The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

臨床心理面接における傾聴度変化の評価ー臨床心理士と初学者の比較ー

Evaluation of active listening in psychotherapy: Comparison of clinical psychologists and students

花田里欧子[*1] 中島隆太郎[*2] 井上 雅史[*3] 古山 宣洋[*4] 入野 俊夫[*5]
Ryoko Hanada[*1], Ryutaro Nakajima[*2], Masashi Inoue[*3], Nobuhiro Furuyama[*4], Toshio Irino[*5]

[*1] 東京女子大学 [*2] 東京大学大学院 [*3] 東北工業大学 [*4] 早稲田大学 [*5] 和歌山大学
[*1]Tokyo Woman's Christian University, [*2]Graduate school of the University of Tokyo, [*3]Tohoku Institute of Technology, [*4]Waseda University, [*5]Wakayama University

Active listening is one of the indispensable axes in evaluating the dialogue of psychotherapy. Although there have been discussions about it in the area of clinical psychology, the method for evaluating active listening has been missing. It is thus necessary to establish it to improve the quality of listening in the interview. The authors have proposed a measurement method of the degree of active listening with a device we originally developed to evaluate emotion (EMO system). This paper reports on the experiment conducted to compare the evaluations of a psychotherapy by expert clinical psychologists with those by undergraduate students as one of the coursework tasks for a clinical psychology course. A new experimental setup was proposed including a multiresolutional analysis to detect the change of active listening evaluation.

1. はじめに

臨床心理面接という対話の評価において，傾聴は欠くことのできない重要な軸の一つである．これまで傾聴を構成する要素に関する知見はあるが[アイビイ 1985]，面接全体として傾聴が真に成立したかどうかについて評価測定するための手法は，まだ十分確立していない．特に傾聴に際してそうしているつもりでも相手はそう感じていないといったすれ違いがつきまとうことはこの課題をいっそう困難にしている．しかし面接の質の把握や向上のためには，傾聴の評価測定をすすめていく必要がある．

本研究ではそのためのひとつのアプローチとして次の手順で実験を実施し，評価の実際と課題について明らかにする．(1)第三者の臨床心理士が面接ビデオを視聴し，感情評価値入力手法により傾聴度を時系列入力，(2)傾聴度時系列データの変化点（上昇／下降）を多重解像度分析によって自動検出，(3) 同じ臨床心理士によって，なぜ変化点をそのように評価したのかの記述，(4) その変化点が上昇か下降かの判断を初学者の集団で実験，(5)両者の評価の比較と統計的分析．

2. 実験設定

臨床面接における傾聴の度合いを判断する際，臨床心理士と心理コースを受講する学生においてどのような違いがあるかを調べるため，傾聴評価変化点に着目して実験を企画した．ここでは，臨床心理面接コーパス，実験使用データ，傾聴度評価，変化点自動抽出について説明し，それを元にした実験設定について述べる．

2.1 臨床心理面接コーパス

我々は臨床心理面接における，対話音声・身振り・表情といった社会的信号から，その特性に迫るために 44 対話からなる臨床心理面接コーパスを構築している[Inoue2012]．このコーパスは，生データとトランスクリプト及びアノテーションからなる．生データは，(1) 音声付き対話ビデオ，口元マイクによる高品位音声，頭部加速度データ等の生データ，(2) 発話内容の書き下しと ELAN による時間対応づけ，感情推移観測システム(EMO)による「傾聴度」評価データ(2.3 節参照)[花田 2017a]，コンプリメントラベル[井上 2018]，マイクロカウンセリングラベル[花田 2017b]等，多様なアノテーションが付与されている．現在も進行中でまだ全てが終了しているわけではないが，さらに必要に応じ別のアノテーション付与にも対応できる柔軟性を持たせてある．なお，このコーパス作成にあたっては，参加者から自身が記録されたビデオ映像・音声・身体動作データを，研究目的で分析し発表することへの同意を得ている．

2.2 実験刺激となる面接映像

臨床心理面接コーパスから，カール・ロジャーズによる「Miss Mun」[ロジャーズ 2007]の内容を元に模擬面接を行った 9 対話のうちランダムに一つを選出し実験に使用した(図 1，付録参照)．面接にはクライエント(Cl)として臨床心理士養成大学院に所属する 20 代女子大学院生 1 名が参加した．Cl の大学院生は臨床心理学の実習の一環として参加を希望した．これに対し，心理臨床歴 5 年の 30 代男性臨床心理士(Th)1 名が面接担当した．なお Th の面接における流派は家族療法／短期療法である．

図 1 評定対象の面接ビデオと，同時進行で入力される傾聴度評定値の表示．傾聴度は 100～-100 の値で保存．

連絡先：花田里欧子，東京女子大学現代教養学部，〒167-8585 東京都杉並区善福寺 2-6-1，電話番号&Fax 番号 03-5382-6419，hanada@lab.twcu.ac.jp

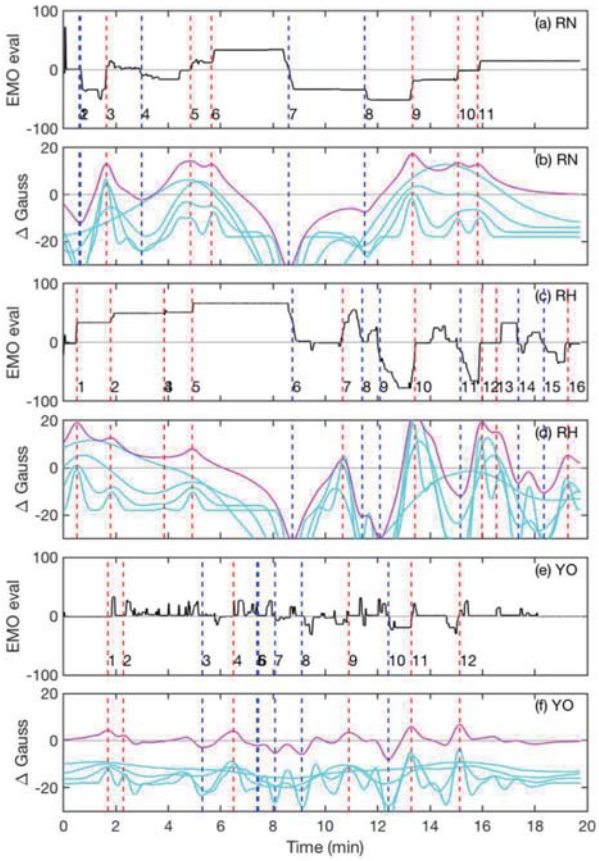

図 2 EMO による傾聴評価時系列データ (a)(c)(e)（黒線）と，変化点検出のための多重解像度分析結果(b)(d)(f)（マジェンタ:統合結果．シアン:解像度ごと）．横軸はビデオの時間．縦破線は検出された上昇変化点(赤)，下降変化点(青)．数字は変化点番号．

2.3 傾聴度評価

音声付き対話ビデオデータから傾聴度評価を行なった．まず，異なる臨床経験（5 年，10 年，5 年）を持つ 3 名の臨床心理士（RN，RH，YO）が，対象のビデオを見ながら，面接を行なっている臨床心理士(Th)がどの程度クライエント(Cl)に「傾聴」しているかを，第三者評価した．

このため，感情推移観測システム(EMO system)[八木 2013]を用いて，ビデオ映像と同期して評定値を時系列データとして入力した．EMO は，ハードウェアとして Arduino に接続したスライド抵抗のインターフェースを持ち，Processing で処理することにより，その時点ごとのつまみ位置に対応した評定値を図 1 のようにビデオ画面の右側に実時間で表示するシステムである．臨床心理士は，Th の傾聴度を判断し，中央をニュートラルとして傾聴～不傾聴の値をスライドバーの位置で入力する．この値は 100～-100 の間の数字で保存される．表示には，程度が分かるように正負両方に「非常に」，「かなり」，「やや」の補助語が入れてある．それぞれ(±100，±66，±33)の値に相当する．

図 2(a)(c)(e)に，それぞれ RN，RH，YO が付けた EMO 評価時系列データを示す．これを見ても分かるように，臨床心理士により評定の仕方が異なることが分かる．RN，RH は一定のところから大きく変動する場合が多く，YO はおそらく発話単位ごとに評定をしているようである．このようにお互いの傾向が異なることから，評定を平均化するのではなく，特定の評定者を基準として分析を行うことと，3 人のアノテータの時系列の生データから何らかの共通点／相違点を抽出することが必要となる．ここでは，傾聴評価値の変動が，対話の転換点等重要な質的変化と関連性があると仮定して，その変化点に着目することとした．

表 1 図 2(a)の RN の変化点の時間ならびに上昇(↑)下降(↓)と変動理由

#	時間	上下	変動理由
1	0’36”	↓	うなずきが小さすぎる上に途中からしなくなる．かつ視線が下を向いており聞いているのか分かりにくい上に寝ているようにも見える．
2	0’39”	↓	うなずきが小さすぎる上に途中からしなくなる．かつ視線が下を向いており聞いているのか分かりにくい上に寝ているようにも見える(2 と同じ理由)
3	1’38”	↑	目を見てジェスチャーを交えながら相手の発言に則して話そうとする様子から，話を受け取って返そうとする様子がうかがえる．
4	2’59”	↓	うなずきが曖昧で声が小さすぎる様が CL の発言を受け取っている様子に見えがたい．うなずきも CL の発言のペースに合っていないように感じる．
5	4’51”	↑	徐々にうなずきが CL の発言のタイミングと一致し始め，自身の発言では明瞭な語り口調で展開される様子から CL へ向き合う態度がうかがえる．
6	5’39”	↑	質問の中で CL が母親に対する対応について要約ができており，それを正確に伝え返していたため傾聴していた様子がうかがえる．
7	8’36”	↓	一方的に具体的なところまで介入の方針が展開されてしまい，もう少し「母親との間にリミットを設ける」という事に対してまず CL が行動していきたいのか，それを踏まえて展開した方が良かったのではないか．
8	11’30”	↓	回数が少なくて「焦る」ということは言っているが，時間かかるか否かは特に CL は言っておらず，「時間かかるだろうから焦る」というニュアンスで伝え返して質問してしまい一方的な問題意識が展開されてしまう．
9	13’20”	↑	CL が Th の発言に対して同意の言葉を述べ(確かに)，話が展開する様子から，ちゃんと Th が Cl の意図を汲めていたようにうかがえた．
10	15’04”	↑	リミットを設ける頻度について，CL が決めても良いし，Th との話し合いで決めても良いといった感じに決定における自由を保証してくれたことが良かった．
11	15’49”	↑	CL の問題意識を要約して今後の方針に結びつけて提案している様子が傾聴的な態度を感じさせた．

2.4 変化点自動検出

変動の解像度が事前には明確でない時系列データからの変化点検出のために，多重解像度分析を行った．ここでは，ガウス関数 G を一回微分した ΔG を，wavelet フィルタの母関数とした．このフィルタを EMO 評価時系列データに畳み込み，中心周波数ごとの平滑化微分を行なった．ここで，ΔG-wavelet 長を 64 秒，128 秒，256 秒，512 秒の 4 解像度とし， 1 分から 8 分程度の変動を抽出できるようにした．これは，傾聴評価時の時間遅れや過去の経緯による判断を含めて大局的な変化が見られるようにしたためである．64 秒より短い解像度を含めることも試したが，変化点の数が多くなりすぎてしまい，実験や評価の枠組に乗らないと判断した．

図 2(b)(d)(f)に各臨床心理士の ΔG-wavelet フィルタ出力を示す．シアン色が各解像度の出力で，それらを加算した結果がマジェンタの曲線である．この統合曲線の局所最大時刻を上昇変化点(赤破線)，局所最小時刻を下降変化点(青破線)に対応するとした．この結果，図 2 (a)(c)の目視で分かる特徴的な変化点を精度良く検出できていることが分かる．また，図 2 (e)では，傾聴評価入力値は細かく変化しているが，この多重解像度分析によって大局的な評価変動を取れているものと考えられる．

2.5 実験の目的と対象設定

前節の手順で，臨床心理士がつけた傾聴評価時系列データから自動抽出された変動の方向を，初学者がどのように判断す

		耳を傾けている評価
1)	36秒	↑ ↓
2)	38秒	↑ ↓
3)	1分38秒	↑ ↓
4)	2分59秒	↑ ↓
5)	4分51秒	↑ ↓
6)	5分39秒	↑ ↓
7)	8分36秒	↑ ↓
8)	11分30秒	↑ ↓
9)	13分20秒	↑ ↓
10)	15分4秒	↑ ↓
11)	15分49秒	↑ ↓

図 3　質問紙

るかを調べることが，今回の実験の目的である．また，この際，第三者の EMO 時系列データ作成者が，なぜそのように評価を変動させたのかの情報とも対比して検討したい．そこで，今回は表 1 のようにその理由の記述が得られた図 2(a)の RN の変化点に関して実験を行うことにした．変化の方向は上昇が 6 箇所(3,5,6,9,10,11)で下降が 5 箇所(1,2,4,7,8)であった．1,2 は時間の開きが比較的短く違いが見られないため変動理由は同じとなる．

2.6 実験方法

首都圏の女子大学心理系学科における臨床心理の授業の一環として，学部 2 年生 73 名(以下，初学者)を対象に実験を行なった．教室前方のスクリーンに，今回分析した動画(図 1)を再生し，質問紙(図 3)を用いて図 2(a)の変化点番号 1～11 の箇所で，上昇「↑」か下降「↓」で傾聴の変化方向の評定を求めた．この際，変化点時刻より 30 秒後ろまで再生したところで一時停止を行ない，評定時間を確保した．変化点番号(1～11)と時刻が記されており，その次に評定の選択肢が用意されている．上昇の場合は「↑」，下降の場合は「↓」に○をつけて評定する．

3. 実験結果

臨床心理士と初学者の評定の比較を行い(表 2)，2.6 節で得られた初学者の評定結果をクラスタ分析によって分類，デンドログラムを表示し(図 4)，類似変動があるかどうか図 2 と対比する．

3.1 臨床心理士と初学者の評定の比較

臨床心理士と初学者による評定を示す(表 2)．2.4 節で述べた，変化点の自動検出によって割り出された時刻の前 15 秒間から後 15 秒間にかけてどのように評定値が変動したかを臨床心理士による変動の方向性として定義した．また，上昇する際には前 15 秒間の中から最低値を，後 15 秒間の中から最高値を，下降する際には最低値と最高値を逆にして参照した．

初学者による評定について，該当箇所の 1～11 で上昇を選んだのは全回答中平均 63%(min=25%, max=87%)，下降を選んだのは全回答中平均 37%(min=13%, max=75%)であり，上昇の割合が高かったのは 11 問中 8 問(4, 5, 6, 7, 8, 9, 10, 11)だった．このことから全体的に比較的上昇を選ぶ傾向にあることが分かった．なお，上昇と評定される割合が高かった変化点は，1 位から順に 4(87%), 10(80%), 6(79%)，下降と評定される割合が高かった変化点は，1 位から順に 1(77%), 2(68%), 3(51%)と，前半部分で下降を選ぶ割合が比較的高いことが分かった．

臨床心理士が評定した変化の方向と，初学者の過半数の評定結果が一致したものは 1,2,5,6,9,10,11 で，一致する割合が高いのは順に 10(↑, 80%), 6(↑, 79%), 5(↑, 76%)であった．逆に過半数が不一致だったものは 3,4,7,8 で，不一致になった割合が高いのは順に 4(↓, 87%), 7(↓, 67%), 8(↓, 56%)であった(カッコ内の矢印は臨床心理士による評定)．一致率が高いものはいずれも上昇であり，低いものはいずれも下降であった．

表 2　臨床心理士と初学者による評定まとめ

		臨床心理士				初学者	
設問	時間	前	後	差分	変動	↑ a)	↓ a)
1	0'36"	1	-37	-37	↓	25 b)	75 b)
2	0'39"	1	-37	-37	↓	32	68
3	1'38"	-50	15	66	↑	49	51
4	2'59	4	-10	-14	↓	87	13
5	4'51"	-1	17	18	↑	76	24
6	5'39"	12	34	21	↑	79	21
7	8'36"	35	-33	-68	↓	67	33
8	11'30"	-34	-51	-17	↓	56	44
9	13'20"	-51	-18	33	↑	67	33
10	15'04"	-18	-1	17	↑	80	20
11	15'49"	-1	15	16	↑	73	24

a) ↑: 上昇, ↓:下降　b) 単位は%

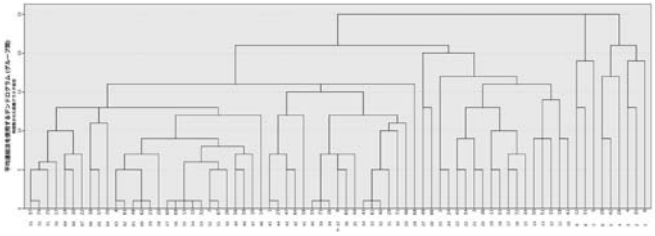

図 4　グループ間平均連結法によるデンドログラム．横軸は初学者，縦軸は評定パターン．

3.2 初学者のクラスタリング結果

初学者 73 名の全 11 問の評定をグループ間平均連結法によりクラスタリングし，おおよそどのような評定のパターンが見られるか検証し，さらに臨床心理士の評定がどれに類似するか比較する．なお，欠損値のある 1 名は分析から除いた．SPSS を使用してデンドログラムを作成し，距離 24 を閾値とすると，3 つのクラスタが得られたが，12,31,5 を含むクラスタと 39,42,28,4,35,6 を含むクラスタは人数が少ないため一つのクラスタに統合し，2 つのクラスタを得た(図 4)．第 1 クラスタは 64 名，第 2 クラスタは 9 名の初学者が含まれた．人数比の偏りを見るために χ^2 検定を行った．結果，有意な人数比の偏りが見られた(χ^2=41.44, df=1, p<.01)．

得られた 2 つのクラスタにおける評定の特徴を見るために，χ^2 検定と残差分析を用いて各設問における評定の偏りを検証した．結果，1(χ^2=4.65, df=1, p<.05), 2(χ^2=5.31, df=1, p<.05), 4(χ^2=5.27, df=1, p<.05), 6(χ^2=6.79, df=1, p<.01)，7(χ^2=14.59, df=1, p<.01)，9(χ^2=4.79, df=1, p<.05)，10(χ^2=7.71, df=1, p<.01)，11(χ^2=22.80, df=1, p<.01)で有意な差が見られたため，2 つのクラスタの特徴は以上の設問で検討した．

各クラスタにおける評定の特徴を人数比の偏りから見た残差分析の結果と，臨床心理士の評定を表 3 に示す．

表 3 臨床心理士と初学者クラスタにおける評定の傾向まとめ

設問	臨床心理士	初学者	
		第 1 クラスタ	第 2 クラスタ
1	↓	↓*	↑*
2	↓	↓*	↑*
4	↓	↑*	↑*
6	↑	↑**	↓**
7	↓	↑**	↓**
9	↑	↑*	↓*
10	↑	↑**	↓**
11	↑	↑**	↓**

※ ↑: 上昇, ↓: 下降, **: $p<.01$, *: $p<.05$

第 1,2 クラスタ間で差は 4 以外で見られた. 第 1 クラスタは 1,2 で下降し, 6,7,9,10,11 で上昇した. 第 2 クラスタは逆に 1,2 で上昇し, 6,7,9,10,11 で下降した. 臨床心理士の結果は第 1 クラスタに近く, 一致する設問がより多いことが分かった(1,2,6,9,10,11).

3.3 臨床心理士同士での相違

RN の変化点は, 第 1 クラスタのパターンと比較的類似していた. 初学者のほとんどが第 1 クラスタに含まれていたことからも(73 名中 64 名), このパターンは本面接の傾聴評価としておおまかな特徴のひとつを捉えているものと考えられる. 一方一定のところから大きく変動するという点では RN と共通する場合が多い RH については, 本実験では扱っていないために初学者との比較はただちにできないが, RN の変化点とおおよそ同位置の RN1-RH1, RN10-RH11 について見てみると, 評定結果は真逆になっている. つまり, 第 2 クラスタは RN とは反対の評定であるが, RH とは傾向が一致する.

このように専門家間でも評価はばらついている. 質的な理由の一つは, 面接を行なった Th との心理的距離もあるかもしれない. RH は指導的な立場にいた者で, RN は一度も直接会ったことがない者である. RH は Th の言動に対し, より積極的に「このようにすれば良いのに…」という内的な評価が働いた可能性がある. 面接終盤 RN は RN11 以降フラットに評定しているが, RH は所々下降させている. 傾聴に関して理論的, 技法的側面では一定の評価軸を持っていると想定される臨床心理士においても内的な影響を受ければ実際の評定はぶれてくる.

4. 本研究の成果と課題

本研究は, 臨床心理面接における傾聴度変化の評価における臨床心理士と初学者の比較を行った. このため, 従来に無い新しい手法を開発した. 実験の結果, 2 つの初学者の評定パターンが得られ, そのうち一方は臨床心理士との類似点が示された. 本結果は一人の臨床心理士による評定との差異を検討したものであり, 他の臨床心理士も同様ではない. このように, 傾聴評価の場合でも単純ではなく, 予測等の社会的信号処理を行う場合も, 表層的ないわゆる信号だけではなく, 内的な状態を定式化することは必須であろう. そのため今後, 基準となる評価者を複数人設定した比較検討が必要である. 感情評価は主観的であり, 社会的信号処理をする際, 参照する知識体系が異なる. たとえば本研究が示すように, 臨床心理面接の熟練のみが評価を一義的に規定するものではなく, それは正解を定めにくいということである. 感情評価の観点からエージェントを検討するにあたっては(たとえば傾聴対話エージェント[石田 2018]等), 振る舞いの個人差を考慮した上ですすめていく必要がある.

謝辞

本研究は JSPS 科研費 26380925, 15K13143, 16K00202, 人工知能研究振興財団および電気通信普及財団の助成を受けたものです.

参考文献

[アイビイ 1985] アレン・E・アイビイ, 福原真知子, 椙山喜代子, 国分久子, & 楡木満生(訳編): マイクロカウンセリング―“学ぶ-使う-教える”技法の統合, 川島書店, 1985.

[花田 2017a] 花田里欧子, 入野俊夫, 古山宣洋, 井上雅史, & 中島隆太郎: 臨床心理面接コーパスと感情推移観測システム(EMO system)を用いた傾聴学習支援, 信学技報, 116(436), 5-10, Jan. 2017.

[花田 2017b] 花田里欧子, 入野俊夫, 古山宣洋, 井上雅史, & 中島隆太郎: 感情推移観測システム(EMO system)による傾聴評価とマイクロカウンセリングのタグ付けとの関連, 信学技報, 116(524), 113-118, Mar. 2017.

[Inoue2012] M. Inoue, R. Hanada, N. Furuyama, T. Irino, T. Ichinomiya, & H. Massaki: Multimodal corpus for psychotherapeutic situation, The 8th Workshop on multimodal corpora LREC workshop on multimodal corpora for machine learning: How should multimodal corpora deal with the situation?, pp.18-21, Istanbul, Turkey, May 2012.

[井上 2018] 井上雅史, 中島隆太郎, 花田里欧子, 古山宣洋, & 入野俊夫: コンプリメントのアノテーション, 信学技報, 117(509), 11-15, Mar. 2018.

[石田 2018] 石田信也, 井上昂治, 中村静, 高梨克也, 河原達也:共感表出と発話促進のための聞き手応答を生成する傾聴対話システム, 人工知能学会研究会資料 SIG-SLUD-B509, pp. 7-12, Mar. 2018.

[ロジャーズ 2007] カール・ロジャーズ, 加藤久子 & 東口千津子(共訳), 畠瀬稔(監修): ロジャーズのカウンセリング(個人セラピー)の実際, コスモス・ライブラリー, 星雲社 , 2007.

[八木 2013] 八木みゆき, 森田礼子, 中井正人, 西村竜一, 河原英紀, & 入野俊夫: 説明対話における韻律情報および頭部運動と感情評価値の間連性分析の試み, 電子情報通信学会／音響学会音声研究会, SP2013-67, pp.15-20, Sept. 2013.

付録:面接の概要

面接の進行と対応する変化点(表 1)を()内に示す. Cl が人の期待に応えることについて, 自分のできないことはやらないよう「リミット」を設けた方が良いのではないかという気づきが語られ, Th は小さく頷きながら聞き始める(1,2). Th が「リミット」についてさらに説明を求めると(3), Cl の具体例を交えた説明を再び小さく頷きながら聞き始める(4). Cl の発言にタイミングを合わせて頷き初め明瞭な口調でまとめ返し(5), さらに Cl が語った母親との「リミット」に関するエピソードをまとめながらその実現可能性の是非を問う(6). Th から「リミット」を設ける頻度について助言がありつつ(7), そのためには時間がかかることへの懸念も Th が口にする(8). 急に「リミット」を設けて行動様式を変えることへの危惧を Th が語ると Cl はそれに同調し(9), そのあり方を Cl だけでも Th との話し合いでも決めるのはどちらでも良いという決定の自由を Th が伝え(10), 最後にそれを方針に面接を進める旨を共有する(11).