The linguistic anatomy of individual differences in Japanese monologues: Focusing on particles and interjections

Dr. Shunichi Ishihara
The Australian National University, Canberra
shunichi.ishihara@anu.edu.au

Abstract. This is a linguistic study on idiosyncrasy manifested through language use in Japanese monologues. For this purpose, we use speaker classification techniques as analytical tools. Focusing on Japanese particles, the subcategories of these particles, and interjections, we aim to find out to what extent Japanese speakers are idiosyncratic in selecting certain words above others in monologues. We are interested in how differently or similarly the individualising information of speakers is manifested between the subcategories of these particles, and also between particles and interjections. The genres of the monologues in this study vary from conference presentations on various topics covering humanities, social sciences, natural sciences and engineering to mock public speeches on a variety of general topics, such as "most pleasant memory," "about your community," etc. We demonstrate in this study that Japanese particles and interjections carry different degrees of individualising information. We also discuss what contributes to the identified differences between them.

Keywords. individual differences, particles, interjections, Japanese, speaker classification

Proceedings of the 42nd Australian Linguistic Society Conference – 2011

M Ponsonnet, L Dao & M Bowler (eds)

1. Introduction[1]

We intuitively know that different people talk and write differently, even when they try to convey the same message. We also know that people tend to use their individually selected preferred words despite the fact that, in principle, they can use any word at any time from the vocabulary built up over the course of their lives—given that their word choice falls within the constraints arising from their topic, the register, the audience, etc. Every speaker of a given language has their own distinctive and individual version of language, which is often referred to as their *idiolect* (Halliday *et al.* 1964, Coulthard & Johnson 2007). This idiolect manifests itself in various aspects of communication, such as the choice of words and expressions, grammar, morphology, semantics and discourse structure. The focus of the current study is idiosyncratic word choice, by means of particle and interjection usage in spoken Japanese monologues.

In the domain of written language, in contrast to spoken language, linguistic idiosyncrasy has been mainly studied as authorship attribution. A large number of studies have been conducted on this topic (Burrows 1987, Baayen *et al.* 1996, Fung 2003). Authorship attribution concerns the task of identifying the author of a text. Studies in authorship attribution first emerged as stylometric studies[2], with many of the pioneering studies based on literary texts (Mendenhall 1887, Thisted & Efron 1987, Mosteller & Wallace 1984, Holmes 1992).

Various techniques have been proposed to model authorship attribution, such as those based on syntactic or grammatical features (Baayen *et al.* 1996, Stamatatos *et al.* 2001) and on probabilistic language models (Keselj *et al.* 2003, Peng *et al.* 2003). Many of them are based on the unique lexical usage of authors (Holmes *et al.* 2001, Juola & Baayen 2005), assuming that the selection of words is unique to each author and that their preferred selection is consistent over time (Mendenhall 1887, Holmes 1992). Indeed, it has been demonstrated that word category usage is very stable across time and writing topics (Pennebaker & King 1999).

[1] This study was financially supported by the ANU Research School of Asia and the Pacific. The author thanks anonymous reviewers for their valuable comments.

[2] Stylometry is the science of measuring literary style.

In particular, function words are often used as an individualising feature to quantify the unique lexical usage of individual authors, which has been attested in many previous studies (Burrows 1987, Holmes 1992, Holmes *et al.* 2001, Binongo 2003, García & Martín 2007). Function words are closed class words, therefore having little contextual meaning. As such, the selection of function words is considered to be less influenced by the content of a text than by that of lexical words. Mosteller & Wallace (1964) were the first to demonstrate the effectiveness of frequently occurring function words (e.g. *the*, *if*, *to*) in addressing the issues of the so-called *Federalist Papers*. Burrows (1987) also successfully used 30-50 function words for his authorship analysis work. Previous studies have inferred that the use of function words has large variation between authors, but little variation within a single author, which is ideal for authorship classification (Baayen *et al.* 1996, Burrows 1987, Mosteller & Wallace 1964).

In contrast to written language, studies on the idiosyncratic choice of words in spoken language are relatively few. However, the concept of idiolect in the selection of function words has been incorporated into automatic speaker recognition systems in order to enhance their performance (Doddington 2001, Weber *et al.* 2002). In addition to function words, fillers (such as English *um*, *you know*, *like*), which are unique to spoken language, have also been reported to carry idiosyncratic speaker information. Weber *et al.* (2002) reported that the inclusion of fillers, as well as functions words, as a speaker individualising feature in automatic speaker recognition systems improves their performance. In Japanese, Ishihara (2010) and Ishihara & Kinoshita (2010) demonstrated that Japanese fillers bear speaker idiosyncratic information to the extent that the accuracy of speaker classification based solely on fillers can be as high as 85% for male speakers. For these studies, speech samples collected from Japanese monologues across various genres were used.

Previous studies on idiosyncratic word choice have centred on English as the target language, and, as mentioned earlier, have mainly concerned the written domain. Thus, in the current study, we look into the idiosyncratic selection of particles and interjections in spoken Japanese, as found in spoken monologues. More precisely, the current study investigates:

- To what extent Japanese speakers are idiosyncratic in selecting certain particles or interjections over others;
- How many particles and interjections need to be included for the most accurate speaker classification results;
- Whether there are any differences between particles and interjections in the degree of idiosyncrasy; and,
- Whether there are any differences between the subcategories of particles in the characteristics of individual differences.

In this study, we focus on particles and interjections. Particles are function words, while interjections are content words. As such, there are distinctive differences in the type of information they provide, as is explained in §2. As a result of these differences, the idiosyncratic information that they carry about speakers may also be different.

In order to answer the aforementioned research questions, we conducted a series of speaker classification tests based solely on particles or interjections. The hypothesis is that the more consistent the individual speaker's selection and use of these words is, and the more strongly the selection and use by one speaker differs from that of another, the more accurate the speaker classifications. We would like to emphasise here that the purpose of the current study is not to improve the accuracy of the speaker classification system, but to investigate the nature of idiosyncrasy in word selections, and to what extent and how the idiosyncrasy of speakers is manifested in word selection for the case of particular particles and interjections.

The current study aims to contribute not only to a better understanding of speaker idiosyncrasy in language use, but also to the advancement of language and speech technologies such as automatic speaker recognition systems (Doddington 2001), plagiarism detection systems (Woolls 2003), and automatic authorship identification systems (Burrows 1987, Baayen *et al.* 1996, Fung 2003). The current study is also relevant to the forensic investigation of linguistic data (Ishihara 2010, Ishihara & Kinoshita 2010).

2. Particles and interjections in Japanese

In this section, the linguistic nature and functions of particles *(jyoshi)* and interjections *(kantanshi)* in Japanese is explained.

There are many different ways of classifying Japanese particles, *jyoshi*, into subcategories, with Japanese linguists forever arguing about what words need to be considered as particles. As a consequence, in Japanese, the term 'particle' is used in a variety of contexts, though generally referring to small, uninflected grammatical words that follow items such as nouns, verbs, adjectives or sentences (Backhouse 1993). In the database we use for this study (cf. §3.1), particles are classified into the subcategories of case particles *(kaku-jyoshi)*, focus particles *(kakari-jyoshi)*, adverbial particles *(fuku-jyoshi)*, conjunctive particles *(setsuzoku-jyoshi)*, final particles *(shu-jyoshi)* and nominal particles *(jyuntai-jyoshi)*. However, in the current study, we combine case and focus particles as case-focus particles because only one item *(-wa)* is subcategorised as a focus particle in the database, and the location in which the focus particle *(-wa)* appears is the same as that of case particles. We do not consider nominal particles, often called *nominalisers*, in this study because there is only one item *(-no)* classified in this subcategory and there is no other category into which nominal particles can be sensibly included. Thus, as shown in Table 1, we investigate case-focus, adverbial, conjunctive and final particles.

	Database subcategories	**Target subcategories**
Particles in Japanese	• case particles • focus particles • adverbial particles • conjunctive particles • final particles • nominal particles	1. case-focus particles 2. adverbial particles 3. conjunctive particles 4. final particles

Table 1. The particle subcategories used in the database and the target subcategories for the current study.

According to Ameka (1992:101), interjections are well recognised by people, but are a neglected part of speech in theoretical linguistics. Ameka (1992:113-114) classifies interjections into three categories: expressive, conative and phatic interjections. Expressive interjections are vocal gestures that indicate the speaker's mental state, for example, *Yuk!* 'I feel disgust' and *Aha!* 'I now know this'. Conative interjections are those expressions that are uttered at an auditor, such as

Sh! 'I want silence here'. Phatic interjections are those expressions that are used to establish and maintain communicative contact, including backchanneling and fillers.

In the following subsections, we provide more detailed information about the target subcategories of particles and interjections.

2.1 Case particles

Case particles (*kaku-jyoshi*) provide the grammatical relationship between the predicate of a sentence and the noun phrases appearing in the sentence. In (1), the case particles, *-ga*, *-de* and *-o* indicate that the immediately preceding noun phrases serve as the subject, instrument and direct object of the predicate of the sentence, respectively.

(1) *ani* ***-ga*** *boo* ***-de*** *watashi* ***-o*** *tataita*
elder.brother-**SUBJECT** stick-**INSTRUMENT** I -**DIRECT.OBJECT** hit.PAST
'My elder brother hit me with a stick.'

2.2 Focus particles

Focus particles focus on, or emphasise, the noun to which they are attached. In (2), the noun that is followed by the focus particle *-wa* serves as the topic of this sentence. Note that the location in which the focus particle appears is the same as that of case particles, though the function is significantly different. Another difference between the focus particle, *-wa* and case particles, is that *-wa* follows some of the case particles.

(2) *watashi-**wa*** *sore-o* *tabenakatta*
I -**FOCUS** it -DIRECT.OBJECT eat.NEGATIVE.PAST
'As for me, I did not eat it.'

As explained earlier, case and focus particles in this study are treated as one group of case-focus particles.

2.3 Conjunctive particles

As the name indicates, conjunctive particles are used to join clauses in a variety of contexts. In sentences (3) and (4), the two verbs are joined with the conjunctive

particles *-kedo* and *-nagara*, which provide the meanings of *but* and *while* in English, respectively.

(3) *ringo -o katta* ***-kedo*** *tabenakatta*
apple-DIRECT.OBJECT buy.PAST**-but** eat.NEGATIVE.PAST
'I bought an apple, but I did not eat it.'

(4) *ringo -o aruki* ***-nagara*** *tabeta*
apple-DIRECT.OBJECT walk **-while** eat.PAST
'I ate an apple while walking.'

2.4 Adverbial particles

Adverbial particles are attached to clauses, and modify the predicate of a sentence, as can be seen in (5). They are adverbial in behaviour (Matsumura 1969). As illustrated in (6), some adverbial particles can be attached to nouns (also adjectives and adverbs) (Kaiser *et al.* 2001).

(5) *watashi-wa ringo -o tabeta* ***-dake*** *-da*
I -TOPIC apple-DIRECT.OBJECT eat.PAST **-only** -COPULA
'I ate only an apple.'

(6) *watashi* ***-dake*** *ringo -o tabeta*
I **-only** apple-DIRECT.OBJECT eat.PAST
'Only I ate an apple.'

2.5 Final particles

Final particles appear in sentence-final position. These particles show in various ways how the speaker appeals to the listener, and with what sort of interactional attitude (Kaiser *et al.* 2001). The example sentences given in (7), (8) and (9) are of the same construction, except for the final particles *-ka*, *-yo* and *-ne*, respectively. The final particle *-ka* is a question particle. The final particle *-yo* is used to indicate that the sentence expresses what the speaker knows or believes, while the final particle *-ne* is used to indicate that the sentence expresses what the speaker believes that the hearer knows or believes (Katagiri 2007:1315). However, as Katagiri (2007) argues, amongst other things, intonation plays an important role in the interpretation of the meaning of the final particle (Davis 2011, Venditti 1995).

(7) *kaigi -wa rokuji -kara desu* ***-ka***
meeting-TOPIC 6.oclock-from COPULA**-KA**
'Is the meeting from 6 o'clock?'

(8) *kaigi -wa rokuji -kara desu* ***-yo***
meeting-TOPIC 6.oclock-from COPULA**-YO**
'The meeting is from 6 o'clock (I believe).'

(9) *kaigi -wa rokuji -kara desu* ***-ne***
meeting-TOPIC 6.oclock-from COPULA**-NE**
'The meeting is from 6 o'clock, isn't it (I believe that you believe so).'

There are well-reported gender differences in the use of final particles (Martin 2004, Kinsui 2007). For example, *-ze* and *-zo* are fairly crude expressions, and thus are exclusively used by (young) males while *-wa* tends to be used by females to express femininity (Martin 2004, Matsumura 1969).

2.6 Interjections

According to Martin (2004:1041), interjections function to A) express the speaker's emotional reactions, such as pleasure, relief, surprise, hesitation, or disgust; B) call attention; C) respond to a question, a command, or a social transaction; and D) hold the floor when fluency fails and the speaker is searching for a desired expression (e.g. fillers).

Since the target utterances in the current study are monologues, the majority of tokens categorised as interjections are in fact fillers, which belong to group D. However, there are some occurrences that belong to A, such as *ara* 'oh', *ee* 'eh' and *yoisho* 'oof' and to C, such as *hai* 'yes' and *un* 'yep'.

2.7 Differences between particles and interjections, and also between the subcategories of particles

As explained in §2.1 to §2.5, particles are non-conjugated function words. They follow items such as nouns, verbs, adjectives or sentences, and they prosodically merge into the preceding material (Backhouse 1993). On the other hand, interjections can be used by themselves as independent free-standing units, grammatically like sentences (Tokieda 1950). Like the four functions of

interjections summarised in §2.6, interjections are more related to higher level information (e.g. para-/extra-linguistic information, such as emotions) than particles, which mainly serve to carry linguistic information such as syntactic relationships and minor modifications of meaning. It is interesting to see if there is any difference in the manifestation of speaker idiosyncrasies between particles and interjections. Furthermore, the nature and function of the subcategories of particles are also very different. For example, final particles provide the speaker's attitude towards the listener, which is beyond simple syntactic information. Thus, it is also of interest how the idiosyncratic information of speakers is carried by the different categories of particles.

3. Methodology

This is a linguistic study on idiosyncrasy using speaker classification techniques as analytical tools. The more consistent the individual speaker's selection of certain words is, and the more significantly those words selected by the speaker vary from those selected by another, the more accurately the speaker classification is performed.

Two kinds of comparisons are involved in speaker classification tests. The first is called *Same Speaker Comparison* (SS comparison) in which two speech samples produced by the same speaker need to be correctly identified as the same speaker. The other is, *mutatis mutandis*, *Different Speaker Comparison* (DS comparison).

The series of speaker classification tests that we conducted can be categorised into two experiments: Experiment 1 investigates how well we can classify speakers based on each of the different subcategories of the particles (cf. §5.1). Experiment 2 investigates the overall performance of all particles and interjections in speaker classification (cf. §5.2). Although the target words for Experiments 1 and 2 are different, the experimental methodology is identical for both of them.

3.1 Database and speakers

For speech data, we used the Corpus of Spontaneous Japanese (CSJ) (Maekawa *et al.* 2000), which contains recordings of various speaking styles such as sentence

reading, monologue, and conversation. For this study, we used only the monologues, categorised as either Academic Presentation Speech (APS) or Simulated Public Speech (SPS). APS was mainly live-recorded academic presentations, between 12-25 minutes long. For SPS, 10-12 minute mock speeches on everyday topics were recorded. We selected our speakers from this corpus based on three criteria: availability of multiple and non-contemporaneous recordings, spontaneity (e.g. not reading) of the speech, and standard modern Japanese speech. The spontaneity of the language and the extent to which it conforms to standard modern Japanese were assessed on the basis of the rating the CSJ provided. Thus, only those speech samples which were high in spontaneity and uttered entirely in Standard Japanese were selected for this study. This gave us 416 speech samples for inclusion (= 208 speakers: 132 male and 76 female speakers x 2 sessions).

3.2 Basic statistics

Table 2 provides the basic statistics of the target particles and interjections. In this study, we decided to use those particle types that appeared three times or more in the selected speech samples for the speaker classification experiments. As seen in Table 2, 50% of all particle types belong to case-focus particles. Final particle types account for only 10% of all particle types.

	Occurrences (% in all particle types)	**N ≥ 3 (% in all particle types)**
Case-focus particles	88 (49%)	64 (50%)
Conjunctive Particles	29 (16%)	20 (15%)
Adverbial particles	39 (22%)	31 (24%)
Final particles	21 (11%)	13 (10%)
All particles	177	128
Interjections	123	70

Table 2. Basic statistics of the target particle and interjection types.

70 different interjections are used in this study. The number of different types of interjections is very similar to the number of different types of case-focus particles, 64.

Table 3 contains the ten most frequently used particle types listed in descending order, separately for the subcategories and all together for all particle types.

	C-F	N	Conj	N	Adverb	N	Final	N	All	N	Type
1	*-no*	49,206	*-te*	19,344	*-mo*	16,327	*-ne*	8,289	*-no*	49,206	Case
2	*-wa*	30,823	*-keredo*	8,541	*-toka*	4,566	*-ka*	6,350	*-wa*	30,823	Focus
3	*-ga*	30,646	*-ga*	5,303	*-tte*	4,156	*-na*	2,005	*-ga*	30,646	Case
4	*-o*	30,623	*-to*	5,255	*-kurai*	2,860	*-yo*	1,211	*-o*	30,623	Case
5	*-ni*	29,603	*-node*	3,701	*-made*	1,737	*-no*	56	*-ni*	29,603	Case
6	*-to*	20,033	*-ba*	1,541	*-tari*	1,580	*-zo*	38	*-to*	20,033	Case
7	*-toiu*	19,438	*-kara*	1,464	*-dake*	1,567	*-wa*	29	*-toiu*	19,438	Case
8	*-de*	16,167	*-shi*	912	*-ya*	1,248	*-ke*	23	*-te*	19,344	Conj
9	*-kara*	4,711	*-demo*	906	*-nado*	916	*-ya*	21	*-mo*	16,327	Adverb
10	*-toshite*	2,233	*-nagara*	535	*-hodo*	906	*-kashira*	13	*-de*	16,167	Case

Table 3. The ten most frequently used particle types for each subcategory of the particles. C-F = case-focus particles; Conj = conjunctive particles; Adverb = adverbial particles; Final = final particles; All = all particles; N = occurrences; Type = type of particles appearing in all particles.

Table 3 is also referred to when we discuss the results of the speaker classification experiments in §5.

Mirroring the fact that case-focus particle types account for 50% of all particle types, the occurrences of the ten most frequently used case-focus particles are significantly greater than those of the other particles. Consequently, eight of the ten most frequently used particles are case-focus particles, as can be seen in the rightmost column of Table 3. Note that the *-no* particle presents as the most frequently used particle. This is the case despite the fact that the genitive particle as the nominaliser particle *-no* is excluded in this study.

The different types of interjections listed in Table 4 are all fillers.

	Interjections	N
1	*e-*	27776
2	*e*	12046
3	*ma*	8816
4	*ano-*	7213
5	*ano*	6988
6	*ma-*	5990
7	*sono*	2533
8	*e-to*	2479
9	*a*	2364
10	*n*	1924

Table 4.
The ten most frequently observed interjection types. N = occurrence. '-' indicates long vowel length.

3.3 Vector space model

In this study, we compare many sets of paired speech samples. Using the occurrences of the identified words, each speech sample is modelled as a real-valued vector[3]. If n different words are used to represent a given speech sample S, the dimensionality of the vector is n. That is, S is represented as a vector of n dimensions ($\vec{S} = (F_1, F_2 \ldots Fn)$, in which F_n represents the nth component of $\vec{S}$ and F_n is the frequency of the nth word). For example, if 5 words (e.g. *ah*, *like*, *OK*, *yes*, *all right*) are used to represent a speech sample (x), and the frequency counts of these words in the speech sample are 3, 10, 4, 18 and 1, respectively, the speech sample x is represented as given in (1).

$$(1)\ \vec{x} = (3,10,4,18,1)$$

The speech samples in this study are modelled using different vector dimensions (e.g. using the first 20 most frequently used fillers). This is to see how the performance of the speaker classification system is influenced by the number of dimensions.

3.4 Term frequency-inverse document frequency weighting

The usefulness of particular words for the purposes of speaker classification is determined by their uniqueness. This is based on the number of different speech samples in which they occur, as well as how frequently they are used in a particular speech sample. For instance, if a given word is used by many speakers many times, this particular word is not as useful as a word which is used by a smaller number of people in many instances. Different weights are therefore given to different words depending on their uniqueness in the pooled data. The $tf{\cdot}idf$ (term frequency-inverse document frequency) weight (cf. Formula (2)) is used to evaluate how unique a given word is in the population. A corresponding weight is given to that word to reflect its importance in speaker classification (Manning & Schütze 2000).

[3] Readers with little background in mathematics and statistics are advised to read chapter five of (Manning & Schütze 2000), in which they explain the statistics that are available and how they can be used for the analysis of word usages.

$$(2)\ w_{i,j} = tf_{i,j} * log(\frac{N}{df_i})$$

In Formula (2), term frequency ($tf_{i,j}$) is the number of occurrences of word i (w_i) in the document (or speech sample) j (d_j). Document frequency (df_i) is the number of documents (or speech samples) in the collection in which that word i (w_i) occurs. N is the total number of documents (or speech samples).

3.5 Cosine similarity measure

The similarity between two speech samples, which are represented as vectors ($\vec{x}$, $\vec{y}$), is calculated based on the cosine similarity measure. This is indicated in (3) (Manning & Schütze 2000). This particular method was selected in order to normalise the different durations of the speech samples. The cosine similarity measure is based on the assumption that the direction of a vector should be constant if the speech sample is long enough.

$$(3)\ similarity(\vec{x},\vec{y}) = cos(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}||\vec{y}|} = \frac{\sum_{i=1}^{n} x_i * y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 * \sum_{i=1}^{n} y_i^2}}$$

The range of difference between the two vectors ($\text{similarity}(\vec{x},\vec{y})$) is between 1.0 ($=cos(0°)$) for two vectors pointing in the same direction—e.g. speech samples which are identical—and 0.0 ($=cos(90°)$) for two orthogonal vectors—two speech samples which are completely different, because weights are by their definition not negative[4]. Note that in the experiments of this study, the length (number of dimensions) of the vectors was standardised by only looking at the X most frequent particles and interjections (X = (5, 10, 15, 20, 25, 30, 35, 40 … N); N = the maximum number of dimensions), since the cosine similarity measure requires vectors of equal length (number of dimensions).

[4] Note that the range of cosine similarity measure, which is between 0 for two orthogonal vectors and between 1 for two vectors pointing in the same direction, is counter-intuitive. Readers need to be reminded that 0 stands for two speech samples being completely different and 1 for those being identical.

4. Method for speaker classification

The performance of speaker classification is assessed on the basis of the probability distribution functions (PDFs) of the difference between two contrastive hypotheses. One is the hypothesis that two speech samples were uttered by the same speaker (the same speaker (*SS*) hypothesis) and the other is that two speech samples were uttered by different speakers (the different speaker (*DS*) hypothesis). These probabilities can be formulated as $P(E|H_{ss})$ and $P(E|H_{ds})$ respectively, where E is the difference, H_{ss} is the *SS* hypothesis and H_{ds} is the *DS* hypothesis. In this study, the *PDF* of the difference assuming the *SS* hypothesis is true is called the *SS PDF* (PDF_{ss}), and the *PDF* assuming the *DS* hypothesis is true is the *DS PDF* (PDF_{ds}). Specific to this study, the difference between two speech samples refers to the cosine difference between the two vectors representing the two speech samples. Each *PDF* was modelled using the kernel density function (KernSmooth library of R statistical package). Examples of PDF_{ss} and PDF_{ds} are given in Figure 1. In Figure 1, the PDF_{ss} and PDF_{ds} do not conform to a normal distribution, which is the motivation for the use of the kernel density function in this study.

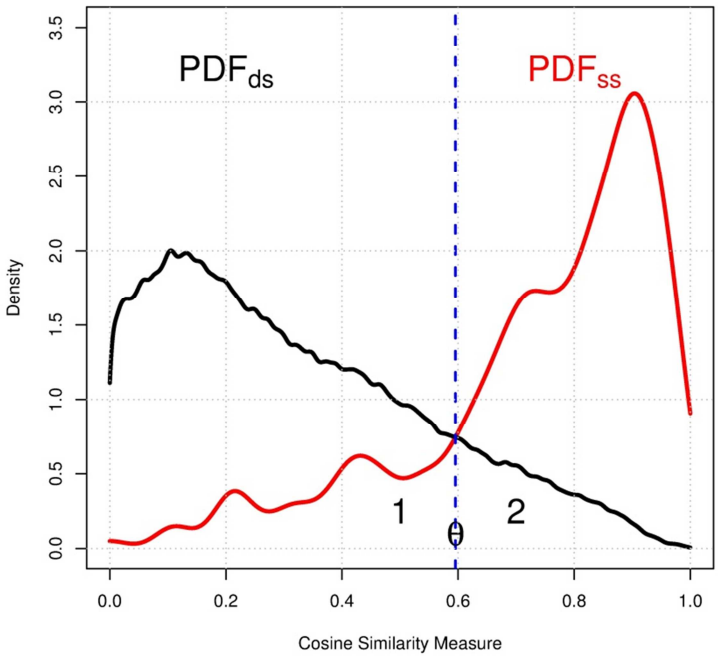

Figure 1. An example of PDF_{ss} (red curve) and PDF_{ds} (black curve). The x-axis is the cosine similarity measure (c) and the y-axis is the probability density (d). The blue vertical dotted line (ɵ) is the crossing point between PDF_{ss} and PDF_{ds}. Area 1 is the area surrounded by the red curve (PDF_{ss}), d = 0 and c = ɵ. Area 2 is the area surrounded by the black curve (PDF_{ds}), d = 0 and c = ɵ.

As can be seen from Figure 1, PDF_{ss} and PDF_{ds} are not always monotonic. This may result in more than one crossing point (which is not shown in Figure 1, particularly when the dimension of a vector is less than 5. Thus, the performance of the system with the dimension of a vector less than 5 is not given. These two *PDF*s also show the accuracy of this particular speaker classification system. If the crossing point (θ) of the PDF_{ss} and the PDF_{ds} is set as the threshold, we can estimate the performance of this particular speaker classification system from these *PDF*s. Area 1 in Figure 1—the area surrounded by the red line (PDF_{ss}), the vertical dotted line of $c = \theta$ and the line of $d = 0$—is the predicted error for the SS comparisons. Area 2 of Figure 1—the area which is surrounded by the black line (PDF_{ds}), the vertical dotted line of $c = \theta$ and the line of $d = 0$—is the predicted error for the *DS* comparisons. Therefore, the accuracy (%) of the *SS* ($ACCURACY_{ss}$) and *DS* ($ACCURACY_{ds}$) comparisons can be calculated by (4) and (5), respectively.

$$(4)\ accuracy_{ss}(\%) = \left(\frac{\int_0^{\theta} PDF_{ss}(x)dx}{\int_0^1 PDF_{ss}(x)dx}\right) * 100$$

$$(5)\ accuracy_{ds}(\%) = \left(\frac{\int_{\theta}^1 PDF_{ds}(x)dx}{\int_0^1 PDF_{ds}(x)dx}\right) * 100$$

The accuracy of a speaker classification system (both in *SS* and *DS* comparisons) was estimated in this way.

For the selected 416 speech samples obtained from 208 speakers, 208 *SS* and 86,112 *DS* comparisons are possible. In the speaker classification tests, spatial vectors of different dimensions (5, 10, 15, 20 … N, where N is the maximum number of dimensions) are used to see how the number of vector dimensions affects the performance of speaker classification. That is, for the adverbial particles, which include 31 different kinds, we applied the vector sizes (number of dimensions) of 5, 10, 15, 20, 25, 30 and 31.

5. Test results and discussions

In this section, the classification performance of the different subcategories of the particles is closely investigated in §5.1, followed by comparison between the

performance of all particles and that of interjections in §5.2. In §5.3, the speaker-individualising characteristics of the particles belonging to the different subcategories will be scrutinised in terms of between- and within-speaker differences.

5.1 Experiment 1: subcategories of particles

The respective speaker classification performances of the different particle subcategories (case-focus, adverbial, conjunctive, and final particles) are presented first. The differences between them in terms of performance are described before discussing possible reasons for the identified differences.

In Figure 2, the average speaker classification accuracy between the same speaker (*SS*) and different speaker (*DS*) comparisons is plotted separately for the different subcategories of the particles as a function of the number of vector dimensions.

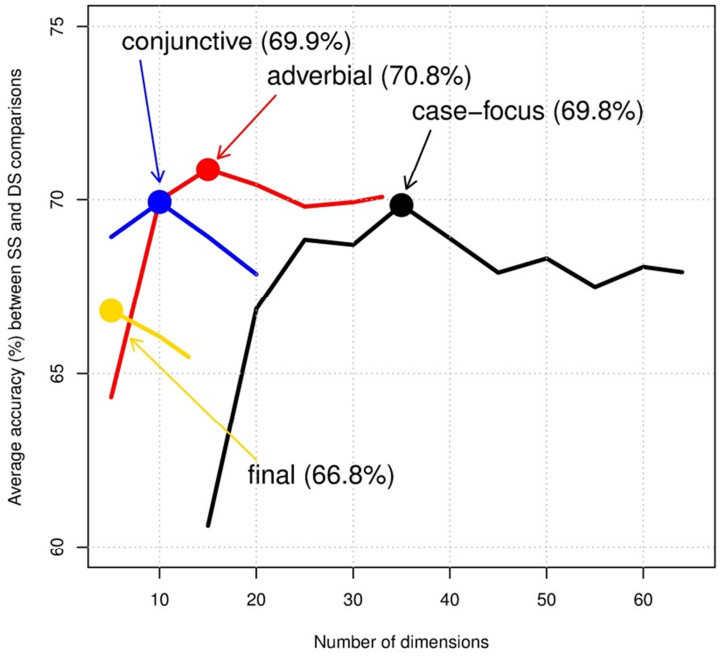

Figure 2. The average accuracy (y-axis) between the SS and DS comparisons is plotted separately for case-focus particle, conjunctive particle, adverbial particle and final particle as a function of the number of dimensions (x-axis) used in the speaker classification tests. The circles indicate the best accuracy for each type.

As can be seen from Figure 1, the speaker classification accuracy reaches as high as approximately 70% for case-focus, adverbial and conjunctive particles. Adverbial and conjunctive particles reach their highest accuracy points with a fewer number of dimensions (15 and 10 dimensions, respectively) than case-focus particles (35 dimensions). The reader is reminded that, for example, 15

dimensions indicates that the speaker classification test was conducted using the 15 most frequently used particles in the subcategory. For case-focus particles, the speaker classification accuracy considerably improves from 15 dimensions (60.6%) to 25 dimensions (68.8%). A similar jump in accuracy can be observed with fewer dimensions (from 5 dimensions: 64.3% to 15 dimensions: 69.9%) for adverbial particles. The classification accuracy of conjunctive particles is as high as 69.9% with as few as only 5 dimensions.

The observation that more dimensions (or particle types) need to be included for case-focus particles to reach the same level of accuracy (approximately 70%) as adverbial and conjunctive particles is probably because the first 15-20 most frequently used case-focus particles are so ubiquitous. Hence, there is not much room for them to bear the individualising information of the speakers. This frequent occurrence of case-focus particles can be seen from Table 3, in which the occurrence of the top ten case-core particles is substantially higher than those of the other particles. Please also note that the curve of the case-focus particles in Figure 2 starts with 15 dimensions because the PDF_{ss} and the PDF_{ds} with less than 15 dimensions become non-monotonic, having multiple crossing points between them[5]. Sensible results therefore cannot be obtained with less than 15 dimensions.

Case particles (in particular, those which are frequently used) are the backbone of the syntactic structure of Japanese utterances. It would be impossible for the speaker to accurately convey the intended message were it not for case particles. Since case particles serve as the dominant carrier of information, which is directly connected to the propositions of the messages, it is likely that less idiosyncratic individual speaker information is encoded in case particle usage. Consequently, more case-focus particles need to be included to get the same level of accuracy as adverbial and conjunctive particles.

After case-focus particles reach their highest accuracy of 69.8% with 35 dimensions, the classification accuracy continues to marginally decrease with some minor ups and downs as the number of dimensions increases. However, this trend is not surprising. The feature vectors are based on the frequency of a

[5] In Figure 1, for example, the PDF_{ss} and the PDF_{ds} have only one crossing point which is aligned with c = ɵ. However, with fewer than 15 dimensions, the PDF_{ss} and the PDF_{ds} start having two or more crossing points.

given particle word; we picked those with a higher frequency first to be included in the feature. As such, vectors in the later orders have low frequencies. This means that the latter part of longer vectors tends to include very similar low numbers across speakers, introducing noise into the assessment of between-speaker difference and thereby making them look more similar. The same trend cannot be clearly observed for adverbial and conjunctive particles; this is most likely due to the fact that the number of dimensions of the feature vectors for adverbial and conjunctive particles is not as high as that of the case-focus particles.

The speaker classification accuracy is notably lower for final particles in comparison to the other particles. This is contrary to our conjecture that the gender difference in the use of final particles would work in favour of speaker classification. Two possible reasons can be noted for the poor performance of final particles. One is due to the speech style of the monologue samples (conference presentation and mock speech), both of which are fairly formal. Gender and speaker differences in the use of final particles may be more salient in informal colloquial speech, as many final particles are related to interaction rather than monologue-style speech. Another reason may be due to the fact that the length of the feature vector is far shorter (only 13) for final particles than for the other particles.

5.2 Experiment 2: particles and interjections

The following section compares the classification performance with all particles together versus that of interjections. In Figure 3, the average speaker classification accuracy between the same speaker (*SS*) and different speaker (*DS*) comparisons is plotted as a function of the number of vector dimensions. These functions are shown separately for all of the particles and interjections. Figure 3 (next page) also includes the results presented in Figure 2.

There is a notable sudden improvement in accuracy in both all particles and interjections: a substantial improvement can be observed between 15 dimensions (74.8%) and 25 dimensions (79.4%) for all particles, and between 5 dimensions (75.6%) and 15 dimensions (81.5%) for interjections. As for the highest accuracy, it is 80.5% for all particles with 45 dimensions, while it is 82.7% for interjections

with 25 dimensions. The observation that all particles need more dimensions than interjections to reach the highest accuracy point can be attributed to the fact that, as can be seen from Table 2, the earlier order vectors of all particles contain many of the frequently occurring case-focus particles. It was previously discussed in §5.1 that these case-focus particles do not have much individualising information.

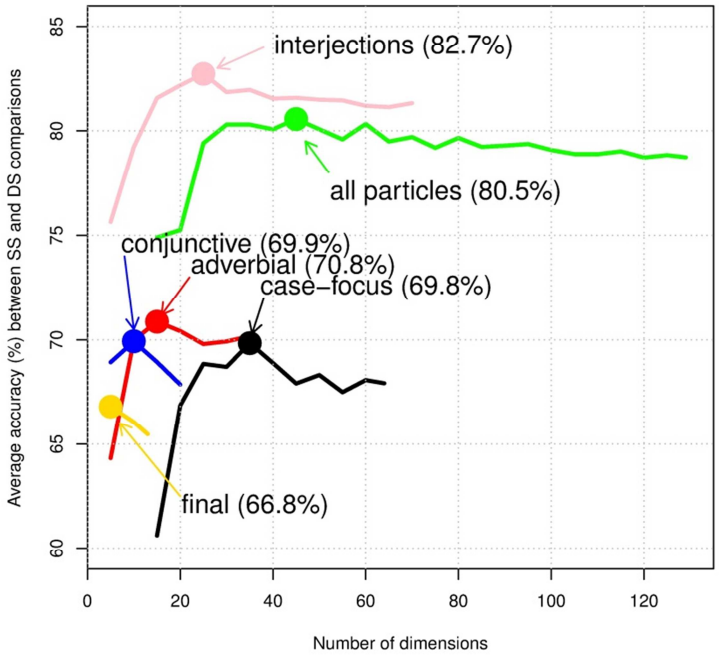

Figure 3. The average accuracy (y-axis) between the SS and DS comparisons is plotted separately for all particles and interjections as a function of the number of dimensions (x-axis) used in the speaker classification tests (top half). The circles indicate the best accuracy. The results presented in Figure 2 are also included as a reference (bottom half).

It is evident from Figure 3 that the performance of speaker classification is consistently better for interjections than for all particles, indicating that interjections carry more individually identifying information than particles do. As explained earlier, an interjection is a word used to express an emotion or a sentiment on the part of the speaker. Communication has been traditionally viewed as an intentional act of transferring information. However, independent of the mode of communication (spoken or written), paralinguistic or extralinguistic information is also conveyed along with the symbolic content of the intended message. Paralinguistic information is information about the speaker or writer, such as their age, gender, social background, psychological state, or health. This latter sort of information is often called paralinguistic or extralinguistic information (Abercrombie 1967, Nolan 1983, Rose 2002).

A large portion of the words classified as interjections in the database are fillers. It has been argued based on empirical data that fillers manifest the cognitive process

that the speaker is undergoing (Sadanobu & Takubo 1995), and also reflect the speaker's difficulty in conceptual planning and linguistic encoding (Watanabe *et al.* 2008). The cognitive process is a well-known source of individual differences (Cooper 2002). Fillers therefore transfer more than linguistic information encoded in written messages; fillers do not appear in written texts. On the other hand, particles (except for final particles) are directly involved in transmitting linguistic information such as the syntactic relationship between a noun phrase of a sentence and the predicate of the sentence, or the logical relationship between two clauses. These usages of case particles show that they are more directly relevant for transferring the content information encoded in messages as accurately as possible than interjections are.

Despite the fact that each subcategory of particles has only approximately 66.5-71.0% accuracy (cf. §5.1), the speaker classification result drastically improves by approximately 10% when all particles are included in the tests. This indicates that the individualising information of the speakers is encoded differently in the uses of the different subcategories of particles. If the individual characteristics of the speakers had been encoded in the different subcategories of particles in the same manner, the inclusion of all particles would not have had any effect on the performance of the speaker classification. This point is explored in §5.3 in terms of the degree of between- and within-speaker differences.

5.3 Differences between particle subcategories

It was pointed out that individualising information of speakers is manifested differently in the uses of different subcategories of particles. That is, the different subcategories of particles carry different aspects of individual speaker idiosyncrasies. In this subsection, we investigate how differently different types of particles possess speaker individualising information.

The performance of speaker classification is mainly determined by two factors: 1) the degree of between-speaker differences, and 2) that of within-speaker differences. We explained earlier that the more consistent the individual speaker's selection of words is, and the more significantly the selected words of one speaker differ from those selected by another, the more accurately the speaker classification can be performed. In other words, the greater the between-speaker

differences are, and concurrently, the smaller the within-speaker differences are in terms of the selection of words, the more accurately speakers can be classified.

Having said that, with the degree of within-speaker differences being constant, the performance of speaker classification will improve as the degree of between-speaker differences becomes greater. Equally, with the degree of between-speaker differences being constant, the performance will also improve as the degree of within-speaker differences becomes smaller. Although the speaker classification accuracy appears to be comparable between the case-focus, adverbial and conjunctive particles, the results presented in §5.2 show that their configurations in terms of the degree of between- and within-speaker differences are distinct from one another.

The degree of between-speaker differences and that of within-speaker differences are manifested as the shape of the PDF_{ds} and PDF_{ss}, respectively. How they are derived is explained using Figure 4, a modified version of Figure 1.

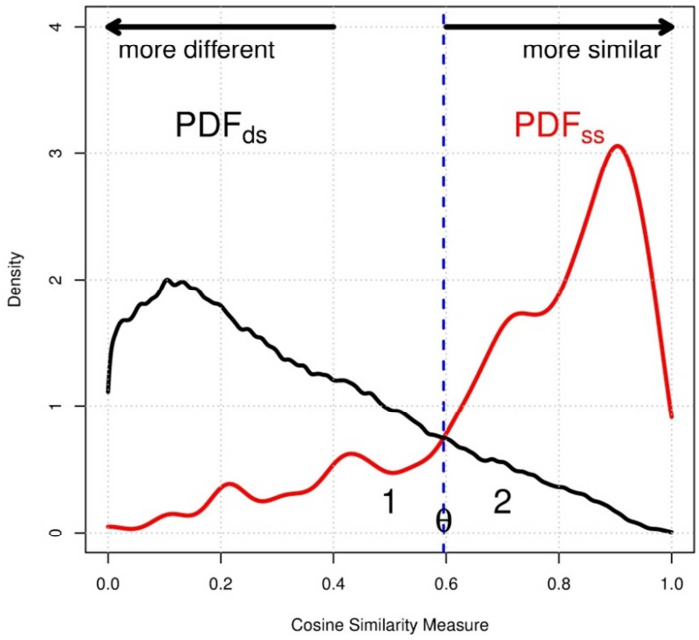

Figure 4. A modified Figure 1 is given to demonstrate that the degree of between-speaker differences and that of within-speaker differences are manifested as the shape of the PDF_{ds} and PDF_{ss}, respectively. The x-axis is the cosine similarity measure (c) and the y-axis is the probability density (d). The blue vertical dotted line (θ) is the crossing point of PDF_{ss} and PDF_{ds}. Area 1 is the area surrounded by the red curve (PDF_{ss}), d = 0 and c = θ. Area 2 is the area surrounded by the black curve (PDF_{ds}), d = 0 and c = θ.

The PDF_{ds} becomes more skewed towards the cosine similarity measure $c = 0$ as the degree of between-speaker differences increases (i.e. the particles used by different speakers are more different), but towards $c = 1$ as the degree of between-speaker differences decreases (i.e. the particles used by different people are more

similar). Likewise, the PDF_{ss} becomes more skewed towards $c = 0$ as the degree of within-speaker differences increases (i.e. the particles used by the same speaker vary more), but towards $c = 1$ as the degree of within-speaker differences decreases (i.e. the particles used by the same speaker are more consistent). In order to quantify the shape of the *PDF*s, two measurements were taken: one is the mean value of the cosine similarity values which constitute each of the PDF_{ds} and PDF_{ss}, and the other is the skewness[6] of the PDF_{ds} and PDF_{ss}. These two measurements were made for each of the different subcategories of particles: case-focus, adverbial, conjunctive and final particles, and also for all particles and interjections, as they are plotted in Figure 5 (next page).

Figure 5 clearly demonstrates that the different subcategories of particles have different characteristics with respect to the degree of between- and within-speaker differences. The characteristics that can be viewed from the two panels (mean and skew) of Figure 5 are essentially the same. Thus, the differences between the different subcategories of particles are described by reference to the mean values (the top panel of Figure 5).

As can be seen in Figure 5, case-focus particles (3) have greater between- and within-speaker differences, with their mean values located closer to cosine similarity measure $c = 0$ than the other subcategories of particles. Final particles (4), however, exhibit less between- and within-speaker differences, with their mean values located closer to $c = 1$. That is, in comparison to the other subcategories of particles, the selection of different case-focus particles is highly idiosyncratic across speakers, yet the selection of case-focus particles is not consistent within the same speaker. The behaviour of final particles is completely opposite to that of case-focus particles. The same speaker uses the same type(s) of final particles more consistently than the other subcategories of particles, while the selection of different types of final particles is less variable than that of the other types of particles across different speakers. Conjunctive particles (2) are similar to final particles. Adverbial particles (1) occupy an intermediate position compared to the other subcategories.

[6] Skewness was quantified by cubing the deviations from the mean, and dividing the average cubed distance by the cube of the standard deviation.

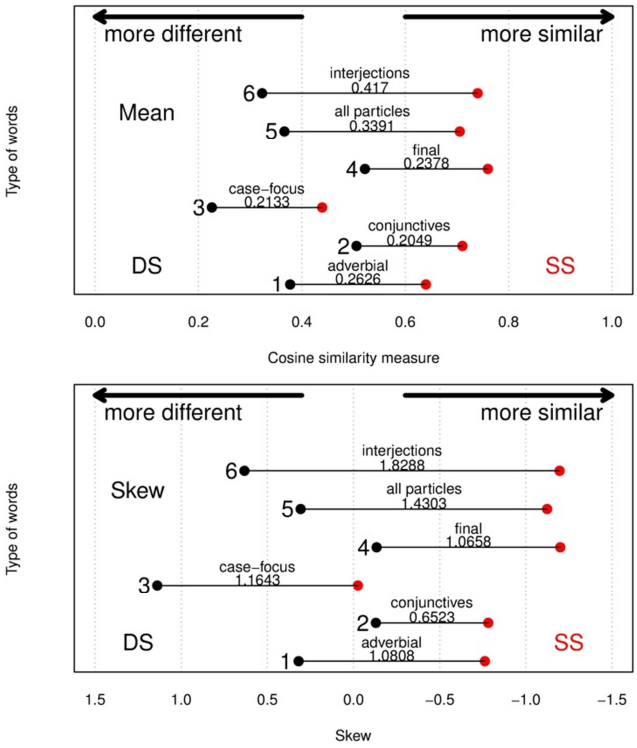

Figure 5. The mean (top panel) and skew (bottom panel) values of the cosine similarity measures of the PDF_{ds} (black circles) and PDF_{ss} (red circles), plotted separately for adverbial particles (1), conjunctive particles (2), case-focus particles (3), final particles (4), all particles (5) and interjections (6). The numerical values are the distances between the measurements for PDF_{ss} and PDF_{ds}.

As for all particles (5) and interjections (6), it can be seen from Figure 5 that interjections perform better than all particles because the former has greater between-speaker differences and smaller within-speaker differences than the latter.

6. Summary and conclusions

We investigated the following research questions in Japanese monologues:

- To what extent are Japanese speakers idiosyncratic in selecting certain particles and interjections rather than others;
- How many particles and interjections need to be included for the best speaker classification results;
- Whether there are any differences between particles and interjections in the degree of idiosyncrasy; and
- Whether there are any differences between the subcategories of particles in the characteristics of individual differences.

It has been demonstrated that particles and interjections carry idiosyncratic speaker information to the extent that the average speaker classification accuracy of the same and different speaker comparisons is about 80.5% and 82.7%, respectively. We suggested that interjections carry more idiosyncratic information about speakers than particles do because of the different levels of information that they denote. Namely, particles mainly handle a linguistically lower level of structural information, which is directly relevant to the content of messages, whereas interjections assume the task of conveying paralinguistic and extralinguistic information. These types of information have a stronger relevance to the speakers' cognitive processes and are highly diverse on an individual level. We also demonstrated that in comparison to interjections, particles require the inclusion of more dimensions in order to reach the highest accuracy point.

We showed that the different subcategories of particles (case-focus, adverbial, conjunctive and final particles) exhibit distinctive characteristics in terms of the degree of between-speaker and within-speaker differences. Due to these differences, although the speaker classification performance was only approximately 70% accurate for each subcategory of case-focus, adverbial and conjunctive particles, the classification performance substantially improved when all particles were combined together.

Particles and interjections account for merely a small part of our entire word usage. Despite this, we may say that they carry a substantial amount of speaker idiosyncratic information. If we are able to exploit all the word usage information as speaker classification features, it is likely that speaker classification can be performed with a high level of accuracy. This can lead to the interpretation that language usage is fairly individualised—even more so than we tend to think. Thus, linguistic studies on individual differences deserve more attention, perhaps as much as the more common studies which focus on the invariant aspects of language use.

References

Abercrombie D 1967 *Elements of General Phonetics* Edinburgh: Edinburgh University Press.

Ameka F 1992 'Interjections: the universal yet neglected part-of-speech' *Journal of Pragmatics* 18(2-3):101-118.

Baayen H, H Van Halteren & F Tweedie 1996 'Outside the cave of shadows: using syntactic annotation to enhance authorship attribution' *Literary and Linguistic Computing* 11(3):121-132.

Backhouse AE 1993 *The Japanese Language: An introduction* Oxford: University Press Oxford.

Binongo JNG 2003 'Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution' *Chance* 16(2):9-17.

Burrows JF 1987 'Word-patterns and story-shapes: the statistical analysis of narrative style' *Literary and Linguistic Computing* 2(2):61-70.

Cooper C 2002 *Individual Differences* 2nd ed. Arnold; New York; London: Oxford University Press.

Coulthard M & A Johnson 2007 *An Introduction to Forensic Linguistics: Language in Evidence* London: Routledge.

Davis CM 2011 *Constraining Interpretation: Sentence Final Particles in Japanese* Unpublished PhD thesis, University of Massachusetts.

Doddington G 2001 'Speaker recognition based on idiolectal differences between speakers' *Proceedings of 2001 Eurospeech*. Pp. 2521-2524.

Fung G 2003 'The disputed federalist papers: SVM feature selection via concave minimization' *Proceedings of the 2003 Conference on Diversity in Computing*. Pp. 42-46.

García AM & JC Martín 2007 'Function words in authorship attribution studies' *Literary and Linguistic Computing* 22(1):49-66.

Halliday MAK, A Macintosh & PD Strevens 1964 *The Linguistic Sciences and Language Teaching* London: Longmans.

Holmes DI 1992 'A stylometric analysis of Mormon scripture and related texts' *Journal of the Royal Statistical Society Series a-Statistics in Society* 155:91-120.

Holmes DI, M Robertson & R Paez 2001 'Stephen Crane and the New York Tribune: a case study in traditional and non-traditional authorship attribution' *Computers and the Humanities* 35(3):315-331.

Ishihara S 2010 'Variability and consistency in the idiosyncratic selection of fillers in Japanese monologues: gender differences' *Proceedings of the Australasian Language Technology Association Workshop 2010*. Pp. 9-17.

——— & Y Kinoshita 2010 'Filler words as a speaker classification feature' *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*. Pp. 34-37.

Juola P & RH Baayen 2005 'A controlled-corpus experiment in authorship identification by cross-entropy' *Literary and Linguistic Computing* 20(Suppl):59.

Kaiser S, S Butler, N Kobayashi & H Yamamoto 2001 *Japanese: A Comprehensive Grammar* London: Routledge.

Katagiri Y 2007 'Dialogue functions of Japanese sentence-final particles 'yo' and 'ne'' *Journal of Pragmatics* 39(7):1313-1323.

Keselj V, F Peng, N Cercone & C Thomas 2003 'Ed)' *Computational Linguistics* 3.

Kinsui S 2007 'Kindai nihon manga no gengo [The language of modern Japanese manga]' in S Kinsui (ed) *Yakuwarigo kenkyuu no chihei [The Horizon of Role Language Research]* Kuroshio publisher Tokyo. Pp. 176-186.

Maekawa K, H Koiso, S Furui & H Isahara 2000 'Spontaneous speech corpus of Japanese' *Proceedings of the 2nd International Conference of Language Resources and Evaluation*. Pp. 947-952.

Manning CD & H Schütze 2000 *Foundations of Statistical Natural Language Processing* 2nd ed. Cambridge, Mass.: MIT Press.

Martin SE 2004 *A reference grammar of Japanese* Honolulu: University of Hawai'i Press.

Matsumura A (ed) 1969 *Nihon bunpo daijiten [A Comprehensive Dictionary of Japanese Grammar]* Tokyo: Meiji shoin.

Mendenhall TC 1887 'The characteristic curves of composition' *Science* (214S):237-246.

Mosteller F & DL Wallace 1964 *Inference and Disputed Authorship, The Federalist* Addison-Wesley series in behavioral science Quantitative methods Reading, Massachusets: Addison-Wesley.

——— 1984 *Applied Bayesian and Classical Inference: The Case of the Federalist Papers* 2nd ed. New York: Springer-Verlag.

Nolan F 1983 *The Phonetic Bases of Speaker Recognition* Cambridge: Cambridge University Press.

Peng F, D Schuurmans, S Wang & V Keselj 2003 'Language independent authorship attribution using character level language models' *Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics* 1:267-274.

Pennebaker JW & LA King 1999 'Linguistic styles: Language use as an individual difference' *Journal of Personality and Social Psychology* 77(6):1296-1312.

Rose P 2002 *Forensic Speaker Identification* London: Taylor & Francis.

Sadanobu T & Y Takubo 1995 'The monitoring devices of mental operations in discourse: a case of 'eeto' and 'ano (o)' *Gengo kenkyu [Language Studies]* (108):74-93.

Stamatatos E, N Fakotakis & G Kokkinakis 2001 'Computer-based authorship attribution without lexical measures' *Computers and the Humanities* 35(2):193-214.

Thisted R & B Efron 1987 'Did Shakespeare write a newly-discovered poem?' *Biometrika* 74(3):445-455.

Tokieda M (ed) 1950 *Nihongo bunpoo koogohen [Spoken Japanese Grammar]* Nihongo bunpoo koogohen [Spoken Japanese Grammar] Tokyo: Iwanami Shoten.

Venditti JJ 1995 *Japanese ToBI Labelling Guidelines*, Unpublished Manuscript, Ohio State University

Watanabe M, K Hirose, Y Den & N Minematsu 2008 'Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners' *Speech Communication* 50(2):81-94.

Weber F, L Manganaro, B Peskin & E Shriberg 2002 'Using prosodic and lexical information for speaker identification' *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 1:141-144.

Woolls D 2003 'Better tools for the trade and how to use them' *Forensic Linguistics – the International Journal of Speech Language and the Law* 10(1):102-112.