小津作品に見る物語映画における談話の時間特性

佐藤大和
東京工業大学　大学院情報理工学研究科

本論文は、小津安二郎監督の映画作品である「秋日和」を素材として、その中で交わされる会話の時間特性に関して分析したものである。映画は、まずショット毎に、また発話単位毎に区分化され、それぞれの時間長とアノテーション・タグから成る時間構造データが作成される。このデータに基づいて統計的分析を行った結果、発話開始時の Leading voice のタイミング区間、 Major juncture におけるポーズ長、および発話句の平均長はいずれも約1秒となった。 小津映画の談話においては、1秒の長さが談話リズムの基本的な役割を果たしている。また、Minor juncture におけるポーズ長や、話者間の Turn-taking に伴うポーズ長は0.5〜0.6秒程度が多かった。後者のうち、連続する二つのショットに渡る話者間ポーズは、フィルムの編集によって作りだされること、また話者交代に伴うポーズの0.5秒からのズレとモダリティー等との関連が議論される。

Temporal Characteristics of Discourse in One of Ozu's Narrative Movies

Hirokazu Sato
Graduate School of Information Science and Engineering
Tokyo Institute of Technology

This paper describes temporal characteristics of discourse taking the narrative movie film "Late Autumn" directed by Yasujiro Ozu. The film was segmented into shots and spoken units with durational information and other annotation tags. Various statistical analyses were conducted on the temporal data. The temporal interval of the leading voice at the beginning of an utterance, the interval of a major juncture pause and the average length of a spoken continuum are approximately 1.0 second in duration, which indicates that intervals of one second play important rhythmic roles within discourses of Ozu's film. The analyses also reveal that both average minor juncture pauses and inter-speaker utterance pauses are around 0.5 seconds. In the latter case, the short pause is frequently created by editing together two portrait shots. The functions of deviations from the 0.5 second inter-speaker rhythm are also discussed from the viewpoint of modality.

1．はじめに

1895 年、ルミエール兄弟がパリで「映画」を公開して以来、わずか 100 年ほどの間に、映画をはじめとする映像文芸は、世界の国々の一大文化資産となるに至った。映画文化の浸透によって、人間は「言語」とはまた異なった「映像によって語る」手段を手に入れたのである。このことは、人間の頭脳が、映像要素の時間系列から、ひとつの統合された物語を構成し、これを理解する能力を手に入れたということでもある。

映画は、娯楽あるいは芸術作品として楽しむために作られるものではあるが、それ自身また膨大なる知識資源でもある。それは、その映画が制作された古い時代の文化や生活の記録という意味ばかりでなく、「映像による物語」を支えている映画制作技術の集積体でもあるからである。映画フィルムは、一千あるいはそれ以上の離散的ショットが連結されて、それが「連続した流れ」として構成されている。ショット系列による映像ナレーションの技術は、カメラワークやフィルム編集など多くの原理に基づいている[1]。このような映画フィルム構成上の状況は、分節的単位が韻律などの超分節的特徴によって統合されるという音声言語の分析的スキームとよく似たところがある[2]。そのため、これらの原理を「映像文法」と呼ぶ研究者達もあるが、「文法」という用語が適切なものであるかどうかは議論のあるところであろう[3,4]。

また一方、映画は、大量の話し言葉コーパスでもある。これらの話し言葉は、シナリオ作家や監督らによって創作されたものであって、現実の発話資料とは異なるものであるけれども、実際の会話と比べると、注意深く創作された well-formed な会話ともなっており、話し言葉の研究や教育の資料として有用であると考えられる。

著者は、映画をひとつの「知識資源」と見て、そこから映画製作上の知見や原理を取り出すことを試みている。これまで、小津安二郎監督の映画作品を素材として、彼の映画スタイルの一断面を明らかにするため、ショットの時間特性と画面の空間構成の方法に関して報告した[5,6]。本報告は、その続編をなすものであって、映画の中の談話の時間構造や発話のリズムのなかに、小津映画の特性の一面を見つけ出し、加えて日本語の談話の時間的な特性をもそ

表１　小津映画における会話の時間構造データ例

(a) ポートレート(バスト)ショット系列の例

時間（秒）	ショットNO.	区間長（秒）	A1	A2	A3	A4	発話テキスト
1513.83	210	4.16	U	A	PA		
1518.00	210	1.12			S1		気がつきませんでしたか
1519.12	210	0.94			P	##	
1520.05	210	1.40			S1		さっきエレベータのとこで、
1521.45	210	1.60			S1		僕に挨拶したやつなんですがね
1523.05	210	0.26			P	<	
1523.31	211	0.30	U	D	P	>	
1523.61	211	1.07			S1	LV	ええ、
1524.68	211	2.00			S1		何ですかわたくしボンヤリしてて
1526.69	211	0.32			P	<	
1527.01	212	0.11	U	D	P	>	
1527.12	212	1.37			S1		僕よりチョッと低くて
1528.49	212	0.53			P	#	
1529.03	212	1.72			S1		髪はこうチョッと垂らしてて
1530.75	212	0.40			P	<	
1531.14	213	0.22	U	D	P	>	
1531.37	213	1.06			S1	LV	さあ
1532.42	213	1.62			S1		私ほんとにボンヤリで
1534.05	213	0.46			P	<	

(b) ミディアム-ロングショットの例

1559.16	216	0.10	AU	D	P	>	
1559.26	216	1.08			S1		面白すぎますよ
1560.35	216	1.17			P	##	
1561.51	216	0.85			S1		あいつにかかると、
1562.36	216	0.53			P	#	
1562.90	216	1.30			S1		何でも面白くなっちまう
1564.20	216	1.00			PL	< >	
1565.20	216	1.13			S2		いいですわ、
1566.33	216	0.63			P	#	
1566.96	216	1.79			S2		でもああいう方がいらっしゃると
1568.75	216	0.67			P	< >	
1569.42	216	1.67			S1		学校はどこを出たのか
1571.08	216	1.20			S1		よく覚えてませんけどね
1572.28	216	1.10			P	##	
1573.38	216	2.17			S1		うちの会社に入って四年かな、
1575.55	216	0.50			P	#	
1576.05	216	0.58			S1		五年かな
1576.63	216	1.17			P	##	
1577.80	216	2.63			S1		見たところそうガッチリした男でもありませんけどね
1580.43	216	1.50			P	##	
1581.93	216	2.34			S1		うちのバスケットのキャプテンしてて
1584.27	216	1.07			P	##	
1585.34	216	0.39			S1		どうです、
1585.73	216	0.61			P	#	
1586.34	216	0.59			S1		そんな奴
1586.93	216	0.37			P	<	

の中から探ることを目的としている。本論では、研究のため作成された映像データベースの内容と、物語映画における会話の時間特性を分析したこれまで得られた結果について報告する。

2．映画のおける会話とリズム

小津映画では、登場人物達の一定の調子の会話を通して物語が進行・展開していくので、彼の映画は、一口で言えば「会話のドラマ」という表現も可能であろう。一連の談話は、発話の音声部とポーズ（無音）部の交代によって成り立っており、その交代が会話にリズム的感覚をもたらす。このような談話の時間的な特性は、いわゆる「小津調」と呼ばれる彼の映画スタイルと密接に関係しているのではないかと考えられる。さらに、映画における発話は、監督の指示のもとで注意深くなされているので、日本語における談話の特徴的な一断面を示してくれるものと思われる。

一方、映画における談話の時間構造は、ショットの cutting-and-pasting の編集作業と不可分である。例えば、対話シーケンスにおいて、二つのポートレート（バスト）ショットが連結される際、先行人物の発話の終わりからそのショットの終わりまでのポーズと、次の登場人物のショットの発話開始までのポーズの和が、ショットを挟んでの会話のポーズ長となる。このようなポーズは、実際の発話におけるポーズではなく、フィルム編集者によって「人工的」に挿入されたものである。会話における時間的な不自然さを避け、滑らかな会話となるように調整されている。

本報告は、日本語の談話に関して、物語映画において見られる時間特性にのみ焦点を当てて分析したものである。映画の研究面からみても、これまで会話の時間特性の分析研究は見当たらない。本論は、そのひとつの試みである。

3．映像データベースの構築

小津作品の分析資料として、「秋日和」（1960 年公開）を取り上げた。これは、小津監督の後期のスタイルが確立された典型的作品のひとつと見なされるためである。この作品の映像・音声データは PC に取り込まれ、まずショット毎に区分化される。区分化されたショットは、映像と音声波形を同時表示しながら、さらに細かい音声区間とポーズ（無音）区間に分割される。各ショットと発話区間には、その区分の時間長と区分の特性を表すタグ情報が付与された。表 1 (a)(b)に、区分毎の時間構造データを示す。(a)は、正面向きの短いポートレートショットの例であり、(b)は、二人の場面を捉えたミディアムーロングの 1 ショットの例である。(a)では談話の turn-taking はショットの切り替えと同期するが、(b)では談話は同一ショット内で進行する。

表中、1 列目、2 列目、3 列目は、それぞれ経過時間、ショット番号、および当該区分の時間長である。横棒の実線で区切られた区分が 1 ショットに相当する。4～7 列目は、各区分のアノテーション用領域であり、各列の内容は以下のとおりである。

- 第 4 列(A1)：ショット種別
  （U：会話、A：動作、O：事物・風景、AU：動作→会話、等）
- 第 5 列(A2)：二つのショットの接続形式
  （U：会話による接続、A：アクションによる接続等（「立つ」「座る」などの動作は、しばしばショットの境界でなされる））
- 第 6 列(A3)：区分の種別（P：ポーズ、PA：動作を伴うポーズ、Si：当該ショットの i 番目の発話者の音声、等）
- 第 7 列(A4)：ポーズ等の種別を示す欄
  （内容は後述）

・第 8 列は発話内容のテキスト化部。

なお、今回の報告は、映画全体の前半、1 時間強の部分について分析したものである。小津の映画は全篇一定のテンポで作成されており、前半部のみでも彼の映画の特徴は把握できるであろう。

4．持続時間の基本統計量

最初に、映画における談話の音声部分とポーズ部分等の基本統計を示す。図 1 に示されているように、ポーズで区分化された音声部分の長さは、発話内容の長さに依存するが、3 秒程度までの範囲に分布し、0.5～1.5 秒の時間長が多い。平均は 1.2 秒である。一般に発話の音声連続は、二つ程度の音調区分（イントネーション句と呼ぶことにする）から成っている場合がある。このような音調区分は、ポーズを伴ってはいないが、発話の単位とみなされる。

（例）
「私の場合だったら母と一緒ですし」(2.16 秒)
→「私の場合だったら」＋「母と一緒ですし」
(1.36 秒)　　(0.81 秒)

音声部を、こうしたイントネーション句に再区分化して平均値を求めると、発話単位の平均は 1.1 秒程度となる。

句を構成するモーラ（拍）の長さは、6 モーラ以上の句で求めると、その平均は 0.11 秒／モーラである。それ故、この映画における発話単位の平均長は 1 秒程度であるから、10 モーラ／句程度の発話速度で話されていることになる。

ポーズ長の分布を図 2 に示す。この分布では、明確な動作を伴う無音区間は除外している。つまり言葉のやり取りのみに伴うポーズに限定した結果である。ポーズ長は、0.3～2.0 程度の範囲に分布するが、平均は 0.74 秒である。これは長いポーズと短いポーズの平均的な値となっている。このことについては、あらためて以下に述べる。

音声区間とポーズ区間の時間長の上記の結果は、表 2 にまとめて示す。

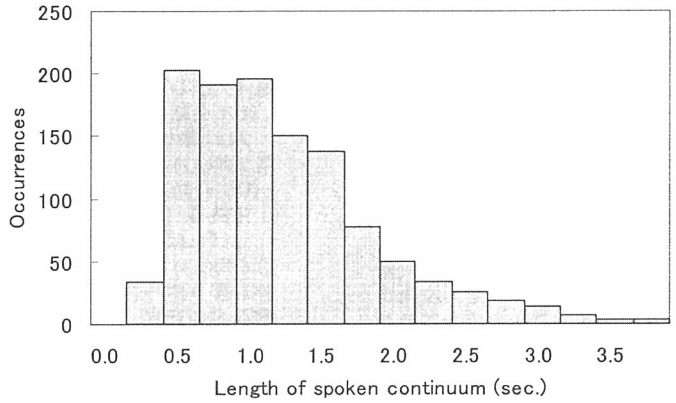

図１　会話における音声区間の時間長分布

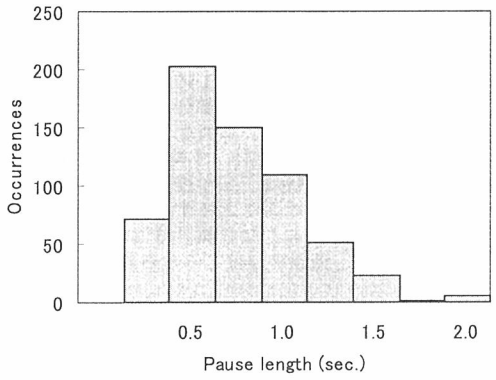

図２　ショット内発話のポーズ長分布

表２　談話における音声区間とポーズ長の統計量

(sec.)

	音声区間			ポーズ
	ポーズ間発話	イントネーション句	モーラ (6モーラ以上の句)	
平均長	1.20	1.13	0.113	0.74
標準偏差	0.67	0.57	0.028	0.34

５．ポーズ長等の分析

ポーズ分析をより詳細にするために、以下のような分類でポーズを分析した。カッコ内の記号は、表１の第７列(A4)に表示されている記号である。

（１）　話者内ポーズ
(1-1) Major juncture ポーズ：MJ (##)
(1-2) Minor juncture ポーズ：mj (#)
（２）　話者間ポーズ(Turn-taking ポーズ)
(2-1) ショット内ポーズ：(< >)
(2-2) ショット間ポーズ：
(<：ショット末尾、>：ショット開始部)

同一話者の発話のポーズが、Major juncture(MJ)であるか Minor juncture(mj)であるかは、ポーズを挟む前後の句が、まとまった発話として緊密性を持つか否かで判断された。原則的には、文末境界は Major juncture、句間境界は Minor juncture と判断されるが、実際の会話では、連続した短い二つの文が、一発話のように発声される場合がある。また、不完全な句のまま発話が終了する場合などもしばしば見られる。そのため、juncture の判断は、実際の２文、あるいは２句の発話の連続性等を考慮してなされた。

話者間ポーズは、話者間における発話権の移動(Turn-taking)の時間長である。ショット内の話者間ポーズ(< >)は、実際の会話での発話の受け渡しの時間であるが、ショット間に渡るポーズは、既に述べたように、ショット末ポーズ(<)と、次に続くショット開始時のポーズ(>)の和によって作られる。これらポーズ記号の実際の例は、表１に示されている。

図３と図４それぞれに、Major juncture ポーズと Minor juncture ポーズの長さの分布を示す。また、図５と図６は話者間ポーズ長の分布であるが、図５は同一ショット内の話者間ポーズ、図６は連続するショット間に跨る話者間ポーズである。なお、これらポーズの平均値と標準偏差を表３にまとめて示す。

表３に示すように、MJ のポーズは約 １ 秒であるが、mj のポーズはちょうどその半分の 0.5 秒であった。話者が交代する際(Turn-taking)の平均ポーズは、

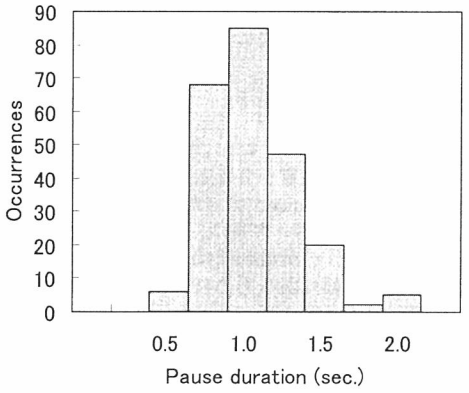

図３　Major juncture ポーズの時間長分布

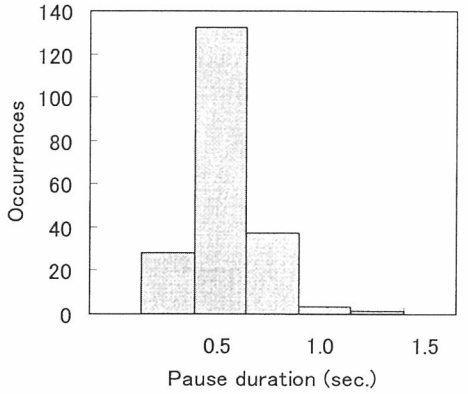

図４　Minor juncture ポーズの時間長分布

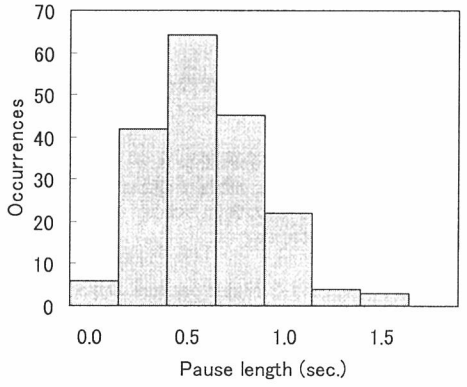

図５　ショット内話者交代に伴うポーズの時間長分布

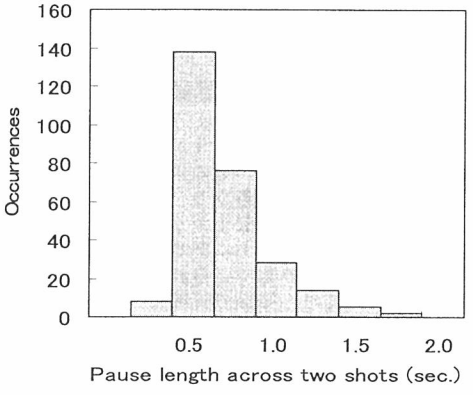

図６　ショット間に渡る話者交代に伴うポーズの時間長分布

表３　各種ポーズ長およびLeading Voiceタイミング長の統計量

(sec.)

	Major juncture	Minor juncture	ショット内 Turn-taking	ショット間 Turn-taking	Leading voice タイミング長
平均長	1.04	0.53	0.58	0.67	1.02
標準偏差	0.29	0.14	0.29	0.26	0.15

同一ショット内で話者が交代する場合 0.58 秒であり、ショット間に渡る場合は 0.67 秒であったが、どちらも 0.5 秒のピーク値を中心に分布しており、これは Minor juncture のポーズ長と一致している。表２の全体のポーズ長の平均 0.74 秒は、ショット内における略 1 秒長の長いポーズと、0.5〜0.6 秒程度の短いポーズすべての平均である。

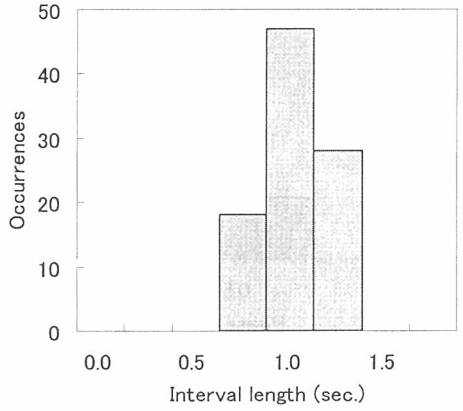

図７　Leading voice と後続句開始までの時間長分布

６．Leading Voice と話頭タイミング

会話における音声のどのような単位が談話のリズムと密接に関連するであろうか。まず、発話の音声連続区間とポーズ区間がその単位であり、二つの交代がリズムを生むであろうことは既に述べた。これとは別に、発話区分の立上がり部の時間間隔もまたリズム感覚と密接な関係にある。音声のタイミングは、発話の立上がり部でとられるからである。自然な会話においては、発話の冒頭で「えー」「あの」「じゃー」など、フィラー、呼びかけ、返答等の短い音声がしばしば使用される。これを Leading voice (LV)と呼び、その音声の開始時から、短い間（ま）を挟んで次の main phrase の発話の開始部に至る時間もまた会話音声のリズムに与っていると考えられる。小津映画における会話においても、以下のように、しばしば Leading voice が使われている。

（返答）
「ええ、高等学校もおんなじ寮でしてね」
「そうね、あしたも早いんですから」
（呼びかけ）
「ねえ、わたしキャラバンシューズ買いたいの」
「やあ、よく来たね」
（拒絶）
「いや、おじ様ふざけてらっしゃる」
（躊躇）
「え、あの、よろしいんです」
（フィラー）
「うむ、俺のみたところではね」

このような Leading voice は、発話のはじめにおいて、時間的調子（リズム）を整える役割があると考えることができる。表１における記号（LV）は、Leading voice と、それに続く発話句までの無音を含む区間の例を示している。

区間 LV の時間長の分布を、図７に示す。LV 区間長は、0.8～1.2 秒の範囲に分布しており、平均は 1.02 秒であった（表３）。

７．考察

小津映画「秋日和」における会話の発話長とポーズ長の統計的な性質を分析した。その結果、発話の最初の Leading voice、いったん発話を終える際の Major juncture ポーズ長は、約１秒であった。また、音声区間（句）の継続時間は、発話内容に依存するので時間長に幅はあるものの、これら二つの時間長と同様に平均は約１秒であった。小津映画では、このような約１秒の区分の繰り返しが、会話および映画のリズム形成上基本的な役割を担っていると推定される。このことは、映画全体に渡ってこうしたリズム構造が常時観測されるということではなく、１秒のリズムが、小津映画の基本特性としてその根底にあり、それがしばしば現れると見るべきであろう。

表４に、会話の時間特性の典型的な一例を示す。近似的に １秒長の音声やポーズの役割が見出せるであろう。
なお、筆者は既に小津映画のショット長等に関して報告したが、その中で、映画で使われているテーマ音楽に関しても、１拍の時間長が約 １秒であったことを触れておきたい[5,6]。

同一話者内における Major juncture のポーズ長は約１秒であったが、一方 Minor juncture でのポーズ長は、その半分の 0.5 秒である。1 秒のポーズは、リズムの単位となるとともに、1 発話をまとめる信号となっている。また、聞き手にとっては、話し手の発話の終了のサインともなっているであろう。一方、0.5 秒程度のポーズでは、間が挿入されて発話が切れたという印象はなく、むしろ引き続いて生ずる二つの発話に連続性を与え、両者を「繋ぐ」役割がある。つまり、連続して流れる会話を、1 秒の単位で切りつつ、0.5 秒の単位で繋げている。
次に、例えば 2 秒程度のさらに長いポーズの場合は、談話の流れを中断させてしまう。一般に、このように長いポーズの場合、登場人物に何らかの動作が伴うケースが多く、会話は途切れても動作によって映像の連続性が維持されている。以上説明したように、ここで分析された物語映画の会話におけるポーズは、会話の連続性に着目すると、大略以下のような三種類に分類できるであろう。

（１）　～0.5 秒：　　連続性無音区間
（２）　～1.0 秒：　　タイミング生成発話区分
（３）～1.5 秒以上：　中断（分離）ポーズ

上記は、Brown & Yule によって提案されている会話におけるポーズの 3 分類、short pause, long pause , extended pause に近い分類である[7]。

表4 映画の会話における音声長とポーズ長の例

ショット NO.	時間長（秒）	セグメントタグ		発話テキスト
321	0.4	P	>	
	1.1	S1		あそこの課長、
	1.0	S1		意地が悪いのよ
	0.2	P	<	
322	1.1	PA	>	
	0.8	S1		じゃ七人ね
	0.5	P	<	
323	0.2	P	>	
	0.2	S1		そう
	1.0	P	##	
	1.1	S1	LV	ねえ、
	1.1	S1		あたしキャラバンシューズ買いたいの
	0.4	P	#	
	0.8	S1		帰りに付き合ってよ
	0.4	P	<	
324	1.4	PA	>	
	0.6	S1		今日はだめ
	0.4	P	#	
	0.9	S1		約束しちゃったから
	0.3	P	<	
325	0.2	P	>	
	1.0	S1	LV	へええ
	0.5	S1		デート？
	1.0	P	##	
	1.1	S1		あんたでもそんあことあんの？
	0.3	P	<	

ここで述べた時間特性に基づいて、小津映画の会話の時間構造の例を図式的にまとめたものを、図８に示す。

この 0.5 秒程度の間は、話者の Turn-taking に際して生ずる間の時間長とほぼ同一であった(〜0.6 秒)。このことは、滑らかに、連続的に談話が進行するためには、Minor juncture でのポーズと同じ程度の Turn-taking 時間が適当であるということを示している。

ショット間に跨る話者間ポーズは、編集者によって人工的に「挿入された」ポーズであり、0.4〜0.6 秒程度が作り出されている。図６からも分かるように、0.4 秒よりも短いポーズは少なく、0.5 秒以上の長さに偏った非対称な分布となっているが、0.5 秒程度の長さが最も多い。画面フレームが順次切り替わっていく場合でも、この長さのポーズを与えることによって、観客に談話の流れをスムースに感じさせることができると考えられる。

映画における 0.5 秒の長さは、フレーム数でいうと 12 フレームである。小津映画の編集を担当していた浜村へのインタビューによれば、小津は彼に発話の終から 6〜8 フレーム後ろで、フィルムをカットして編集するように指示していたという[8]。ポーズ長が 0.5 秒と仮定すれば、この場合次に続くショットの頭の無音部は、4〜6 フレーム長で調整されていることになる。

話者交代に伴うポーズ長は、話者内の Minor juncture ポーズ長と比べて、平均値はほぼ同じでも、バラツキ（標準偏差）は倍程度大きい（SD=0.29 秒）。登場人物による対話の状況によって、Turn-taking の時間長が伸ばされたり、あるいは逆に短かくされたりしているためである。Turn-taking の時間が、平均的な一定の長さで談話が推移する場合は、話し手と聞き手の間で、特別な意図や感情の表現はない。一方、話者交代に伴うポーズの平均長 0.5 秒からのズレは、登場人物の感情や発話のモダリティと深く関係する。例えば、(Yes/No 疑問文)や（強い意図をもった発話権の取得）などが関わるとポーズ長は短くなり、（躊躇）や（怪訝、疑い）などが関わるときは長くなる。以下に例を示す。（下線を施した時間長が、話者間ポーズである）

(Yes/No 疑問文)
（S1）お前、ほんとうにあるのか？(0.1 秒)
（S2）あるんだ。
(意図的、反論的 Turn-taking)
（S1）本当にいいのよ、家柄だって。(0.2 秒)
（S2）家柄より本人よ。

（S1）ああ、長かった。(0.1 秒)
（S2）お前言うことないよ。遅く来やがって。
(躊躇)
（S1）どっかちょいと出かけましょうか(0.5 秒)
うまいものもないけど。（0.8 秒）
（S2）でもわたくし（0.5 秒）2 時からまた…。
(怪訝)
（S1）お薬、何をお買いになったの？（1.2 秒）
（S2）何？

以上の例から分かるように、談話の参加者間における発話権の遷移に伴う時間間隔は、発話者の感情や意図の表現と深く関わって変動するが、平均的には Minor juncture におけるポーズ長と同程度の値となっている。

我々はよく「間」という言葉を使用する。「間」は狭義ではポーズ長などの時間間隔のことであり、特に、能、歌舞伎、落語、舞踊などの伝統芸能や演芸に対して用いられている。小津映画では、発話ポーズや発話時間長、話頭タイミング等に、ある一定の時間的調子が見られるが、これらは彼の映画における固有の時間特性であり、ある種の「間」を反映したものであろう。ただこれは、平均的時間間隔の系列が生み出す neutral な特性にすぎない。既に述べたように、平均的時間特性からのズレが、登場人物の感情や意図の表現と関わっており、これが本当の「間」の問題であるのかもしれない。今後、会話だ