk dengan baik oleh *shallow parsing*, dan nilai *Recall* sebesar 0.88 yang menunjukkan hasil potongan frasa *shallow parsing* yang terbentuk menghasilkan frasa yang relevan berdasarkan struktur kalimat. Serta memperoleh nilai akurasi *F-Measure* sebesar 0.88 sehingga keberhasilan *shallow parsing* dalam menghasilkan frasa berdasarkan struktur kalimat sudah baik.

b. Pengujian *Recall* dan *Precision* Pada Frasa Jeda Kalimat Tunggal

Pada pengujian *Recall* dan *Precision* frasa jeda kalimat tunggal dibandingkan untuk mendapatkan kecocokan antara potongan frasa jeda hasil *shallow parsing* dan potongan frasa jeda oleh penutur. Kalimat yang digunakan untuk pengujian kalimat tunggal berjumlah 47 kalimat Hasil perhitungan *recall* dan *precision* untuk kalimat tunggal dapat dilihat nilai akurasi *Recall* dan *Precision* frasa jeda berdasarkan perbandingan terhadap frasa jeda penutur pada tabel berikut:

Tabel 6 Hasil Nilai Pengujian *Recall*, *Precision* dan *F-measure* Kalimat Tunggal

Tipe kalimat tunggal	Jumlah Kalimat	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
S-P	10	0.77	0.70	0.73
S-P-O	8	0.63	0.77	0.69
S-P-O-K	11	0.65	0.75	0.69
S-P-k	11	0.75	0.78	0.76
S-P-O-Pel	7	0.91	0.91	0.91
Jumlah	47			
Nilai Akurasi		0.74	0.78	0.75

Hasil perhitungan *recall* dan *precision* kita dapatkan nilai *precision* sebesar 0.74, nilai *recall* sebesar 0.78 dan nilai *F-Measure* sebesar 0.75. Hal ini menunjukkan bahwa frasa jeda dengan metode *shallow parsing* cukup baik menghasilkan potongan jeda pada kalimat tunggal yang sesuai dengan penutur Bahasa Melayu Pontianak. Tetapi nilai akurasi dari frasa jeda dari penutur mengalami penurunan dari nilai akurasi frasa jeda berdasarkan pola kalimat terstruktur hal ini disebabkan oleh penutur memberi jeda tidak berdasarkan struktur kalimat tetapi berdasarkan pemberhentian ucapan dari penutur tersebut.

c. Pengujian *Recall* dan *Precision* Pada Frasa Jeda Kalimat Majemuk dan gabungan kalimat

Hasil perhitungan *recall*, *precision* dan *F-measure* untuk kalimat tunggal dan kalimat majemuk dapat dilihat pada tabel berikut:

Tabel 7 Hasil Nilai Pengujian *Recall*, *Precision* dan *F-measure* Kalimat Tunggal dan kalimat Majemuk

Tipe kalimat	Jumlah Kalimat	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
Kalimat Tunggal	47	0.74	0.78	0.75
Kalimat Majemuk	121	0.57	0.67	0.61
Jumlah	168			
Nilai Akurasi		0.60	0.70	0.64

Pada tabel 7 merupakan hasil perhitungan *recall*, *precision* dan *F-measure* frasa jeda dari kalimat tunggal dan kalimat majemuk untuk kecocokan dalam potongan frasa satu kalimat penuh. Hasil perhitungan *precision*, *recall* dan *F-measure* kalimat tunggal lebih tinggi dengan nilai 0.74, 0.78, dan 0.75 dari hasil akurasi kalimat majemuk yang hasil *precision*, *recall* dan *F-measure*-nya memperoleh nilai 0.57, 0.67, dan 0.61. Hal tersebut sangat dipengaruhi oleh batas jeda yang ditandai oleh penutur. Kalimat tunggal lebih sederhana dan terstruktur kalimatnya sehingga potongan jeda yang ditandai oleh penutur lebih mirip dengan struktur kalimat tunggal. Sedangkan pada kalimat majemuk dalam satu kalimat penuh kebanyakan frasa jeda dari penutur tidak sepenuhnya sama dengan potongan *shallow parsing*. Sedangkan untuk gabungan kalimat tunggal dan kalimat majemuk dengan total kalimat uji 168 kalimat memperoleh *precision*, *recall* dan *F-measure* sebesar 0.60, 0.70, dan 0.64. Akurasi pada *precision* menunjukan nilai yang paling kecil karena banyaknya potongan *shallow parsing* ditemukan tidak sesuai dengan potongan jeda oleh penutur.

F. Analisis Hasil Pengujian

Dari pengujian dengan menghitung akurasi kecocokan frasa jeda satu kalimat penuh dan perhitungan *precision* & *recall* serta *F-measure* yang telah dilakukan, analisis yang dapat disimpulkan sebagai berikut:

1. Pengujian pada prediksi frasa jeda dengan menggunakan *precision* & *recall* serta *F-measure* berhasil dilakukan, dimana hasil prediksi frasa jeda yang relevan dapat dibandingkan antara frasa jeda yang ditandai oleh penutur dan frasa jeda yang dihasilkan oleh *shallow parsing*.
2. Prediksi jeda dalam kalimat dapat dilakukan dengan cara memotong kalimat menjadi frasa-frasa dengan menggunakan metode *shallow parsing* yang sesuai dengan aturan grammatikal yang dibuat, meski ada beberapa frasa jeda dari hasil *shallow parsing* yang tidak sesuai dengan jeda yang diberi oleh penutur.
3. Presiksi jeda dengan menggunakan *shallow parsing* belum sepenuhnya relevan. Hal ini dapat dilihat pada nilai kecocokan frasa jeda dalam satu kalimat penuh dimana pada kalimat tunggal dari 47 kalimat yang diuji menghasilkan nilai 63.8% frasa jeda yang relevan, sedangkan pada kalimat majemuk dengan 121 kalimat uji menghasilkan nilai kecocokan frasa jeda 31.4% yang relatif lebih rendah karena jeda yang ditandai oleh penutur mengacu pada jeda berdasarkan frasa ucapan sipenutur sehingga ada perbedaan dengan frasa jeda yang

dihasilkan berdasarkan pola struktur kalimat. Perbedaan tersebut menyebabkan dalam satu kalimat terdapat potongan yang tidak relevan yang dapat mempengaruhi kecocokan frasa jeda yang dihasilkan dalam satu kalimat. Begitu juga hasil nilai *precision* dan *recall* frasa jeda yang dihasilkan oleh *shallow parsing*, pada kalimat tunggal didapatkan nilai *precision* 0.74 dan nilai *recall* 0.78 yang menunjukkan kecocokan frasa jeda hasil *shallow parsing* dan frasa jeda dari penutur dengan nilai yang baik namun mengalami penurunan dari hasil uji *precision* dan *recall* berdasarkan frasa yang terstruktur dengan pola S-P-O-K-Pel, dengan nilai uji *precision* 0.89 dan *recall* 0.88. Sedangkan pengujian pada kalimat majemuk didapatkan nilai *precision* 0.57 dan nilai *recall* 0.67 yang relatif lebih kecil dari hasil kalimat tunggal.

4. Presiksi jeda dengan menggunakan *shallow parsing* belum sepenuhnya akurat, karena nilai *F-measure* yang didapat relatif kecil. Hal ini dapat dilihat pada tabel 4.16, dimana nilai *F-measure* kalimat tunggal mendapatkan nilai 0.75 tetapi untuk nilai *F-measure* kalimat majemuk memperoleh nilai 0.61 yang relatif lebih kecil. Sehingga diperoleh nilai *F-measure* untuk gabungan kalimat tunggal dan kalimat majemuk yang berjumlah 168 kalimat dengan nilai *F-measure* 0.64. Sehingga keakuratan frasa jeda yang dihasilkan oleh *shallow parsing* belum cukup baik dalam menghasilkan potongan-potongan jeda yang sesuai dengan jeda yang didapatkan dari penutur yang memang menggunakan Bahasa Melayu Pontianak dalam percakapan mereka sehari-hari.
5. Nilai akurasi kecocokan dan nilai *precision & recall* serta *F-measure* kalimat majemuk yang kecil dikarenakan *rule grammars shallow parsing* dirancang berdasarkan aturan struktur kalimat S-P-O-K-dan Pel, sehingga potongan-potongan frasa jeda kalimat majemuk mengacu pada struktur kalimat tersebut, sedangkan frasa yang ditandai oleh penutur tidak memandang struktur kalimat dan mereka menandai jeda ucapan berdasarkan keinginan mereka untuk berhenti pada satu kata dalam kalimat, yang memang memungkinkan mereka untuk berhenti jeda jika mengucapkan kalimat tersebut.

#### IV. KESIMPULAN/RINGKASAN

Berdasarkan hasil analisis dan pengujian terhadap prediksi frasa jeda ucapan kalimat Bahasa Melayu Pontianak dengan menggunakan potongan-potongan *shallow parsing*, maka dapat ditarik kesimpulan sebagai berikut:

1. Metode *shallow parsing* dapat menghasilkan frasa jeda pada kalimat bahasa Melayu Pontianak sesuai dengan *Rule grammars* yang disiapkan.
2. *Rule grammars shallow parsing* yang dibentuk belum sepenuhnya relevan menghasilkan frasa jeda yang sesuai terhadap penandaan jeda dari penutur. Data frasa jeda hasil pengujian menghasilkan nilai *recall* 0.78 untuk kalimat tunggal dan 0.67 untuk kalimat majemuk yang menunjukkan bahwa sebagian besar frasa jeda yang relevan tidak terbentuk oleh *shallow parsing*. Nilai *precision* 0.74 untuk kalimat tunggal dan 0.57 untuk kalimat majemuk yang menunjukkan nilai yang cukup baik namun masih banyak terdapat hasil potongan frasa jeda yang lain selain hasil potongan frasa jeda yang relevan yang ikut terbentuk oleh *shallow parsing*.
3. Potongan frasa jeda *shallow parsing* belum mampu menghasilkan nilai akurasi kecocokan pada frasa jeda penutur dengan nilai yang baik. Dari 168 kalimat yang merupakan gabungan dari kalimat tunggal dan kalimat majemuk menghasilkan nilai kecocokan dengan jeda penutur sebesar 40.4% atau 68 kalimat.
4. Potongan frasa jeda *shallow parsing* tidak sepenuhnya mampu melakukan proses pemrediksian jeda pada kalimat Bahasa Melayu Pontianak secara akurat karena hasil perhitungan *F-measure* untuk kalimat tunggal 0.75 sedangkan *F-measure* kalimat majemuk sebesar 0.61, sehingga untuk nilai gabungan dari kedua jenis kalimat tersebut dihasilkan nilai *F-measure* sebesar 0.64.

5. Hasil frasa jeda dari *shallow parsing* ini dapat membantu dalam pengembangan sistem *text to speech*. Dimana potongan-potongan frasa jeda tersebut dapat digunakan sebagai informasi untuk memproses prosodi generasi pada sistem TTS Bahasa Melayu Pontianak.

#### DAFTAR PUSTAKA

- [1] Na'im, Akhsan., Syaputra, Hendri. 2010. *Hasil Sensus Penduduk 2010: Kewarganegaraan, Suku Bangsa, Agama, dan Bahasa Sehari-Hari Penduduk Indonesia*. Jakarta: Badan Pusat Statistik..
- [2] Novianti,Evi. 2011. *Menilik Nasib Bahasa melayu Pontianak*. Balai Bahasa Propinsi Kalimantan Barat.
- [3] Abney, S. P., 1991.*Parsing By Chunk*. In Robert Berwick, Steven Abney, and Carol Tenny (eds), *Principle-Based Parsing*. Kluwer Academic Publishers.
- [4] Arman A.A, Bijaksana Putra Negara, Arif, Purwariantia A, Kuspriyanto. 2013. *Syntactic Phrase Chunking for Indonesian Language: The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)*.
- [5] Chowdhury, G. 2003 *Natural language processing. Annual Review of Information Science and Technology*.
- [6] Jurafsky, Daniel, dan James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Edisi kedua. Prentice-Hall.
- [7] Indrayana, Danny. 2016. *Meningkatkan Akurasi Pada Mesin Penerjemah Bahasa Indonesia Ke Bahasa Melayu Pontianak Dengan Part Of Speech*. Pontianak: Skripsi Teknik Informatika Universitas Tanjungpura.
- [8] Wicaksono, A F dan Purwarianti, A.( 2010). *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*. Proceeding of the Fourth International MALINDO Workshop (MALINDO2010).jakarta,Indonesia.
- [9] Bijaksana Putra Negara, Arif dan Arman A.A. 2013. *Syntax Based Prosody Modeling Using HMM for Bahasa Indonesia*. Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International Conference.