

要約筆記品質評価システムの概要

Summary of Quality Evaluation System of Summary Transcript

高尾 哲 康

Takao Tetsuyasu

聴覚障害者・難聴高齢者に対する情報保障手段である要約筆記について、要約筆記者のスキルや要約品質向上のために定量的評価ができるシステムの研究開発を行なった。発話内容テキストと要約筆記者が入力したテキストをもとに定量的な評価ができるシステムを試作した。要約評価となる評価値計算には重み付き編集距離単位（主に形態素基本形と品詞）列の編集距離計算に基づく方式を提案し、個々の文や段落など局所的範囲での評価手法の有効性を確認した。要約筆記テキスト間の関連性や要約筆記者のタイプやくせを見つけ出すため、多次元尺度法や潜在意味解析を利用した評価を導入した。また、要約評価計算の過程で得られた情報からよりよい要約テキストとなるような書き換え候補を提示する機能を試験的に導入した。これにより、要約筆記者がよりよい要約ができるようになり、要約筆記者の養成支援になることを目指している。

キーワード：要約筆記、聴覚障害者支援、情報保障、ユニバーサルデザイン

1. はじめに

聴覚障害者や高齢者への情報保障手段のひとつに要約筆記がある。要約筆記には「PC 要約筆記」と「手書き要約筆記」があり、いずれも要約筆記者が講演や番組などを聞き取り、リアルタイムで要約を行ない、キーボードや手書きで入力する。一般に日本語の発話速度は200～400文字/分であり、要約筆記者による入力量はPCの場合で100～200文字/分、手書きの場合で40～80文字/分となっている。要約筆記者は「速く」、「正確に」、「読みやすく」の3原則をもとに、技術の向上を目指してさまざまな研修プログラムで訓練を重ねる。個々の研修プログラムでは要約筆記の品質の尺度として、要約筆記利用者からのフィードバックや意見・要望を受けることが多い[1][2]。これらのフィードバックは個々の事例として受けることが多く、定量的な品質評価を受けることはほとんどなかった。そのため、長期間の研修を経ても要約筆記の品質向上の実感が得られにくくなっていた。これまで筆者らは講演者の発話内容のテキストと要約筆記者が入力したテキストをもとに定量的な評価ができるシステムを試作した[3][4][5]。要約評価となる評価値計算に

は重み付き編集距離単位（主に形態素基本形と品詞）列の編集距離（Levenshtein Distance）計算に基づく方式を提案した。これにより個々の文や段落など局所的範囲での評価に有効性を確認できた。さらに、要約筆記されたテキストどうしの関連性や要約筆記者それぞれのタイプやくせを見つけ出すために多次元尺度法や潜在意味解析などの評価も行なえるようにした[4]。本論文では、要約評価値計算の過程で得られた情報をもとに要約筆記者が書き下したテキストについて、よりよい要約テキストとなるような書き換え候補を提示し[5]、さらによりよい要約になっていることを確認できる機能を実験的に導入した。これにより、要約筆記者が指針やアドバイスを受けることでよりよい要約ができるようになり、要約筆記者の養成支援になることを目指している。また、コンピュータによる自動要約システム[6]への知見が得られると考える。

2. 要約筆記データ

今回、要約筆記データ収集のために使用した発話テキスト（各4分の観光ガイド3コースの音声データを文字化、T1、T2、T3 とする）と要約筆記者がリアルタイム要約筆記（同一発話について休憩をはさんで2回要約筆記）した要約筆記テキスト（手書き要約筆記8名(H01～H08)、PC要約筆記11名(P01～P11)）を利用した。

表1. 要約筆記テキストと要約評価(1回目と2回目)

発話	筆記者	文字数		入力速度 文字数/分	要約率 (%)	1回目			2回目		
		1回目	2回目			編集距離		要約評価	編集距離		要約評価
		文字列	形態素			文字列	形態素				
手書き	H01	286	292	74.1	29.7%	0.3628	0.4420	0.4067	0.3418	0.4218	0.3664
	H02	238	238	61.0	24.5%	0.2714	0.3569	0.3318	0.2253	0.3747	0.3015
	H03	238	235	60.6	24.3%	0.3153	0.4131	0.3529	0.2968	0.4312	0.3605
	H04	297	290	75.3	30.2%	0.3214	0.4062	0.3888	0.3323	0.4144	0.3997
	H05	285	288	73.5	29.5%	0.2824	0.3234	0.3046	0.3229	0.4446	0.3744
	H06	270	260	67.9	27.3%	0.2879	0.3574	0.3224	0.2949	0.3723	0.3222
	H07	229	230	58.8	23.6%	0.2414	0.2471	0.2382	0.2612	0.3208	0.2941
	H08	286	275	71.9	28.9%	0.3111	0.3348	0.3174	0.2948	0.4327	0.3566
PC	P01	533	582	142.9	57.4%	0.5884	0.6456	0.6346	0.6022	0.6967	0.6598
	P02	492	553	134.0	53.8%	0.5544	0.5919	0.5719	0.6155	0.6465	0.6063
	P03	370	413	100.4	40.3%	0.4011	0.5093	0.4606	0.4130	0.5418	0.5011
	P04	497	492	126.8	50.9%	0.5854	0.5867	0.5947	0.5831	0.6592	0.6246
	P05	592	721	168.3	67.5%	0.6185	0.6301	0.6224	0.6668	0.6942	0.7136
	P06	411	421	106.7	42.8%	0.4419	0.5038	0.4661	0.5015	0.5463	0.5055
	P07	457	451	116.4	46.7%	0.4466	0.5091	0.4680	0.5186	0.5615	0.5527
	P08	413	349	97.7	39.2%	0.4269	0.5386	0.4941	0.4291	0.5130	0.4764
	P09	360	402	97.7	39.2%	0.3920	0.4648	0.4189	0.3987	0.4963	0.4622
	P10	462	519	125.8	50.5%	0.5104	0.5083	0.4831	0.5342	0.5592	0.5354
	P11	756	827	202.9	81.4%	0.6953	0.6947	0.6935	0.7079	0.7855	0.7618
手書き平均		266.1	263.5	67.9	27.2%	0.2992	0.3601	0.3329	0.2963	0.4016	0.3469
PC平均		485.7	520.9	129.1	51.8%	0.5146	0.5621	0.5371	0.5428	0.6091	0.5818

T1 の発話のみについて、表1に発話テキストと要約筆記テキストの詳細データ(入力文字数、入力速度、要約率)を載せた。また、文字列および形態素の編集距離に基づく要約評価結果および本アルゴリズムによる要約評価を表の右側に示した(要約評価の詳細は後述)。文字数には句読点や記号(矢印記号「→」、項目を表わす中黒「・」)、繰り返し記号(「〃」など)、削除記号(訂正線)などを含めている。手書き要約筆記データを計算機可読テキストにする際には二次元的な表現や複数行にわたる括弧記号などの意図がわかるようにXMLタグ付きテキストにした。図1に1

回目と2回目の要約評価の変化を示す。おおむね1回目よりは2回目のほうがよく要約できているが、要約筆記者19名のうち3名については逆に低下している。別の発話T2、T3の場合の要約評価を調べてみたところ、手書き要約筆記、PC要約筆記にかかわらずこのようなことが起こる可能性があることがわかった。つまり、同じ発話を繰り返して要約筆記した場合、2回目以降はおおむねよくなる方向であるが、場合によっては低下する場合もあるということである。

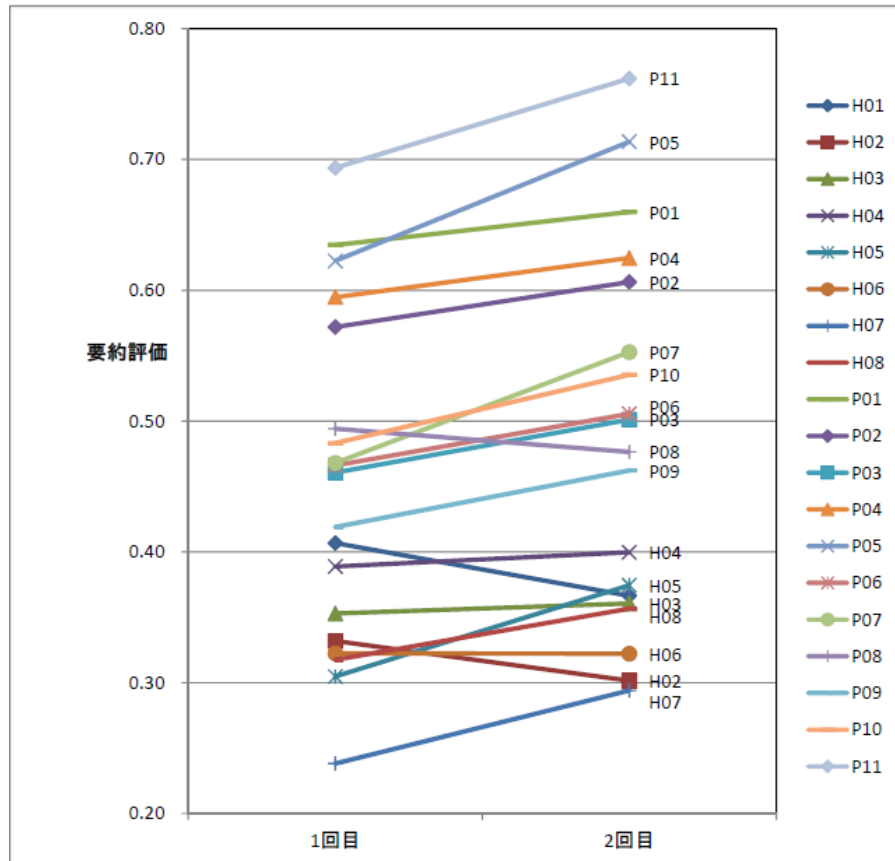
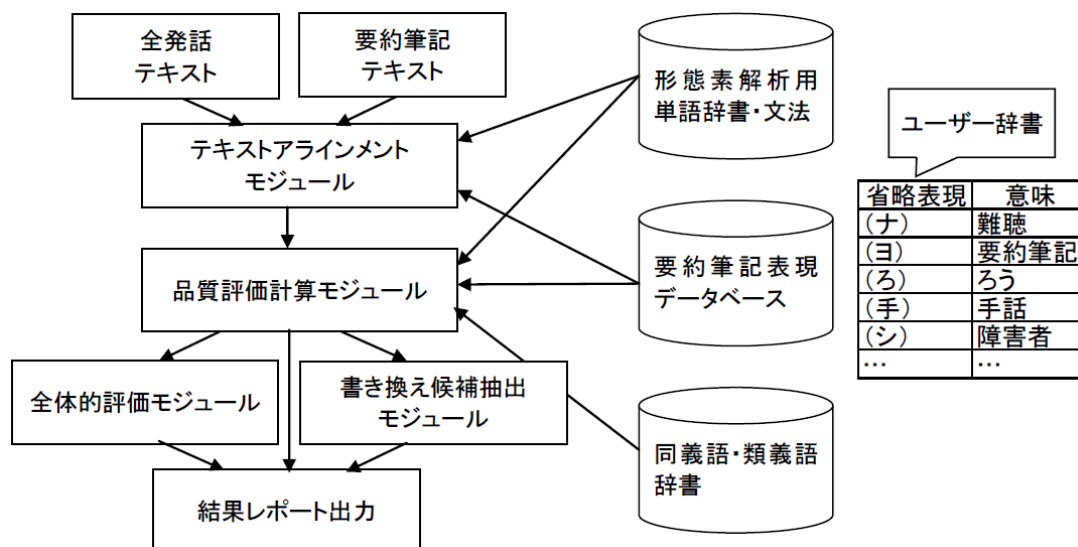


図1. 要約筆記1回目と2回目の要約評価の変化(発話T1)

3. 要約筆記品質評価システム

本システムのメインはテキストアライメントモジュールと品質評価計算モジュールから構成される(図2システム構成図参照)。テキストアライメントモジュールは全発話テキストと要約筆記テキストを入力とし、統計情報と言語情報をもとに動的計画法を利用して対応する文や段落を対応づけるモジュールである(m文対n文)[7]。アライメント単位ごとに発話テキストと要約筆記テキストのペアが作成される。これにより品質評価計算対象範囲を狭くすることにより、後段の品質評価計算モジュールにおける評価計算精度を高めることができる。品質評価計算モジュールは、表記ゆれや要約筆記特有の省略表現などを吸収して正規化した形態素解析結果の形態素列に対し、単語コスト、品詞コスト、単語間接続コスト、重複出現コスト(出現のたびに単調減少)を統計処理することにより要約の品質評価の計算を行なう。形態素解析ツールにはChasenとMeCabを利用した。



省略表現	意味
(ナ)	難聴
(ヨ)	要約筆記
(ろ)	ろう
(手)	手話
(シ)	障害者
...	...

図2. システム構成図

また、発話テキスト・要約筆記テキストについて、テキスト全体およびアラインメント単位で形態素列の N-gram に基づいた多次元尺度法と潜在意味解析を行なう。これらの処理には、統計解析用ソフトウェア R とそのライブラリ群を利用した。

多次元尺度法は、個体間の類似度をそれぞれの個体から得られる多変量の数値をもとに個体間距離を計算し、類似性の高いものが近くに配置するようになる方式である。多元尺度法では、個体間の距離座標値を求める方法として計量的方法と非計量的方法がある。本実験では、単語頻度や単語接続頻度の数値から個体間の距離を計算するので、非計量的方法に基づいて発話テキスト・要約筆記テキストの座標を求めた。

潜在意味解析は単語の持つ多義性と多様性に対処した統計的技法であり、多変量ゆえに高次元となる単語文書行列を低次元に圧縮して近似的に表現する。これにより関連性の強い単語どうしが強調され、文書間の意味的類似度が明確になる。潜在意味解析では、ユニグラムをベースに TF (Term Frequency、索引語頻度)・IDF (Inverse Document Frequency、文書頻度逆数) の重み付けと正規化を行なった単語文書行列をもとに特異値分解によって次元を圧縮している。ここでは 3 次元に圧縮して発話テキスト・要約筆記テキストの座標値を求めることにした。

要約評価計算では、テキストを形態素列と見なし、0~1 の重み付き編集単位要素の編集距離 (Edit Distance または Levenshtein Distance) を利用した。編集距離とは列 A と列 B について、A を編集操作(削除、挿入、置換)して B にするときの必要最低限の操作数のことである。操作数を、文字列編集距離の場合は比較元テキストの文字列長、形態素編集距離の場合は比較元テキストの形態素数で割り、数値の範囲を 0~1 に正規化した数値とした。0 に近ければ元のテキストとの相違が多く、1 に近ければ相違が少なくなる。要約評価には重み付き操作数を比較元テキストの形態素数で割り、数値の範囲を 0~1 に正規化した数値をベースにした。表 2 に要約評価の評価値計算の例を示す。表において、列(横)方向は発話テキストの各形態素、行(縦)方向は要約筆記テキストの各形態素を配置する。コスト値の欄は、形態素解析用辞書に格納されている形態素コストが初期値となり、これが編集距離における重み付き操作数に相当する。各セル値 E_{ij} の計算は次の

計算式で行なう。

$$E_{i,j} = \min(E_{i-1,j} + C_{i-1}/C, E_{i,j-1} + C_{j-1}/C, E_{i-1,j-1} + A)$$

$$A = \begin{cases} 0 & : i-1 \text{ と } j-1 \text{ の位置の形態素がマッチ} \\ & \text{(表記基本形、品詞、同義語) した場合} \\ (C_{i-1} + C_{j-1})/C & : \text{上記以外 (C:コスト値の総和)} \end{cases}$$

全セルについての計算の結果、表の最右下のセル値を1から引いた値が要約筆記テキスト全体の要約評価となり、この値が1に近いほどよい要約になっていることになる。

要約候補テキストの抽出は、要約評価を算出するマトリクス (表2) において、最右下のセルから最左上のセルまで評価値が最も小さくなる方向 (上方、左方、左上方のいずれか) に順次たどることで発話テキストと要約筆記テキストとの対応セルを求める。次に、発話テキストと要約筆記テキストとの対応関係のうち相互にマッチしないもの (前後のセル間で評価値の差が大きい場合) を抽出する。表2では、

- ・「急性中耳炎は」(発話) と文頭 (要約筆記)
- ・「で起こる。」(発話) と「で」(要約筆記)

が該当する。この際、直前のセルの評価値との差が大きいもの、発話テキスト内の形態素コストの大きいものから優先的に要約候補テキストを提示する。なお、候補テキストの提示における修正箇所は自立語を含む文節単位とし、文章としての形態素間のつながりが保たれるようにした。文節境界の判定には構文解析ツール CaboCha を利用した。

表2. 要約評価計算と要約候補抽出

i \ j		急性	中耳炎	は	風邪	が	原因	で	起こる	。	
		コスト値	3135	4000	10	2914	10	2326	10	2206	188
コスト値	0.000	0.156	0.356	0.356	0.502	0.502	0.618	0.618	0.728	0.738	
カゼ	2914	0.145	0.302	0.501	0.502	0.356	0.357	0.473	0.473	0.583	0.593
が	10	0.146	0.302	0.502	0.502	0.357	0.356	0.472	0.473	0.583	0.592
原因	2326	0.262	0.418	0.617	0.618	0.473	0.472	0.356	0.357	0.467	0.476
で	10	0.262	0.419	0.618	0.618	0.473	0.473	0.357	0.356	0.466	0.476

4. 実験結果

発話テキスト T1 について要約筆記者 (H01~H08, P01~P11) ごとの要約筆記テキストを本システムに適用した結果を表1の右側部分に示した(1回目の要約筆記と2回目の要約筆記)。手書き要約筆記より PC 要約筆記のほうが全体的に要約評価が高いことがわかる。また、全体的評価として多次元尺度法と潜在意味解析の結果をそれぞれ図3と図4に示す。

両方の図から、手書き要約筆記(H01~H08)より PC 要約筆記(P01~P11)のほうが発話テキスト T1 に近いので、要約筆記には PC 要約筆記が有利なことがわかる。図3では手書き要約筆記グループと PC 要約筆記グループに境界線を付加してみた。多次元尺度法でははっきりわからないが、潜在意味解析結果からは手書き要約筆記より PC 要約筆記のほうがグループ化しやすい傾向にあることがわかる。これは PC 要約筆記においては PC による日本語入力システムの性能にある程度

依存した表現になりがちになることから納得できる。実際、図4では要約評価が高めのグループ(P01、P02、P04、P05、P10、P11)とそうでないグループ(P03、P06、P07、P08、P09)に分かれる傾向にあった。

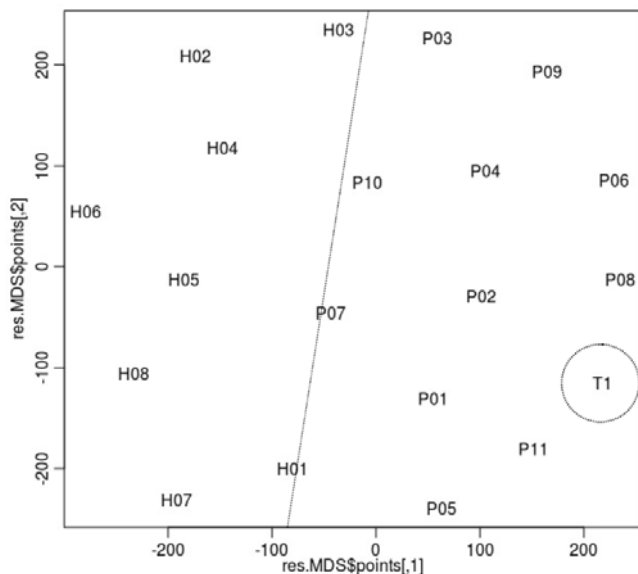


図3. 多次元尺度法による解析結果

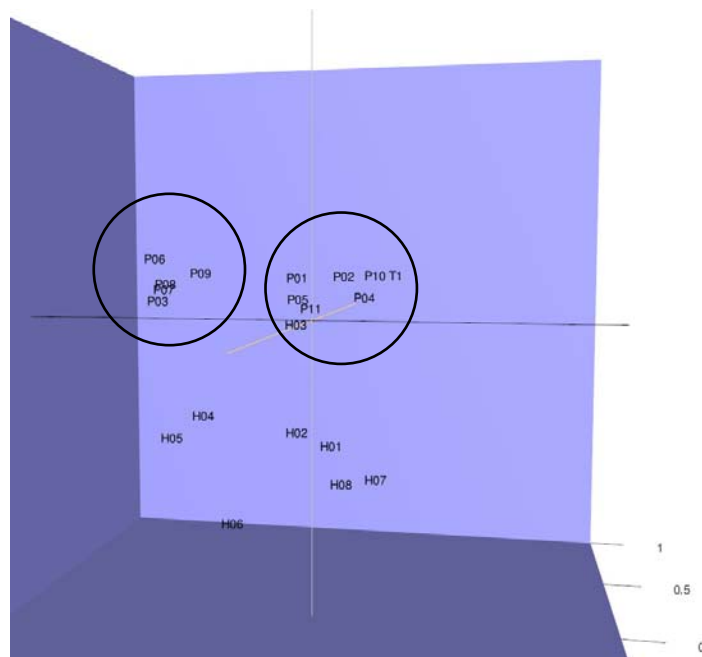


図4. 潜在意味解析による解析結果

5. 要約候補テキスト提示機能の試み

筆者ごとの要約筆記テキストを本システムに適用した結果を表3の右側に示す。ここでは予備調査時に収集した要約筆記データ（エッセイ文（約6分）、発話テキストS、要約筆記テキストK1～K6とする）について行なっている。各筆者が書き下したままの要約筆記テキスト（要約筆記原データ）の要約評価とともに、アラインメント単位ごとの要約候補テキストの提示にしたがい、1箇所および2箇所の修正をほどこした場合の要約評価の向上率も示した。要約評価向上率の変化を図5に示した。要約筆記原データの要約評価が低い場合（要約筆記に十分習熟していない場合）ほど向上率が高くなることがわかる。要約候補テキスト提示による効果は、要約筆記の初心者ほど効果が高いことになり、学習効果や訓練効果が大きいことがわかる。

表3. 要約筆記テキストと要約評価の向上

	文字数	文字数/分	要約率 (%)	要約評価の変化(全発話との比較)			
				要約筆記原データ	各単位最大1箇所修正	各単位最大2箇所修正	向上率
全発話(S)	1808	226.0					
筆者1(K1)	545	68.1	30.14%	0.6649	0.7485 (10)	0.7718 (3)	0.1069
筆者2(K2)	434	54.3	24.00%	0.6635	0.7417 (8)	0.7889 (3)	0.1254
筆者3(K3)	539	67.4	29.81%	0.5625	0.6867 (8)	0.7549 (6)	0.1924
筆者4(K4)	665	83.1	36.78%	0.7783	0.8328 (7)	0.8496 (4)	0.0713
筆者5(K5)	395	49.4	21.85%	0.5101	0.5624 (4)	0.6491 (4)	0.1390
筆者6(K6)	443	55.4	24.50%	0.4512	0.5454 (6)	0.6067 (6)	0.1555

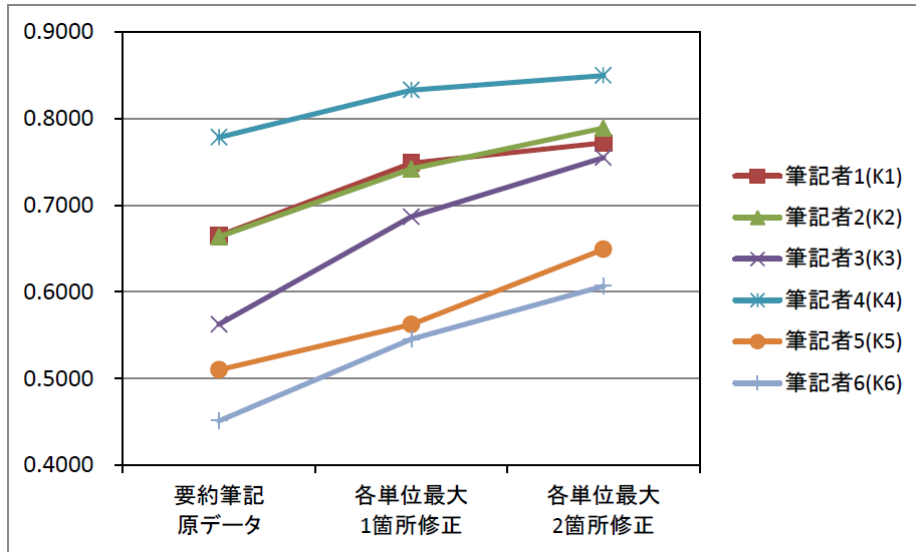


図5. 要約候補テキスト選択時の要約評価

要約筆記文修正候補の例を下記に示す。要約筆記文に対して、下線部分が候補として提示された（下線部分が第1候補、二重下線部分が第2候補）。

発話文：「だれかに自分の話を聞いてもらうことを望んでいます。」

要約文：「聞いてもらうことを望む。」

候補文1：「自分の話を聞いてもらうことを望む。」

候補文2：「だれかに 自分の話を聞いてもらうことを望む。」

発話文：「えー少しでも気になる症状があれば、ぜひ一度、耳鼻咽喉科を受診なさってください。」

要約文：「気になる時は受診を。」

候補文1：「気になる時は耳鼻咽喉科を受診を。」

候補文2：「気になる症状時は耳鼻咽喉科を受診を。」

要約筆記文修正候補失敗例を下記に示す。二重否定文や形容動詞・副詞の誤用などがある。

発話：「人の話を聞くのは、そう簡単なことではないのです。」

要約：「人の話を聞くのは難しい」

候補1：「人の話を聞くのは簡単なこと」(×)

候補2：「人の話を聞くのは簡単なこと ではない」

発話：「話をよく聞いてくれる人が意外に少ないことに気がつくはずです。」

要約：「そんなに多くはいない。」

候補1：「そんなに多くはいない、気がつくはず。」

候補2：「そんなに多く意外にはいない、気がつくはず。」(△)

次に向上率が高かった K3 と K6 について、改良による効果を多次元尺度法にて図 6 に示した (例えば、K3 は要約筆記原データのプロットを示し、K31 は原データに対して 1 箇所を修正、K32 は 2 箇所を修正した後のプロットを示す)。1 箇所の修正で発話文 S に大きく近づき、2 箇所の修正でさらに近づくことが明確になった。さらに、2 箇所の修正前後の各要約筆記者の類似度の潜在意味解析による結果を図 7 に示した。これも 2 箇所の修正後は発話文 S の近傍によりまとまりつつあることがわかる。

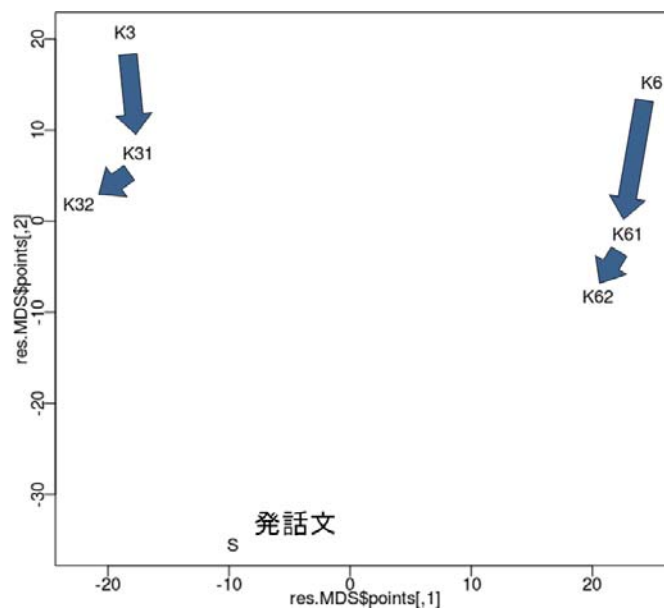


図 6. 多次元尺度法による解析結果 (1 箇所および 2 箇所修正時の変化)

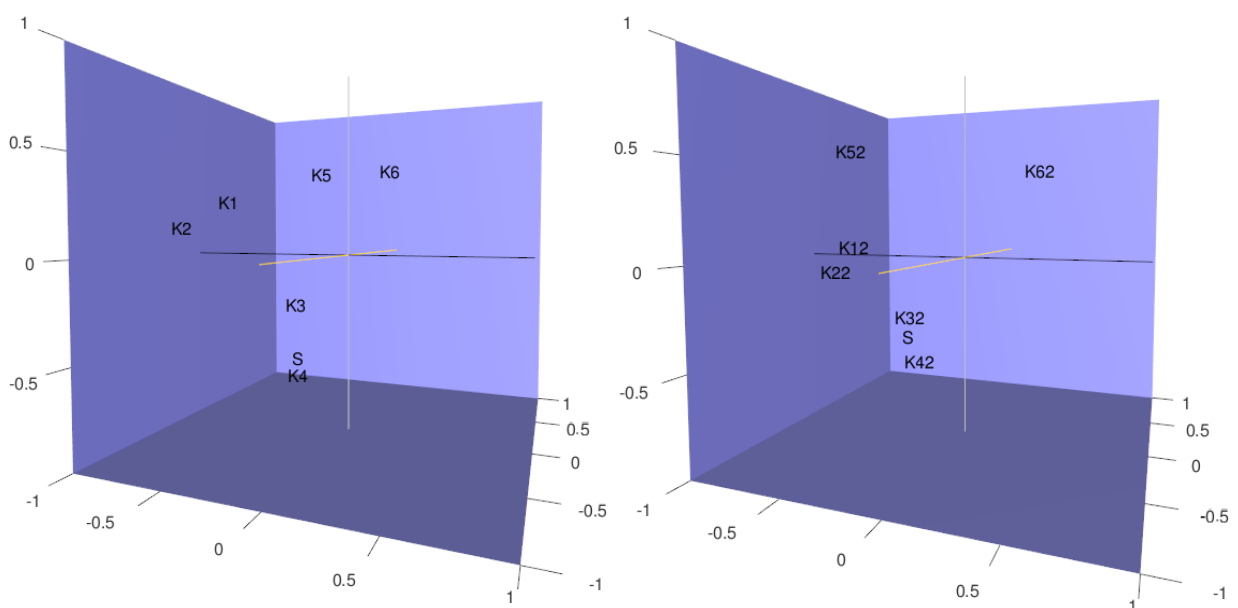


図 7. 潜在意味解析による解析結果 (右側は 2 箇所修正時の解析結果)

5. まとめ

今後は、さまざまな要約筆記データについて、要約評価精度の向上や失敗箇所についての分析を進めていくとともに、下記の実例（発話テキスト→要約筆記テキスト）のように、柔軟な文節配置に対応するために構文情報を利用した評価計算の改良などを行なっていく。「読みやすさ」の判定については、テキスト解析コストの数値を利用していく。さらに、複数人での要約筆記テキストを自動マージしてよりよい要約筆記テキストを構築するなどの応用も検討していく。

- ・「急性中耳炎は風邪が原因」→「原因はかぜ」
- ・「話すことより聞くことが重要」→「聞くことが話すことより重要」
- ・「2つの耳と1つの口」→「1つの口、2つの耳」
- ・「人の話を聞くのは簡単なことではない」→「人の話を聞くのは難しい」
- ・「この銅像が建てられたのは明治31年です。」→「明治31年、立像。」
- ・「製作者は高村光雲です。」→「高村耕雲作。」
- ・「新聞連載が始まったのは1946年4月です。」→「新聞は1940年4月に連さいスタート。」

参考文献

- [1]要約筆記再履修資料、名古屋市登録要約筆記者の会編 (2007)
- [2]話しことばの要約、三宅初穂、全国要約筆記問題研究会 (2012)
- [3]高尾哲康、要約筆記品質評価システムの改良、FIT2011、3Q-5、(2011)
- [4]高尾哲康、要約筆記品質評価システムにおける書き手のタイプ判別、IPSJ74 全国大会、3F-4、(2012)
- [5]高尾哲康、要約筆記品質評価システムにおける要約候補文提示機能、FIT2012、2M-6、(2012)
- [6]特集 テキスト自動要約、情報処理、Vol.43、No.12 (2002)
- [7]高尾哲康、対訳テキストコーパスからの対訳語情報の自動抽出、情報処理学会自然言語処理研究会、115-8、(1996)