

論文 / 著書情報
Article / Book Information

題目(和文)	自然言語処理を利用したテキストマイニング手法に関する研究
Title(English)	
著者(和文)	寺田昭
Author(English)	
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第5470号, 授与年月日:2003年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第5470号, Conferred date:2003/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

自然言語処理を利用した
テキストマイニング手法に関する研究

寺田 昭

東京工業大学
情報理工学研究科 計算工学専攻

目次

第1章 序論	1
1.1 背景と目的	1
1.2 対象分野	2
1.3 システム概要	4
1.4 本論文の構成	6
第2章 略語の処理	9
2.1 略語の定義	9
2.2 略語の特徴	10
2.3 対象分野	11
2.3.1 「略語の多いコーパス」と「略語の少ないコーパス」	11
2.4 関連研究	12
2.5 システム設計	13
2.5.1 ターム重み付け	15
2.5.2 ランキング	15
2.5.3 復元語の選択	18
2.6 評価	19
2.6.1 訓練セットとテストセット	19
2.6.2 評価	19
2.6.3 先行研究との比較	21
2.7 議論	22
2.8 まとめ	24
第3章 名詞化構文の処理	27
3.1 名詞化構文	27
3.2 名詞化構文の特徴	28
3.2.1 名詞化構文の特徴	29
3.2.2 先行研究	29
3.3 方法	31
3.3.1 略語の復元	32
3.3.2 述語-項構造の抽出	32
3.3.3 名詞化構文の検出と変形	33
3.4 評価	36
3.4.1 名詞化構文変形	37
3.5 議論	39
3.5.1 構文解析器のエラー修正	39

3.5.2	移植性	39
3.6	まとめ	40
第4章	因果関係の検出	41
4.1	因果関係	41
4.1.1	因果関係の表現	41
4.1.2	単純なパターンの因果関係抽出	42
4.2	テキストマイニング	44
4.2.1	PrefixSpan	44
4.2.2	因果関係のマイニング	46
4.2.3	因果関係を表す動詞の発見	48
4.3	まとめ	49
第5章	時間関係の処理	51
5.1	はじめに	51
5.1.1	時間関係の推論	51
5.2	時間関係の処理	52
5.3	実験結果	54
5.4	まとめ	57
第6章	考察	59
6.1	自明な因果関係について	59
6.2	時間関係と節／文について	63
6.3	「有益」なパターン	64
6.4	まとめ	67
第7章	結論	71
7.1	まとめ	71
7.2	今後の課題	72
	謝辞	75
	参考文献	77
付録A	抽出結果	83
A.1	ASRS コーパス1	83
A.2	ASRS コーパス2	92

第1章 序論

1.1 背景と目的

計算機に人間と同じように言葉を理解させようとする試みは、計算機の出現以来行なわれてきたが、今だに成功していない。初期の試みは、人工知能的に人間の思考を模倣しようとするものであったが、そもそも人間の脳の仕組みが解明されていない状況では限界があった。近年、計算機の性能が記憶容量と処理速度の双方において飛躍的に向上し、さらに、計算機で処理できる形で蓄えられたデータが莫大な量になって来た。人間の脳の働きを模倣するのではなく、大量のデータを統計的に処理して、蓄えられたデータの中から自動的に人間が必要な知識や関係を得ようとする研究が盛んに行なわれてきている。

その1つとして情報検索 (IR : Information Retrieval) の研究がある [44, 63]。情報検索は、人間の質問に対して、その質問に近い文書を計算機が回答するものである。比較的簡単な仕組みで精度のよい結果が出力される事と、大量の文書を扱える事から、成功した技術である。ただし、質問する人間はどのような質問をすればよいのかという知識は持っていなければならない。また、検索システムの回答は文書そのものであるので、システムがデータ間の関係を自動的に取り出すことはできない。

関連する研究分野として情報抽出 (IE : Information Extraction) がある。情報抽出は分野ごとに抽出すべき情報を定義したテンプレートに従って自動的に情報を文書から取り出す技術である。情報抽出に関する国際会議 MUC (Message Understanding Conference) が過去に数回開催され、この技術の性能が競われた [35, 36, 37]。比較的分野が限定されていること、抽出する情報があらかじめ決められていることから、情報の抽出精度はかなり高いものである。しかしながら、文書から取り出せる関係はテンプレートで定義された関係に限られたものである。

データマイニングの研究は、Agrawal らが相関ルールを高速に求める Apriori というアルゴリズムを提案して以来、盛んに行なわれて来ている [1]。データマイニングとは、大量のデータの中から有益で新しいデータ間の関係を求める技術である。相関ルールは、

$$X \Rightarrow Y : c, s \quad X, Y \text{ はアイテムの集合}$$

で表され、 X を含むレコードは Y を含む可能性が大きいと解釈される。ここで、確信度 c (confidence) は、 X を含むレコードの中での Y を含むレコードの割合、サポート s (support) は、全体のレコードの中で X と Y を同時に含むレコードの割合と定義される。確信度とサポートがユーザの指定した閾値より大きな X と Y のペアを有益なパターンとして出力する。

自然言語で記述されたテキストにデータマイニングの技術を適用するテキストマイニングが最近注目され期待されている。データマイニングの技術は、構造化されたデータ（データベースなど）を対象にしているので、自然言語で記述されたテキストに直接適用することはできない。したがって、テキストマイニングは、何らかの前処理をテキストに施し、テキストを構造化されたデータに変換後、データマイニングの技術を適用することになる。自然言語で記述されたテキストを如何にしてデータマイニングの技術が適用できる構造化されたデータに変換するかが大きな

問題であり、それを行なう前処理によってシステムの処理能力や対象分野が大きく異なってくる。最も簡単な処理法は、前処理を単語レベルで行なう方法で、レコードとして接続した単語をとった場合は複合語の出力、文をとった場合には共起する語のペアの出力となるはずであるが、実際には意味のないパターンが大量に出力される。この原因は、単語の掛かり受けなどのテキストの持つ意味構造を全く考慮していないからである。初期の研究では、この単語レベルの研究を発展させ熟語（Phrase）を抽出するようなタスクや [3, 4], Feldman らのキーワード間の関係性を出力させる FACT というシステムが開発された [9]。FACT では、キーワード間の関係や属性をユーザが定義することが可能であった。

情報抽出の技術をテキストに適用して、あらかじめ決められたテンプレートの情報を取り出し、それにデータマイニングの技術を適用する手法もある。情報抽出の精度は比較的高いので抽出する関係が前もって決まっているような処理には適しているが、前に述べたように、テンプレートで定義された関係しか取り出せないという点と、扱うことのできる動詞がテンプレートで定義されたものだけで、その動詞とテンプレートに関するイベントしか抽出できないという点に限界がある [58]。情報抽出の技術そのものが、分野に依存するものなので、他の分野に移植するには、情報抽出の部分を作り直さなければならないという欠点もある。

自然言語処理技術が進歩した結果、テキストから単語の掛かり受けや句構造をある程度の精度で抽出できるようになって来ている。自然言語処理技術を利用してテキストを意味のあるまとまりに分割して、それに対してマイニングを行なう手法も考案されている [62]。

長いテキストでは、テキスト全体を対象にしないで、話題の中心部分を特定し、その部分から情報を抽出することが考えられる [14, 23]。この手法は、語の頻度、分布などの統計的な情報から話題の中心部分を特定しその部分から情報を抽出する技術である。

本論文では、汎用の構文解析器を使用して、英文のテキストを構文解析した解析結果から述語-項構造を自動的に抽出する。その結果から、イベントやイベント間の意味関係を抽出する。抽出する意味関係（Semantic Relation）としては重要な意味関係である因果関係（Cause-Effect Relation）を対象とする。イベント間の因果関係については、事故解析、医学、生物などの分野では重要なものであり、計算機による研究も盛んに行なわれている [30, 53, 57]。

1.2 対象分野

本論文では、テキストマイニングの対象として英文で記述された航空分野の安全レポートを取り上げるが、その背景について以下に述べる。

現在の航空機の事故率は、1992 年から 2001 年の 10 年間の定期航空旅客便の平均で 100 万出発あたり 1.18 回と他の輸送手段に比較して非常に低い¹。しかしながら、1992 年から 2001 年の 10 年間で航空機の出発回数が約 30% 増加していることを考慮すると、現在の事故率のままでは、将来、世界中で事故が頻繁に起こり、航空輸送が到底世の中に受け入れられないものになってしまうのではないかと危惧されている。

事故原因の約 66% は、乗務員が原因だとされている。したがって、事故の発生を減少させるには、人間に関わる要因を取り除かなければならない。1 つの事故の裏には、事故には至らなかったものの重大なトラブルが発生したインシデントが 100 あり、1 つのインシデントの裏には結果として何も起こらなかったが当事者が危険を感じたもの（ヒヤリハットと呼ぶ）が 100 あると云われている（図 1.1）。事故、インシデント、ヒヤリハットには共通の要因が含まれているので、

¹ Boeing の資料より (<http://www.boeing.com/>)。事故の定義は、飛行機的全損／大破又は飛行機による人間の死亡／重傷である。

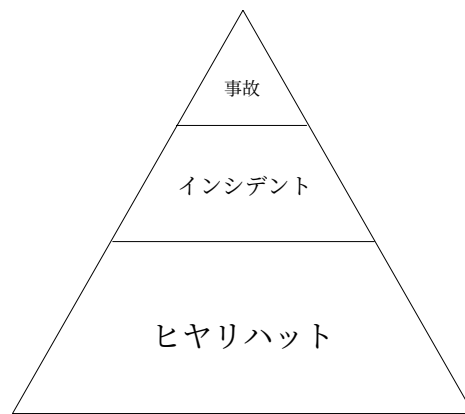


図 1.1: 事故／インシデント／ヒヤリハット

インシデントやヒヤリハットの原因を調べて対策をたてることで事故を防止する上で重要である。しかしながら、ヒヤリハットの情報は、事故報告書等には通常出て来ないものである。

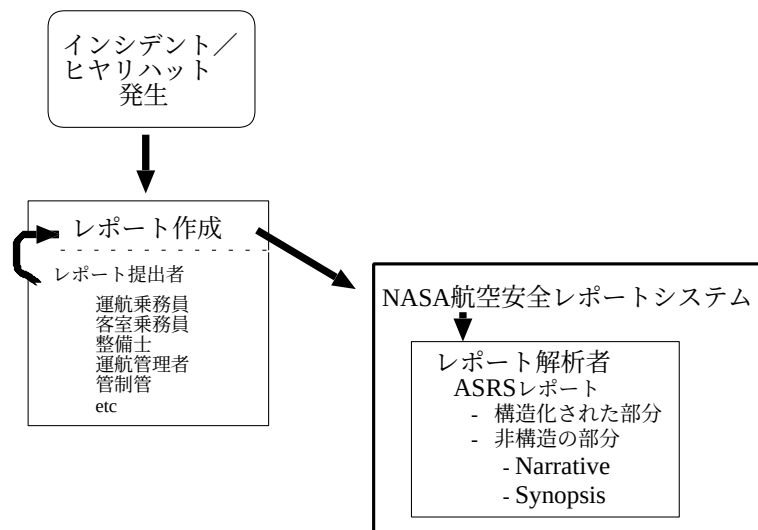


図 1.2: ASRS レポートの流れ

航空会社、政府機関等では、インシデント／ヒヤリハット情報を収集するためのシステムが作られており、航空安全レポートシステムと呼ばれている。その中でも、米航空宇宙局（NASA : National Aeronautics and Space Administration）の航空安全レポートシステム（ASRS² : Aviation Safety Reporting System）は最も規模の大きなものである。NASA の航空安全レポートシステムの ASRS レポートは、パイロット、管制官、客室乗務員、整備士や地上作業員などが自発的に提出したインシデント／ヒヤリハットレポートを収集したものである。このレポートには、提出者が経験した不安全事態の発生や危険な状況が記述されており、その内容は将来の航空安全向上の為に役立つ

² <http://asrs.arc.nasa.gov/>

と考えられている。このレポートシステムは、提出者の自発的提出、レポート内容の秘匿性、提出者の免責という特徴があり、それだけに重要な情報が記載されていると考えられる。現在月に2,900件以上のレポートが提出され、25年間でトータル46万件以上のレポートが収集されている。提出された安全レポートは、少なくとも2人のASRS航空安全解析者によって解析され、必要があれば提出者に質問をして事実を確認することもある（図1.2）。

このシステムは、提出者の秘匿性が守られるようになっているので、提出者が特定できるような情報（例えば、飛行便名など）は、安全レポートから削除されている。このようなレポートを解析するのは、インシデント／ヒヤリハット情報を解析し、将来の事故の原因の芽を摘みとるという意味で重要であるが、計算機によるアプローチは少ない [24, 29, 43]。

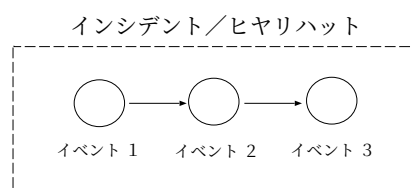


図 1.3: イベント連鎖

本論文は、このASRSレポートを対象にして、テキストマイニングを行なうことを目的とする。インシデント／ヒヤリハットは、単独の要因で起こることは少なく、通常いくつかのイベントの連鎖（Chain of Events）によるものである（図1.3）。イベント連鎖は、原因と結果という因果関係によって結びついていると考えられる。したがって、インシデント／ヒヤリハットを防止するには、このイベント連鎖の解析をし、その要因を断ち切ることが必要になる。

ASRSレポートは、構造化された部分と非構造化の部分から構成されている。構造化された部分には、レポート番号、インシデント／ヒヤリハット発生時のローカル時間、天気などが記述されている。一方、非構造化の部分は平文で記述されており、提出者が記述したNarrative部分と解析者がそれを要約したSynopsis部分に分かれている。重要な情報はASRSレポートのテキスト部分（非構造化の部分）に記述されていると考えられるので、ASRSレポートのテキスト部分から因果関係を抽出する技術が必要になる。

ASRSレポート24,626件に対する1件のレポート当たりの平均語数は、Narrative部分が約38語、Synopsis部分が約15語である。図1.4にASRSレポートの例を示す。

1.3 システム概要

本論文では、ASRSレポートのSynopsis部分をテキストマイニングの対象とする。1.1節で述べたようにテキストマイニングを行なうには、

- テキストの話題の中心をどのようにして見つけるか
- 前処理でどのような構造を抽出するか
- どのようなマイニング技術を使用するか

が重要な問題である。

構造化された部分

General Information

Report Number: 80094
Local Date(YrMon): 198801
Light Condition: DAYLIGHT
Primary Problem Area: AIRCRAFT AND THEIR SUBSYSTEMS

...

...

非構造の部分

<Narrative>

JUST AFTER LEVEL OFF AT CRS, CABIN BEGAN TO CLB AT APPROX 3000 FPM. AUTO FAIL AND STAND BY LIGHTS CAME ON. PRESSURE CTLR SET TO MANUAL. AC AND DC AND GND WERE SELECTED AND USED. VALVE INDICATOR SHOWED CLOSED ENTIRE TIME. PACKS SWITCHED TO HIGH-REDUCED CLB TO 1500'. DURING ABOVE CALLED ATC AND CLRD TO FL240, ...

<Synopsis>

PRESSURIZATION CTL ON ACR MLG LOST AS ACFT REACHED CRUISE ALT OF FL350. ARTCC AMENDED ALT ASSIGNMENT TO 10000' ON REQUEST. OXYGEN MASKS DROPPED BEFORE PRESSURIZATION CTL REGAINED. ACFT RETURNED TO DEP ARPT.

図 1.4: ASRS レポートの例

1 番目の問題については、対象とした Synopsis 部分は、人間が Narrative 部分を要約したものであるため、これ以上内容を絞る必要はないと仮定してそのまま使用する。

2 番目の問題については、意味関係 (Semantic Relations) の中で重要なものの 1 つであり、イベント連鎖を記述していると考えられる因果関係を取り上げることとする。因果関係は、ある状態／行動から別の状態／行動への原因と結果を表すので、テキスト中から状態／行動と因果関係を示す手掛かりを発見しなければならない。そのために、テキストから述語-項構造を抽出する。構文解析を行ない、構文解析された結果から、述語-項構造を抽出する際には、主語、動詞、目的語、前置詞句 (動詞の右の一番目の前置詞句のみ使用) を抽出する。また、名詞句はそのまま抽出する。なお、格構造の一部分の省略や照応関係についての処理は行なわない。因果関係については、自明な因果関係を示す表現を表層的なパターンから抽出し、さらに、時間関係の表現、節／文の連接からも因果関係を抽出する。

3 番目のデータマイニングの問題に関しては、シーケンシャルなデータを扱うことのできる PrefixSpan を用いることとする。

前処理の要素技術として、略語の復元と名詞化構文の述語-項構造への変形を提案する。両提案手法とも辞書の代わりに、コーパスから自動的に知識を獲得して処理する手法を採用する。略語の復元は、構文解析の前の段階で適用する。2 種類のコーパス (略語の含まれているコーパス (略語の多いコーパス) と復元語を探すコーパス (略語の少ないコーパス)) を用意して、略語と復元語の文脈情報と人が略語を使用する際の規則性を利用して略語の復元を行なう。名詞化構文とは、本来節で表現されるものが名詞化され名詞句になったものである。この処理を行なうには、システムが名詞句を認識する必要があるため、構文解析の後で行なう。名詞化構文から述語-項構造へ

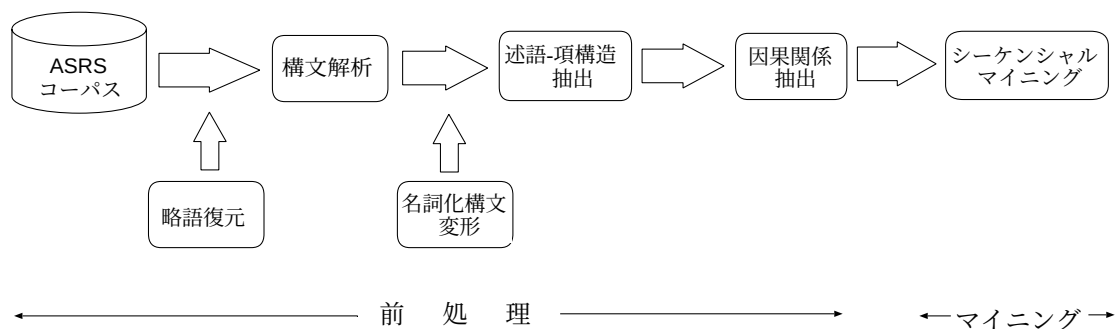


図 1.5: システム概観

の変形は、コーパス中の動詞派生名詞に対応する節表現の格構造を抽出して、その統計情報を利用して、名詞化構文を自動的に述語-項構造に変形する。

システムの概観を図 1.5 に示す。

1.4 本論文の構成

本論文の構成は以下の通りである。

第 2 章では、コーパスを用いて単一語からなる英文の略語を元の語に復元する手法について述べる。略語は、技術的な文献、車、飛行機の修理記録、写真の見出しなどで頻繁に使用されるが、略語の計算機による取り扱いについてはこれまであまり研究されていない。先行研究では、復元候補語の探索に汎用の辞書を用いる手法が多いが、本論文では、辞書の代わりに同じ分野のコーパスを利用する手法を提案する。第 2 章の前半では、略語の特徴や先行研究について述べ、第 2 章の後半では、辞書の代わりに同じ分野のコーパスから略語の復元候補語を探索し、人間が略語を使用する際の規則性と略語と復元語の文脈情報を統合して、復元語を決定木により絞り込み、文脈情報により最終的に復元語を決定する手法及びその実験結果について述べる。

第 3 章では、名詞化構文に対応する節に変形する処理について述べる。名詞化構文は、英語では頻繁に使用される用法である。テキストマイニングや情報抽出では、同じイベントが異なる形式で表現されたとき、それらが同一のイベントであることを同定する必要がある。本論文では、名詞化構文も節表現も同じ述語-項構造に変形することにより両者の同定を行なう。第 3 章の前半では、名詞化構文の特徴及び先行研究について述べる。第 3 章の後半では、コーパスから節表現の格構造を抽出して、その統計情報を利用して、名詞化構文から自動的に述語-項構造へ変換する手法について述べる。ASRS コーパスで頻度上位 20 位までの動詞派生名詞について提案手法の評価を行なう。

第 4 章では、テキスト中の自明な因果関係を表す表現を利用して、因果関係を表層的なパターンから抽出する手法について述べる。第 4 章の前半では、まず、因果関係を抽出するための 4 種類のパターンについて述べ、次に、本論文で使用するテキストマイニング手法である PrefixSpan の説明をする。第 4 章の後半では、4 種類のパターンによるマイニング結果を考察する。

第 5 章では、時間関係を表す表現、文／節の接続から抽出されるイベント間の関係を因果関係に加えた場合について述べる。第 5 章の前半では、時間関係を表す表現についての一般的な特徴を述べ、ASRS コーパスでの時間関係の表現について考察する。第 5 章の後半では、時間関係を表す表現、文／節の接続を抽出パターンに加えた場合の実験結果について述べる。時間関係による表現と節／文の接続は、明示的に因果関係を表現していないが、イベントを時間推移順に並べ、その中で頻出するパターンは暗黙の内に因果関係を表していると仮定して因果関係を抽出する。

第 6 章では、マイニングの結果、抽出したパターンの具体的内容の考察を行なう。自明な因果関係を表す表現及び時間関係を表す表現、文／節の接続による抽出パターンについて検討する。

第 7 章では、本論文のまとめと今後の課題について述べる。

第2章 略語の処理

本章では、略語の処理について述べる。略語は分野毎に用法が異なり、またその用法も厳密には定まっていないのが現状である。したがって、略語辞書を作成してもその登録略語がカバーできる範囲は限られたものになる。ところが、略語はそのままにしておくと、構文解析において未知語として処理され、解析の性能に悪影響を及ぼすと考えられる。つまり略語は、構文解析器の辞書に登録されていないので、未知語として処理され、結果として出力される構文木も間違っただけのものになる可能性が大きくなる。また、同じ語が異なる略語で表現されることも多いので、それらを同定するためにも略語の復元は必要である。本章では、単一語からなる略語を対象とし、人間が略語を使用する際の規則性と略語と元の語（以後、復元語と呼ぶ）の文脈情報を統合して、略語復元を行なう手法を提案する [49, 50, 52]。復元語の探索には、辞書を使用する代わりに略語と同じ分野の「略語の少ないコーパス」を用いる。「略語の少ないコーパス」とは、略語があまり使用されず、対象としている略語の復元語が十分多く出現するようなコーパスのことを云う。それに対して、「略語の多いコーパス」とは、略語が多く出現するようなコーパスのことを云う。復元候補語の絞り込みは、略語を使用する際の規則性から略語と復元候補語の文字情報を比較することにより行なう。復元候補語の妥当性については、略語と復元候補語の文脈の類似性から計算する。類似度はベクトル空間法を用いて計算する。

2.1 略語の定義

略語と似た用語で用いられるものに頭字語がある。略語と頭字語の定義については、明確に定まっていないのが現状であるが、本章では、略語と頭字語を次のように定義する。

- 頭字語は複数語からなり、その名の通り、一部または全部の構成語の頭文字あるいは頭文字に続く文字を並べて1語としたものである（例：“USA - United States of America”）。
- 略語は1個以上の語からなり、構成語の文字を短縮して1語にしたものの内、頭字語でないものである（例：“TOWT - Takeoff Weight”）。

頭字語は以前はアルファベットをそのまま発音するもの（例：“USA - United States of America”）が多かったが、後に単語のように一括して発音するものが一般化した（例：“UNESCO - United Nations Educational Science and Cultural Organization”，“RADAR - Radio Detecting And Ranging”）。TNT（Trinitrotoluence）は頭字語のように思われるが、本章の定義では略語になる。ただ、語源を考えると3語から出来ている。略語と頭字語の厳密な区別は難しいし、厳密な区別をするのは本論文の趣旨でもない。

略語／頭字語の使用に関しては、テキスト中で略語／頭字語リストを明示的に定義しているような場合や最初に略語／頭字語が出現する前後の位置に定義をしている場合もある。その一方で、略語／頭字語を定義なしに使用している場合も多い。

本章では、略語の定義をしていない場合で、単一語からなる英語の略語を対象とする。

固有名詞、略語、頭字語、ミススペル等の未知語は、計算機でテキスト処理（情報検索やテキストマイニング）を行なう場合の障害になる。特に略語は、分野毎に特殊なものが多く、またその種類も増え続けているのでそのリストを前もって作成しておいても現実的でない場合が多い。

2.2 略語の特徴

略語は、元の語から文字を短縮したものであるが、使用目的としては、見出し語のように印象付けをねらったもの、整備記録で用いられるように時間／スペースの短縮を目的としたものなどがある。しかしながら、その効用の反面、分野毎にその使われ方が異なり、異分野の人間には理解が難しい面がある。例えば、航空分野では“DEP”は“departure”の略語であるが、金融界では“deposit”の略語であるというように同じ文字列の略語でも意味している復元語は全く異なる（以後、略語は大文字、復元語は小文字で表記する）。

人間が略語を見てその復元語を推測する場合を考える。人間が略語を元の語に戻す場合（以後、略語の復元と呼ぶ）、略語の周辺の語を調べたり、略語に含まれる文字を調べる。また、略語の周辺の語とその復元語の周辺の語との類似性からヒントを得る場合がある。本章の提案手法はこれらを計算機上で模擬するものである。

略語には、以下のような特徴がある。

1. 略語に含まれる文字数は、その復元語の文字数より少ない。
2. 略語は、その復元語と同じ文字が同じ順序で含まれる (例：“TKOF-takeoff”).
但し、略語に含まれる文字の内“X”、“-”、“/”は例外とする。
3. 略語は辞書に登録されていない語である。
4. 略語は単一の語に対応する。この仮定は、必ずしも常にあてはまるわけではないが、本論文では略語が単一の語に対応する場合のみを考える。

従来の略語復元の手法は、システムの精度が低いために、複数の復元候補語をシステムが返し、ユーザにその中から選ばせるというものが多い [41]。しかしながら、このようなシステムでは、システムの完全な自動化を必要とするようなタスク、例えばテキストからの音声発話のようなものには適用できない。

従来の略語復元の手法の精度が低い原因は、略語の文字情報のみを使用し、略語の周辺の情報を使用していないことがあげられる [41, 55]。人間が略語を復元する場合を考えると、略語の周辺の文脈情報から復元語を推測している [41]。したがって、略語復元の精度を向上させるには、文脈情報が不可欠であると考えられる。

略語リストを作成すれば、このようなシステムは必要ないとも考えられるかもしれないが、実世界では同じ分野であっても、様々な人間が同じ語を異なる略語で表し、さらに同じ人間でさえ同じ語を異なる略語で表わすことが珍しくない。この現象は整備士が整備記録をつける場合や、時間的／空間的制約のある飛行機から地上への通信などの場合に顕著に見られる現象である。したがって、略語リストの使用には限界がある（例：航空分野で“A/C”及び“ACFT”は、どちらも“aircraft”の略語である）。さらに、略語リストを使用する場合には、文書作成時略語リストを参照する人間が掛かるという問題と略語リストの保守／維持管理という問題がある。

2.3 対象分野

一般に略語／頭字語はそれ程頻繁に使用されるものではない。1989年のWall Street Journalの145,295語が含まれている文書をサンプリングして調べた結果、略語は0.003%、頭字語は0.01%しか含まれていなかった。この略語／頭字語からは、“Nov”や“USA”のような辞書に載っている語は除いてある。一方、技術文書やE-Mailなどでは、略語／頭字語を頻繁に使用している。

航空界では、略語を頻繁に使用している。本章では、ASRS (Aviation Safety Reporting System) コーパスのNarrative部分を使用して実験を行なうこととする。ASRS コーパスのNarrative部分8,000語のサンプルを人手で調べた結果、略語は約11.0%、頭字語は約2.0%含まれていた。

ASRSの航空安全レポートシステムは、提出者の秘匿性を守るようになっているので、提出者を特定できるような情報(例えば、飛行便名など)は、削除している。したがって、“X”や“Aircraft X”などの表現を頻繁に使用しており、これらの表現もシステムにとっては未知語になる。システムは、これらの未知語と略語も区別できなければならない。

2.3.1 「略語の多いコーパス」と「略語の少ないコーパス」

「略語の多いコーパス」は、略語が多く含まれているコーパスを意味し、「略語の少ないコーパス」は、略語の使用頻度が少なく略語に対応する復元語が多く含まれているようなコーパスを意味するものとする。例えば、ASRS コーパスは「略語の多いコーパス」である。提案手法は、「略語の少ないコーパス」を使用して、コーパスから略語に対応する復元語を探索する。「略語の少ないコーパス」は、略語に対応する復元語を含んでいる必要があるため、「略語の多いコーパス」と同じ分野のものである必要がある。

略語に対応する復元語を効率よく正確にランク付けするため、文脈を使用する(文脈は、略語の前後に出現する語を意味する)。略語とその復元語は本来同じ意味のものであるので、前後に出現する語の分布は統計的に類似していると仮定する。したがって、略語とその復元語の類似度は、略語と略語に対応しない復元候補語の類似度に比べ大きくなる筈である。この類似度を使用することにより、復元語の最終的な決定を行なう。

「略語の少ないコーパス」を使用するのは次の2つの理由からである。

- 「略語の少ないコーパス」から復元語とその統計情報が得られる
- 「略語の多いコーパス」には復元語がない可能性がある

この目的のため、「略語の少ないコーパス」として米国家運輸安全委員会(NTSB¹ : National Transportation Safety Board)の事故情報を使用する。NTSB事故情報は航空事故を英文のテキストで記述したものであり、ASRS コーパスと同じ航空分野のコーパスである。

「略語の少ないコーパス」が必要な例をあげる。例えば、ASRS コーパスには“TKOFF”が1,156回出現するが、その復元語である“takeoff”は1回も出現しない。これに対して、NTSB コーパスでは“TKOFF”が28回出現し、“takeoff”は344回出現する。ASRS コーパスとNTSB コーパスにおける“TKOFF”の復元語は、いずれも“takeoff”である。したがって、“TKOFF”を“takeoff”に復元するためには「略語の少ないコーパス」が必要である。

ASRS コーパスとNTSB コーパスの統計情報を表2.1に示す。ASRS コーパスは、10,000件のレポートを用意し、その中でNarrative部分の記載のある2,648件のレポートを使用する。

¹ <http://www.nts.gov/aviation/aviation.htm>

表 2.1: ASRS コーパスと NTSB コーパスの統計情報

	ASRS コーパス	NTSB コーパス
文書総数	2,648	3,937
語の総数	677,725	426,717
語の総数 (タイプ)	18,422	14,940
1 文書の平均語数	260	108
未知語の総数	100,983	33,575
未知語の総数 (タイプ)	6,077	7,190
1 文書の平均未知語数	38.1	8.5
未知語率	14.9%	8.5%

2.4 関連研究

略語の定義をその略語の周囲から探索する手法については、これまでいくつかの研究がなされている。このようなものを対象にして略語／頭字語の定義を自動的に求めるような研究もある [20, 31, 61]。この手法は、略語／頭字語を最初に使用する時にはその復元語を定義するという仮定の元に成り立っている。略語／頭字語を定義する場合には、括弧書きによって定義するという特徴がある。したがって、この手法でその定義を抽出するためのタスクは、パターンマッチングの問題になる。この手法は、略語／頭字語の定義が略語／頭字語の周囲に存在することから、比較的精度がよい。

略語／頭字語の定義は、一般の文書では比較的よくなされているが、特定領域の文書 (例えば、航空安全レポート、整備記録、E-Mail 等) ではあまりなされていない。したがって、このような場合には、全く異なる手法が必要になる。本章では、略語の定義がない場合についての略語の復元法について考察する。

略語復元のタスクは、略語が辞書の登録語でないという点ではスペルエラーの中で辞書の登録語でないものを復元するタスクに類似しているように見える (スペルエラーにより元の語と異なる辞書登録語になる場合があるが、その処理は難しい) [19]。しかしながら、Rowe らが指摘しているように、略語は復元語に比べその文字数が少なく、保持している文字情報はスペルエラー語に比較してかなり少ない [41]。Kukich は、辞書語でないスペルエラー語とそれに対応する正しい語の文字数の差は、ほとんどの場合 2 文字以内であると報告している [19]。本章で使用するコーパスでは、略語とそれに対応する復元語との文字数の差は、平均で 4.3 文字である。このことから、文字情報のみの使用では略語の復元が困難であることが分かる。

しかしながら、既存のほとんどの略語復元システムは、文字情報しか使用していない [54, 55]。Uthurusamy らは、テキスト修正と正規化の過程の中で未知語情報を取得した後、略語復元を行なっている [55]。人間が略語を使用する際の規則性を利用して、略語をいくつかのタイプに分け、そのタイプ別に異なるアルゴリズムを適用し復元候補語のランク付けを行なっている。Rowe らは、略語と復元語のペアからルールを作成して復元語のランク付けを行なっている [41]。文脈情報と Rowe らが呼んでいるのは、文字情報そのものである。Toole は、略語復元を略語の検出と復元という 2 つのタスクに分けている [54]。Toole は、文脈情報として 2 値を使用している (略語の検出と復元両方の過程で)。復元候補語が略語と同じ文脈に現れた場合、スコアが “1” にセットされ、それ以外は “0” にセットされている。例をあげると、“ALT” が “altitude” の略語の場合、

“ALT” が “cruising ALT was” という文脈に出現し、復元候補語 “altitude” が “cruising altitude” 又は “altitude was” という文脈に出現した場合、スコアが “1” になる。もし、“altitude” の前後にこれらの語が出現しなければ、スコアは “0” である。しかしながら、Toole は、この特徴は予想にくいものだと結論している。その理由は、Toole は文脈情報として前後 1 語しか使用していないのと、前後の語との完全一致のみを使用しているためだと考えられる。提案手法は、このアイデアを拡張し、文脈として前後数語のウィンドウ内の語を使用し、スコアは 2 値ではなく語の出現頻度の余弦を使用する。

本章の提案手法は、略語復元のタスクを、略語候補語の検出／復元候補語の選択／ランキングという 3 つのサブタスクに分ける。その中で、略語候補語の検出は最も簡単で、形態素解析器が未知語のタグ付けしたものを使用する。復元候補語の選択に関しては、Rowe らや Toole の手法では、汎用の辞書を用いているので、略語と関係のない復元候補語が大量に選択され、システムの性能低下をもたらしていると考えられる [41, 54]。提案手法では、「略語の少ないコーパス」から復元候補語を選択することにより、復元候補語の選択範囲を絞る。さらに、略語使用時の規則性により復元候補語を絞る。ランキングには、略語候補語と復元候補語双方の文脈情報を用いる。2.5 節ではその詳細について述べる。

2.5 システム設計

2.2 節の略語の特徴 (3) により形態素解析器が「未知語」のタグを付与したものを略語の候補語とする。未知語は略語である可能性はあるが、まだ略語かどうかは決定できない。その決定は復元候補語の選択の段階で行なう。

復元候補語は、未知語に対して 2.2 節の略語の特徴 (1) 及び (2) を使用して、「略語の少ないコーパス」(NTSB コーパス) から探索する。さらに、略語使用時の規則性により復元候補語を絞り込む。この段階で、全ての復元候補語を略語の復元語ではないと決定する場合がある。これは、システムが未知語を略語ではないと決定する場合である。

その後、略語候補語及びその復元候補語の文脈情報を収集する。この情報を利用して、どの復元候補語が最も略語復元語に近いかを調べる。

以上述べたように、システムの略語復元のタスクは、次の 3 つのサブタスクからなる。

- 略語候補語の検出
- 復元候補語の選択
- ランキング

ASRS コーパスと NTSB コーパスの形態素解析には、TreeTagger を使用する [45, 46]。TreeTagger は、形態素解析器の辞書に記載されていない語が現れると未知語のタグを付与する。未知語であっても形態素解析器は、周囲の単語情報から品詞情報を付与する。この品詞情報を利用して、名詞、動詞、形容詞、形容動詞、前置詞のタグを付与された未知語を略語候補語とする (品詞フィルター)。

文脈情報として、対応する語 (略語候補語及び復元候補語) の前後に出現する語を使用する。同じ語でも語尾変化により表層の形は異なる場合があるが、そのようなものは語の原形を用いて同一のものとする (例: “Aircraft landed ...”, “After aircraft landing, ...” で “landed” および “landing” は、どちらもその原形の “land” に変形する)。TreeTagger は、単語の品詞情報とその原形を付与するので、stemming を行なう必要はない。

2.2 節の略語の特徴 (4) により、略語は単一の語に対応することから、略語をタイプとして捉えることとする。略語候補語／復元候補語の前後にウィンドウを設定し、コーパスに出現する同じタイプの略語候補語／復元候補語のウィンドウ内の情報を全て収集する。文脈情報を収集するウィンドウ幅は、略語候補語／復元候補語とも前後 3 語を用いることとする。予備実験でウィンドウ幅 3 とウィンドウ幅 5 を比較したが、顕著な違いは発見できなかった。語の統語的な範囲は、設定したウィンドウ幅より小さいかもしれないが、文脈情報としてはもう少し大きなものが適当であると考えられる。

略語の復元候補語を探索するのに 2.2 節での略語の特徴 (1) と (2) を使用する (再掲)。

1. 略語に含まれる文字数は、その復元語の文字数より少ない
2. 略語は、その復元語と同じ文字が同じ順序に含まれる (例：“TKOF-takeoff”)

但し、略語に含まれる文字の内 “X”，“-”，“/” は例外とする。

例えば、略語候補語が “FLT” である場合、その復元候補語は、上記条件を満たすものを NTSB コーパスから探索して、“flight”，“filters”，“flat”，“float”，“difficult”，“flights”，“felt”，“floated” を得る。復元候補語の場合も品詞として名詞，動詞，形容詞，形容動詞，前置詞のタグを付与されたもののみを復元候補語とする (品詞フィルター)。

更に、復元候補語はルール (人間が略語を使用する際の規則性) を使用してその数を絞り、その後文脈情報を使用してランク付けする。

略語復元システムの概観を図 2.1 に示す。図 2.1 の例では、“TWR” が TreeTagger により未知

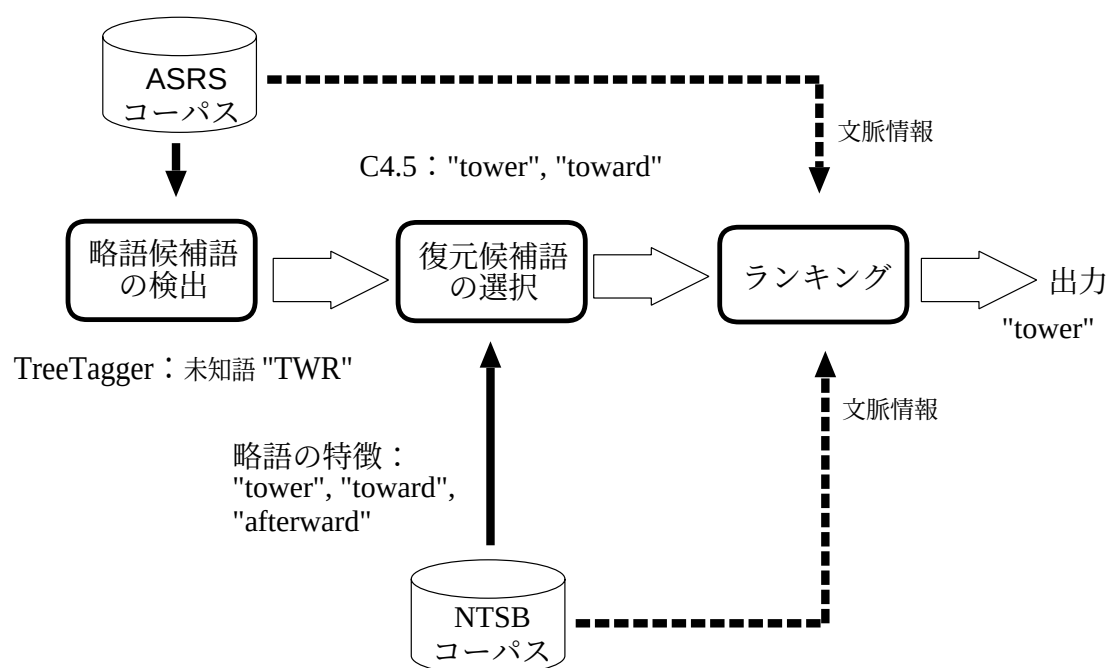


図 2.1: 略語復元システム概観

語と認識され、略語候補語となる (略語候補語の検出)。略語の特徴により、NTSB コーパスを探索して “tower”，“toward”，“afterward” を復元候補語として選択する。さらに、C4.5 決定木により復元候補語を絞りこみ、“tower”，“toward” を復元候補語とする (復元候補語の選択)。略

語候補語は ASRS コーパスから、復元候補語は NTSB コーパスから文脈情報を抽出してランキングを行ない、最終的に復元語として “tower” を出力する (ランキング)。

図 2.2 は、略語復元の流れを示したものであり、図 2.3 は、文脈情報の具体例である。

2.5.1 ターム重み付け

文脈として収集する語の重み付けについて述べる。収集する語は、略語候補語／復元候補語ともにその周辺のウインドウ内の語とする。情報検索では、質問文に含まれる語と文書に含まれる語を基礎として質問文と文書の類似度を計算する。本章のシステムの略語候補語のウインドウ内の語を質問、各復元候補語を文書、各復元候補語のウインドウ内の語を各文書の内容とすると、情報検索の問題に帰着できる。

情報検索では、“the” や “or” のように文書内で頻出する機能語はそれ自体意味を持たないので、通常、索引語から除外する [10]。一方、本章のシステムでは、略語と復元語の周辺には沢山の機能語が現れ、略語／復元候補語を特徴付けている。これは、略語／復元語が文書でなく語であるからである。例えば、“CLR D – cleared” の後には “to” 或いは “for” が頻繁に出現する (例: “cleared to land”, “cleared for takeoff”)。したがって、提案手法での機能語の削除は、システムに悪影響を与えると考えられる。予備実験の結果もこの考えに沿うものであったので、機能語は削除しないこととする。

情報検索では、 tf (term frequency) 又は $tf-idf$ (term frequency-inverse document frequency) が文書／質問に対する重み付けとしてよく用いられる。 tf は再現率向上に貢献し、 idf は精度向上に貢献すると云われている。 tf と idf は、組み合わせ、 $tf-idf$ として使用する事が多い。本章のシステムでの tf と $tf-idf$ の性能を予備実験で比較した結果、特に顕著な差異はなかった。したがって、重み付けとして tf を使用することとする。Hearst も、テキストの近接単語を比較するのに $tf-idf$ は適当でなく、 tf が最も効果的であると報告している [14]。 idf を使用すると文書全般にわたり出現頻度の低い語が大きな重み付けをされるが、これは、略語と復元語の性質を正確に表現するものではないと考えられる。更に、タイプとして文脈情報を収集することにより、出現頻度の低い語の影響は低くなると考えられる。

2.5.2 ランキング

復元候補語のランキングにはベクトル空間法を用いる [44]。ベクトル空間法は、文書と質問をそれを特徴付けるタームのベクトルで表すもので、ベクトルの次元はタームの異なり数を表す。ベクトルの類似度 ($score$) を求めるのに、次式のような余弦を用いる。

$$score = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

ここで、 x_i と y_i は、それぞれ略語候補語と復元候補語のウインドウ内に現れる語の重みであり、 n はウインドウ内に現れる語の異なり数を表す。

Kukich は、スペルエラーを修正するのに使用するランキング手法として余弦が最もすぐれていると報告している [18]。

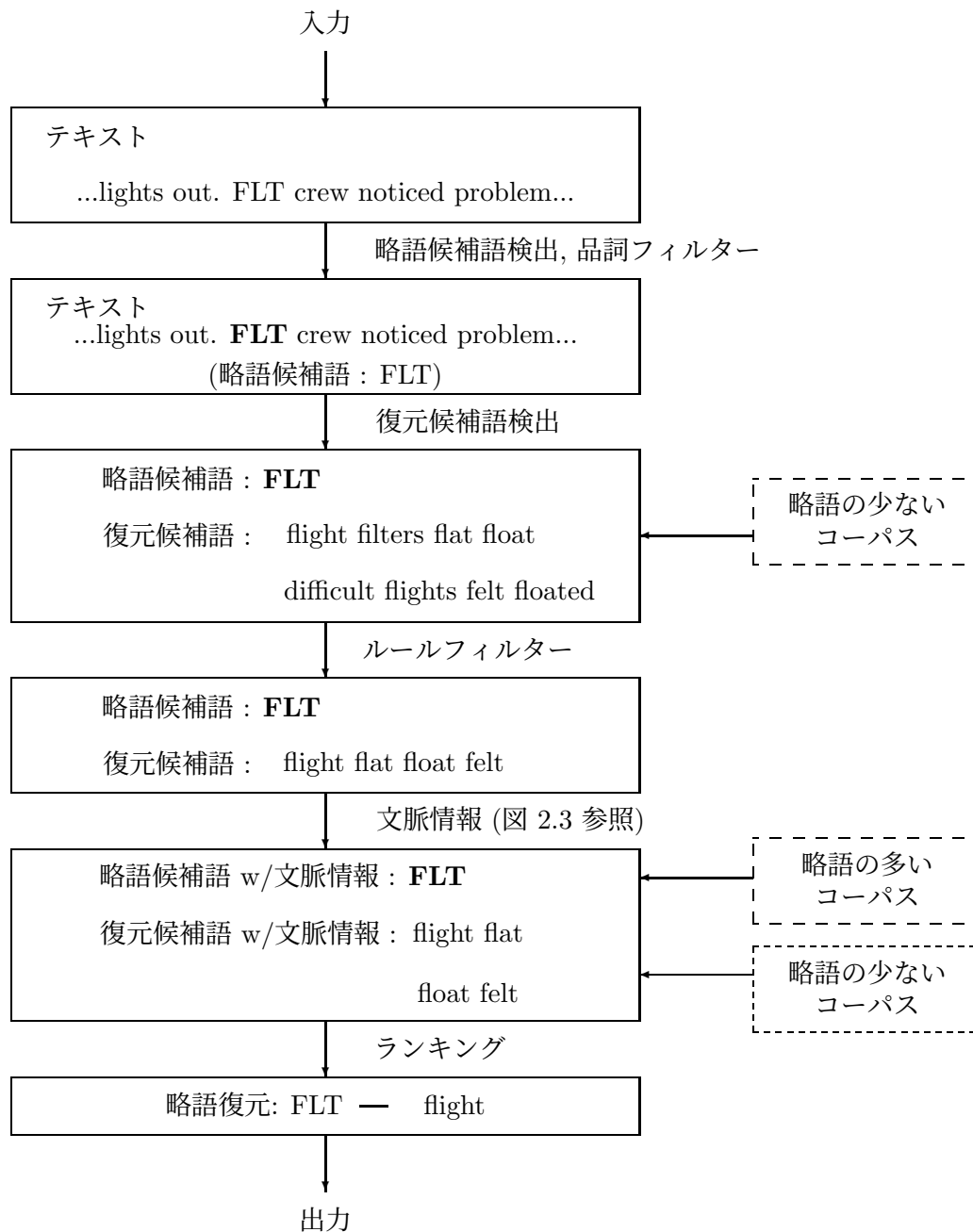


図 2.2: 略語復元の流れ

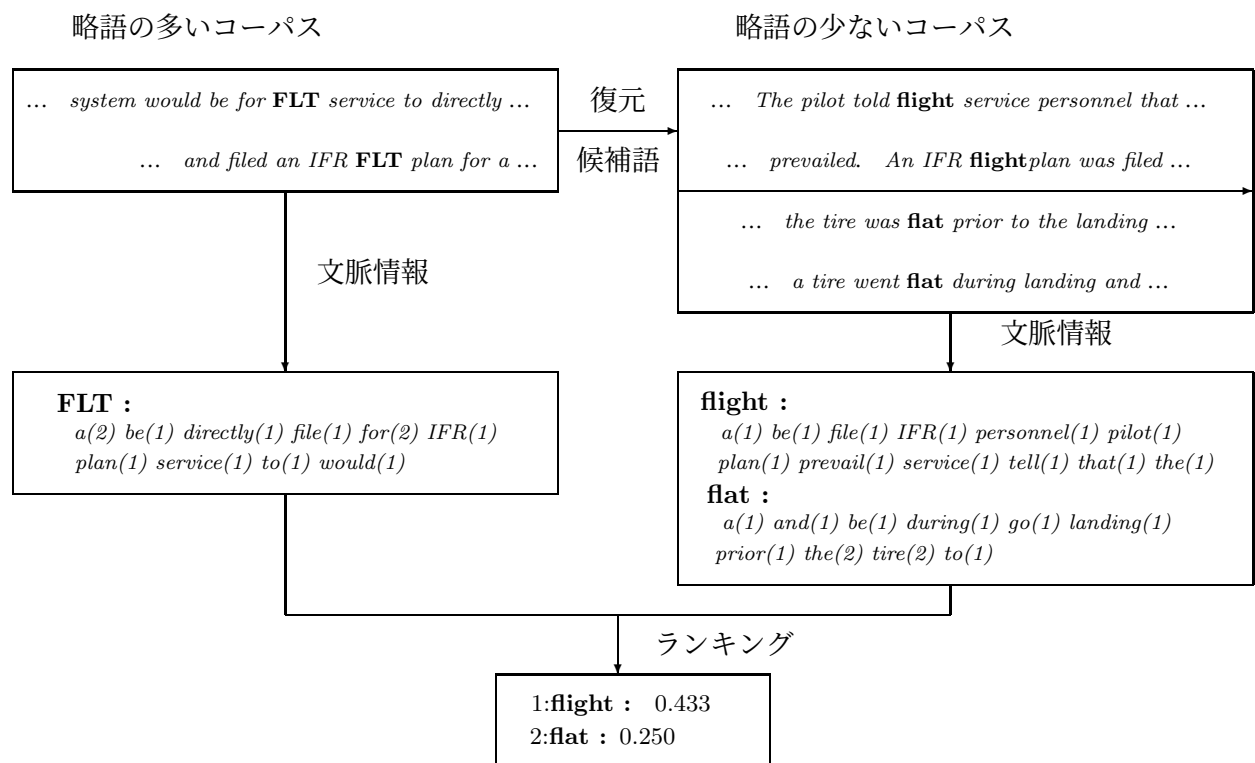


図 2.3: 文脈情報の例

2.5.3 復元語の選択

システムは、最終的に復元候補語の中から最も略語の復元語に相応しい語を選ぶか、あるいは、復元候補語の中には略語の復元語に相応しい語がないと決定する。この場合には、略語候補語が略語でない場合も含まれる。そのために、システムは様々な方法で復元候補語の中から略語の復元語には相応しくない語を取り除き、最もスコア (*score*) の大きなものを最終結果としてユーザに返す。提案手法の性能を以下の5つの方法について検討する。

方法 1：最もスコアの高い復元候補語を略語の復元語として選ぶ。この方法は、人間の略語使用時の規則性を使用しないで、文脈情報のみを使用する方法である。

方法 2：この方法は、人間の略語使用時の規則性を使用して、復元候補語の絞りこみを行なう。Uthurusamy らが指摘しているように、英語で略語を使用する場合には次の2つの形が多様される [55]。それは、truncation (語の接尾文字列の省略) と contraction (語の内部の文字の省略) である。方法2では、以下に述べる truncation, contraction を細分化した特徴、ASRS コーパスと NTSB コーパスの頻度を使用して、C4.5 決定木により復元語であるか否かのクラス分けを行なう [38]。最初に決定木のルールを満足する復元候補語を絞り込み、次にその中で最もスコアの高い復元候補語を略語の復元語として選ぶ。C4.5 の訓練は、2.6 節で述べる訓練セットで行なう。以下は、C4.5 で使用する特徴である。

- 先頭の1文字の一致：
略語と復元語の最初の1文字が一致する。この特徴は、略語によく現れる特徴である。但し、“X” で始まる略語は例外とする (例：“XING – crossing”)。
- 接尾文字列の一致：
略語と復元語が最後の文字から何文字一致するか。この特徴も、略語によく現れる特徴である。
- Truncation：
略語と復元語の接頭文字列が一致して、接尾文字列が略語から削除されているもの (例：“CAPT – captain”)。
- 母音の省略：
略語の復元語からの母音の省略 (例：“DSCNT – descent”)。
- 母音と子音を含む内部文字列の省略：
(例：“CLB – climb”)
- 子音のみを含む内部文字列の省略：
(例：“CLIB – climb”)
- 略語の ASRS コーパスでの頻度
- 復元語の NTSB コーパスでの頻度

上記8つの特徴に加え、略語の文字列長は略語と略語でない語を区別するのによい指標と考えられる。しかしながら、本章のシステムは、略語の少ないコーパスに復元候補語がないという理由で、文字数の長い未知語 (固有名詞や、2語が“ー”で連結した語など) を略語候補語から排除する。また、大文字／小文字の違いもよい指標と考えられるが、本章で使用するコーパスは全て大文字で記述されているので、この特徴は使用できない。上記理由により、本章のシステムではこれらの特徴は使用しない。

方法 3：最初の文字が“X”ではじまる略語候補語を“TRANS”，“CROSS”及び“OUT”に変換した後方法 2 を適用する。

方法 4：頭字語は，航空分野では比較的頻繁に使用され，頭字語は表現の揺れが少ないという特徴がある（例：“GS – glide slope”）．方法 4 では，方法 3 を適用する前に 114 個の頭字語リストを使用して略語と頭字語の区別を行なう．この頭字語リストは航空分野で一般に使用されるものを用いる．

方法 5：航空分野では，一般に，国際航空輸送協会（IATA：International Air Transport Association）の定めた空港の 3 文字コードを使用する．方法 5 では，方法 4 を適用する前にこの 3 文字コードを使用して略語と空港名を区別する．

2.6 評価

最初に訓練セットとテストセットの作成法について述べ，次に 2.5.3 項で提案した 5 種類の方法でシステムを評価する．

2.6.1 訓練セットとテストセット

形態素解析器が未知語のタグを付与した語を使用して，ASRS コーパス全体に含まれる略語を調べる．ASRS コーパスの中で形態素解析器が未知語のタグを付与した語を手で調べた結果，265 種類の略語を確認した．しかしながら，形態素解析器が未知語のタグを付与していない語の中にも略語があることに注意しなければならない．

訓練セットとテストセットの作成方法について述べる．テストセットは，次のようにして作成する．ASRS コーパスの中から 20 レポートを選び，その中に含まれる略語のタイプをテスト用の略語タイプとする．この方法で 106 種類の異なるタイプの略語を抽出した．システムは，この 106 タイプの略語の中で，84 タイプを略語候補語として認識する．残り 22 タイプについては，形態素解析器が未知語のタグを付与しないので，略語候補語としては認識できない．システムが略語候補語として認識する 84 タイプの略語の内，略語“MI”は，“mile”と“minute”両方の復元語に対応する（これは，2.2 節での略語の特徴 (4) に反する）．したがって，テストセットとして合計 107 タイプの異なるタイプの略語を選ぶこととする（“MI”は 2 タイプとして計算する）．システムが略語候補語として認識できない 22 タイプは，265 タイプの略語には含まれない．したがって，訓練セットとして残り 181 タイプの略語を使用することとする（ $181 = 265 - (106 - 22)$ ）．このようにして，訓練セットとテストセットに同じタイプの略語が含まれないようにする．訓練セットに含まれる 181 タイプの略語を用いて，C4.5 の訓練を行なう．

2.6.2 評価

評価は，精度（precision）と再現率（recall）を使用する．精度と再現率は，次式のように定義する：

$$\text{精度} = \frac{w}{w + x + y}$$
$$\text{再現率} = \frac{w}{w + x + z}$$

ここで、 w は正しく復元された略語のタイプ数、 x は間違って復元された略語のタイプ数、 y は実際には略語でない未知語が間違って復元されたタイプ数、 z は略語として認識されなかった略語のタイプ数である²。

表 2.2: テストセット（20 レポート）での結果

	精度 (%)	再現率 (%)
方法 1	40.4 (55/136)	51.4 (55/107)
方法 2	63.3 (57/90)	53.3 (57/107)
方法 3	64.4 (58/90)	54.2 (58/107)
方法 4	70.7 (58/82)	54.2 (58/107)
方法 5	81.4 (48/59)	44.9 (48/107)

システムを評価するのに、2.5.3 項で提案した 5 つの方法を用いる。実験結果によると、方法 1 から方法 4 までは性能が向上しているが、方法 5 では逆に性能が低下している。方法 5 で性能が低下する原因は、空港の 3 文字コードと略語に同じ文字列を使用している場合があり、方法 5 では、空港の 3 文字コードと一致した文字列は、略語候補語から除かれ再現率が低下するからである。例えば、“SVC” は、単語の “service” 及び空港名に対応している。表 2.2 にテストセット（20 レポート）での結果を示す。

次に形態素解析器が未知語のタグを付与しない語の影響を排除した場合のシステムの性能を調べるために、形態素解析器が未知語のタグを付与しない 22 タイプの略語のタグを人手で未知語に変更したものをを用いる。その結果を表 2.3 に示すが、精度は表 2.2 の結果とほぼ同じであるが、再現率は約 10% 向上している。この結果から、形態素解析器が略語を正しく未知語と判定できれば、精度の低下なしに再現率を向上できることが分かる。次に ASRS コーパス全てでの評価を次

表 2.3: 22 タイプの略語のタグを未知語に修正した後のテストセット（20 レポート）での結果

	精度 (%)	再現率 (%)
方法 1	40.5 (64/158)	59.8 (64/107)
方法 2	60.9 (67/110)	62.6 (67/107)
方法 3	61.8 (68/110)	63.4 (68/107)
方法 4	66.3 (67/101)	62.6 (67/107)
方法 5	73.2 (52/71)	48.6 (52/107)

のように行なう。システムが未知語タグを付与しない略語を全て調べることはできないので、再現率*を次式で定義する。

$$\text{再現率*} = \frac{w}{w + x}$$

² 20 年以上航空分野で経験のある著者が正解リストを作成した。

もし、形態素解析器が全ての略語を正しく未知語と判定できれば、“再現率*”は、“再現率”と等しくなるはずである。表 2.4 は、この結果を示す。

表 2.4: レポート全体での結果

	精度 (%)	再現率* (%)
方法 1	23.9 (102/426)	38.3 (102/266)
方法 2	51.9 (110/212)	41.4 (110/266)
方法 3	54.1 (111/205)	41.7 (111/266)
方法 4	58.1 (111/191)	41.7 (111/266)
方法 5	70.7 (94/133)	35.3 (94/266)

2.6.3 先行研究との比較

次に先行研究と提案手法の比較を行なう。比較には、精度、再現率と F-measure を用いる。F-measure は、精度と再現率を 1 つの式で表し比較する指標で、本章では精度と再現率の重み付けの等しいものを使用することとする [56]。

$$F = \frac{2PR}{P + R}$$

ここで、 F は F-measure、 P は精度、 R は再現率を表す。

Toole は、ASRS コーパスを用いて略語の復元を行なっている [54]。Toole の使用したコーパスは、本論文で使用するのと同じ ASRS 航空安全レポートである。同一のテストセットを使用していないので性能の直接の比較はできないが、参考にはなると思われる。Toole は、略語復元のタスクを 2 ステップに分けている。1 つは略語の検出で、もう 1 つは略語の復元である。略語の検出ステップで 188 個の異なるタイプの略語を検出し、その内の数個は略語でないものを誤って検出したと報告している。システムが復元語を 1 語を出力する場合の略語の復元ステップでの再現率は、58.5%であると報告している。

表 2.5: 先行研究との比較 (1)

	精度 (%)	再現率 (%)	F (%)
Toole のシステム	< 58.5	58.5	< 58.5
提案手法 (方法 3)	64.4	54.2	58.9
提案手法 (方法 4)	70.7	54.2	61.4
22 タイプの略語のタグを未知語に修正した後			
提案手法 (方法 3)	61.8	63.4	62.6
提案手法 (方法 4)	66.3	62.6	64.4

Toole はシステム全体の精度を計算していないが、略語の検出ステップで数個の誤りがあることを考慮すると、数パーセント再現率よりも低い値であると考えられる。提案手法と Toole の結果を比較すると、提案手法の再現率は約 4～5% 低いが、精度は約 6～12% 高く、F-measure で比較すると約 0～3% 高い結果となっている (表 2.5)。

Rowe らは、ユーザとのインタラクティブなシステムを提案している [41]。ユーザは、システムが出力する復元語が正しいかどうかの判断を行ない、誤っていると判断すれば次の復元語をシステムに要求する。システムは最大 5 個までの復元語をユーザに提示することができる。

Rowe らは複合語に関しては 2 語まで認め、写真の表題に使用されている略語を復元するタスクを行なっている。Toole は、Rowe らの実験結果と比較できるように再計算を行なっている [41]。本章のシステムも同様の再計算を行なう。本章のシステムの結果は、先行研究と比較して、再現率で約 10% 低いものの、精度では約 15～60% 高く、F-measure では約 10～45% 高い結果となっている (表 2.6)。

表 2.6: 先行研究との比較 (2)

	精度 (%)	再現率 (%)	F (%)
Rowe らのシステム	14.9	71.1	24.6
Toole のシステム	32.8	69.6	44.6
提案手法 (方法 3)	49.2	59.8	54.0
提案手法 (方法 4)	76.8	58.9	66.7
22 タイプの略語のタグを未知語に修正した後			
提案手法 (方法 3)	46.9	71.0	56.5
提案手法 (方法 4)	73.1	69.2	71.1

形態素解析器が略語を正しく未知語と判定できれば、再現率は Toole の結果と同様の値が得られるものと考えられる (表 2.6 の下の部分)。

次に人間と提案手法の比較を行なう。提案手法のテストに使用したものと同一テストセット (20 レポート) を使用し、航空分野に従事している 3 人の専門家による略語復元との比較を行なう。専門家による略語復元の結果の精度は 85.7%～94.9% で、再現率は 50.4%～78.5% の範囲となっている。F-measure は、65.8%～85.5% で平均 F-measure は 77.7% となっている。本章のシステムで最も結果のよい方法 4 の場合の F-measure 値は、最も低い人の F-measure 値より約 4% 低く、平均の F-measure 値よりも約 16% 低い結果となっている。この結果から、提案手法は専門家には劣るものの、かなり高い性能が得られたと云える。

2.7 議論

本章では、ASRS コーパスを用いて、略語の復元を行なっている。航空分野では略語の使用頻度が高く、サンプリングテストの結果、ASRS コーパスには出現語に対して略語が約 11.0% 含まれている。それに対して、新聞の記事を調査した結果では、略語はほとんど使用されていない。車の修理記録、飛行機の整備記録、E-Mail などには比較的多くの略語が使用されていることから、分野によっては略語の使用頻度が高いといえる。

提案手法は、辞書を使用する代わりに「略語の少ないコーパス」を使用して復元語を探索している。「略語の少ないコーパス」は、常に存在するとは限らないのも事実である。その場合には、以下の2つの解決法が考えられる。1つは、ある期間略語を使用しないようにして文書を作成することである。そうすることにより、「略語の少ないコーパス」を自動的に作成することができる。もう1つの方法は、1つのコーパスに略語とその復元語が両方とも含まれているような場合に、このコーパスを「略語の少ないコーパス」及び「略語の多いコーパス」両方として使用することである。この場合頻度情報がどの程度使用できるかについては、略語と復元語の分布によるので、コーパスの頻度情報の特徴は使用できるかどうかは分からない。

表 2.2 の方法 4 のシステムの出力誤りについて、詳しく解析をしてみる。82 タイプのシステム出力の内、58 タイプは正解、24 タイプは不正解となっている。24 タイプの不正解の内、13 タイプは間違っ異なる略語に復元した略語であり、11 タイプは、略語でない未知語を間違っ復元したものである。間違っ異なる略語に復元した 13 タイプの中で、2 タイプは語尾変化による誤りで、例えば、“DEG”を“degrees”というように復元している。一方、“DEGS”は正しく“degrees”に復元できている。語尾変化による誤りは、意味的には正しい語を選んだものの、語尾変化までは正しく判別できていないものである。このような誤りについては、例えば、略語の原形を抽出し、この情報を利用することにより誤り率を減少できるのではないかと考えられる(この例の場合には、“DEG” – “degree”を抽出する)。略語でない未知語を間違っ復元した 11 タイプの内訳は、5 タイプは頭字語との誤り、5 タイプは空港コードとの誤り、1 タイプは形態素解析器に登録されていない語となっている (RETUNED)。

システムが略語として認識しない略語の内、4 タイプは「略語の少ないコーパス」に出現しないもので、他の 2 タイプは復元語は存在するものの、その出現頻度が低いためにシステムが略語として認識しないものである。提案手法のテストセットに対する略語とそれに対応する復元語の出力結果を表 2.8 に示す。

ウインドウ内に出現する語の文脈情報を略語の復元候補語のランク付けに使用する有効性について調べる。そのために、表 2.2 の方法 3 で文脈情報を使用する場合と文脈情報を使用しない場合との結果の比較を行なう (表 2.7)。文脈情報を使用する場合には、文脈情報を使用しない場合に比べて、再現率は約 6%低い、精度は約 39%高く、F-measure で比較すると約 23%高い。この結果により、文脈情報は精度向上に有効であると結論できる。

表 2.7: 文脈情報を使用する場合と文脈情報をしない場合の方法 3 の結果の比較

	精度	再現率	F
	(%)	(%)	(%)
方法 3 - 文脈情報あり	64.4	54.2	58.9
方法 3 - 文脈情報なし	25.9	59.8	36.1

未知語のタグ情報については不正確なので、フィルターとしては使用するが、特徴としては使用していない。

形態素解析器が、未知語のタグを付与しない 22 タイプの語を人手で未知語に修正している (AFT, ALT, APPROX, CA, COM, COMP, CRS, DES, INFO, LAV, MI, MIN, MINS, N, NE, NW, PAX, POS, REF, S, VIS, W)。これらの内いくつかは、汎用辞書に登録されている語であるが、すべての可能な場合についてシステムの性能をテストするために、22 タイプの語すべ

てを使用している。また、この中である語は辞書に登録され、他の語は辞書に登録されていないというようにバラツキがあり、略語と辞書登録語は明確に境界を定められないものである。

方法4と方法5では、分野特有の知識を使用している。実際にこのシステムを使用する場合には、最初に分野特有の知識から略語でないものを排除し、その後方法1から方法3を適用する方法も考えられる。

2.8 まとめ

提案手法は、復元候補語をコーパスから選別し、その後文脈情報を利用し、ベクトル空間法を使用して類似度を計算している。

従来の略語の復元手法は、主に略語と復元候補語の文字情報のみを使用しているが、精度的に限界がある。例えば、略語“S”の復元語は“south”であるが、“S”は文字を1個しか含まないので、含まれている文字情報は最小である。したがって、文脈情報を利用しない方法では略語復元は不可能である。提案手法では、文脈情報を利用することにより2,725の復元候補語の中から正しく“south”を選択できている。

略語の復元候補語を探索するのに辞書の代わりにコーパスから探索する手法を提案している。コーパスは、略語が多く含まれる「略語の多いコーパス」と復元語が多く含まれる「略語の少ないコーパス」を使用している。「略語の少ないコーパス」を辞書の代わりに使用しているが、形態素解析器の辞書登録語として約7,700タイプの語が含まれている。他方 Toole の使用している辞書には約35,000語 [54]、Rowe らの使用している辞書には約29,000語が登録されている [41]。提案手法で使用している7,700タイプの語を辞書登録語と考えると、先行研究で使用している辞書登録語と比較して、登録語を約22～27%に減らすことができている。この結果、復元候補語の範囲が絞られ高い精度と再現率が得られたと考えられる。つまり、略語と同じ分野の「略語の少ないコーパス」を使用することにより、分野特有の辞書を用意したのと同じ効果が得られたと考えられる。

略語の使用は、E-Mailのように比較的形式にとらわれないようなもので頻繁に使用され、略語も多種多様なものが使用されている。また、飛行機と地上間を電文で通信するACARS (Aircraft Communications Addressing and Reporting System) では、飛行機の機内で電文を作成する環境上の制限から、文を短く簡潔にするために略語が頻繁に使用され、また略語表を参照することも現実的でない。このような略語の使用法も多く、今後ますます略語の自動復元の需要は大きくなると考えられる。

本章では、単一語からなる略語の復元を行なったが、複数の語からなる略語についても同一の手法を適用できると考えられる。複数の語からなる略語の場合に、復元語を探索する場合数が膨大になる点については、複数の語からなる略語は品詞が名詞であるので、名詞フィルターを使用して復元候補語の探索数を絞ることができるのではないかと考えられる。

文脈情報の使用は分野／言語によらない。提案手法の方法1から方法3までは、英語の一般的な特徴のみを使用している。したがって、他の分野への移植は簡単にできると考えられる。

また、本手法の応用については、日本語で外来語を表すのにカタカナをよく使用するが、本章での手法は、カタカナを区別するのにも有効であると考えられる。阿玉は、カタカナとそのカタカナに対応する英語の表現が混在する文書において、カタカナを英語に変換するタスクを行なうのに本章で提案した手法を使用している [59]。カタカナに対応する英語がコーパスに含まれる場合に高い精度 (約84%) が得られたと報告している。

同じ略語が複数の復元語に対応する問題に関しては、各略語をタイプとしてではなく、個別に

処理する方法が考えられる．ただし，その場合には略語の周囲から得られる文脈情報が少なくなるので，タイプとして得られた文脈情報とかけ離れた文脈情報の未知語についてのみ個別に処理するような方法が考えられる．

表 2.8: 表 2.2 -手法 4 の結果と正解リストとの比較 (*は、間違って復元されたもので括弧内が正しい復元語, **は、略語の復元語が出力されなかったもので右側が正しい復元語を示す.)

1: **A/C aircraft	2: ACFT aircraft	3: **AFT after
4: AIRSPD airspeed	5: **ALT altitude	6: APCH approach
7: APCHING approaching	8: **APPROX approximately	9: ARPT airport
10: ARPTS airports	11: ATTN attention	12: *BOS bolts (boston)
13: **CA carifornia	14: CAPT captain	15: CHK check
16: CHKLIST checklist	17: CLB climb	18: CLRD cleared
19: CLRED cleared	20: CLRLY clearly	21: **CLRNC clearance
22: **COM communication	23: **COMP company	24: *COMS components (communications)
25: *COORD colored (coordination)	26: COPLT co-pilot	27: **CRS cruise
28: CTL control	29: **CTTABLE controllable	30: CTLED controlled
31: CTLR controller	32: CTLRS controllers	33: *DEG degrees (degree)
34: DEGS degrees	35: DEP departure	36: **DES descent
37: DEST destination	38: **DIR direct	39: DSCNT descent
40: EBND eastbound	41: EMER emergency	42: *FAC factors(facility)
43: **FDRS feeders	44: FLT flight	45: **FREQ frequency
46: FREQS frequencies	47: GND ground	48: HDG heading
49: *IDENT accident (identify)	50: INBND inbound	51: **INFO information
52: INOP inoperative	53: INTXN intersection	54: KT knot
55: KTS knots	56: **LAV lavatory	57: LCL local
58: LNDG landing	59: *LOC location (localizer)	60: *MAINT maintained (maintenance)
61: **MAX maximum	62: **MI mile	63: **MI minute
64: **MIN minute	65: **MINS minutes	66: MGMNT management
67: **N north	68: NAV navigation	69: *NBOUND inbound (northbound)
70: **NE northeast	71: **NW northwest	72: **OPP opposit
73: OVCST overcast	74: **PAX passenger	75: PLT pilot
76: PLTS pilots	77: **POS position	78: PROB problem
79: PROX proximity	80: *PVT prevent(private)	81: PWR power
82: **REF reference	83: RPT report	84: RPTED reported
85: RPTS reports	86: RWY runway	87: RWYS runways
88: **S south	89: *SEP sheep(separation)	90: SPD speed
91: **SSW south-southwest	92: **SUPVR supervisor	93: **SVC service
94: *SW snow(southwest)	95: SYS system	96: TFC traffic
97: TKOF takeoff	98: TURB turbulence	99: TWR tower
100: *TWRS towards (towers)	101: **VIS visibility	102: **W west
103: **W/O without	104: WKS weeks	105: **WX weather
106: XING crossing	107: XPONDER transponder	

第3章 名詞化構文の処理

第2章では、テキスト処理の障害となる略語の処理を述べた。本章では、もう1つの問題の名詞化構文 (Nominalization) の処理について述べる。本章での提案手法は、意味辞書を使用しないで、コーパスから名詞化構文に対応する節表現の格構造を抽出して、その統計情報を利用して名詞化構文を自動的に述語-項構造に変形するものである [51]。

3.1 名詞化構文

名詞化構文とは、名詞句の主要部の名詞 (Head Noun) と節表現の動詞が対応しているものをいう。英語では名詞中心の表現が多用される傾向がある。Lees は、名詞化構文を以下の6つに分類し、名詞句と節表現の統語的、内部構造上の類似性を記述している [60]。

1. 事実名詞化 - 例: That he came was obvious.
2. 動名詞名詞化 - 例: His eating vegetable is surprising.
3. 不定詞名詞化 - 例: For him to go there is strange.
4. 行為名詞化 - 例: The committee objects to John. ⇒ The committee's objection to John.
5. 行為者名詞化 - 例: She baked this cake. ⇒ She was the baker of this cake.
6. 抽象名詞化 - 例: ... the man is clever to go ⇒ ... the man's cleverance in going ...

上の分類1～分類3は、THAT 節、WH 節、TO 不定詞節、ING 節が関与する変形で、節表現の動詞を名詞句の主要部の名詞に変形させるものではない。分類4～分類6は、動詞を名詞に変形する過程が含まれている。Lees の名詞化は、分類1～分類3が名詞句と節表現の分布上の類似性を示すものであり、分類4～分類6は内部構造上の類似性を示すものである。後者を語い論的問題という。

Chomsky が対象とした名詞化は、語い論的問題の名詞句で次の2つである [60]。

7. 動名詞名詞句 - 例: John criticized the book. ⇒ John's criticizing the book.
8. 派生名詞句 - 例: John criticized the book. ⇒ John's criticism of the book.

動名詞名詞句は、Lees の動名詞名詞化と同じであるが、派生名詞句は、Lees の行為名詞化の内接尾辞が “ing” でないものと抽象名詞化を含んだものである。Chomsky は、動名詞名詞句と派生名詞句との違いについて、生産性、主要部に対応する動詞の間の外形的／意味的關係、内部構造の違いについて注目し、派生名詞句はそれぞれの項目につき生産的でなく、個別の派生名詞で異なり、名詞句の内部構造を持つと指摘している。生産性では、動名詞名詞句は、ほとんど制限なしに生成できるが、派生名詞句の生成には多くの制限がある。

9. John amused the children with his stories.
10. John's amusing the children with his stories.
11. *John's amusement the children with his stories.

どうして例 9 から例 10 が適格であり、例 11 が不適格であるかの判断は難しい。

本章では、イベント表現が名詞句で表現される現象を取り扱う。Chomsky の分類では派生名詞句の中で抽象名詞化を除いたものである。テキストを構文解析してその解析結果から述語-項構造を抽出するタスクを考える。名詞化構文は、構文解析すると名詞句として出力されるので、同じ意味の節表現に変換するか、あるいは、対応する節表現を名詞句に変換する必要がある。

例えば、例 12 は、動詞 “destroy” によりイベントを表現している。

12. The army destroyed the city.

例 12 と同じ内容を表現するのに名詞句を使用すると、例 13、例 14 のような複数の名詞句表現が可能である。

13. The destruction of the city by the army.

14. The army's destruction of the city.

テキストから自動的にイベントを抽出するようなタスク（例：情報抽出、テキストマイニング）の場合には、これらの名詞句と節表現を同定できなければならない。上記例で見られるように、名詞化構文では、複数の表現が可能である。本章ではテキスト表現の正規化を目的としているので、名詞化された表現を節表現に対応する述語-項構造に変形することとする。

イベント表現の抽出では、節表現の部分を構文解析し、解析結果から述語-項構造を抽出するという方法を取る。このようなタスクは、通常行なわれるものである [6]。名詞化構文からも節表現と同様に述語-項構造を抽出する技術、或いは、名詞化構文を節表現に変換して、その後述語-項構造を抽出する技術が必要である。本章では、後者の技術について述べる。

本章では、コーパスを使用して自動的に名詞化構文を述語-項構造に変形する手法を提案する。この手法は、名詞化構文の動詞派生名詞に対応する動詞の節表現をコーパスから探し、その統計情報を使用して、名詞化構文を述語-項構造に変形するものである。コーパスの節表現を使用するので、動詞派生名詞と対応する動詞の間の意味的關係は自動的に織り込まれると考えられる。

名詞化構文に対応する可能な節表現の格要素を数えあげ、その頻度の合計の最も大きなものを名詞化構文に対応する節表現とする。なお、この処理は名詞句の同定が必要なので、テキストを構文解析した後に行なう。

コーパスは、第 2 章と同じ ASRS コーパスを使用する。第 2 章では、Narrative を使用したが、本章では Narrative の要約である Synopsis を使用する。

以下の節では、3.2 節で名詞化構文の特徴と関連する先行研究について述べる。3.3 節では、本章で提案する名詞化構文から述語-項構造に変形する手法について述べる。3.4 節では、実験とその評価について述べ、名詞化構文から述語-項構造に変形する場合の精度について述べる。3.5 節では、本手法の応用について、3.6 節では、まとめと今後の課題について述べる。

3.2 名詞化構文の特徴

一般に「名詞」は、持続するもの、具体的なもの、抽象的なもの等の静的状態を表すものと特徴付けられる。一方「動詞」は、動作、行動、状態変化等の動的な状態を表すものと特徴付けられる [39]。Quirk らによると、名詞化構文は動詞を含む節表現に体系的に対応する名詞句で、名詞句の主要部は統語的に動詞に関係する [39]。英語では名詞化は頻繁に起こるので、その解釈はテキストの解析には重要である。

3.2.1 名詞化構文の特徴

本章で対象とする名詞句は、次のようなものである。

「名詞句」：Aircraft takeoff without clearance.
限定部 主要部 補部

名詞句は、次の2つに分けられる。

単純名詞句 = 限定部 + 主要部
複雑名詞句 = 単純名詞句 + 補部

名詞化構文は、節表現と以下の様に様々な点で異なっている [21, 22].

- 動詞の同じ格は1回しか使用できない（唯一性の制限）.
- 名詞句の限定部では、主語と目的語はこの順でなければならない（順序制限）.
- 動詞の格の出現は任意で順序も制限が緩い.
- 名詞句の限定部で表される所有格は主語になる可能性が大きい.

次の例を考えてみる（簡便のため冠詞は省略してある）.

15. Aircraft penetrated space.

例 15 に対応する可能な名詞化構文は以下の様なものがある.

- 16. Aircraft penetration of space.
- 17. Aircraft space penetration.
- 18. Space penetration by aircraft.
- 19. Aircraft penetration.
- 20. Space penetration.
- 21. Aircraft's penetration of space.

例 16 において、“space” は目的語になっているので、“aircraft” は目的語になれない（唯一性の制限）. 例 17 と例 18 では、主語と目的語が倒置されている. 例 19 では目的語が省略され、例 20 では主語が省略されている. このように、動詞の格の出現は任意で順序も制限が緩い. 例 21 において、“aircraft” は所有格になっているので、主語になっている.

例 22 の名詞化構文は順序制限のために許されない（限定部では、主語 “aircraft” と目的語 “space” はこの順序でなければならない）.

22. *Space aircraft penetration.

これ以外にも各動詞派生名詞により、生産的制限や目的語が必須である等の制限がある.

3.2.2 先行研究

次に、先行研究について述べる. 名詞化構文を計算機で取り扱う手法は、知識ベースのものとコーパスベースのものに分けられる.

```

(defend
  ((object
    ((if% is-a obj location)
      (verbal-concept-is defend-physical-thing))
    ((if% is-a obj idea)
      (verbal-concept-is defend-idea))))
  (prep
    (in((if% equal obj-of-prep court)
      (verbal-concept-is legal-defend)))))
  (end-of-clause
    ((if% is-a obj championship)
      (verbal-concept-is defend-championship))
    (t(verbal-concept-is defend)))))

```

図 3.1: Hull らの意味辞書

Dahl らは、海軍の艦船の故障通報の解析に知識ベースの名詞化処理を使用している。Dahl らは、統語解析を行なってテキストを述語-項構造に変換し、その後意味的解析を行なっている [8]。例えば、動詞 “fail” を、“become inoperative with patients” のように変換している。名詞化構文の処理は、最初に動詞派生名詞に対応する動詞の格の処理を行ない、名詞化構文特有の名詞句内の処理と照応の処理を行なっている。意味的な役割は、統語的な要素もしくは文脈から補完している。

Hull らは、動詞を中心にして作られた辞書を名詞化構文に適応できるようにしている。同じ名詞でも動詞的用法とそうでないものがあるので、名詞句内の限定部と補部を調べている [15]。名詞句の主要部の名詞化の可能性については、主要部の名詞の概念が WordNet の上位概念に “actions” または “events” として記載されているかにより調べている [25]。もしこれに該当する場合には、その名詞の用法に名詞化があるといえる。本章の手法も Hull らのこの手法を参考にしてしている。その後、名詞句の処理、前置詞句の処理、最終処理を行なっている。各処理には、名詞化処理を含まない意味辞書を用い、動詞の格要素の対応を調べている。例えば、“defend” には図 3.1 のような記述がある。この手法では、図 3.1 の例で分かるように各動詞に意味記述をしておく必要がある。

Macleod らは、NOMLEX と呼ばれる英語名詞化辞書を作成している [21]。NOMLEX には、動詞の補語とその制限が記述されている。NOMLEX は IE の精度を向上させる目的で節表現から全ての可能な名詞化構文のパターンを出力できるように人手で作成した辞書である。節表現とそれに対応する名詞化構文のために、1,015 の動詞派生名詞の登録がある。例えば、“deviation” の登録内容は図 3.2 のようになっている。動詞の主語の要素を “VERB-SUBJ” に、名詞に動詞的用法以外の概念があるかを “NOUN ((EXISTS))” に、動詞の下位範疇を “VERB-SUBC” にというように記述している。これらの情報を使用することにより、対象とする動詞を含む節表現を名詞化構文に変形することが可能である。

知識ベースとは異なるコーパスベースの手法もある。Grefenstette らは、名詞化構文に付随するサポート動詞（例：“*make a proposal*”）を抽出するタスクにコーパスベースの手法をとっている。

```

(NOM :ORTH "deviation"
  :PLURAL "deviations"
  :PLURAL-FREQ "not rare"
  :VERB "deviate"
  :NOUN ((EXISTS))
  :NOM-TYPE ((VERB-NOM))
  :VERB-SUBJ ((DET-POSS)
    (N-N-MOD)
    (PP :PVAL ("by" "of"))))
  :VERB-SUBC ((NOM-INTRANS :SUBJECT ((DET-POSS)
    (N-N-MOD)
    (PP :PVAL ("of" "by"))))
    :REQUIRED ((SUBJECT)))
    (NOM-PP :SUBJECT ((DET-POSS)
      (N-N-MOD)
      (PP :PVAL ("of" "by"))))
      :PVAL ("from"))))

```

図 3.2: NOMLEX エントリーの例

る [12]. Grefenstette らは、動詞派生名詞とそれ以外の名詞を区別するために、名詞化構文と節表現の類似性を利用している。そのために、名詞句の主要名詞の後ろの前置詞の中で各主要名詞で頻出する 3 つの前置詞を対象にし、その分布から類似性を計算している。しかしながら、Grefenstette らの使用したのは、前置詞のみである。

知識ベースの手法では、辞書を作成／維持し、新しい分野に移植する場合には辞書の修正／変更が必要になる。このような辞書の修正／変更は、人手と時間が掛かる作業である。コーパスベースの手法では、コーパスから如何にして知識を得るかが難しく、名詞化構文の処理に本格的にコーパスの情報を使用したものはない。

本章ではコーパスベースの手法を採用し、コーパスからの情報を最大限に利用できるよう試みる。そのために、動詞派生名詞に対応する動詞の格要素をコーパスから収集し、その分布を調べる。簡便のため、節表現において、主語、目的語及び動詞の右に最初に出現する前置詞句を対象にする。名詞化構文の要素と節表現の対応を取るために、3.2 節で述べた名詞化構文の特徴を使用する。

3.3 方法

3.3 節では、名詞化構文から述語-項構造への変形アルゴリズムについて述べる。最初に前処理について述べ、その後変形アルゴリズムの詳細を述べる。コーパスとして、ASRS コーパスの Synopsis を使用する。ASRS コーパスの Synopsis の統計情報を表 3.1 に示す。

表 3.1: ASRS コーパスの Synopsis 部分の統計情報

ASRS コーパス	
文書総数	24,626
1 文書の平均語数	15.4
1 文の平均語数	10.9
1 文書の平均文数	1.4

3.3.1 略語の復元

第 2 章で述べたように，ASRS レポートには出現語に対して約 11%の略語が含まれている．略語の復元には第 2 章の手法を適用し，略語のタイプ別に復元語を生成し，その後人手で修正を加える．その結果，293 個の略語リストを作成した．このリストを使用して略語を復元する．例えば，“Flight crew was *clred* for a visual approach.” は，“Flight crew was *cleared* for a visual approach.” に復元する．略語は，そのまま残しておいて構文解析を行なうと，構文解析器の辞書に登録されていない未知語として処理され，構文解析の結果に誤りが多くなると考えられる．したがって，構文解析を行なう前の段階で略語の復元を行なう．

頭字語については，出現語に対して約 6.4%含まれている．そのタイプ数は比較的少なく 198 個である．頭字語もそのまま残しておいて構文解析を行なうと，構文解析の結果に誤りが多くなると考えられるが，頭字語の品詞は全て名詞なので，頭字語の復元は行わず，構文解析器の辞書に頭字語の品詞情報だけを与えることとする．頭字語は，複数の語が集まってある概念を表すことが多く，構成語に分解しない方がよいと考える．また，後述するように名詞化構文の処理過程で限定部の名詞句を分解して主語，目的語，前置詞句の補語の可能性を調べるが，頭字語を構成語に分解すると調べる要素数が増加し誤り率も増えると考えられる．

3.3.2 述語-項構造の抽出

構文解析器として Apple Pie Parser を使用してテキストを構文解析する [47]．Apple Pie Parser は，ボトムアップチャートパーザで，最良優先探索法により最も確率の高い構文木を 1 つ出力する．文法と辞書は，PennTreeBank から自動的に生成されている [27]．

解析結果から名詞句が出力されると，本アルゴリズムが起動する．まず，テキストの正規化として，略語の復元を行ない，その後構文解析を行なう．語の語尾変化を基本形に変形するのは構文解析の後で，略語の復元で使った TreeTagger の情報を使用する [45, 46]．冠詞は，除去する．解析結果から，“主語”，“動詞”，“直接目的語”と動詞の右の最初に出現する“前置詞句”を認識し，述語-項構造を抽出する．もし，文が名詞句のみからなっている場合には，そのまま保持する．

文とそれに対応する述語-項構造の例を次にあげる．

23. 文：Aircraft took off without clearance.

24. 述語-項構造：(Aircraft)_{subj} (without clearance)_{pp} (take-off)_v

25. 名詞化構文：Aircraft takeoff without clearance.

本章の目的は，例 25 のような名詞化構文からも対応する節表現（例 23）と同じ述語-項構造（例 24）を得ることである．

3.3.3 名詞化構文の検出と変形

名詞化構文の変形アルゴリズムは、名詞句の主要部の名詞に動詞派生名詞の可能性があることにより起動する。そのために、主要部の名詞が動詞派生名詞の可能性があるかどうかの判別が必要になる。この目的のために、名詞に複数の意味がある場合でも、その中に1つでも節表現の動詞に対応する動詞派生名詞の意味があれば、その名詞をリストにする。このリストを「動詞派生名詞候補リスト」と呼ぶことにする。

動詞派生名詞を探索するため、ASRS コーパスに出現する全ての名詞に対して、WordNet の上位概念に “actions” 或いは “events” の意味があるかを調べる [25]. もし WordNet の上位概念に “actions” 或いは “events” の意味があれば、この名詞を「動詞派生名詞候補リスト」に登録する. このリスト作成の背景には、名詞の WordNet での上位概念に “actions” 或いは “events” の意味があるということは、その名詞には動詞派生名詞の意味が少なくとも 1 つあるという仮定がある. このアイデアは、Hull らのものと同じである [15]. この「動詞派生名詞候補リスト」に登録されている名詞には、一般に動詞派生名詞とそうでない名詞の両方の意味が混在するが、その判別は後の処理で行なう. 動詞派生名詞に対応する動詞形は、手作業で作成している. このようにして、「動詞派生名詞候補リスト」に 776 個の動詞派生名詞を登録した. NOMLEX の登録語と比較してみると 314 語が重複している. このことは、動詞派生名詞は分野依存しており、分野が異なれば辞書の再構築が必要であることを示唆している.

述語-項構造に変形するのに、名詞句の限定部と補部の中に前置詞句があれば最初に出現した前置詞句だけを対象とする。システムは、名詞句を入力として受けとり、述語-項構造にスコアを付けて出力する。もし、入力が節表現に対応する名詞化構文でなければ、何も出力しない。システムの概観を図 3.3 に示す。

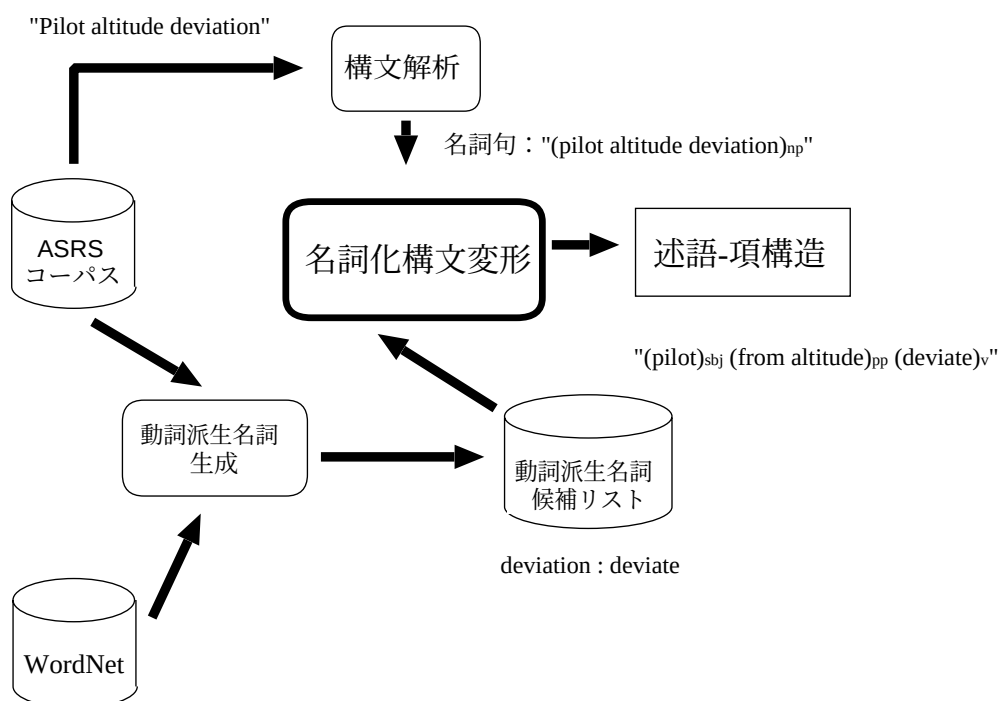


図 3.3: 名詞化構文変形システム概観

名詞化構文変形アルゴリズムを以下に示す。

ステップ 1 名詞句の主要部にマークを付ける。マークされた名詞が「動詞派生名詞候補リスト」にあれば、ステップ 2 に進む。なければ、何も出力しないで終了する。

ステップ 2 主要部の直前に所有格があれば、それを *sbj* とマークする。

ステップ 3 前置詞 “by” を主辞とする前置詞句があり、ステップ 2 で主要部に *sbj* とマークされていなければ、その補語の名詞句を *sbj* とし、限定部は *obj* とする仮説をたて、ステップ 6 に進む。

ステップ 4 前置詞 “of” を主辞とする前置詞句があれば、その補語の名詞句を *sbj* として、限定部を *obj* にする仮説と、逆にその補語の名詞句を *obj* として、限定部を *sbj* にする仮説をたてる。もし、ステップ 2 で限定部の名詞句が *sbj* とマークされていれば、補語の名詞句を *sbj* にする仮説を捨てる。ステップ 6 に進む。

ステップ 5 前置詞 “by” 或いは “of” 以外を主辞とする前置詞句があれば、その前置詞句を後の処理のため保持しておく。

(a) 限定部に名詞が 1 個あり、それが所有格でなければ、以下の 3 つの仮説をたてる。名詞に *sbj*, *obj* 或いは *pp* のマークをつける。もし名詞にステップ 2 で既に *sbj* のマークがされていれば、*sbj* のマークをつける仮説を捨てる。

(b) 限定部に名詞が 2 個以上ある場合には、その名詞を 1 個又は 2 個に分割する数を数え上げる。例えば、限定部の名詞が $n_1n_2n_3$ であった場合、可能な分割は次のようになる： $n_1n_2n_3$, n_1/n_2n_3 , n_1n_2/n_3

限定部の名詞に所有格があれば、所有格の直後で分割する。

分割した後の 1 つ又は 2 つの要素がコーパスに 1 度も出現しなければ、その分割仮説は捨てる。

もし分割した結果 1 個の要素になった場合（上の例では、 $n_1n_2n_3$ ）ステップ 5(a) と同様の処理をする。

分割した結果 2 個の要素になった場合、次の 3 つの仮説をたてる：2 個の要素に *sbj-obj*, *sbj-pp* と *obj-pp* という 3 つの仮説をたてる。

obj-sbj という仮説は、3.2 節での順序制限に反するために排除する。所有格があり、その格がステップ 2 の仮説に反する場合には、本ステップでの仮説を捨てる。

ステップ 6 に進む。

ステップ 6 この時点で何も仮説がなければ、何も出力しないで処理を終る。それ以外の場合、マークされた要素の出現頻度をコーパスから数え上げる。出現頻度を数え上げる場合に、主要部の名詞に対応する動詞を「動詞派生名詞候補リスト」から探し、その格要素がこれまでのステップでマークされているかを調べる。要素が複合名詞の場合には、最右名詞の出現頻度を数え上げる。

前置詞句 *pp* を数える場合には、各前置詞について個別に出現頻度を数え、その最大頻度を前置詞句 *pp* の出現頻度とする。もし、前置詞句がなくステップ 5 で保持された前置詞句がある場合には、それを前置詞句 *pp* とする。仮説の中で要素がコーパスで出現しないものは排除する。

ステップ 7 仮説の各要素の和を計算し、スコアが最大値の仮説を出力する。

本章のアルゴリズムの根拠は、名詞化構文はその主要部の名詞に対応する動詞の節表現の格要素を参照することにより、述語-項構造に変換できるというものである。つまり、主要部の名詞に対応する動詞の節表現の格に出現する要素と名詞化構文内での分割した要素と統計的に共通なものを選択する。場合数を増やすために比較には最右名詞を使用している。ただし、あり得ない限定部での分割を排除するために、コーパスに1度も出現しないような要素は排除している。個別の動詞派生名詞の意味的制限は、コーパスの節表現を参照することにより自動的に獲得できると考える。仮説が複数存在する場合には、最大値の仮説を選ぶ。

いくつかの例をあげながら、アルゴリズムを説明する。

26. 名詞句：Commuter smt¹ propeller strike.

名詞化構文変形アルゴリズムは、例 26 が名詞句であることから起動し、名詞句の主要部の“strike”が「動詞派生名詞候補リスト」に載っているのを確認して始まる（ステップ 1）。補部に前置詞句がないのでステップ 5 に進み、限定部の分割を試みる。例 26 の場合の可能な分割は、“commuter smt propeller”, “commuter smt/propeller” と “commuter/smt propeller” になる。上記の分割から可能な仮説は次のようになる（*p* は、省略された前置詞を示す）。

仮説 a：(Commuter smt propeller)_{subj} (strike)_v

仮説 b：(Commuter smt propeller)_{obj} (strike)_v

仮説 c：(Commuter smt propeller)_{pp} (strike)_v

仮説 d：(Commuter smt)_{subj} (propeller)_{obj} (strike)_v

仮説 e：(Commuter smt)_{subj} (*p* propeller)_{pp} (strike)_v

仮説 f：(Commuter smt)_{obj} (*p* propeller)_{pp} (strike)_v

仮説 g：(Commuter)_{subj} (smt propeller)_{obj} (strike)_v

仮説 h：(Commuter)_{subj} (*p* smt propeller)_{pp} (strike)_v

仮説 i：(Commuter)_{obj} (*p* smt propeller)_{pp} (strike)_v

“Commuter smt propeller” と “smt propeller” は、コーパス中に1回も出現していないので、仮説 a, 仮説 b, 仮説 c, 仮説 g, 仮説 h と仮説 i をステップ 5(b) の段階で棄却する。仮説 d は、動詞“strike”に対して、“propeller”が目的語として1回、“smt”を最右名詞とする名詞句“acr smt”が主語として1回出現しているので、スコア 2 を獲得する。“propeller”は、前置詞句の補語の最右名詞として1回も出現していないので、仮説 e と仮説 f を棄却する。最終的に、1つだけ残った仮説 d を出力する。3.3.1 項で頭字語は構成語に復元しないこととしたが、もし構成語に復元していた場合、“smt”が“small transport”になり、ステップ 5(b) での限定部の分割数が多くなるという欠点がある。

2 つ目の例は、次のようなものである。

27. 名詞句：Pilot altitude deviation.

¹ “Smt” は、“Small Transport” の頭字語

例 27 が名詞句であり, “deviation” が「動詞派生名詞候補リスト」に載っているのを確認して名詞化変形アルゴリズムが起動するのは例 26 と同様である (ステップ 1). ステップ 5(b) において, 限定部の分割は, “pilot altitude” 及び “pilot/altitude” になる. “pilot altitude” は, 分割した結果が 1 個の要素になっているので, ステップ 5(a) において, 次のような仮説をたてる.

仮説 j : (pilot altitude)_{subj} (deviate)_v

仮説 k : (pilot altitude)_{obj} (deviate)_v

仮説 l : (p pilot altitude)_{pp} (deviate)_v

“pilot altitude” は, コーパス中に 1 回も出現しないので, これらの仮説 j, 仮説 k, 仮説 l を棄却する. “pilot/altitude” は分割数が 2 つなので, ステップ 5(b) において次のような仮説をたてる.

仮説 m : (pilot)_{subj} (altitude)_{obj} (deviate)_v

仮説 n : (pilot)_{subj} (p altitude)_{pp} (deviate)_v

仮説 o : (pilot)_{obj} (p altitude)_{pp} (deviate)_v

“pilot” は動詞 “deviate” の主語の最右名詞として 22 回出現する (“pilot”:13, “ga student pilot”:1, “reporter pilot”:1, “distracted pilot”:1, “smt pilot”:1, “air taxi pilot”:1, “sma pilot”:2, “ga sma pilot”:1, “commuter pilot”:1). “altitude” は動詞 “deviate” の目的語の最右名詞として 2 回出現する (“altitude”:2). したがって, 仮説 m のスコアは 24 になる. “pilot” が動詞 “deviate” の目的語の最右名詞としては, 1 回も出現しないので, 仮説 o のスコアは 0 になる. “altitude” が動詞 “deviate” の前置詞 “from” の補語の最右名詞として 21 回出現する (“clearance altitude”:3, “assigned altitude”:15, “assigned crew altitude”:1, “altitude”:2). “altitude” が動詞 “deviate” の前置詞 “at” の補語の最右名詞として 2 回出現する (“same altitude”:1, “low altitude”:1). したがって, 最大頻度の前置詞は “from” になり, 仮説 n のスコアは 43 になる. 最終的に仮説 n のスコアが最大となり仮説 n を出力する.

次の例は, 前置詞句を含んだものである.

28. 名詞句 : Ga² sma penetration of airspace.

例 28 は, ステップ 4 において, 前置詞 “of” の補語 “airspace” に *subj* 又は *obj* を割り当て, 限定部の “Ga sma” には *obj* 又は *subj* を割り当てる. ステップ 4 では, *subj* の *obj* どちらかは, 前置詞句の補語で与えられるので, 限定部は分割しない. したがって, 仮説は次のようになる.

仮説 p (ga sma)_{subj} (airspace)_{obj} (penetrate)_v

仮説 q (airspace)_{subj} (ga sma)_{obj} (penetrate)_v

仮説 p の方がスコアが高くなり, 仮説 p を出力する.

3.4 評価

名詞化構文変形アルゴリズムの精度について調べる.

² “Ga” は, “General Aviation” の頭字語

3.4.1 名詞化構文変形

「動詞派生名詞候補リスト」に登録された動詞派生名詞の中で出現頻度が上位 20 位までの動詞派生名詞について名詞化構文変形アルゴリズムの評価を行なう。表 3.2 に結果を示すが、「ランク (派生名詞)」は動詞派生名詞候補リストの中での動詞派生名詞の出現頻度のランク、「頻度」は動詞派生名詞の出現頻度、「ランク (全体)」はコーパスの全名詞に対する動詞派生名詞の出現頻度のランクを示す。表 3.2 に示すように出現頻度上位 20 位までの動詞派生名詞は、全体の名詞の中の出現頻度で 50 位に入っている。したがって、名詞化構文の処理は重要であるといえる。

ほとんどの「動詞派生名詞候補リスト」に登録された名詞には、動詞的意味と非動詞的意味がある。動詞的意味とは、節表現において動詞派生名詞に対応する動詞が動詞派生名詞と同じ意味を持つものである。非動詞的意味とは、そのような対応がないものである。例えば “operational deviation” の “deviation” は非動詞的意味で、対応する節表現がない。言い換えると、3.3.3 項で述べたアルゴリズムは、この対応を探しているのである。

「動詞派生名詞候補リスト」の上位 20 位にランクされた動詞派生名詞について、各々 20 個ずつサンプルを取り、名詞化構文変形アルゴリズムを評価する。「用法 動詞的意味:非動詞的意味」は、各動詞派生名詞の動詞的意味と非動詞的意味の用法の配分を示し、その違いは人間が判断した。表 3.2 は、動詞派生名詞の動詞的意味が平均 54%の精度で正しく述語-項構造に変形でき、また、非動詞的意味の名詞の同定については、平均 85%の精度であることを示す。

名詞化構文変形アルゴリズムの誤りの主な原因は 2 つある。1 つは、動詞派生名詞に対応する節表現がコーパスに存在しないというものである。例えば、“Atx takeoff” は正しく述語-項構造 “(Atx)_{subj} (take off)_v” に変形出来ないが、その理由は、“atx” を最右名詞に持つ名詞句が動詞 “take off” の主語としてコーパスに存在しないからである。これは、本章の提案アルゴリズムの限界を示すものである。

もう 1 つの誤りの原因は、形容詞が動詞派生名詞を修飾している場合に、名詞句に対応する節表現において、必ずしも形容詞が副詞に対応しないというものである。例を以下に示す [7, 60]。

29. 名詞化構文: John’s sudden refusal.

30. 節表現: John refused suddenly.

31. 名詞化構文: John’s curious resemblance to Bill.

32. 節表現: *John resembles Bill curiously.

例 29 において動詞派生名詞 “refusal” を修飾している形容詞 “sudden” は、例 29 に対応する節表現の例 30 では、副詞 “suddenly” に対応しているが、例 31 において動詞派生名詞 “resemblance” を修飾している形容詞 “curious” は、例 31 に対応する節表現の例 32 では、副詞 “curiously” に対応していない。この原因は、動詞派生名詞を修飾している形容詞とそれに対応する節表現での形が各形容詞毎に決まっているからであると考えられる。そこで、その対応リストを人手で表 3.3 のように作成する。このコーパスでは、動詞派生名詞を修飾する形容詞の数は少ない。

動詞派生名詞を修飾している形容詞が存在し、その形容詞が表 3.3 のリストにある場合にはその形容詞に対応する表現をステップ 5 での *subj*, *obj*, *pp* の要素に割り当てないで、前置詞句とし、もし前置詞句が既にある場合にはそれを削除するように 3.3.3 項のアルゴリズムを変更する。名詞化構文変形アルゴリズムを修正した場合の精度は表 3.2 の括弧で示される値である。この修正により、動詞派生名詞の動詞的意味が平均 71%の精度になり、動詞的意味、非動詞的意味全体で平均 78%まで精度を上げることができている。

上記 2 つ以外の誤りには、構文解析器のエラーがある。構文解析器のエラーの修正については、

表 3.2: 名詞化変形の結果

ランク (派生 名詞)	頻度	ランク (全体)	動詞派生名詞	用法 動詞の意味: 非動詞の意味	精度 動詞的 意味	非動詞的 意味	計
1	5,232	2	flight	3 : 17	67%	65%	65%
2	3,518	5	pilot	0 : 20	N.A.	90%	90%
3	3,439	6	approach	12 : 8	42%(58%)	88%	60%(70%)
4	2,729	9	landing	14 : 6	43%(86%)	100%	60%(90%)
5	2,299	11	deviation	7 : 13	86%	100%	95%
6	2,288	12	takeoff	18 : 2	56%(72%)	100%	60%(75%)
7	2,024	13	separation	0 : 20	N.A.	100%	100%
8	1,204	17	error	0 : 20	N.A.	100%	100%
9	1,182	18	descent	12 : 8	58%(92%)	88%	70%(90%)
10	1,129	19	departure	5 : 11	40%	36%	38%
11	799	27	control	0 : 10	N.A.	80%	80%
12	777	28	climb	13 : 7	31%(38%)	71%	45%(50%)
13	696	32	land	17 : 3	53%	100%	60%
14	560	39	track	N.A.	N.A.	N.A.	N.A.
15	552	40	failure	20 : 0	70%	N.A.	70%
16	535	41	restriction	0 : 11	N.A.	100%	100%
17	521	43	loss	19 : 1	73%(84%)	100%	75%(85%)
18	513	44	penetration	20 : 0	40%(85%)	N.A.	40%(85%)
19	489	47	communication	3 : 10	66%	90%	85%
20	455	50	turn	1 : 8	0%	50%	44%
計				48% : 52%	54%(71%)	85%	70%(78%)

表 3.3: 動詞派生名詞を修飾する形容詞の節表現での対応表

形容詞	節表現で対応する表現
unauthorized	without authorization
late	late
premature	prematurely
unstabilized	without stabilization
uncoordinated	without coordination
momentary	momentarily
partial	partially
possible	possibly
superfluous	superfluously

3.5 節で述べる.

3.5 議論

3.5.1 構文解析器のエラー修正

実験の解析結果を見ると、テキストの中で名詞句の後ろに主語が続いているような場合に名詞句と主語が1つの名詞句を形成してしまい、本来区切られるべきところで区切られないというエラーがよく起こっている.

33. Acr mlg³ altitude deviation crossing restriction not met.

構文解析器は、例 33 を解析して次のような解析結果を返す.

34. 解析結果：(acr mlg altitude deviation crossing restriction)_{sbj} (not met)_v

しかしながら、正しい解析結果は解析結果 35 である.

35. 正しい解析結果：(acr mlg altitude deviation)_{np}
(crossing restriction)_{sbj} (not met)_v

正しい解析結果と比較してみると、解析結果 34 は動詞派生名詞とその後ろの名詞が結合しているのが分かる. したがって、このようなエラーを修正するために、3.3.3 項のアルゴリズムを修正して、名詞化変形アルゴリズムの起動を名詞句の主要部の名詞だけでなく、限定部の中の名詞でも「動詞派生名詞候補リスト」に載っている場合に起動するように変更する. 限定部の中の動詞派生名詞の場合には、誤った仮説をたてる可能性が大きいため、動詞派生名詞かどうかを決定する閾値であるスコアを 20 に設定する. もし閾値をスコアを 1 のままにしておくと、“acr flight crew path deviation” のような名詞句を名詞化構文変形アルゴリズムに入れると、“flight” が動詞派生名詞の候補で、(acr)_{sbj}(fly)_v のように変形されてしまう.

動詞派生名詞とその後ろの名詞が結合して正しい述語-項構造が得られない 20 例についてテストを行なう. その結果、14 例については、正しい結果が得られる (精度 70%). 3 例に関しては、名詞化構文変形アルゴリズムの仮定と矛盾するものである.

36. 文：Ga sma unauthorized penetration tca.

例 36 では、“penetration” が動詞派生名詞であるが、その対応する動詞 “penetrate” の目的格 “tca” が前置詞なしで動詞派生名詞の後ろに来ている. この原因は、おそらく前置詞 “of” を “penetration” の後ろに付け忘れたものと思われる. 2 例のエラーは、構文解析器によるエラーである. 1 例のエラーは、2 つの動詞派生名詞を含む名詞句が等位接続詞 “and” により結合しているもので、提案アルゴリズムでは処理出来ないものである (“nose gear collapse and wing tip damage”).

3.5.2 移植性

本章の提案手法の移植性を調べるために、1989 年の WSJ (Wall Street Journal) コーパスについて実験を行なう. WSJ コーパスの構成は、541,910 語から成り、容量は約 2.9 メガバイトで

³ “Mlg” は、“Medium Transport” の頭字語

表 3.4: WSJ コーパスの名詞化変形

動詞派生	精度
loss	40%
decline	10%
decision	20%
proposal	20%

ある。WSJ コーパスで頻出する動詞派生名詞 “loss”, “decline”, “decision” と “proposal” について、10 例ずつテストを行なう。

表 3.4 にその結果を示す。平均の精度は、約 23% である。精度が低い理由は、主に構文解析器によるエラーである。約 55% の誤りが構文解析器によるエラーで、約 20% はコーパスが名詞化構文に対応する節表現を含んでいないためである。その他の原因は提案アルゴリズムが関係代名詞を取り扱わないためだと考えられる (例: “Mr. Spoon said ... that has reported *decline* in operating profit ...”).

3.6 まとめ

本章では、コーパスを利用して名詞化構文から述語-項構造への変形手法を提案した。航空分野における安全レポートを対象にして、頻度上位 20 個の動詞派生名詞について実験したところ、平均で約 78% の精度が得られた。名詞句の限定部の要素数が多い場合でも、限定部の分割情報に対応する節表現の格構造の統計情報から得ることにより比較的正しく変形できている。先行研究の多くは、動詞派生名詞の動詞に対応する格構造を決定するのに意味辞書を使用している。しかしながら、意味辞書はその作成／維持管理に人手と時間が掛かる上、他の分野に移植する場合に意味辞書を再構築しなければならないという欠点がある。本章で提案する手法によりコーパスからの情報のみでもある程度の精度を得ることができることが判明した。故障等を記述した文書については、本手法を実際に使用できる程度の精度が得られていると推測されるが、WSJ コーパスでの実験結果の精度は低い。解決策として、精度のよい構文解析器の使用、関係代名詞に含まれる情報の利用について検討する必要がある。意味辞書が既に構築されているような分野では、本手法と統合することで、精度を向上できる可能性があると思われる。

第4章 因果関係の検出

因果関係は、ある状態／出来事の原因と結果を表すもので、様々な意味関係 (Semantic Relations) の中でも重要なものの1つである。本章では、テキスト中に記述されている因果関係を表層的な手掛かりから抽出する手法について述べ、その手法を使用して ASRS レポートから因果関係を抽出する手法について述べる [51]。その後、シーケンシャルマイニング手法である PrefixSpan の説明を行ない、PrefixSpan を用いて ASRS レポートから抽出した因果関係をマイニングした結果を示す。第3章で提案した名詞化構文の変形がテキストマイニングにどのように影響するのかについても検討する。

4.1 因果関係

テキストに表現されている因果関係は、

- 1文／複数の文
- 自明な表現／暗黙の表現

のように分類できる。暗黙の表現から因果関係を抽出するのは難しく、現在の技術では対象分野を絞り分野独特の知識ベースを用意しなければ不可能である。

知識ベースを使用しないで、自明な表現の因果関係を表層上のパターンから抽出する研究もある [16, 17]。この手法は、知識ベースは必要なく他の分野に適応するのも比較的簡単である。英語での因果関係の表現について 4.1.1 項で述べる。

4.1.1 因果関係の表現

Girju らによると、因果関係の表現には、次のようなものがある [11]。

1. 自明な表現

(a) 因果接続詞

- i. 副詞的リンク：2つの節を繋ぐもので次のようなものがある。

例：for this reason, the result that

- ii. 前置詞：例：because of, thanks to, due to

- iii. 従属節

A. 結果接続詞：例：because, as, since, for, so, so that

B. 構造的リンク：ing-clause

C. 相関比較級：so...that...

- iv. 節統合リンク：例：“The reason was ...”, “...is due to...”

(b) 因果動詞

- i. 単純因果動詞：例：cause, lead to, bring about, generate, make, force, allow など

ii. 結果因果動詞：例：kill, melt, dry, break, drop など

(c) 条件：例：if..., then...

(d) 形容詞，副詞：例：fatal, fatally など

2. 暗黙の表現

(a) 複合名詞句：例：cold tremble

自明な表現であっても表層上の手掛かりから因果関係を抽出するのは，自明な表現が常に因果関係を表すとは限らず曖昧性があるので，知識ベースなしでは難しい。

4.1.2 単純なパターンの因果関係抽出

本章で述べる因果関係の抽出パターンは，自明な表現で比較的単純なパターンを使用するものである．その理由は，ASRS レポートにはそのような表現が多用されているのと，このようなパターンは，常に因果関係を表すと考えられるからである．使用した表層上のパターンは，次のようなものである．

X <clue> Y

ここで，<clue> には，次のものが該当する．

- 前置詞句：“because of”，“due to”
- 因果動詞：“cause”，“causing”，“caused by”，“result in”，“resulting in”，“result from”，“lead to”

例えば，例 37 は “NP1 *caused* NP2” というパターンにマッチする．

37. Aircraft loss of oil pressure caused engine shutdown.

したがって，表層上のパターンを利用して，例 37 から，

$$(\text{Aircraft loss of oil pressure})_{np} \rightarrow (\text{engine shutdown})_{np}$$

という因果関係を抽出できる．

(ここで，“X $\rightarrow$ Y” は，X が原因で Y が結果である因果関係を示す.)

38. Aircraft lost oil pressure causing engine shutdown.

例 38 は，節表現を含んでいるので，

$$(\text{Aircraft})_{subj} (\text{oil pressure})_{obj} (\text{lose})_v \rightarrow (\text{engine shutdown})_{np}$$

という因果関係を抽出できる．例 37 と例 38 は，表層上の表現は異なるものの，意味する内容は同じであるので，同一の述語-項構造を抽出できなければならない．これが本章の提案手法の目的である．

次にイベントと状態について述べる．

39. Aircraft altitude deviation caused less than standard separation.

例 39 から次のような因果関係を抽出できる．

$(\text{Aircraft altitude deviation})_{np} \rightarrow (\text{less than standard separation})_{np}$

ここで、名詞句 “Aircraft altitude deviation” はイベントを表し、名詞句 “less than standard separation” は状態を表している。

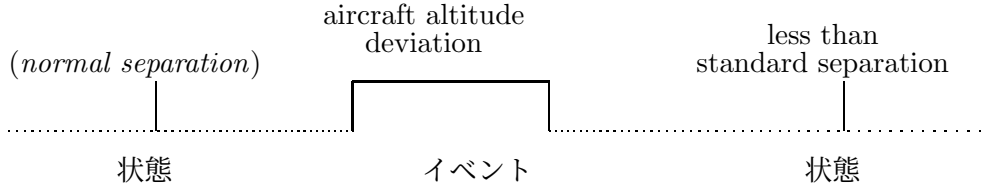


図 4.1: イベントと状態

イベントは新たな状態やプロセスを発生するプロセスを表し、状態はある時間持続するものを云う [32]. 例 39 は、文中には明確に表現されていない “normal separation” という状態が, “less than standard separation” という状態に移ったと解釈することができる. その原因は, “aircraft altitude deviation” である (図 4.1). 例 39 により, 重要な事実がテキスト中に状態やイベントとして表現されていることが分かる. 本章では, 状態とイベントの区別を行なわないで, テキスト中に表現されている状態やイベントは重要なものとして全て収集する. その中で頻度が少なく重要でないものは, マイニング段階で除去されると仮定する.

自明な因果関係表現から以下に述べる 4 種類の方法で因果関係を抽出して評価を行なう. 表層パターン

$X \langle \text{clue} \rangle Y$

から X と Y として以下のような異なった要素を抽出する.

- (1) **NP** : X, Y が名詞句の場合, X, Y としてベース NP のみを抽出する.
(ベース NP とは, NP が限定部と主要部からなる単純名詞句である.)
- (2) **NP+PP** : X, Y が前置詞句を含んだ名詞句の場合, X, Y としてベース NP 及びそれに続く最初の前置詞句を抽出する. 前置詞句がなければベース NP を抽出する.
- (3) **P/A** : NP+PP に加え, もし節表現があれば, それを構文解析して述語-項構造に変換したものを抽出する.
- (4) **P/A+** : P/A の NP, NP+PP に対して名詞化構文変形アルゴリズムを適用する. 名詞化構文変形アルゴリズムを適用できなければ, P/A と同じものを抽出する.

上記 4 種類の方法による例 37 と例 38 の因果関係の抽出結果を以下に示す:

(1)' **NP** :

37: $(\text{aircraft loss})_{np} \rightarrow (\text{engine shutdown})_{np}$

38: N.A. (節表現は, NP では表現できない)

(2)' NP+PP :

37: (aircraft loss)_{np} (of oil pressure)_{pp} → (engine shutdown)_{np}

38: N.A. (節表現は, NP+PP では表現できない)

(3)' P/A :

37: (aircraft loss)_{np} (of oil pressure)_{pp} → (engine shutdown)_{np}

38: (aircraft)_{subj} (oil pressure)_{obj} (lose)_v → (engine shutdown)_{np}

(4)' P/A+ :

37: (aircraft)_{subj} (oil pressure)_{obj} (lose)_v → (engine)_{obj} (shut_down)_v

38: (aircraft)_{subj} (oil pressure)_{obj} (lose)_v → (engine)_{obj} (shut_down)_v

例 37 に名詞化構文変形を適用することで, (4)' P/A+のみが例 37 と例 38 を同一の述語-項構造として抽出できる. P/A や P/A+においても名詞句で残るものはそのまま残し, 述語に対する補語 (主語, 目的語) が埋まらないものについては, 照応解析は行なわないでそのまま残す. このような処理を実施するには, 意味辞書の構築が必要であるからである.

ここまでの全体の流れを見ると図 4.2 のようになる.

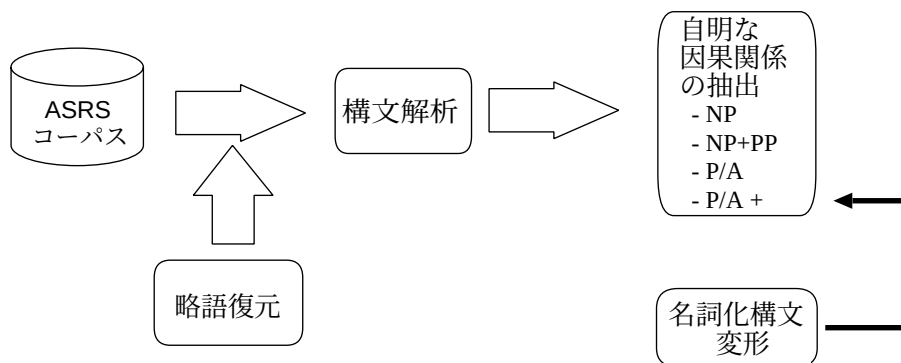


図 4.2: 自明な因果関係の抽出

4.2 テキストマイニング

4.1.2 項で抽出したパターンから有益なパターンとして頻出するパターンを抽出する. マイニングには, シーケンシャルパターンのマイニングツールである PrefixSpan¹ を使用する [33, 34]. 4.2.1 項では, PrefixSpan の簡単な説明を行なう.

4.2.1 PrefixSpan

シーケンシャルパターンマイニングとは, 系列データベースから頻出する部分パターンを抽出することである. ほとんどのシーケンシャルパターンマイニングは, Apriori をベースとしたものである [2]. Apriori は, 部分パターンが低頻度であればそれを含むパターンは多頻度になりえな

¹ PrefixSpan のプログラムは次の URL から入手できる. <http://cl.aist-nara.ac.jp/~taku-ku/software/prefixspan/>

いという考えに基づくものである。アイテム数が少ないパターンから数えあげていくので、アイテムの集合が大きくなった場合に、多くの候補パターンが生成され効率が悪くなるという欠点がある。

PrefixSpan の簡単な説明を行なう。 $I = \{i_1, i_2, \dots, i_n\}$ をアイテム集合とする。アイテムセットとは、アイテム集合の部分集合である。系列とは、アイテムセットの順序列で、系列 s は、 $\langle s_1 s_2 \dots s_l \rangle$ と表記される。ここで、 s_j はアイテムセットである ($s_j \subseteq I$ for $1 \leq j \leq l$)。 s_j は、系列エレメントとも呼ばれ、 $(x_1 x_2 \dots x_l)$ と表記される。ここで、 x_k はアイテムである ($x_k \in I$ for $1 \leq k \leq m$)。系列のアイテムの数は、系列の長さと呼ばれる。 $1 \leq j_1 < j_2 < \dots < j_n \leq m$ という整数が存在し、 $a_1 \subseteq b_{j_1}, a_2 \subseteq b_{j_2}, \dots, a_n \subseteq b_{j_n}$ である場合、系列 $\alpha = \langle \alpha_1 \alpha_2 \dots \alpha_n \rangle$ は、系列 $\beta = \langle \beta_1 \beta_2 \dots \beta_n \rangle$ のサブ系列、 β を α のスーパー系列と云い、 $\alpha \sqsubseteq \beta$ と表記する。

系列データベース S とは、系列 id である sid と系列 s での tuple $\langle sid, s \rangle$ の集合である。 tuple $\langle sid, s \rangle$ は、系列 α が系列 s のサブ系列の場合に、系列 α を含むと云い、 $\alpha \sqsubseteq s$ と表記する。系列 α の系列データベース S におけるサポートとは、系列 α を含む tuple の数で表される。

$$support_S(\alpha) = |\{ \langle sid, s \rangle \mid (\langle sid, s \rangle \in S) \wedge (\alpha \sqsubseteq s) \}|$$

サポートの閾値 ξ に対して、 $support_S(\alpha) \geq \xi$ であるような系列データベース S は、頻出系列パターンと呼ばれる。

Pei らは、Apriori のような候補生成-テストとは異なり、prefix-postfix から深さ優先で頻出系列パターンを抽出する手法を考案した [33]。

ある系列 $\alpha = \langle e_1 e_2 \dots e_n \rangle$ に対して、 $\beta = \langle e'_1 e'_2 \dots e'_m \rangle$ ($m \leq n$) は、以下の場合に α の prefix という。

1. $e'_i = e_i$ (for $i \leq m - 1$)
2. $e'_m \subseteq e_m$
3. $(e_m - e'_m)$ がアルファベット順に並んでいる

$\beta \sqsubseteq \alpha$ であるような系列 $\alpha \beta$ があり、 $\alpha' \sqsubseteq \alpha$ は以下の 場合、 α' は α の β に対する projection という。

1. α' は、prefix β をもつ。
2. $\alpha' \sqsubseteq \alpha''$ であるが、 $\alpha' \neq \alpha''$ であるようなスーパー系列 α'' が存在しない。

系列 $\alpha' = \langle e_1 e_2 \dots e_n \rangle$ を α の $\beta = \langle e_1 e_2 \dots e_{m-1} e'_m \rangle$ ($m \leq n$) に対する projection とする。

$\gamma = \langle e''_m e_{m+1} \dots e_n \rangle$ ($e''_m = (e_m - e'_m)$) を α の β に対する postfix といい、

$$\gamma = \alpha / \beta$$

と表記する。 β が、 α のサブ系列でなければ、 α の β に対する projection, postfix とも空である。 project データベースとは、次の様に定義される。

$$S|\beta = \{ \langle sid, \gamma \rangle \mid (\langle sid, \alpha \rangle \in S) \wedge \gamma = \alpha / \beta \wedge \gamma \neq \epsilon \}$$

表 4.1 で最小サポート値を 2 とした場合について prefix - projection の手法を説明する。

ステップ 1: 長さ 1 のシーケンシャルパターンを探す。

$\langle a \rangle : 4, \langle b \rangle : 4, \langle c \rangle : 4, \langle d \rangle : 3, \langle e \rangle : 3, \langle f \rangle : 3$

ステップ 2: ステップ 1 の 6 つ prefix により、6 つの部分系列に分割する。

表 4.1: シーケンシャルデータベース ([33] より引用)

系列 id	系列
10	$\langle a(abc)(ac)d(cf) \rangle$
20	$\langle (ad)c(bc)(ae) \rangle$
30	$\langle (ef)(ab)(df)cb \rangle$
40	$\langle eg(af)cbc \rangle$

ステップ 3: project データベースを作成して、再帰的に上記ステップを繰り返す。

表 4.2 に project データベース及びシーケンシャルパターンを示す。

表 4.2: project データベース及びシーケンシャルパターン ([33] より引用)

prefix	project(postfix) データベース	シーケンシャルパターン
$\langle a \rangle$	$\langle (abc)(ac)d(cf) \rangle, \langle (-d)c(bc)(ae) \rangle, \langle (-b)(df)cb \rangle, \langle (-f)cbc \rangle$	$\langle a \rangle, \langle aa \rangle, \langle ab \rangle, \langle a(bc) \rangle, \langle a(bc)a \rangle, \langle aba \rangle, \langle abc \rangle, \langle (ab) \rangle, \langle (ab)c \rangle, \langle (ab)d \rangle, \langle (ab)f \rangle, \langle (ab)dc \rangle, \langle ac \rangle, \langle aca \rangle, \langle acb \rangle, \langle acc \rangle, \langle ad \rangle, \langle adc \rangle, \langle af \rangle$
$\langle b \rangle$	$\langle (-c)(ac)d(cf) \rangle, \langle (-c)(ae) \rangle, \langle (df)cb \rangle, \langle c \rangle$	$\langle b \rangle, \langle ba \rangle, \langle bc \rangle, \langle (bc) \rangle, \langle (bc)a \rangle, \langle bd \rangle, \langle bdc \rangle, \langle bf \rangle$
$\langle c \rangle$	$\langle (ac)d(cf) \rangle, \langle (bc)(ae) \rangle, \langle (b) \rangle, \langle bc \rangle$	$\langle c \rangle, \langle ca \rangle, \langle cb \rangle, \langle cc \rangle$
$\langle d \rangle$	$\langle (cf) \rangle, \langle c(bc)(ae) \rangle, \langle (-f)cb \rangle$	$\langle d \rangle, \langle db \rangle, \langle dc \rangle, \langle dcb \rangle$
$\langle e \rangle$	$\langle (-f)(ab)(df)cb \rangle, \langle (af)cbc \rangle$	$\langle e \rangle, \langle ea \rangle, \langle eab \rangle, \langle eac \rangle, \langle each \rangle, \langle eb \rangle, \langle ebc \rangle, \langle ec \rangle, \langle ecb \rangle, \langle ef \rangle, \langle efb \rangle, \langle efc \rangle, \langle efc b \rangle$
$\langle f \rangle$	$\langle (ab)(df)cb \rangle, \langle cbc \rangle$	$\langle f \rangle, \langle fb \rangle, \langle fbc \rangle, \langle fc \rangle, \langle fcb \rangle$

4.2.2 因果関係のマイニング

PrefixSpan を使用して、4.1.2 項で抽出した因果関係のマイニングを行なう。エレメントのアイテム数が1である場合で、例えば、4.1.2 項で抽出した因果関係が表 4.3 のような場合に $event1 \rightarrow event2$ のようなパターンを抽出するものである（最小サポート値：2）。第 3 章で使用したものと同一 ASRS コーパスの Synopsis を使用して、4.1.2 項で解説した 4 つの抽出パターンについて実験する。航空会社で航空安全の仕事に携わっている 3 人のパイロット／航空機関士が、以下の基準により実験結果の評価を行なった。

パターンの評価の基準：「有益」、「曖昧」、「無意味」

有益： 抽出したパターン／イベントは、将来の事故／インシデントを防ぐ対策を立てるのに有益な情報を含んでいる（例：“altitude deviation”）。

表 4.3: 因果関係抽出の例

系列 id	系列
10	<i>event1 event2</i>
20	<i>event1 event3</i>
30	<i>event1 event2</i>

曖昧：抽出したパターン／イベントは、内容が広すぎて分かりにくい、あるいは、解析結果が完全でないで内容を明確に掴めない。しかしながら、抽出したパターン／イベントは、参考になる（例：“fatigue”，“loss”. 一方 “loss of oil pressure” は、「有益」である）。

無意味：抽出したパターン／イベントは、何を表しているのかよく分からない。または、意味している内容が広すぎる（例：“landing”）。

長さ 2 以上で、最小サポート値 2 の抽出パターンに対する 3 人の評価結果の平均を表 4.4 に示す。最小サポート値を 2 にした理由は、構文解析器のエラー、スペルエラーなどの誤りを除くことと、2 回以上起きているパターン／イベントは「有益」であるという仮定からである。なお、3 人の評価の κ 値は 0.49 であり、中程度の一致であった。 κ 値は、複数人の評価の一致度を表す指標である [42]。

表 4.4: 因果関係の評価

	有益	曖昧	無意味	合計
NP (平均数)	5.3	4.7	1	11
(%)	48.5	42.4	9.1	
NP+PP (平均数)	3.7	0.3	0	4
(%)	92.0	8.3	0	
P/A (平均数)	13.3	9	0.6	23
(%)	58.0	39.1	2.9	
P/A+ (平均数)	16.3	6.3	0.3	23
(%)	71.0	27.5	1.4	

P/A により得られるパターン数は、NP により得られるパターン数に比べ約 2 倍となっている。これは、因果関係を表す表現が名詞句だけでなく、節表現にも多いためである（例：“Aircraft rejected takeoff due to engine problem”）。

NP+PP から得られるパターンは、わずか 4 パターンである。この理由は、前置詞句により名詞句が限定され、最小サポート値 2 を満たすようなパターンが少なくなったためである。P/A と P/A+ からは、それぞれ 23 パターン得られている。P/A+ から得られる「有益」なパターン数は、P/A に比べ平均で 3 パターン多い。名詞句から名詞化構文アルゴリズムにより変形されたパターンが 3 パターンある。例えば、“(distraction)_{np} → (altitude deviation)_{np}” は、“(distraction)_{np} → (from altitude)_{pp} (deviate)_v” に変形されている。内容を検討すると、これら 3 パターンは、名詞化構文変形を適用する前でも後でも「有益」と判断されており、名詞化構文変形の有効性は

確認できない。P/A と P/A+ の差を検討すると、同じパターンに対する評価者の判断が P/A と P/A+ で揺れており、第 3 章で述べた名詞化構文変形がマイニングに有効に働いたとは結論できない。

次に、長さ 1 以上のパターンについて最小サポート値を 2 として同じ基準で調べる。長さ 1 のパターンは、イベントを表す。イベントの因果関係だけでなくイベントも抽出するので、抽出パターン数は表 4.4 の抽出パターン数と比較して約 10 倍になっている。結果を表 4.5 に示す。表 4.5 では P/A+ が最もよいことを示している。P/A と P/A+ の詳細な抽出パターンの解析結果も名詞化構文変形がマイニングに有効に働いたことを示している。例えば、“oil quantity loss” と “loss of oil quantity” は、ASRS コーパスに 1 度ずつ出現する。この 2 つの表現は同じ意味を表しているが、出現回数が 1 であるので、P/A では抽出されない。この 2 つの表現は、P/A+ では名詞化構文変形により同じ述語-項構造 “(oil quantity)_{obj} (lose)_v” となり、出現回数が 2 となり、「有益」と判断されている。このような例が 5 例あり、名詞化構文変形がマイニングに有効に働いていることを確認できた（例：“(propeller)_{obj} (damage)_v”, “(without authorization)_{pp} (land)_v”, “(nose gear)_{subj} (collapse)_v”, “(passenger)_{subj} (injure)_v”）。

表 4.5: 長さ 1 以上のパターンの評価

	有益	曖昧	無意味	合計
NP (平均数)	57	21	14	92
(%)	62.0	22.8	15.2	
NP+PP (平均数)	55.7	9.7	7.7	73
(%)	76.3	13.2	10.5	
P/A (平均数)	151.7	43.7	25.7	221
(%)	68.6	19.8	11.6	
P/A+ (平均数)	166.3	43.3	26.3	236
(%)	70.5	18.4	11.2	

4.2.3 因果関係を表す動詞の発見

4.2.2 項での結果によると、最も成績のよい P/A+ でも「有益」なパターンは 16 パターン抽出できるだけである。4.1.2 項で提案した因果関係表現以外にも、X <clue> Y を満たすような動詞 <clue> が存在すると考えられる。Giru らは、因果関係を表す動詞を発見する手法を提案している [11]。Giru らは、1 文で表される最も一般に使用される自明な因果関係の表現は、次のような表現であるとしている。

NP1 <clue> NP2

Giru らは、NP1 と NP2 のペアを WordNet の “CAUSE-TO” という意味関係から抽出し、429 個のペアを得ている [25]。このペアから

NP1 <verb/verb_expression> NP2

を満たす <verb/verb_expression> を抽出している。

我々も、抽出する因果関係を増やすために 4.1.2 項で抽出した原因と結果を表す表現から新しい動詞の発見を試みる。場合数を増やすためにペアではなく、原因と結果を表す表現はどの組合せでもマッチするようにする。NP, NP+PP 及び P/A+について実験したところ、NP については 61 パターン発見できその内 30 パターンが正解、NP+PP では 16 パターン発見できその内 8 パターンが正解、P/A+については 21 パターン発見できその内 10 パターンが正解であった。この実験では、名詞化構文変形が有効には働いていない。その理由は、X, Y が節表現での X (verb/verb_expression) Y の表現の頻度が少ないためと考えられる。NP+PP は、前置詞句が名詞句を限定するので、抽出できるパターンは NP の約 4 分の 1 になり、精度も向上していない。NP は、最も成績がよい。因果関係を表す動詞として新たに、“create”, “force”, “dictate”, “make”, “induce” などを発見できた。しかしながら、このような動詞は、常に因果関係を表すとは限らないことに注意しなければならない。

40. Mlg *made* go around.

41. Ga sma² loss of all electrical *makes* emergency landing.

例 40 では、“made” は因果関係を表さない意味を持たない動詞であるが、例 41 では“makes” は因果関係を表わしている。このような曖昧性があるので、新たに発見した動詞は自明な因果関係抽出パターンの動詞には加えないこととした。

4.3 まとめ

本章では、自明で簡単な因果関係の表現を表層的な表現から抽出し、シーケンシャルマイニング手法である PrefixSpan を用いてマイニングを行なった。因果関係を抽出する表現として、4 種類の抽出方法を試みた。長さ 2 以上のパターンでは、P/A と P/A+の結果がよかった。P/A と P/A+の差は確認できなかったが、抽出パターン数が少ないので確認できなかったものと思われる。長さ 1 以上のパターンでは、P/A+の結果が最もよかった。このことから、名詞化構文変形を適用することにより、長さ 2 以上のパターンでは確認できなかったものの、長さ 1 以上のパターンでは「有益」として抽出されたパターン数が増加することが確認できた。

テキストマイニングにおける評価は、IR の精度と再現率のように共通に使用される評価尺度は存在しない。マイニングを行なう分野によって、精度は低くても 1 件でも有益なパターンを得ることができれば十分な場合もあるであろうし、逆にある程度の精度やパターン数を要求される場合もあるかもしれない。本論文では、有益として抽出されたパターン数と精度によって比較した。

² “sma” は、“Small Aircraft” の頭字語

第5章 時間関係の処理

第4章では、自明な表現を用いて因果関係の抽出を行なった。しかしながら、予想に反して十分な数のパターンを抽出することはできなかった。本章では、時間関係の表現から因果関係を抽出する手法を述べる。さらに、節／文の接続からも同様に因果関係を抽出する手法を述べる。

5.1 はじめに

因果関係は、第4章で用いた自明な表現以外に“after”や“before”などの時間関係の表現や単に節、文の繋がりや暗黙の内に表現される場合がある。そこで、本章では第4章で扱わなかったこのような表現から因果関係を抽出する手法について述べる。

5.1.1 時間関係の推論

時間関係の推論については、これまで多くの研究がなされてきた。Allen は、時間の記述を参照時間（Reference Time）と時間間隔（Temporal Interval）をもとに説明している [5]。Allen は、表 5.1 のような 13 種類の時間関係を提案している。

表 5.1: Allen の 13 の時間関係（[5] より引用）

Relation	Symbol	Symbol for Inverse	Pictorial Example
X <i>before</i> Y	<	>	XXX YYY
X <i>equal</i> Y	=	=	XXX YYY
X <i>meets</i> Y	m	mi	XXXYYY
X <i>overlaps</i> Y	o	oi	XXX YYY
X <i>during</i>	d	di	XXX YYYYYY
X <i>starts</i> Y	s	si	X XX YYYYY
X <i>finishes</i> Y	f	fi	XXX YYYYY

Guillén らは、Allen のアイデアを使い、自動翻訳（MT）に応用している [13]。Passonneau は、時間関係の構造として、状態、プロセス、推移イベントを記述し、Allen の時間間隔を適用している [32]。我々も基本的には Allen の時間関係の考えに沿うが、イベント／状態の起点を問題とし、

時間関係の包含関係などは考慮しないこととする。つまり、時間の流れの中でイベント列を考える場合に、イベントの起点を重要視し、それがどの時点まで継続しているかは考慮しない。例えば、

イベント 1 エンジン火災

イベント 2 緊急降下

イベント 3 着陸

という3つのイベントがこの順序で発生した場合を考える。各イベントの起点で考えると、

イベント 1 $\Rightarrow$ イベント 2 $\Rightarrow$ イベント 3

になる。ここで“ $\Rightarrow$ ”は、起点を基準にした時間関係を表し、Allenの記法では“*before*”, “*meets*”, “*overlaps*”のいずれかにあたる。終点で考えるとイベント1はイベント3の後になるかもしれないが、本論文での因果関係は、起点のみを考慮することとする。

Moensらは、例42のような例を挙げて時間関係の記述は必ずしも時間概念での線形的な意味を表していないと指摘している[26]（この例は元々Ritchieにより提案されている[40]）。

42. When they built the 39th street bridge ...

- (a) ... a local architect drew up the plans.
- (b) ... they used the best materials.
- (c) ... they solved most of their traffic problems.

例42を見れば分かるように、接続詞 *when* が導く従属節と主節の関係は、必ずしも時間的な関係だけでなく、主節の内容が従属節に対して42aでは準備を、42bでは道具を、42cでは結果を表す。Moensらは、時間関係で説明するのではなく、原因-結果を説明するオントロジーを提案している。Nakhimoveskyは、イベントに関する知識を準備、初期段階、主要部、最終段階、結果に分類している[28]。Singhらは、事実を状態とイベントで記述し、計算機で自動的に両者を区別する方法を提案している[48]。

5.2 時間関係の処理

例42から分かるように接続詞 *when* が導く従属節と主節の関係は、必ずしも時間関係を表すものではない。しかしながら、その曖昧性を解消するには、意味解釈が必要である。本論文では意味辞書などを作成しないで行なうアプローチを目指している。もし、

- テキストを時間的に線形な順序に並べ替える
- 述語-項構造を抽出する

ことが可能であれば、テキストから時間的に線形なイベント列を抽出できる。

1番目の問題に関して、時間関係の接続詞について常に表5.2のような時間関係があるものと仮定する。テキストは、表5.2の規則により、並べ替えを行なう。さらに、等位接続詞“and”, “but”や“, ”で繋がる節を抽出して、出現した順序に並べる。文自体も出現した順序に並べる。表5.3は、上記仮定がASRSコーパスでどの程度正しいかを各20例について人手で検査した結果である。表5.2の仮定に対する時間関係の接続詞“then”, “until”, “after”, “as”, “while”及び等位接続

表 5.2: 時間関係の接続詞

X “時間関係の接続詞” Y	
X < Y	X > Y
X before Y	X after Y
X then Y	X as Y
X until Y	X when Y
	X while Y
X < Y は, X の起点が Y の起点より前であることを示す.	
X > Y は, X の起点が Y の起点より後であることを示す.	

表 5.3: 時間関係の接続詞, 等位接続詞及び文の各 20 例の正解率

時間関係の接続詞	before	then	until	after	as	when	while
正解率	90%	100%	100%	100%	100%	80%	100%
等位接続詞及び文の接続	and	but	comma	sentence			
正解率	100%	85%	85%	80%			

詞 “and” の正解率は 100% で, “before” の正解率は 90% である. 不正解の理由は, 従属節の内容が実際には起こっていない場合である (例: “Aircraft stopped before hitting the gate.” この例では, 実際には飛行機はゲートに衝突していない). 等位接続詞 “but” の正解率は 85% で, 15% の誤りは, 仮定との時間的順序が逆の場合である. 時間関係の接続詞 “when” の正解率は 80% で, 20% の誤りは, 時間的に前後関係のものではなく, 同時に起こったものである. “文の接続” と “コンマ” の正解率はそれぞれ 80% と 85% である. 主な誤りの原因は, 後ろに続く文／節が前の文／節の説明や要約になっているものである (例: “Less than standard separation between 2 aircrafts. Operational error.”).

表 5.3 から明らかなように, 表 5.2 の仮定は常に正しくはないが, 仮定に反するようなものは数が少なくマイニング段階で除かれると仮定する.

次のような例を考えてみよう.

Pressurization control on ACR MLG lost as aircraft reached cruise altitude of FL350.
 ARTCC¹ amended altitude assignment to 10000 feet on request. Oxygen masks
 dropped before pressurization control regained. Aircraft returned to departure air-
 port.

この例には, 第 4 章で扱ったような因果関係を表す自明な表現は出現しないが, 因果関係は暗黙の内に表現されている. 時間関係の接続詞については表 5.2 の規則を用いて並べ変え, 等位接続詞及び文については出現した順に並べると次のようになる.

イベント 1 Aircraft reached cruise altitude of FL350.

¹ “ARTCC” は, “Air Route Traffic Control Center” の頭字語

イベント 2 Pressurization control on ACR MLG lost.

イベント 3 ARTCC amended altitude assignment to 10000 feet on request.

イベント 4 Oxygen masks dropped.

イベント 5 Pressurization control regained.

イベント 6 Aircraft returned to departure airport.

イベント 1 は、後ろのイベントの原因とはなっていないかも知れないが、例えば、他のレポートの解析結果から、イベント 2、イベント 4、イベント 6 がこの順序で頻出していれば、

イベント 2 $\Rightarrow$ イベント 4 $\Rightarrow$ イベント 6

のような時間推移パターンを抽出できる。時間推移パターンの中で頻出するものは因果関係を表すと仮定すると、上の時間推移パターンは、以下のように表せる。

イベント 2 $\rightarrow$ イベント 4 $\rightarrow$ イベント 6

イベント 2、イベント 4 は、自明な因果関係の表現を用いて、“Pressurization control on ACR MLG lost causing drop of oxygen masks.” のようにも表現できる。

5.3 実験結果

第 3 章で使用したのと同じ ASRS コーパスの Synopsis を用いて実験する。最小サポート値を 2、抽出パターン長は 2 以上に設定する。評価については、第 4 章で用いたのと同じく「有益」、「曖昧」、「無意味」の 3 つの基準により行なう。評価者は、今回は 1 人である。まず、時間関係の接続詞について、その影響を調べるために第 4 章で最も性能の良い P/A+ について時間関係の接続詞を抽出パターンに含めない場合 (P/A+) と時間関係の接続詞を抽出パターンに含めた場合 (P/A++) の比較を行なう。

表 5.4: 時間関係の接続詞の因果関係に対する影響

	有益	曖昧	無意味	合計
P/A+	17	5	1	23
(%)	73.9	21.7	4.3	
P/A++	19	5	1	25
(%)	76.0	20.0	4.0	

表 5.4 より、抽出パターンに時間関係の接続詞を含めると抽出されるパターン数が 2 つ増え、2 つとも「有益」と評価されていることが分かる。また、時間関係の接続詞の曖昧性については、マインニング段階で吸収されて結果としては出力されていない。しかしながら、出力されるパターン数は 2 つしか増加しない。このことは、ASRS コーパスでは、時間関係の接続詞の使用頻度は高くなく、この情報により因果関係の抽出数を増加させるのは難しいことが分かる。

表 5.5: 節／文と時間関係の接続詞の因果関係に対する影響

	有益	曖昧	無意味	合計
P/A+++	89	163	13	265
(%)	33.6	61.5	4.9	

次に等位接続詞で表される節及び文の接続と時間関係の接続詞を抽出パターンに含めた場合 (P/A+++) について調べる。

表 5.5 より、節／文の接続と時間関係の接続詞を抽出パターンに含めた場合には、265 パターン得られその内 89 パターンが「有益」と判断されている。しかしながら、節／文の接続まで抽出範囲を広げることにより、以下のような誤りが発生している。

- NASA の解析者の説明や要約がイベントとして検出される
- 節／文の順序が時間推移に対応しない
- 構文解析器のエラー
- 名詞化構文の変形エラー

NASA の解析者の説明や要約：「曖昧」と判断された内 107 パターンについては、前の文や節の説明や要約になっているものである。例を挙げると、“system error”, “pilot deviation”, “potential conflict” や “operational error” で NASA の解析者が付与したものと思われる (例：Acr² x had less than standard separation from acr. System error.)。この表現は、レポートの最後の文に多く現れるが、最初の文や途中の文にも現れる。これらの定型表現を除くと、表 5.6 のような結果になる。

表 5.6: 節／文の接続と時間関係の接続詞の因果関係に対する影響 (定型表現を除いたもの)

	有益	曖昧	無意味	合計
P/A+++	89	57	13	159
(%)	56.0	35.8	8.2	

節／文の順序が時間推移に対応しない：節／文の順序が必ずしも時間的推移に対応しないものがある。

43. Altitude deviation.
44. Undeshoot on descent.
45. Crossing restriction not issued.

例 43～例 45 が 1 つのレポートの内容で、この順序で記述されているが、例 43 がレポートの要約、例 44 と例 45 の時間的順序は逆になっている。

² “Acr” は、“Air Carrier” の略語

構文解析器のエラー：例えば，“Crossing restriction not met.”という文で“Crossing”が動詞と認識され，“(restriction)_{obj} (cross)_v”のように解析される。

名詞化構文の変形エラー：名詞化構文の変形をすることにより，状態を表す名詞が動詞に変形される．例えば，“Pilot deviation”という表現が“(pilot)_{sbj} (deviate)_v”のようになってしまう．

表 5.5 では約 3 分の 1，表 5.6 では，2 分の 1 以上のパターンが「有益」と判断されている．第 4 章で使用した自明な因果関係を示す表現による抽出パターンでは，「有益」なパターンは 17 しか得られていないが，今回は 89 パターンと約 5 倍の「有益」なパターンが得られている．したがって，節／文の接続と時間関係の接続詞を抽出パターンに含めることにより「有益」なパターンを多く抽出できることが分かる．

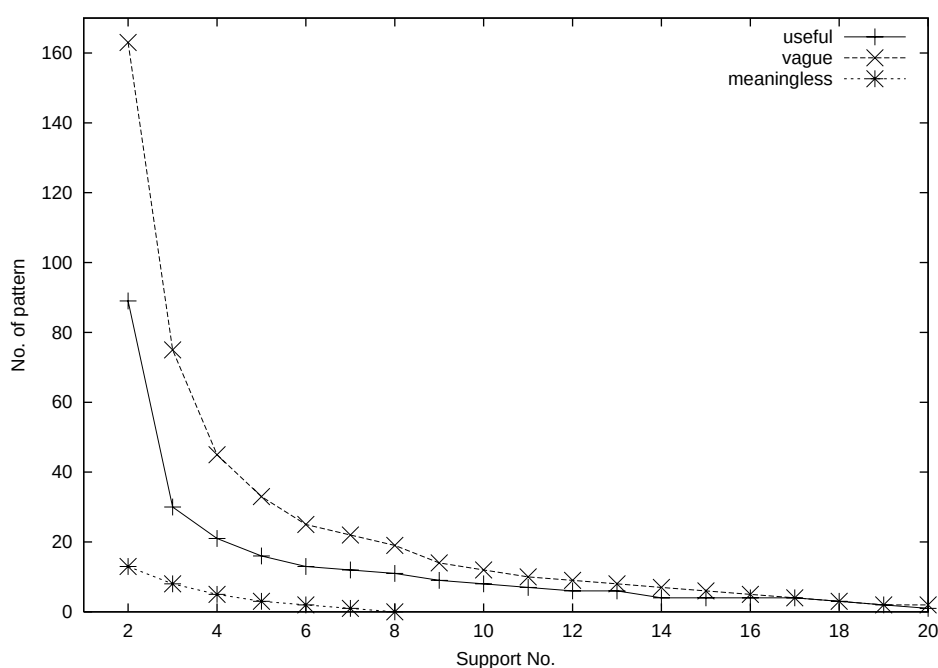


図 5.1: 最小サポート値によるパターン数の変化

長さ 3 のパターンは，17 パターン得られるが，その内「有益」と判断されているものは以下の 2 パターンである．他の 15 パターンは「曖昧」と判断されている．

- (clearance)_{sbj}(intercept)_v → (acr mlg)_{sbj}(from altitude)_{pp} (deviate)_v
→ (excursion)_{np}(from clearance altitude)_{pp}
(解釈：クリアランスが遮られた → ACR MLG が高度逸脱をした → 承認高度から逸脱した)
- (new taxiway centerline light)_{np} → (acr)_{sbj}(from taxi clearance)_{pp} (deviate)_v
→ (active runway)_{obj} (enter)_v
(解釈：新しい誘導路の中心灯 → ACR がタクシー承認から逸脱した → 使用されている滑走路に進入した)

長さ3のパターンでは、「有益」と判断される割合は少ない。

最小サポート値を変化させたときに表5.5の結果がどのように変化するかを調べると図5.1のようになる。図5.1によると最小サポート値を大きくすると、「有益」、「曖昧」、「無意味」全ての項目のパターン抽出数が減り、ほとんど「有益」なパターンが得られないことが分かる。したがって、抽出パターン数が多くなった場合に最小サポート値を大きくすることによって「有益」なパターンだけを抽出する方法は適当でないことが分かる。

次に名詞化構文変形の効果について調べる。表5.7に示すように、名詞化構文変形を行なった場合には名詞化構文変形を行なわない場合と比較して抽出数で約2.2倍、精度で約2.5倍の「有益」なパターンが得られている。このことにより、名詞化変形が有効に働いていることが確かめられる。

表 5.7: 名詞化構文変形の効果

	有益	曖昧	無意味	合計
P/A+++ (名詞化構文変形なし)	40	102	147	289
(%)	13.4	35.3	50.9	
P/A+++ (名詞化構文変形あり)	89	163	13	265
(%)	33.6	61.5	4.9	

5.4 まとめ

本章では、自明な因果関係を表す表現だけでなく、時間関係の接続詞及び節／文の接続から暗黙の内に表現された因果関係を抽出した。結果としては、「有益」であると判断されたパターン数で比較すると、自明な因果関係を表す表現では17であった抽出パターン数が時間関係の接続詞を含めると19になり2つ増えただけであったが、節／文の接続を含めると89になった。このことから、ASRSコーパスで表現されている因果関係は、自明的に表現される数は少なく、暗黙の内に表現される場合が多いことが分かる。

時間関係の接続詞及び節／文の接続を抽出パターンに含めるとエラー率は増加するが、「有益」なパターンを発見するという意味では有効である。

抽出されたパターンを見てみると、例えば、

“(distracton)_{np} → (altitude deviation)_{np}” (解説：注意散漫 → 高度逸脱)

という抽出パターンは「有益」と判断されている。このパターンから、“altitude deviation”は、インシデントを起こす可能性がある。したがって、これを防ぐためには、その原因である“distracton”を起こさないようにすればよいということが分かる。

第6章 考察

第4章ではASRSコーパスで使用されている自明な因果関係を表層的なパターンから抽出した。第5章では第4章の抽出パターンに加え時間関係の接続詞及び節／文の接続から暗黙の内に表現されている因果関係を抽出した。本章では、抽出したパターンについて、具体例をあげて考察する。

6.1 自明な因果関係について

自明な因果関係については、4種類（NP, NP+PP, P/A, P/A+）の方法による抽出を行なった。NPによる抽出パターンを表6.1に、NP+PPによる抽出パターンを表6.2に示す。

頻度は抽出パターンの出現回数を示す。評価は、3人の評価の中で2人以上が一致したものを採用している。評価が「有益」／「曖昧」／「無意味」と分かれたものは、「曖昧」としている。

表6.1と表6.2を比較すると抽出パターン1及び抽出パターン2は一致している。表6.1の抽出パターン6は原因が“loss”だけで内容が明確でないので「曖昧」と判断されているが、それに対応する表6.2の抽出パターン3では“loss of oil pressure”となり内容が明確になっているので「有益」となっている。表6.1の抽出パターン10は結果が“flight crew”だけで内容が明確でないので「曖昧」と判断されているが、それに対応する表6.2の抽出パターン4では“flight crew to climb”となり内容が明確になるので「有益」となっている。この2例については、NP+PPが有効に働いたが、NPで「有益」と判断された他の2例についてはNP+PPでは抽出できていない。この結果から、NP+PPは前置詞句があれば、1番目の前置詞句と一緒に抽出することにより内容を明確化するが、抽出パターン数は減少することが分かる。ただし、「有益」と判断された抽出パターン数は5から4に1つ減少するだけなので、NP+PPの抽出法が優ることが分かる。

P/Aでは、NP+PPと同じNPの抽出パターン4個に加えて、新たに動詞で表現される19個の抽出パターンを抽出できている（表6.3）。その中で10個の抽出パターンが「有益」と判断されている。

P/A+は、P/Aの名詞化構文の中で変形可能なものを変形したものであるため、抽出パターン数はP/Aの抽出パターン数と同じである。変形された箇所をいくつか見てみる。

- 表6.3の17番の (altitude deviation)_{np} の派生名詞 “deviation” が動詞 “deviate” に変形されて、表6.4の17番の (from altitude)_{pp} (deviate)_v のように変換されている。
- 表6.3の19番の (ga sma off airport landing)_{np} の派生名詞 “landing” が動詞 “land” に変形されて、表6.4の19番の (ga sma)_{subj} (at off airport)_{pp} (land)_v のように変換されている。
- 表6.3の21番の (loss)_{np} (of oil pressure)_{pp} の派生名詞 “loss” が動詞 “lose” に変形されて、表6.4の21番の (lose)_v (oil pressure)_{obj} のように変換されている。

名詞句として抽出された (fuel exhaustion)_{np}, (fuel starvation)_{np} の動詞派生名詞 “exhaustion”, “starvation” については変形されないが、原因はASRSコーパスに対応する節表現がないためと

表 6.1: NP の抽出パターン

抽出パターン	評価	頻度
1. (distraction) _{np} → (altitude deviation) _{np} (注意散漫 → 高度逸脱)	有益	2 回
2. (fuel exhaustion) _{np} → (ga sma off airport landing) _{np} (燃料枯渇 → GA SMA 空港以外での着陸)	有益	6 回
3. (fuel exhaustion) _{np} → (off airport landing) _{np} (燃料枯渇 → 空港以外での着陸)	有益	2 回
4. (fuel starvation) _{np} → (landing) _{np} (燃料枯渇 → 着陸)	有益	2 回
5. (loss) _{np} → (emergency descent) _{np} (損失 → 緊急降下)	曖昧	3 回
6. (loss) _{np} → (inflight engine shutdown) _{np} (損失 → 空中でのエンジン停止)	曖昧	3 回
7. (loss) _{np} → (loss) _{np} (損失 → 損失)	曖昧	2 回
8. (loss) _{np} → (runway excursion) _{np} (損失 → 滑走路逸脱)	曖昧	2 回
9. (tcas ra) _{np} → (altitude deviation) _{np} (TCAS RA → 高度逸脱) (TCAS:Traffic Alert and Collision Avoidance System 衝突防止装置, RA:Resolution Advisory)	有益	2 回
10. (tcas ra) _{np} → (flight crew) _{np} (TCAS RA → 乗員)	曖昧	2 回
11. (water leak) _{np} → (safety problem) _{np} (水漏れ → 安全問題)	曖昧	2 回

表 6.2: NP+PP の抽出パターン

抽出パターン	評価	頻度
1. (distraction) _{np} → (altitude deviation) _{np} (注意散漫 → 高度逸脱)	有益	2 回
2. (fuel exhaustion) _{np} → (ga sma off airport landing) _{np} (燃料枯渇 → GA SMA 空港以外での着陸)	有益	6 回
3. (loss of oil pressure) _{np} → (inflight engine shutdown) _{np} (オイル圧損失 → 空中でのエンジン停止)	有益	2 回
4. (tcas ra) _{np} → (flight crew to climb) _{np} (TCAS RA → 乗員による上昇)	有益	2 回

表 6.3: P/A の抽出パターン

	抽出パターン	評価	頻度
1.	(with another aircraft) _{pp} (conflict) _v → (less than standard separation) _{np} (他の飛行機とのコンフリクト → セパレーション不足)	有益	3 回
2.	(acr) _{sbj} (altitude) _{obj} (be assigned) _v → (less than standard separation) _{np} (with another aircraft) _{pp} (ACR が高度アサインされる → 他の飛行機とのセパレーション不足)	有益	2 回
3.	(acr) _{sbj} (climb) _{obj} (through occupied altitude) _{pp} (be issued) _v → (less than standard separation) _{np} (ACR が占有されている高度をアサインされる → セパレーション不足)	有益	3 回
4.	(acr) _{sbj} (through occupied altitude) _{pp} (climb) _v → (less than standard separation) _{np} (ACR が占有されている高度に上昇 → セパレーション不足)	有益	2 回
5.	(acr) _{sbj} (in close proximity) _{pp} (come) _v → (less than standard separation) _{np} (ACR が接近 → セパレーション不足)	有益	3 回
6.	(acr) _{sbj} (through altitude) _{pp} (descend) _v → (less than standard separation) _{np} (ACR が降下 → セパレーション不足)	曖昧	2 回
7.	(acr) _{sbj} (through occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np} (ACR が占有されている高度に降下 → セパレーション不足)	有益	4 回
8.	(acr) _{sbj} (to occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np} (ACR が占有されている高度に降下 → セパレーション不足)	有益	8 回
9.	(acr) _{sbj} (another aircraft) _{obj} (to altitude) _{pp} (descend) _v → (less than standard separation) _{np}	曖昧	2 回
10.	(acr) _{sbj} (climb) _{obj} (through occupied altitude) _{pp} (issue) _v → (less than standard separation) _{np}	有益	2 回
11.	(aircraft) _{sbj} (altitude) _{obj} (be assigned) _v → (less than standard separation) _{np}	曖昧	4 回
12.	(aircraft) _{sbj} (altitude) _{obj} (be assigned) _v → (less than standard separation) _{np} (with another aircraft) _{pp}	曖昧	2 回
13.	(aircraft) _{sbj} (through altitude) _{pp} (descend) _v → (less than standard separation) _{np}	曖昧	2 回
14.	(aircraft) _{sbj} (through occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np}	有益	3 回
15.	(aircraft) _{sbj} (to occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np}	有益	2 回
16.	(clearance) _{sbj} (intercept) _v → (acr mlg altitude devaiation) _{np}	有益	2 回
17.	(distraction) _{np} → (altitude deviation) _{np}	有益	2 回
18.	(fuel exhaustion) _{np} → (ga sma) _{sbj} (landing) _{obj} (force) _v	有益	2 回
19.	(fuel exhaustion) _{np} → (off airport landing) _{np}	有益	6 回
20.	(fuel starvation) _{np} → (landing) _{obj} (force) _v	有益	3 回
21.	(loss) _{np} (of oil pressure) _{pp} → (inflight engine shutdown) _{np}	有益	2 回
22.	(new taxiway centerline light) _{np} → (acr) _{sbj} (from taxi clearance) _{pp} (deviate) _v	有益	2 回
23.	(tcas ra) _{np} → (flight crew) _{np} (to climb) _{pp}	有益	2 回

表 6.4: P/A+の抽出パターン

抽出パターン		評価	頻度
1.	(with another aircraft) _{pp} (conflict) _v → (less than standard separation) _{np}	有益	3 回
2.	(acr) _{sbj} (altitude) _{obj} (be assigned) _v → (less than standard separation) _{np} (with another aircraft) _{pp}	有益	2 回
3.	(acr) _{sbj} (climb) _{obj} (through occupied altitude) _{pp} (be issued) _v → (less than standard separation) _{np}	有益	3 回
4.	(acr) _{sbj} (through occupied altitude) _{pp} (climb) _v → (less than standard separation) _{np}	有益	2 回
5.	(acr) _{sbj} (in close proximity) _{pp} (come) _v → (less than standard separation) _{np}	有益	3 回
6.	(acr) _{sbj} (through altitude) _{pp} (descend) _v → (less than standard separation) _{np}	曖昧	2 回
7.	(acr) _{sbj} (through occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np}	有益	4 回
8.	(acr) _{sbj} (to occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np}	有益	8 回
9.	(acr) _{sbj} (another aircraft) _{obj} (to altitude) _{pp} (descend) _v → (less than standard separation) _{np}	曖昧	2 回
10.	(acr) _{sbj} (climb) _{obj} (through occupied altitude) _{pp} (issue) _v → (less than standard separation) _{np}	有益	2 回
11.	(aircraft) _{sbj} (altitude) _{obj} (be assigned) _v → (less than standard separation) _{np}	曖昧	4 回
12.	(aircraft) _{sbj} (altitude) _{obj} (be assigned) _v → (less than standard separation) _{np} (with another aircraft) _{pp}	曖昧	2 回
13.	(aircraft) _{sbj} (through altitude) _{pp} (descend) _v → (less than standard separation) _{np}	曖昧	2 回
14.	(aircraft) _{sbj} (through occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np}	有益	3 回
15.	(aircraft) _{sbj} (to occupied altitude) _{pp} (descend) _v → (less than standard separation) _{np}	有益	2 回
16.	(clearance) _{sbj} (intercept) _v → (acr mlg) _{sbj} (from altitude) _{pp} (deviate) _v	有益	2 回
17.	(distraction) _{np} → (from altitude) _{pp} (deviate) _v	有益	2 回
18.	(fuel exhaustion) _{np} → (ga sma) _{sbj} (landing) _{obj} (force) _v	有益	2 回
19.	(fuel exhaustion) _{np} → (ga sma) _{sbj} (at off airport) _{pp} (land) _v	有益	6 回
20.	(fuel starvation) _{np} → (landing) _{obj} (force) _v	有益	3 回
21.	(oil pressure) _{obj} (lose) _v → (inflight engine) _{np} (shut down) _v	有益	2 回
22.	(new taxiway centerline light) _{np} → (acr) _{sbj} (from taxi clearance) _{pp} (deviate) _v	有益	2 回
23.	(tcas ra) _{np} → (flight crew) _{np} (to climb) _{pp}	有益	2 回

思われる．本論文での目的は，全ての動詞派生名詞を変形するのではなく，テキストの正規化であるので，その目的は達成されている．

パターンではなく，単独のイベントとしては，“(altitude)_{sbj} (bust)_v” や “(altitude excursion)_{np}”（どちらも高度逸脱を意味する）という表現を発見できた．情報検索システムを使用して「高度逸脱」が記述されている文書を検索する場合，検索語として，“altitude deviation” しか入力しなければ，“altitude deviation” と上の 2 つの用語が同じ文書内で共起している場合を除き，上の 2 つの用語が出現する文書は抽出できない．つまり，“altitude deviation” という検索語では，“altitude bust” や “altitude excursion” というタームを含んだ文書を検索できない．日本人には，“(altitude)_{sbj} (bust)_v” や “(altitude excursion)_{np}” は，あまり馴染みのない表現なのでこの用語が得られたのは「有益」だといえる．

6.2 時間関係と節／文について

時間関係の接続詞による抽出を P/A+に加えた結果，抽出パターン数は P/A+に比べて 2 つ増えている．

表 6.5: P/A++の抽出パターン

抽出パターン	評価	頻度
1. (runway) _{obj} (turn off) _v → (aircraft) _{obj} (conflict) _v (滑走路から出る → 飛行機と衝突)	有益	2 回
2. (cabin pressurization system) _{sbj} (fail) _v → (lgt) _{sbj} (emergency descent) _{obj} (make) _v (客室の与圧システム故障 → LGT 緊急降下)	有益	2 回

表 6.5 に示すように，新たに抽出されたパターンは 2 つとも「有益」と判断されている．時間関係の接続詞の曖昧性は表面に現れないが，抽出パターンは 2 つしか増加していない．

さらに，時間関係の接続詞と節／文の接続を抽出パターンに加えると，265 パターンが抽出され，その中で 89 パターンが「有益」と判断されている．抽出パターンの結果を A.1 節に示す．主な特徴を以下に示す．

抽出パターン P/A+，P/A++，P/A+++の比較： 自明な因果関係での抽出パターン P/A+，時間関係の接続詞を含めた抽出パターン P/A++，節／文の接続を含めた抽出パターン P/A+++を比較すると，

$$P/A+ \subset P/A++ \subset P/A+++$$

という関係にある．しかしながら，同じ抽出パターンの頻度が P/A++，P/A+++で増加することはない．P/A+での抽出パターン：

(acr)_{sbj} (climb)_{obj} (through occupied altitude)_{pp} (be issued)_v
→ (less than standard separation)_{np}

と同種類の抽出パターンが P/A+++で

(acr mlg)_{sbj} (climb)_{obj} (through altitude)_{pp} (climb)_v

→ (less than standard separation)_{np}

として抽出されることから、自明な因果関係で表現されるのと同じ種類の因果関係が節／文においても表現されることが分かる。

構文解析器のエラー： 構文解析器のエラーにより、主語となるべき名詞が動詞なしで名詞句として抽出されたり、主語と目的語が逆になる場合がある。

6.3 「有益」なパターン

抽出されたパターンの内で、筆者が特に「有益」と判断したものをその理由とともに以下にあげる。

- フラップセット誤り → 離陸中止

この抽出パターンは、飛行機が離陸するのに必要なフラップセットが確実にされていなかった為に離陸を中止したことを意味している。離陸前のフラップセットは、フラップセットが正確に行なわれないと飛行機の性能を満足しないので、非常に重要なものである。したがって、これまでに、何重もの防止策がとられてきている。それにもかかわらず、このような事例が発生しているということは、新たな防止策が必要かどうか改めて考慮する必要がある。

- 誤った気圧高度補正 → 高度逸脱

飛行機は、低高度ではその地域の気圧に応じた気圧補正をする必要がある。この気圧高度補正を行なわないと、他の飛行機とのセパレーションが不足したり、着陸する飛行場からの高度が間違っただけのものになる。この抽出パターンでは、この気圧高度補正が誤ってされた為に高度逸脱をしたことを示している。気圧高度補正の情報は、音声通信により得られることが多く、忙しい時にセットする機会が多い。ミスを防ぐためチェックリストで確認することになっているが、それでもこのようなミスが起きるということは、詳しい原因の調査をして新たな対策をたてる必要があるかどうかの検討が必要である。

以下に、筆者が「有益」と判断したパターンをあげる。

- 燃料無くなる → 空港外への着陸
- 飛行機制御不能 → 滑走路からの逸脱
- オイル圧低下 → 空中でのエンジン停止
- 高度つき抜け → セパレーション不足
- エンジン停止 → 緊急事態表明
- 占有高度に降下 → セパレーション不足
- 与圧システム故障 → 緊急降下
- 注意散漫 → 高度逸脱
- 緊急事態宣言 → 出発空港着陸

- 緊急事態宣言 → 代替空港着陸
- エンジン停止 → 強制着陸
- 燃料無くなる → 強制着陸
- エンジン停止 → 空港外への着陸
- エンジン停止 → 目的地へ
- 新しい誘導路灯 → 使用滑走路進入
- 新しい誘導路灯 → クリアランスからの逸脱 → 使用滑走路進入
- 衝突防止装置作動 → 上昇
- クリアランス誤解 → 高度逸脱
- 緊急事態宣言 → 燃料投棄
- 客室ドア OPEN → 緊急事態宣言
- クリアランスに従わない → 高度逸脱
- 通過高度を守らない → 高度逸脱
- 早い降下開始 → 高度逸脱
- 管制塔にコンタクトなし → クリアランスなしに着陸
- 衝突防止装置作動 → セパレーション不足
- 着陸復行指示 → 空中衝突の危険
- TCAS RA → 高度逸脱
- タービュランス → 客室乗務員負傷
- 空中衝突の危険 → TCAS RA
- 空中衝突の危険 → 迅速な対応
- TCAS RA → 迅速な対応
- ダイバート → 問題なし
- 飛行機システムのトラブル → 緊急事態宣言
- 飛行機システムのトラブル → 高度逸脱
- 飛行機システムのトラブル → 出発空港に戻る
- 飛行機システムのトラブル → エンジン停止
- 飛行機システムのトラブル → 目的地へ飛行

- 飛行機システムのトラブル → 与圧低下
- 飛行機システムのトラブル → 火災報知器作動 → エンジン停止
- 飛行機システムのトラブル → 発電機故障
- 飛行機システムのトラブル → 離陸中止
- 間違った高度セット → 高度逸脱
- 高度制限に従わない → 高度逸脱
- クリアランスを確認しない → 高度逸脱
- アプローチコントローラー干渉 → 高度逸脱
- オートパイロット故障 → 高度のオーバーシュート
- 客室火災 → ダイバート
- 機長側の窓がスライド → 離陸中止
- クリアランスを邪魔される → 高度逸脱
- 周波数ミス → クリアランスなしに着陸
- コンプレッサーストール → 離陸中止
- コントローラー訓練 → 迅速な対応が必要
- エンジン故障 → 空港外着陸
- エンジン故障 → 出発空港に戻る
- エンジン故障 → 緊急事態宣言 → 出発空港に戻る
- エンジン火災 → エンジン停止
- 空域の誤った通過 → TCA に接近
- 指定高度でのレベルオフミス → 高度逸脱
- 滑走路進入 → 着陸復行
- ハードランディング → ギア崩壊
- 客室ドア OPEN → 緊急事態宣言
- 複数滑走路使用 → パイロット逸脱
- 複数滑走路使用 → クルー混乱
- 複数滑走路使用 → エラー発生
- 複数滑走路使用 → 高度逸脱

- 客室に煙 → エンジン停止
- 燃料無し → エンジン停止
- フラップ出し忘れ → 離陸中止
- 離陸警報 → 離陸中止
- 高度維持不能 → 緊急事態宣言
- 間違った ALT セット → TCAS RA
- スラットセット間違い → 離陸中止

6.4 まとめ

本章では、抽出されたパターンの具体例をあげて考察した。今後の課題をあげる。

表現の統合 抽出パターンを見ると同じ内容が異なる表現でされている場合がある。例えば，“... fail to comply ...”と“... not comply ...”であるが、意味的な辞書を使用していないので、これらの表現は統合されていない。しかしながら、同じ意味を異なる表現でされているために、統合されず、頻度が2未満のために抽出されないパターンがたくさんあると思われる。意味辞書を作成しない場合は、ユーザが同じ意味表現を指定できるようにする方法も有効かもしれない。しかしながら、同じ意味表現といっても以下に示すように単なる同意語から、単純な置き換えで解決できない表現まであるのでその解決法は今後の課題である。

- 同じ意味の単語
 - “engine shutdown”, “engine quit”, “engine failure”
 - “fuel exhaustion”, “fuel starvation”
 - “deviate”, “transgress”
- 同じ意味の表現
 - “engine shutdown”, “lose power”

サポート動詞の処理 表現の統合と関係するが、自明な因果関係の表現では、例えば、

46. ACR conflicted with another aircraft resulting in less than standard separation.

の抽出パターンは、

$(acr)_{sbj} (with\ another\ aircraft)_{pp} (conflict)_v \rightarrow (less\ than\ standard\ separation)_{np}$

となる。ところが、例 46 の結果部分を表現した、

47. ACR had less than standard separation.

の抽出パターンは、

$(acr)_{sbj} (less\ than\ standard\ separation)_{obj} (have)_v$

となる。“have”は、意味を持たない動詞で「サポート動詞」と呼ばれるが、「サポート動詞」があることにより、例 46 と例 47 で抽出パターンが異なっている。したがって、「サポート動詞」を含んだ抽出パターンと含まないものを同定する処理が必要になる。

上位下位概念 表現の中には、次の例のように概念として包含関係にあるものがある。その処理法についても今後の課題である。

- “altitude deviation”
 - － “altitude undershoot”, “altitude overshoot”
- “engine fail”
 - － “compressor stall”, “engine surge”

名詞句 名詞句の中には次の例のように同じ意味を逆の順序で表す場合がある。このような名詞句の処理についても今後の課題である。

例：“pilot of mlg”, “mlg pilot”

抽出パターンの統合 ある原因に起因する全ての結果を検索したい場合や、逆にある結果をもたらす全ての原因を検索したい場合がある。そのような場合の表示法の開発が必要である。以下に、抽出パターンの「緊急事態宣言」と「高度逸脱」について、人手で原因と結果を統合した例を示す。

例えば、「緊急事態宣言」の結果を見ると、「出発飛行場に戻る」、「代替空港に着陸」、「燃料投棄」になっている。さらに「燃料投棄」の結果を調べると「出発飛行場に戻る」になっている。逆に「緊急事態宣言」の原因を見ると、「エンジン停止」、「客室ドア OPEN」、「システムトラブル」、「客室火災」、「高度維持不能」となっている。さらに、「エンジン停止」の原因は、「オイル圧低下」、「火災警報」、「システムトラブル」となっている（図 6.1）。

もう1つの例「高度逸脱」の結果は、全て「セパレーション不足」になっている。逆に「高度逸脱」の原因は、「オートパイロット故障」、「注意散漫」、「クリアランスに従わない」、「クリアランス誤解」、「通過高度を守らない」、「早い降下開始」、「異常接近装置作動」、「間違った気圧補正」となっている。さらに「異常接近装置作動」の原因は「異常接近」となっている。この例には、「クリアランスに従わない」の下位概念として、「クリアランス誤解」、「通過高度を守らない」、「早い降下開始」が含まれている（図 6.2）。

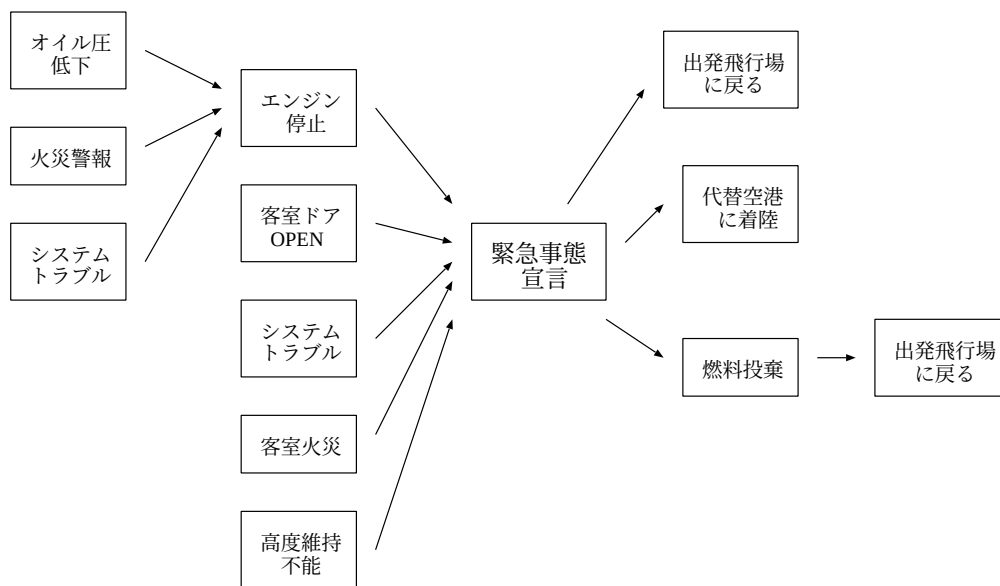


図 6.1: マイニングの表示（1）

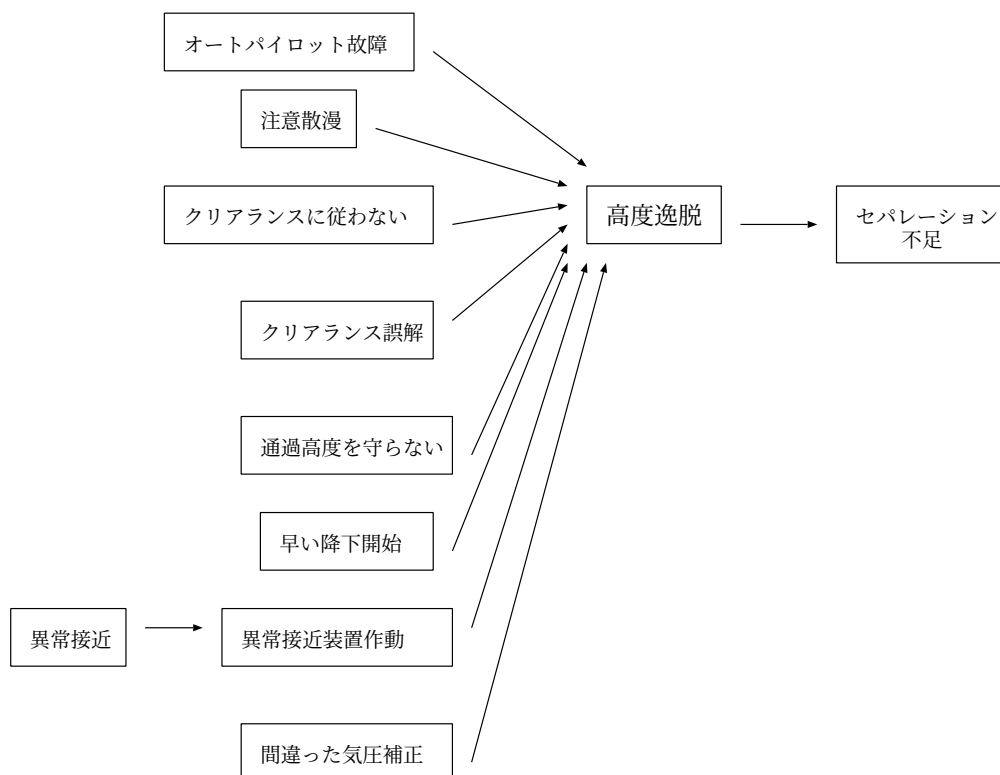


図 6.2: マイニングの表示（2）

第7章 結論

7.1 まとめ

本論文では、航空安全レポートを対象にして自然言語処理技術を利用したテキストマイニング手法を提案した。テキストマイニングは、前処理でテキストを構造化されたデータに変換する必要がある。その処理の中で、同じ意味のもので異なる表現がされているものを同定する必要がある。このような処理として、略語の復元と名詞化構文変形の2つの要素技術を提案した。いずれの要素技術も辞書を使用しないで、コーパスから知識を得る方法を用いている。コーパスとしてはNASAのASRSコーパスを使用して開発／実験を行なったが、他の分野への移植性についても考慮した。

第2章略語の復元では、単一単語からなる略語を対象にし、人間が略語を使用する際の規則性と文脈から得られる情報を融合して、決定木により略語の復元を行なった。テキストから前後の語を文脈として使用する手法は新しいものではないが、人間が省略する際の規則性と組み合わせることで効果的な結果が得られた。復元語の探索には、辞書を使用する代わりに略語が使用されているのと同じ分野に属する「略語の少ないコーパス」を使用した。同じASRSコーパスを使用した先行研究とF-measureで比較して、システムが1語を出力する場合には約6%、システムが複数語を出力する場合には約26%の精度向上があった。本手法は、言語依存性が少ないので移植性にも優れていると考えられる。

第3章名詞化構文の処理では、名詞化構文に対応する節に変形する処理について述べた。本来節で表現されるものが名詞化されたものを名詞化構文という。テキストマイニングや情報抽出では、同じイベントが異なる形式で表現されたとき、それらが同一のイベントであることを同定する必要がある。本論文では、名詞化構文も節表現も同じ述語-項構造に変形することにより両者の同一性を同定した。まず、名詞化構文の特徴および先行研究について述べ、コーパスから節表現の格構造を抽出して、その統計情報を利用して、名詞化構文から自動的に述語-項構造へ変換する手法について述べた。24,626の航空安全レポートからなるコーパスを用い、頻度上位20位までの名詞化構文について提案手法の評価を行ない、平均で約78%の精度が得られた。

第4章因果関係の検出では、表層的なパターンを利用して、テキスト中の因果関係を抽出する手法について述べた。まず、因果関係を抽出するための4種類のパターンについて述べ、本論文で使用するテキストマイニング手法であるPrefixSpanの説明をした。次に、4種類のパターンによるマイニング結果について考察した。4種類のパターンの中で、長さ2以上で最小サポート値2の抽出結果について比較したところ、節表現まで抽出するパターンとその中の名詞化構文変形を行なうパターンが最も成績が良かった。長さ1以上で最小サポート値2の抽出結果について比較したところ、名詞化構文変形をおこなうパターンが最も成績が良く、名詞化構文変形の有効性を確認できた。

第5章時間関係の処理では、時間関係を表す表現や文／節の接続を利用して因果関係を抽出する手法について述べた。これらの情報は、必ずしも明示的に因果関係が表現しているとはいえないが、イベントを時間推移順に並べた時に頻出するパターンは因果関係を表していると仮定して

因果関係を抽出した。24,626 の航空安全レポートからなるコーパスを用い、提案手法の評価を行った。その結果、時間関係を表す表現や文／節の接続を利用して因果関係を抽出することにより、第 3 章で述べた名詞化構文変形の効果がより顕著になり、名詞化構文変形を適用しない場合に比べて約 2 倍の「有益」な因果関係が抽出できた。

第 6 章では、抽出パターン在具体例をあげ、問題点を考察した。

7.2 今後の課題

今後検討すべき課題について以下に述べる。

略語処理：2 語以上からなる略語の復元について検討する必要がある。本論文で提案した手法を用いて 2 語以上からなる略語の復元への拡張は可能であると考える。

他分野への適用として、カタカナとそれに対応する英語を含む日本語のテキストから自動的にカタカナに対応する英語に変換する実験を行なった。今後、日本語の略語（例：「ナビゲーション」 — 「ナビ」）などへの応用の可能性について検討する必要がある。

名詞化構文の処理：深い構造を持たない技術的なテキストでは本論文の手法が有効であるが、一般のテキストでは、本論文で扱っていない関係代名詞などの処理が必要である。一般のテキストでどの程度本論文の手法が有効であるかを検証する必要がある。

コーパスベースのアプローチ：コーパスから知識を得る手法の利点は、

- 辞書を構築するには時間／人手が掛かるし保守も必要であるが、その必要がない。
- 他の分野への移植が比較的簡単である。

しかしながら、コーパスに出現しないものや出現頻度の低いものは利用できない、意味関係を記述できないなどの欠点がある。これらを解決するためには、もし既存の辞書があれば、その辞書とコーパスを利用した手法を組み合わせたハイブリッドな方法が有効であると考えられる。

因果関係の抽出：本論文では 1 文で表される自明な表現を使用した。が、表層的に得られる因果関係はこれ以外にもある [11, 16, 17]。他分野へ適用する場合には、これらの表現からの因果関係抽出を検討する必要がある。実験は ASRS コーパスの Narrative 部分の要約である Synopsis を使用したが、要約がないテキストを使用する場合には、話題の中心部を見つけるための前処理が必要になる。

テキストマイニングの前処理：略語復元と名詞化構文の変形という 2 種類の前処理を行なったが、これ以外にタイプエラーの訂正、意味的に同じ表現の同定などが必要である。意味的に同じ表現とは、例えば、”Aircraft lost cabin pressurization system” と “Cabin pressurization on aircraft failed” などである。

テキストマイニングの後処理：本論文では、テキストマイニングの後処理は扱わなかったが、前処理と同様重要なものである。抽出パターン数が少ない場合には、ユーザが全ての抽出パターンを確認できるが、抽出パターン数が多くなった場合には不可能である。第 6 章で述べたように同じ意味の表現を同定する処理や、同じ原因、あるいは、同じ結果のパターンについて関係するパターンを全て表示するような技術が必要である。

テキストマイニングの抽出と評価：マイニングの評価については、システムが出力した結果については可能であるが（精度）,「有益」なパターンをどの位の割合で抽出できたかについての評価は不可能である（再現率）。また、マイニングは、統計的に頻度の高いものが「有益」であるという仮定のもとに行なっているが、頻度の低いものの中にも「有益」なものはある筈である。このようなものを如何にして抽出するかについては、今後の課題である。

謝辞

本研究を通して、いろいろな方々の御指導、御支援を賜りました。

本研究を行なうにあたり、終始的確で丁寧な御指導、御助言を頂きました指導教官の東京工業大学大学院情報理工学研究科 徳永健伸助教授に心より感謝致します。東京工業大学大学院情報理工学研究科 田中穂積教授には、終始暖かい励ましや助言を頂きました。深く感謝致します。

また、論文審査員の沼尾正行助教授、奥村学助教授、櫻井成一朗助教授には本研究に関して様々な有益な助言を頂きました。ここに感謝致します。

京都大学時代の恩師であります長尾眞京都大学総長には、今回の研究を始めるにあたり、貴重な助言を頂きました。深く感謝致します。辻井潤一東京大学/UMIST 教授には今回の研究に際して、最近の言語処理の動向などについて丁寧に御指導頂きました。ここに深く感謝致します。

白井清昭氏（現北陸先端科学技術大学院大学助教授）、橋本泰一助手には有益な助言を頂き、計算機環境のことでも大変お世話になりました。Timoth Baldwin 氏には英語の校正をして頂きました。

田中・徳永研究室のメンバーの皆様には研究討論会や輪講だけではなく、研究室での日常生活を通じて大変お世話になりました。特に、IR 研究会の皆様には、研究会を通して貴重な意見や助言を頂きました。阿玉泰宗氏には、略語復元の技術を日本語のカタカナ文から英語に変換する応用の研究をして頂きました。武智峰樹氏には、貴重なコメントを頂きました。

田中研究室歴代の秘書の皆様には事務手続きなどで大変お世話になりました。

田中・徳永研究室の皆様にはスケジュールをはじめいろいろな面で御配慮をいただき、御迷惑をおかけしました。ここに御礼申し上げるとともにお詫び申し上げます。

社会人博士としての留学を認めて頂きました日本航空（株）に深く感謝致します。特に、社会人博士の期間在席しました運航本部乗員技術部、運航技術部の皆様にはいろいろと御支援を頂きました。心より感謝致します。

最後になりますが、暖かく見守ってくれました妻 薫、娘 百合子に感謝します。

参考文献

- [1] R. Agrawal, T. Imielinski, and A. Swami. Mining associations between sets of items in massive databases. In *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, pp. 207–216, 1993.
- [2] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of Int. Conf. on Very Large data Bases (VLDB'94)*, pp. 487–499, 1994.
- [3] Helena Ahonen, Oskari Heinonen, Mika Klemettinen, and A. Inkeri Verkamo. Applying data mining techniques in text analysis. Technical Report C-1997-23, Department of Computer Science, University of Helsinki, 1997.
- [4] Helena Ahonen, Oskari Heinonen, Mika Klemettinen, and A. Inkeri Verkamo. Mining in the phrasal frontier. In *Proceedings of Principles of the third International Conference on Knowledge Discovery (KDD-97)*, 1997.
- [5] James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, Vol. 26, pp. 832–843, 1983.
- [6] A. T. Arampatzis, T. Tsoris, C.H. A. Koster, and TH. P. Van Der Weide. Phase-based information retrieval. *Information Processing & Management*, Vol. 14, No. 6, pp. 693–707, 1998.
- [7] N. Chomsky. *Remarks on nominalization*. 1970.
- [8] Deborah A. Dahl, Martha S. Palmer, and Rebecca J. Passonneau. Nominalizations in pundit. In *Proceedings of the 25th Annual Meeting of the ACL*, pp. 131–137, 1987.
- [9] Ronen Feldman and Haym Hirsh. Mining associations in text in the presence of background knowledge. In *Proceedings of the 2nd International Conference on Knowledge Discovery (KDD-96)*, pp. 343–346, 1996.
- [10] William B. Frakes and Ricardo Baeza-Yates. *Information Retrieval*. Prentice Hall, 1992.
- [11] Roxana Girju and Dan Moldovan. Mining answer for causation questions. In *Proceedings of the AAAI Spring Symposium*, pp. 131–137, 2001.
- [12] Gregory Grefenstette and Simone Teufe. Corpus-based method for automatic identification of support verbs for nominalizations. In *Proceedings of EACL-95*, 1995.
- [13] Rocio Guillén, David Farwell, and Janyce Wiebe. Handling temporal relations in scheduling dialogues for an MT system. In *Proceedings of International Workshop on Temporal Representation and Reasoning*, 1996.

- [14] Marti A. Hearst. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, Vol. 23, No. 1, pp. 33–64, 1997.
- [15] Richard D. Hull and Fernando Gomez. Semantic interpretation of nominalizations. In *Proceedings of AAAI-96*, pp. 1062–1068, 1996.
- [16] Christopher S.G. Khoo, Jaklin Kornfilt, Robert N. Oddy, and Sung Hyon Myaeng. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary and Linguistic Computing*, Vol. 13, No. 4, pp. 177–186, 1998.
- [17] Christopher S.G. Khoo, Sung Hyon Myaeng, and Robert N. Oddy. Using cause-effect relations in text to improve information retrieval precision. *Information Processing & Management*, Vol. 31, No. 1, pp. 119–145, 2001.
- [18] Karen Kukich. Spelling correction for the telecommunications network for the deaf. *Communications of the ACM*, Vol. 35, No. 5, pp. 80–90, 1992.
- [19] Karen Kukich. Techniques for automatically correcting words in text. *ACM Computing Surveys*, Vol. 24, No. 4, pp. 377–439, 1992.
- [20] Leah S. Larkey, Paul Ogilvie, M. Andrew Price, and Brenden Tamilio. Acrophile: An automated acronym extractor and server. In *Proceedings of the ACM Digital Libraries conference*, pp. 205–214, 2000.
- [21] Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barrett, and Ruth Reeves. Using nomlex to produce nominalization patterns for information extraction. In *Proceedings of Coling-ACL-98 Workshop: the Computational Treatment of Nominals*, 1998.
- [22] Catherine Macleod, Adam Meyers, Ralph Grishman, Leslie Barrett, and Ruth Reeves. Designing a dictionary of derived nominals. In *Proceedings of Recent Advances in Natural Language Processing*, 1997.
- [23] Laura A. Mather and Jarrod Note. Discovering encyclopedic structure and topics in text. In *Proceedings of the sixth international Conference on Knowledge Discovery and Data Mining (KDD2000) Workshops on Text Mining*, 2000.
- [24] Michael W. McGreevy. Searching the ASRS database using QUORUM keyword search, phrase search, phrase generation, and phrase discovery. *NASA/TM-2001-210913. AMes research Center, Moffett Field, Calif.*, 2001.
- [25] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. Introduction to WordNet: An on-line lexical database. Technical Report CSL Report 43, Princeton, 1993.
- [26] Marc Moens. Temporal ontology and temporal reference. *Computational Linguistics*, Vol. 14, pp. 15–28, 1988.
- [27] Marcus M.P., Santorini B., and Marcinkiewicz M.A. Building a large annotated corpus of English : The penn treebank. *Computational Linguistics*, Vol. 19 (2), pp. 313–330, 1993.

- [28] Alexander Nakhimovsky. Aspect, aspectual class, and the temporal structure of narrative. *Computational Linguistics*, Vol. 14, pp. 29–43, 1988.
- [29] Zohreh Nazeri, Eric Bloedom, and Paul Ostwald. Experience in mining aviation safety data. In *Proceedings of ACM SIGMOD 2001*, pp. 562–566, 2001.
- [30] Toshihide Ono, Haretsuge Hishigaki, Akira Tanigami, and Toshihisa Takagi. *Automatic Extraction of Information on Protein-Protein Interaction from Scientific Literature*. In *Genome informatics 1999*. Univeral Academy Press, 1999.
- [31] Youngja Park and Roy J. Byrd. Hybrid text mining for finding abbreviations and their definitions. In *Proceedings of EMNP2001*, 2001.
- [32] Rebecca J. Passonneau. A computational model of the semantics of tense and aspect. *Computational Linguistics*, Vol. 14, No. 2, pp. 44–60, 1988.
- [33] Jian Pei, Jiawei Han, Behzad Mortazavi-Asi, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Mei-Chun Hsu. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *Proceedings of Int. Conference of Data Engineering*, 2001.
- [34] Helen Pinto, Jiawei Han, Jian Pei, and Ke Wang. Multi-dimensional sequential pattern mining. In *Proceedings of 10th ACM International Conference on Information and Knowledge Management(CIKM’01)*, 2001.
- [35] MUC-4 Proceedings. *Proceedings of the Fourth Message Understanding Conference(MUC-4)*. Morgan Kaufmann, 1992.
- [36] MUC-5 Proceedings. *Proceedings of the Fifth Message Understanding Conference(MUC-5)*. Morgan Kaufmann, 1993.
- [37] MUC-6 Proceedings. *Proceedings of the Sixth Message Understanding Conference(MUC-6)*. Morgan Kaufmann, 1995.
- [38] J. Ross Quinlan. *C4.5:Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [39] Randolph Quirk, Sydney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. LONGMAN group Inc., 1985.
- [40] G. D. Ritchie. Temporal clauses in English. *Theoretical Linguistics*, Vol. 6, pp. 87–115, 1979.
- [41] Neil C. Rowe. and Kari Laitinen. Semiautomatic disabbreviation of technical text. *Information Processing & Management*, Vol. 31, No. 6, pp. 851–857, 1995.
- [42] Siegel S. and Castellan Jr. *Nonparametric statics for the behavioral sciences*. Second ed. McGraw Hill Co., 1988.
- [43] Barbara Sada. Making effective use of aviation safety narratives. In *GAIN Asia-Pacific Regional Conference*, 2002.

- [44] G. Salton. *Automatic Text Processing*. Addison-Wesley, 1988.
- [45] Helmut Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pp. 44–49, 1994.
- [46] Helmut Schmid. Improvements in part-of-speech tagging with an application to German. In *EAL SIGDAT workshop*, 1995.
- [47] Satoshi Sekine and R. Grishman. A corpus-based probabilistic grammar with only two non-terminals. In *Proceedings of the International Workshop on Parsing Technology*, 1995.
- [48] Eric V. Siegel and Kathleen R. McKeown. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Association for Computational Linguistics*, Vol. 26, pp. 595–628, 2000.
- [49] Akira Terada and Takenobu Tokunaga. Automatic disabbreviation by using context information. 情報処理学会自然言語処理研究会 (2001-NL-144), No. 6, pp. 39–45, 2001.
- [50] Akira Terada and Takenobu Tokunaga. Automatic disabbreviation by using context information. In *Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium Workshop on Automatic Paraphrasing: Theories and Applications*, pp. 21–28, 2001.
- [51] Akira Terada and Takenobu Tokunaga. Corpus based method of transforming nominalized phrases into clauses for text mining application. 電子情報通信学会「情報アクセスのためのテキスト処理」小特集号掲載予定, 2003.
- [52] Akira Terada, Takenobu Tokunaga, and Hozumi Tanaka. Automatic expansion of abbreviations by using context and character information. (to appear) in *Information Processing & Management*.
- [53] James Thomas, David Milward, Christos Ouzounis, Stephen Pulman, and Mark Carroll. Automatic extraction of protein interactions from scientific abstracts. In *5th Pacific Symposium on Biocomputing*, pp. 538–549, 2000.
- [54] Janine Toole. A hybrid approach to the identification and expansion of abbreviations. In *Proceedings of RIAO'2000*, Vol. 1, pp. 725–736, 2000.
- [55] Ramasamy Uthurusamy., Linda G. Means., and Kurt S. Godden. Extracting knowledge from diagnostic databases. *IEEE Expert*, Vol. 8(6), pp. 27–38, 1993.
- [56] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, 1979.
- [57] Akane Yakushiji, Yuka Tateishi, Yusuke Miyao, and Jun-ichi Tsujii. Event extraction from biomedical papers using a full parser. In *Proceedings of Pacific Symposium on Biocomputing 2001*, pp. 408–419, 2001.
- [58] R. Yangarber, R. Grishman, P. Tapanainen, and S. Huttunen. Automatic acquisition of domain knowledge for information extraction. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, 2000.

- [59] 阿玉泰宗. 文脈情報を利用した略語の復元. 卒業論文, 東京工業大学, 2002.
- [60] 井上和子, 山田洋, 河野武, 成田一. 現代の英文法 6 「名詞」. 研究社, 1985.
- [61] 鈴木伸哉, 劉連文, 榊井文人, 河合敦夫, 椎野努. 固有名抽出のための短縮形知識の探索手法. 情報処理学会自然言語処理研究会 (2000-NL-140), No. 13, pp. 91–96, 2000.
- [62] 工藤拓, 山本薫, 坪井祐太, 松本裕治. 言語情報を利用したテキストマイニング. 情報処理学会自然言語処理研究会 (2002-NL-148), No. 10, pp. 65–72, 2002.
- [63] 徳永健伸. 情報検索と言語処理. 東大出版, 1999.

付 録 A 抽出結果

第 5 章で節／文の抽出に使用した長さ 2 以上で、最小サポート値 2 の抽出パターンの出力結果を示す。A.1 節では、本文で使用した ASRS コーパス (ASRS1 と呼ぶ)、A.2 節では、本文で使
用しなかった ASRS コーパス (ASRS2 と呼ぶ) のマイニング結果を示す。ASRS1 コーパスと
ASRS2 コーパスの Synopsis 部分の統計情報を表 A.1 に示す。

表 A.1: ASRS コーパスの統計情報

	ASRS1 コーパス	ASRS2 コーパス
文書総数	24,626	43,896
1 文書の平均語数	15.4	19.1
1 文の語数	10.9	12.5
1 文書の平均文数	1.4	1.5

A.1 ASRS コーパス 1

(controller)_{obj} (approach)_v → (system error)_{np} 頻度 2
 (be instructed)_v → (pilot)_{subj} (deviate)_v 頻度 2
 (with another aircraft)_{pp} (conflict)_v → (less than standard separation)_{np} 頻度 3
 (active runway)_{obj} (without clearance)_{pp} (cross)_v → (no conflict)_{np} 頻度 2
 (emergency)_{obj} (declare)_v → (descend)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (restriction)_{obj} (cross)_v 頻度 13
 (from altitude)_{pp} (deviate)_v → (from track)_{pp} (deviate)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (overshoot)_v 頻度 17
 (from altitude)_{pp} (deviate)_v → (in climb)_{pp} (overshoot)_v 頻度 20
 (from altitude)_{pp} (deviate)_v → (in descent)_{pp} (overshoot)_v 頻度 8
 (from altitude)_{pp} (deviate)_v → (undershoot)_v → (restriction)_{obj} (cross)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (in descent)_{pp} (undershoot)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{subj} (restriction)_{obj} (cross)_v 頻度 5
 (from altitude)_{pp} (deviate)_v → (altitude)_{subj} (overshoot)_v 頻度 14
 (from altitude)_{pp} (deviate)_v → (altitude)_{subj} (in climb)_{pp} (overshoot)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{subj} (in descent)_{pp} (overshoot)_v 頻度 4
 (from altitude)_{pp} (deviate)_v → (altitude)_{subj} (undershoot)_v
 → (restriction)_{obj} (cross)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{subj} (undershoot)_v

→ (altitude)_{sbj} (restriction)_{obj} (cross)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude crossing restriction)_{sbj} (make)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} 頻度 13
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion acr mlg)_{np} (after tcasii ra)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (crossing restriction not)_{sbj} (comply)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (excursion)_{np} 頻度 6
 (from altitude)_{pp} (deviate)_v → (excursion)_{np} (from assigned altitude)_{pp} 頻度 4
 (from altitude)_{pp} (deviate)_v → (excursion)_{np} (from clearance altitude)_{pp} 頻度 6
 (from altitude)_{pp} (deviate)_v → (flight crew)_{np} (of lgt)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (flight crew)_{np} (of mlg)_{pp} 頻度 3
 (from altitude)_{pp} (deviate)_v → (hearback problem)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (potential conflict)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (unable)_{np} 頻度 3
 (from altitude)_{pp} (deviate)_v → (wrong altimeter setting)_{np} 頻度 6
 (from track)_{pp} (deviate)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (in heading)_{pp} (deviate)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (fuel)_{obj} (dump)_v → (without incident)_{pp} (land)_v 頻度 2
 (fuel)_{obj} (dump)_v → (land)_{obj} (return)_v 頻度 2
 (exhaustion)_{obj} (fuel)_v → (off airport)_{obj} (land)_v 頻度 2
 (off airport)_{obj} (land)_v → (no injury)_{np} 頻度 2
 (aircraft control)_{obj} (lose)_v → (runway excursion)_{np} 頻度 2
 (oil pressure)_{obj} (lose)_v → (inflight engine)_{sbj} (shut down)_v 頻度 2
 (pressurization)_{obj} (lose)_v → (emergency descent)_{np} 頻度 3
 (separation)_{obj} (lose)_v → (operational error)_{np} 頻度 3
 (altitude)_{obj} (overshoot)_v → (less than standard separation)_{np} 頻度 2
 (tca)_{obj} (penetrate)_v → (sma)_{np} 頻度 2
 (return)_v → (land)_v 頻度 8
 (concept)_{obj} (see avoid)_v → (evasive action)_{sbj} (be taken)_v 頻度 2
 (engine)_{obj} (shut down)_v → (emergency)_{obj} (declare)_v 頻度 2
 (runway)_{obj} (turn off)_v → (aircraft)_{obj} (conflict)_v 頻度 2
 (undershoot)_v → (restriction)_{obj} (cross)_v 頻度 3
 (altitude crossing restriction)_{obj} (undershoot)_v → (situation)_{np} 頻度 3
 (acr)_{np} → (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v 頻度 2
 (acr)_{np} → (less than standard separation)_{np} 頻度 2
 (acr)_{np} → (pilot)_{sbj} (deviate)_v → (system error)_{np} 頻度 3
 (acr)_{np} → (pilot)_{sbj} (deviate)_v → (tcasii ra evasive action)_{sbj} (be taken)_v 頻度 4
 (acr)_{np} → (recurrent event tcasii)_{np} 頻度 2
 (acr)_{np} → (system error)_{np} 頻度 8
 (acr)_{np} → (tcasii ra evasive action)_{sbj} (be taken)_v 頻度 4
 (acr)_{np} → (unauthorized runway entry)_{np} 頻度 3
 (acr)_{sbj} (altitude)_{obj} (be assigned)_v
 → (less than standard separation)_{np} (with another aircraft)_{pp} 頻度 2

(acr)_{sbj} (through altitude)_{pp} (be climbed)_v
 → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (through occupied altitude)_{pp} (be climbed)_v
 → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (climb)_{obj} (through occupied altitude)_{pp} (be issued)_v
 → (less than standard separation)_{np} 頻度 3
 (acr)_{sbj} (for takeoff)_{pp} (clear)_v → (operational error)_{np} 頻度 2
 (acr)_{sbj} (through occupied altitude)_{pp} (climb)_v
 → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (in close proximity)_{pp} (come)_v → (less than standard separation)_{np} 頻度 3
 (acr)_{sbj} (through altitude)_{pp} (descend)_v → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (through occupied altitude)_{pp} (descend)_v
 → (less than standard separation)_{np} 頻度 5
 (acr)_{sbj} (to occupied altitude)_{pp} (descend)_v
 → (less than standard separation)_{np} 頻度 8
 (acr)_{sbj} (another aircraft)_{obj} (to altitude)_{pp} (descend)_v
 → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (assigned altitude)_{obj} (below assigned altitude)_{pp} (descend)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (from taxi clearance)_{pp} (deviate)_v → (active runway)_{obj} (enter)_v 頻度 2
 (acr)_{sbj} (with altitude crossing restriction)_{pp} (fail to comply)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (with atc clearance)_{pp} (fail to comply)_v
 → (altitude crossing restriction)_{obj} (undershoot)_v → (situation)_{np} 頻度 3
 (acr)_{sbj} (with atc clearance)_{pp} (fail to comply)_v → (operational deviation)_{np} 頻度 2
 (acr)_{sbj} (with atc clearance)_{pp} (fail to comply)_v → (pilot)_{sbj} (deviate)_v 頻度 5
 (acr)_{sbj} (with atc clearance)_{pp} (fail to comply)_v
 → (situation)_{np} 頻度 3
 (acr)_{sbj} (with atc instruction)_{pp} (fail to comply)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (with atc clearance)_{pp} (fail to comply)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (be taken)_v 頻度 4
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v
 → (multiple runway operation)_{sbj} (runway)_{obj} (intersect)_v 頻度 2
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v → (tcasii ra)_{np} 頻度 2
 (acr)_{sbj} (near mid air collision standard separation)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (than standard separation)_{pp} (have)_v → (probable system error)_{np} 頻度 2
 (acr)_{sbj} (airborne conflict)_{obj} (with smt)_{pp} (have)_v
 → (concept)_{obj} (see avoid)_v 頻度 2

(acr)_{sbj} (less than standard radar separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 4
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (acr)_{sbj} (tcasii ra evasive action)_{obj} (be taken)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 15
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (situation)_{np} 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 9
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} → (situation)_{np} 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np}
 → (tcasii ta ra)_{sbj} (action)_{obj} (be evasive)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (tcasii ta ra)_{sbj} (action)_{obj} (be evasive)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from acr y)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np}
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from aircraft y)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from atx)_{pp} (have)_v
 → (system error)_{np} 頻度 4
 (acr)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from fgt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from military)_{pp} (have)_v
 → (system error)_{np} 頻度 5
 (acr)_{sbj} (less than standard separation)_{obj} (from mlt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 12
 (acr)_{sbj} (less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 10
 (acr)_{sbj} (less than standard separation)_{obj} (with acr)_{pp} (have)_v

→ (system error)_{np} 頻度 3
 (acr)_{sbj} (near mid air collision less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (tcasii ra)_{obj} (have)_v → (respond)_v 頻度 7
 (acr)_{sbj} (tcasii ra)_{obj} (have)_v → (with climb)_{pp} (respond)_v 頻度 2
 (acr)_{sbj} (climb)_{obj} (through occupied altitude)_{pp} (issue)_v
 → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (on occupied runway)_{pp} (land)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (acr)_{sbj} (clearance)_{obj} (without clearance)_{pp} (land)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (on occupied runway)_{pp} (land)_v
 → (system error)_{np} 頻度 3
 (acr)_{sbj} (tcasii ra)_{obj} (receive)_v → (respond)_v 頻度 4
 (acr lgt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (crossing restriction)_{sbj} (not be met)_v 頻度 2
 (acr lgt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 22
 (acr lgt)_{sbj} (altitude)_{obj} (overshoot)_v
 → (less than standard separation)_{np} 頻度 2
 (acr mdt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 4
 (acr mlg)_{sbj} (through altitude)_{pp} (climb)_v
 → (less than standard separation)_{np} 頻度 2
 (acr mlg)_{sbj} (from altitude)_{pp} (deviate)_v
 → (crossing restriction)_{sbj} (not be met)_v 頻度 11
 (acr mlg)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 56
 (acr mlg)_{sbj} (from altitude)_{pp} (deviate)_v
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 2
 (acr mlg)_{sbj} (from clearance route)_{pp} (deviate)_v
 → (track)_{sbj} (from heading)_{pp} (deviate)_v 頻度 3
 (acr mlg)_{sbj} (land)_{obj} (return)_v → (takeoff)_{np} 頻度 3
 (acr mlg engine)_{sbj} (fail)_v → (emergency)_{sbj} (declare)_v 頻度 2
 (acr possible less than standard separation)_{np} (from acr)_{pp}
 → (possible system error)_{np} 頻度 3
 (acr potential less than standard separation)_{np} (from acr)_{pp}
 → (potential system error)_{np} 頻度 3
 (acr undershot altitude)_{sbj} (restriction)_{obj} (cross)_v → (situation)_{np} 頻度 3
 (acr wdb)_{sbj} (from altitude)_{pp} (deviate)_v
 → (crossing restriction)_{sbj} (not be met)_v 頻度 2
 (acr wdb)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 10

(acr wdb)_{sbj} (from clearance route)_{pp} (deviate)_v
 → (track)_{sbj} (from heading)_{pp} (deviate)_v 頻度 2
 (acr y)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 18
 (acr y descent)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (aircraft)_{sbj} (altitude)_{obj} (be assigned)_v
 → (less than standard separation)_{np} 頻度 4
 (aircraft)_{sbj} (altitude)_{obj} (be assigned)_v
 → (less than standard separation)_{np} (with another aircraft)_{pp} 頻度 2
 (aircraft)_{sbj} (damage)_v → (no injury)_{np} 頻度 3
 (aircraft)_{sbj} (through altitude)_{pp} (descend)_v
 → (less than standard separation)_{np} 頻度 2
 (aircraft)_{sbj} (through occupied altitude)_{pp} (descend)_v
 → (less than standard separation)_{np} 頻度 3
 (aircraft)_{sbj} (to occupied altitude)_{pp} (descend)_v
 → (less than standard separation)_{np} 頻度 2
 (aircraft)_{sbj} (artcc sector y)_{obj} (from sector)_{pp} (enter)_v
 → (operational deviation)_{np} 頻度 2
 (aircraft)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 4
 (aircraft)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 6
 (aircraft)_{sbj} (adjacent center airspace)_{obj} (without coordination)_{pp} (penetrate)_v
 → (operational deviation)_{np} 頻度 2
 (aircraft)_{np} (at same altitude)_{pp} → (system error)_{np} 頻度 2
 (altitude)_{sbj} (on star)_{pp} (deviate)_v
 → (clearance)_{sbj} (misinterpret)_v 頻度 2
 (altitude)_{sbj} (undershoot)_v → (restriction)_{obj} (cross)_v 頻度 2
 (altitude)_{sbj} (undershoot)_v
 → (altitude)_{sbj} (restriction)_{obj} (cross)_v 頻度 3
 (atx)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (atx smt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 2
 (both ifr)_{np} → (operational error)_{np} 頻度 2
 (cabin pressurization system)_{sbj} (fail)_v
 → (lgt)_{sbj} (emergency descent)_{obj} (make)_v 頻度 2
 (charter ltt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 2
 (clearance)_{sbj} (intercept)_v → (acr mlg)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 2
 (clearance)_{sbj} (intercept)_v → (excursion)_{np} (from clearance altitude)_{pp} 頻度 2

(commuter ltt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 6
 (commuter mdt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 2
 (cpr)_{np} → (pilot)_{sbj} (deviate)_v 頻度 4
 (cpr)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (cpr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (cpr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 11
 (cpr)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (cpr)_{sbj} (less than standard separation)_{obj} (from mlt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (cpr)_{sbj} (less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (cpr ltt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 4
 (cpr smt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 2
 (cpr y)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (distraction)_{np} → (from altitude)_{pp} (deviate)_v 頻度 2
 (emergency)_{sbj} (declare)_v → (and land)_{obj} (return)_v 頻度 3
 (emergency)_{sbj} (declare)_v → (land)_{obj} (return)_v 頻度 2
 (emergency)_{sbj} (declare)_v → (aircraft)_{sbj} (to alternate)_{pp} (be diverted)_v 頻度 2
 (engine)_{sbj} (fail)_v → (landing)_{obj} (force)_v 頻度 3
 (engine)_{sbj} (fire)_v → (warning)_{np} 頻度 4
 (evasive action)_{sbj} (be taken)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (evasive action)_{sbj} (be taken)_v → (system error)_{np} 頻度 2
 (fgt)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (flight crew)_{sbj} (clearance)_{obj} (believe)_v
 → (read back same)_v 頻度 2
 (flight crew)_{sbj} (on transponder)_{pp} (fail to turn)_v
 → (operational deviation)_{np} 頻度 2
 (flight crew)_{sbj} (on transponder)_{pp} (fail to turn)_v
 → (pilot)_{sbj} (deviate)_v → (operational deviation)_{np} 頻度 2
 (fuel)_{sbj} (dump)_v → (land)_{obj} (return)_v 頻度 2
 (fuel exhaustion)_{np} → (ga sma)_{sbj} (landing)_{obj} (force)_v 頻度 2
 (fuel exhaustion)_{np} → (ga sma)_{sbj} (at off airport)_{pp} (land)_v 頻度 6
 (fuel starvation)_{np} → (landing)_{obj} (force)_v 頻度 3
 (ga sma)_{sbj} (from altitude)_{pp} (deviate)_v

→ (excursion)_{np} (from clearance altitude)_{pp} 頻度 15
 (ga sma)_{sbj} (without authorization)_{pp} (penetrate)_v → (tca)_{np} 頻度 5
 (ga sma engine failure)_{np} → (emergency off airport landing)_{np} 頻度 2
 (ga smt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 3
 (hit propeller)_{np} → (recovery 2)_{pp} 頻度 2
 (inflight engine)_{sbj} (shut down)_v → (to destination)_{pp} (continue)_v 頻度 2
 (inflight engine)_{sbj} (shut down)_v → (emergency)_{sbj} (declare)_v 頻度 2
 (less than standard)_{np} → (operational error)_{np} 頻度 8
 (less than standard separation)_{np} → (operational error)_{np} 頻度 8
 (less than standard separation)_{np} → (pilot)_{sbj} (deviate)_v 頻度 4
 (less than standard separation)_{np} → (system error)_{np} 頻度 16
 (less than standard separation)_{sbj} (occur)_v → (system error)_{np} 頻度 2
 (less than standard separation)_{np} (between 2 acr)_{pp} → (operational error)_{np} 頻度 2
 (less than standard separation)_{np} (between 2 acr aircraft)_{pp} → (opal error)_{np} 頻度 3
 (less than standard separation)_{np} (between 2 acr aircraft)_{pp}
 → (operational error)_{np} 頻度 21
 (less than standard separation)_{np} (between 2 ga aircraft)_{pp}
 → (operational error)_{np} 頻度 3
 (less than standard separation)_{np} (between acr)_{pp} → (system error)_{np} 頻度 3
 (less than standard separation)_{np} (between acr and another ifr aircraft)_{pp}
 → (operational error)_{np} 頻度 2
 (less than standard separation)_{np} (between acr and corporate aircraft)_{pp}
 → (operational error)_{np} 頻度 2
 (less than standard separation)_{np} (between acr and corporate aircraft)_{pp}
 → (pilot)_{np} 頻度 2
 (less than standard separation)_{np} (between acr and corporate jet)_{pp}
 → (operational error)_{np} 頻度 2
 (less than standard separation)_{np} (between acr and military aircraft)_{pp}
 → (operational error)_{np} 頻度 5
 (less than standard separation)_{np} (between acr and mlt)_{pp}
 → (system error)_{np} 頻度 2
 (less than standard separation)_{np} (between acr and smt)_{pp}
 → (system error)_{np} 頻度 2
 (less than standard separation)_{np} (between corporate aircraft and acr)_{pp}
 → (operational error)_{np} 頻度 2
 (less than standard separation)_{np} (between corporate jet and acr)_{pp}
 → (operational error)_{np} 頻度 4
 (less than standard separation)_{np} (between two acr)_{pp}
 → (operational error)_{np} 頻度 2
 (less than standard separation)_{np} (between two acr aircraft)_{pp}
 → (operational error)_{np} 頻度 8
 (mlt)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v

→ (system error)_{np} 頻度 4
 (new taxiway centerline light)_{np} → (active runway)_{obj} (enter)_v 頻度 2
 (new taxiway centerline light)_{np}
 → (acr)_{sbj} (from taxi clearance)_{pp} (deviate)_v → (active runway)_{obj} (enter)_v 頻度 2
 (non adherence)_{np} (to atc clearance)_{pp} → (pilot)_{sbj} (deviate)_v 頻度 2
 (operational deviation)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (operational deviation)_{np} → (situation)_{np} 頻度 5
 (operational error)_{np} → (controller)_{sbj} (report)_v 頻度 2
 (parachute operator)_{sbj} (about airport)_{obj} (complain)_v
 → (incident)_{obj} (at same time)_{pp} (cite)_v 頻度 2
 (part)_{sbj} (within component assembly)_{pp} (finite life)_{obj} (reach)_v
 → (past finite life)_{obj} (fly)_v 頻度 2
 (pilot)_{sbj} (deviate)_v → (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (pilot)_{sbj} (deviate)_v → (and or operational error)_{np} 頻度 2
 (pilot)_{sbj} (deviate)_v → (less than standard separation)_{np} 頻度 5
 (pilot)_{sbj} (deviate)_v → (operational deviation)_{np} 頻度 9
 (pilot)_{sbj} (deviate)_v → (operational error)_{np} 頻度 2
 (pilot)_{sbj} (deviate)_v → (potential conflict)_{np} 頻度 2
 (pilot)_{sbj} (deviate)_v → (situation)_{np} 頻度 10
 (pilot)_{sbj} (deviate)_v → (system error)_{np} 頻度 17
 (pilot)_{sbj} (deviate)_v → (tcasii ra evasive action)_{sbj} (be taken)_v 頻度 4
 (possible less than standard separation)_{np} (between acr and acr)_{pp}
 → (possible system error)_{np} 頻度 3
 (potential conflict)_{np} → (less than standard separation)_{np} 頻度 3
 (potential conflict)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (potential loss)_{np} (of separation)_{pp} → (operational deviation)_{np} 頻度 2
 (runway)_{sbj} (transgress)_v → (onto active runway)_{pp} (taxi)_v 頻度 2
 (runway excursion)_{np} → (aircraft control)_{obj} (lose)_v 頻度 2
 (runway incursion)_{np} → (pilot)_{sbj} (deviate)_v 頻度 3
 (sma)_{np} → (pilot)_{sbj} (deviate)_v 頻度 6
 (sma)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v 頻度 7
 (sma)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 7
 (sma)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (sma)_{sbj} (less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 4
 (sma)_{sbj} (less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (sma)_{sbj} (on occupied runway)_{pp} (land)_v → (pilot)_{sbj} (deviate)_v 頻度 2
 (sma y)_{sbj} (near mid air collision)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2

(sma y)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (smt)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (smt)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 8
 (smt)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (smt)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (smt)_{sbj} (less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (smt)_{sbj} (less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (smt y)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (system error)_{np} → (acr)_{sbj} (tcasii ra evasive action)_{obj} (be taken)_v 頻度 2
 (system error)_{np} → (acr)_{sbj} (on occupied runway)_{pp} (land)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 3
 (system error)_{np} → (operational deviation)_{np} 頻度 3
 (system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 33
 (system error)_{np} → (situation)_{np} 頻度 7
 (system error)_{np} → (tcasii ra evasive action)_{sbj} (be taken)_v 頻度 4
 (system error)_{np} → (tcasii ta ra)_{np} 頻度 2
 (system error)_{np} → (tcasii ta ra)_{sbj} (action)_{obj} (be evasive)_v 頻度 4
 (takeoff)_{sbj} (abort)_v → (loud noise)_{sbj} (hear)_v 頻度 2
 (tcasii ra)_{np} → (flight crew)_{np} (to climb)_{pp} 頻度 2
 (traffic advisory)_{sbj} (be given)_v → (traffic)_{sbj} (sight)_v 頻度 2
 (unauthorized runway entry)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (uncoord penetration)_{np} (of airspace)_{pp} → (operational deviation)_{np} 頻度 14
 (uncoordinated penetration)_{np} (of airspace)_{pp} → (operational deviation)_{np} 頻度 4

A.2 ASRS コーパス 2

(abort)_v → (on runway)_{pp} (stop)_v 頻度 2
 (be)_v → (make)_v 頻度 3
 (be disconnected)_v → (be)_v 頻度 2
 (be given)_v → (traffic)_{sbj} (sight)_v 頻度 4
 (be instructed)_v → (without clearance)_{pp} (lands)_v 頻度 2
 (be lost)_v → (flight)_{sbj} (assist)_v 頻度 2
 (through occupied altitude)_{pp} (climb)_v → (system error)_{np} 頻度 2
 (clearance)_{obj} (confuse)_v → (from altitude)_{pp} (deviate)_v 頻度 2

(emergency)_{obj} (declare)_v → (be diverted)_v 頻度 3
 (emergency)_{obj} (declare)_v → (to airport)_{pp} (be diverted)_v 頻度 2
 (emergency)_{obj} (declare)_v → (divert)_v 頻度 3
 (emergency)_{obj} (declare)_v → (land)_{obj} (divert)_v 頻度 3
 (emergency)_{obj} (declare)_v → (fuel)_{obj} (dump)_v 頻度 2
 (emergency)_{obj} (declare)_v → (land)_v 頻度 5
 (emergency)_{obj} (declare)_v → (land enroute)_v 頻度 3
 (emergency)_{obj} (declare)_v → (return)_v 頻度 2
 (emergency)_{obj} (declare)_v → (to departure airport)_{pp} (return)_v 頻度 4
 (emergency)_{obj} (declare)_v → (to land)_{pp} (return)_v 頻度 4
 (emergency)_{obj} (declare)_v → (land)_{obj} (return)_v 頻度 4
 (emergency)_{obj} (declare)_v → (land)_{np} 頻度 2
 (from cle)_{pp} (depart)_v → (emergency)_{sbj} (for aircraft)_{pp} (be declared)_v 頻度 2
 (from cle)_{pp} (depart)_v → (main cabin door)_{sbj} (on beech 1900)_{pp} (open)_v
 → (emergency)_{sbj} (for aircraft)_{pp} (be declared)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (be met)_v 頻度 16
 (from altitude)_{pp} (deviate)_v → (be restricted)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (traffic)_{obj} (climbing)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (of assigned altitude)_{pp} (descend)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (do n)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (at assigned altitude)_{pp} (fail to level)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (factor)_{obj} (fatigue)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (intervene)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (clearance)_{obj} (misunderstand)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (overshoot)_v 頻度 8
 (from altitude)_{pp} (deviate)_v → (altitude)_{obj} (overshoot)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (acr)_{sbj} (altitude)_{obj} (climbs)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (aircraft)_{sbj} (climb)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (aircraft)_{sbj} (lose)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (aircraft equipment problem)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{np} 頻度 5
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (be overshoot)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in climb)_{pp} (be overshoot)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (be overshoot)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (be restricted)_v 頻度 6
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (restriction)_{obj} (cross)_v 頻度 11
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (overshoot)_v
 → (traffic)_{obj} (climbing)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (overshoot)_v
 → (acr)_{sbj} (altitude)_{obj} (climbs)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (overshoot)_v
 → (wrong altimeter setting)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in climb)_{pp} (overshoot)_v

→ (non adherence)_{np} (to atc clearance)_{pp} 頻度 4
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (clearance)_{sbj} (acknowledge)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 5
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (wrong altimeter setting)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (in descent procedure)_{pp} (overshoot)_v 頻度 9
 (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (in non adherence)_{pp} (overshoot)_v 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (excursion)_{obj} (in descent)_{pp} (overshoot)_v 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (undershoot)_v → (altitude)_{sbj} (restriction)_{obj} (cross)_v 頻度 5
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (undershoot)_v
 → (altitude crossing)_{sbj} (be restricted)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (undershoot)_v
 → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 4
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (undershoot)_v
 → (potential conflict)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (in climb)_{pp} (undershoot)_v 頻度 8
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (undershoot)_v
 → (altitude crossing)_{sbj} (be met)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (in descent)_{pp} (undershoot)_v
 → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 8
 (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (in descent procedure)_{pp} (undershoot)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude crossing)_{sbj} (be met)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude crossing)_{sbj} (be restricted)_v 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude crossing)_{sbj} (not)_{obj} (be restricted)_v 頻度 7
 (from altitude)_{pp} (deviate)_v
 → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 16
 (from altitude)_{pp} (deviate)_v → (altitude crossing restriction)_{sbj} (make)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np}
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np}
 → (non adherence)_{np} (to atc clearance procedure)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np}

→ (potential conflict less than standard separation)_{np} 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{sbj} (occur)_v 頻度 8
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} (after tcasii ra)_{pp} 頻度 4
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} (during tcasii ra)_{pp} 頻度 4
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (from assigned altitude)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (in evasive action climb)_{pp} 頻度 4
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (in flight crew pilo response)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (in flight crew response)_{pp} 頻度 21
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} (in non adherence)_{pp} 頻度 3
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (in pilo flight crew response)_{pp} 頻度 3
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (in pilo response)_{pp} 頻度 5
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} (in premature descent)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} (on tcasii ra)_{pp} 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} (to tcasii ra)_{pp} 頻度 14
 (from altitude)_{pp} (deviate)_v
 → (altitude excursion acr)_{np} (in flight crew response)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (altitude excursion altitude)_{sbj} (overshoot)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (altitude excursion pic)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (atc)_{sbj} (intervene)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (cfit)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (clearance)_{sbj} (acknowledge)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (clearance)_{sbj} (intercept)_v 頻度 4
 (from altitude)_{pp} (deviate)_v → (climb)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (controller)_{sbj} (intervene)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (controller)_{sbj} (crew)_{obj} (intervene)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (crossing)_{sbj} (be restricted)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (crossing)_{sbj} (not met)_{obj} (be restricted)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (descent)_{np} 頻度 3
 (from altitude)_{pp} (deviate)_v → (error)_{sbj} (admit)_v 頻度 5
 (from altitude)_{pp} (deviate)_v → (evasive action climb)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (evasive action descent)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (excursion)_{np} 頻度 15
 (from altitude)_{pp} (deviate)_v → (excursion)_{np} (from assigned altitude)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (excursion)_{np} (in flight crew response)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v → (false warning)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (flight crew)_{sbj} (distract)_v 頻度 4

(from altitude)_{pp} (deviate)_v → (heading)_{sbj} (from track)_{pp} (deviate)_v 頻度 3
 (from altitude)_{pp} (deviate)_v → (inflight engine)_{sbj} (shut down)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (less than standard separation)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (miscom altitude clearance)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (night operation)_{np} 頻度 4
 (from altitude)_{pp} (deviate)_v → (non adherence)_{np} (to atc clearance)_{pp} 頻度 18
 (from altitude)_{pp} (deviate)_v
 → (non adherence)_{np} (to atc clearance instruction)_{pp} 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (non adherence)_{np} (to atc clearance procedure)_{pp} 頻度 3
 (from altitude)_{pp} (deviate)_v → (potential conflict)_{np} 頻度 6
 (from altitude)_{pp} (deviate)_v → (potential conflict)_{sbj} (ensue)_v 頻度 2
 (from altitude)_{pp} (deviate)_v
 → (potential conflict less than standard separation)_{np} 頻度 5
 (from altitude)_{pp} (deviate)_v → (reporter)_{sbj} (fly)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (reporter error)_{sbj} (admit)_v 頻度 2
 (from altitude)_{pp} (deviate)_v → (sid)_{np} 頻度 2
 (from altitude)_{pp} (deviate)_v → (tcasii ra)_{np} 頻度 4
 (from altitude)_{pp} (deviate)_v → (wrong altimeter setting)_{np} 頻度 9
 (from altitude)_{pp} (deviate)_v → (wrong altitude setting)_{np} 頻度 2
 (from track)_{pp} (deviate)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (in heading)_{pp} (deviate)_v → (from altitude)_{pp} (deviate)_v 頻度 4
 (to alternate airport)_{pp} (divert)_v → (emergency)_{sbj} (declare)_v 頻度 2
 (do n)_v → (flight crew)_{sbj} (do n)_v 頻度 2
 (do n)_v → (reporter)_{sbj} (say)_v 頻度 2
 (tower)_{obj} (fail to contact)_v → (without clearance)_{pp} (land)_v 頻度 2
 (altitude climb)_{obj} (fail to meet)_v → (pilot)_{sbj} (deviate)_v 頻度 2
 (fly)_v → (believe)_v 頻度 2
 (fly)_v → (from track)_{pp} (deviate)_v 頻度 2
 (tcasii ra)_{obj} (get)_v → (less than standard separation)_{sbj} (occur)_v 頻度 2
 (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v 頻度 7
 (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (be taken)_v
 → (concept)_{obj} (see avoid)_v 頻度 2
 (near mid air collision)_{pp} (have)_v → (pilot)_{sbj} (deviate)_v 頻度 3
 (than standard separation)_{pp} (have)_v → (system error)_{np} 頻度 2
 (less than standard separation)_{obj} (have)_v → (system error)_{np} 頻度 5
 (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (less than standard separation)_{obj} (from acr)_{pp} (have)_v → (system error)_{np} 頻度 10
 (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (less than standard separation)_{obj} (from cpr)_{pp} (have)_v → (system error)_{np} 頻度 2
 (near mid air collision)_{obj} (than standard separation)_{pp} (have)_v

→ (system error)_{np} 頻度 2
 (go around)_{obj} (issue)_v → (near mid air collision)_{pp} (have)_v 頻度 2
 (go around)_{obj} (issue)_v → (pilot)_{sbj} (deviate)_v 頻度 2
 (without authorization)_{pp} (land)_v → (without clearance)_{pp} (land)_v 頻度 2
 (without authorization)_{pp} (land)_v → (without clearance)_{pp} (landing)_v 頻度 9
 (without authorization)_{pp} (land)_v → (clearance landing)_{sbj} (receive)_v 頻度 6
 (without authorization)_{pp} (land)_v → (frequency change)_{sbj} (miss)_v 頻度 2
 (without authorization)_{pp} (land)_v → (wrong airport landing)_{np} 頻度 2
 (without authorization)_{pp} (land)_v
 → (wrong runway)_{sbj} (landing)_{obj} (approach)_v 頻度 2
 (aircraft control)_{obj} (lose)_v → (autopilot)_{obj} (disconnect)_v 頻度 2
 (aircraft control)_{obj} (lose)_v → (aircraft)_{sbj} (damage)_v 頻度 2
 (aircraft control)_{obj} (lose)_v → (runway excursion)_{np} 頻度 7
 (oil quantity and pressure)_{obj} (lose)_v → (to field)_{pp} (return)_v 頻度 2
 (power)_{obj} (lose)_v → (land)_{obj} (return)_v 頻度 2
 (off airport)_{obj} (make)_v → (engine)_{np} 頻度 2
 (off airport)_{obj} (make)_v → (engine)_{sbj} (quit)_v 頻度 3
 (altitude assignment)_{obj} (misunderstand)_v → (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (overshoot)_v 頻度 2
 (altitude assignment)_{obj} (misunderstand)_v → (altitude)_{sbj} (overshoot)_v 頻度 2
 (tcasii ra)_{obj} (receive)_v → (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} 頻度 5
 (tcasii ra)_{obj} (receive)_v → (respond)_v 頻度 3
 (tcasii ra)_{obj} (receive)_v → (altitude excursion)_{np} 頻度 5
 (respond)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (return)_v → (without further incident)_{pp} (land)_v 頻度 2
 (return)_v → (land)_{np} 頻度 2
 (to land)_{pp} (return)_v → (takeoff)_{np} 頻度 2
 (land)_{obj} (return)_v → (aircraft)_{sbj} (damage)_v 頻度 2
 (out of fuel)_{pp} (run)_v → (off airport)_{obj} (make)_v → (engine)_{np} 頻度 2
 (out of fuel)_{pp} (run)_v → (engine)_{np} 頻度 2
 (see avoid)_v → (system error)_{np} 頻度 2
 (concept)_{obj} (see avoid)_v → (evasive action)_{sbj} (be taken)_v 頻度 26
 (concept)_{obj} (see avoid)_v → (evasive action descent)_{np} 頻度 2
 (concept)_{obj} (see avoid)_v → (pilot)_{sbj} (deviate)_v 頻度 2
 (engine)_{obj} (shut down)_v → (land)_{obj} (return)_v 頻度 2
 (engine)_{obj} (shut down)_v → (emergency)_{sbj} (declare)_v 頻度 2
 (state)_v → (reporter)_{sbj} (believe)_v 頻度 2
 (think)_v → (reporter)_{sbj} (say)_v 頻度 2
 (nose gear)_{obj} (to retract)_v → (land)_{obj} (return)_v 頻度 2
 (runway)_{obj} (turn off)_v → (landing)_{np} 頻度 2
 (acr)_{np} → (from altitude)_{pp} (deviate)_v 頻度 9
 (acr)_{np} → (concept)_{obj} (see avoid)_v 頻度 8

(acr)_{np} → (aircraft equipment problem tcasii)_{np} 頻度 2
 (acr)_{np} → (evasive action)_{sbj} (be taken)_v → (concept)_{obj} (see avoid)_v 頻度 3
 (acr)_{np} → (evasive action)_{sbj} (be taken)_v → (pilot)_{sbj} (deviate)_v 頻度 3
 (acr)_{np} → (evasive action)_{sbj} (be taken)_v → (system error)_{np} 頻度 6
 (acr)_{np} → (less than standard separation)_{np} 頻度 3
 (acr)_{np} → (non adherence)_{np} (to atc clearance)_{pp} 頻度 3
 (acr)_{np} → (pilot)_{sbj} (deviate)_v → (system error)_{np} 頻度 7
 (acr)_{np} → (possible less than standard separation)_{np} 頻度 5
 (acr)_{np} → (possible system error)_{np} 頻度 4
 (acr)_{np} → (runway incursion)_{np} 頻度 6
 (acr)_{np} → (system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 5
 (acr)_{np} → (system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 8
 (acr)_{sbj} (descent)_{obj} (through occupied altitude)_{pp} (be issued)_v
 → (less than standard separation)_{np} 頻度 2
 (acr)_{sbj} (in response)_{pp} (climb)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (through occupied altitude)_{pp} (climb)_v
 → (system error)_{np} 頻度 36
 (acr)_{sbj} (less than standard separation)_{obj} (to occupied altitude)_{pp} (climb)_v
 → (system error)_{np} 頻度 18
 (acr)_{sbj} (occupied altitude)_{obj} (to occupied altitude)_{pp} (descend)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (wrong altitude)_{obj} (to wrong altitude)_{pp} (descend)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr)_{sbj} (from altitude)_{pp} (deviate)_v → (without authorization)_{pp} (descend)_v 頻度 2
 (acr)_{sbj} (turbulence)_{obj} (encounter)_v → (cabin attendant)_{sbj} (injure)_v 頻度 2
 (acr)_{sbj} (have)_v → (ra)_{np} 頻度 2
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 4
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (be taken)_v
 → (concept)_{obj} (see avoid)_v 頻度 2
 (acr)_{sbj} (near mid air collision)_{pp} (have)_v
 → (tcasii ra)_{np} 頻度 2
 (acr)_{sbj} (than standard separation)_{pp} (have)_v → (system error)_{np} 頻度 3
 (acr)_{sbj} (acr y)_{obj} (than standard separation)_{pp} (have)_v → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard radar separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (have)_v
 → (system error)_{np} 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 31
 (acr)_{sbj} (less than standard separation)_{obj} (from acr y)_{pp} (have)_v

→ (system error)_{np} 頻度 4
 (acr)_{sbj} (less than standard separation)_{obj} (from acry roll)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 5
 (acr)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (acr)_{sbj} (less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (near mid air collision less than standard separation)_{obj} (from acry)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (near mid air collision less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (near mid air collision less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (tcasii)_{obj} (have)_v → (ra)_{np} 頻度 2
 (acr)_{sbj} (tcasii ra)_{obj} (have)_v → (comply)_v 頻度 2
 (acr)_{sbj} (tcasii ra)_{obj} (have)_v → (respond)_v 頻度 9
 (acr)_{sbj} (tcasii ra)_{obj} (have)_v → (with climb)_{pp} (respond)_v 頻度 2
 (acr)_{sbj} (tcasii ra)_{obj} (to climb)_{pp} (have)_v → (respond)_v 頻度 2
 (acr)_{sbj} (clearance)_{obj} (without clearance)_{pp} (land)_v
 → (without authorization)_{pp} (land)_v 頻度 2
 (acr)_{sbj} (less than standard separation)_{obj} (on occupied runway)_{pp} (land)_v
 → (system error)_{np} 頻度 4
 (acr)_{sbj} (acry)_{obj} (overtake)_v → (less than standard separation)_{obj} (have)_v
 → (system error)_{np} 頻度 2
 (acr)_{sbj} (acry)_{obj} (overtake)_v → (system error)_{np} 頻度 2
 (acr)_{sbj} (radio station interference)_{obj} (on runway 18l localizer frequency)_{pp} (receive)_v
 → (ongoing problem)_{sbj} (several acrs)_{obj} (be reported)_v 頻度 2
 (acr climb)_{sbj} (less than standard separation)_{obj} (from acry)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr climb)_{sbj} (to occupied altitude tcasii ra)_{pp} (less than standard separation)_{obj}
 (have)_v → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr climb)_{sbj} (to occupied altitude tcasii ra)_{pp} (less than standard separation)_{obj}
 (have)_v → (system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr crew)_{sbj} (report)_v → (error)_{sbj} (admit)_v 頻度 2
 (acr descent)_{sbj} (less than standard separation)_{obj} (from acry)_{pp} (have)_v
 → (system error)_{np} 頻度 24
 (acr descent)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 9
 (acr descent)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v

→ (system error)_{np} 頻度 7
 (acr descent)_{sbj} (less than standard separation)_{obj} (from military)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr descent)_{sbj} (less than standard separation)_{obj} (from mlt)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr descent)_{sbj} (less than standard separation)_{obj} (from mlt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr descent)_{sbj} (less than standard separation)_{obj} (from smt)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (acr descent)_{sbj} (near mid air collision less than standard separation)_{obj} (from acr)_{pp}
 (have)_v → (system error)_{np} 頻度 2
 (acr descent)_{np} (from acr)_{pp} → (system error)_{np} 頻度 2
 (acr flight)_{sbj} (communication)_{obj} (per icao rule)_{pp} (be established)_v
 → (controller)_{sbj} (control)_{obj} (of to airspace boundary)_{pp} (assume)_v 頻度 2
 (acr flight crew)_{sbj} (mdt)_{obj} (with undisclosed and unsigned)_{pp} (accept fly)_v
 → (aircraft maintenance discrepancy)_{np} 頻度 2
 (acr lgt)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 6
 (acr mlg)_{sbj} (from altitude)_{pp} (deviate)_v → (excursion)_{np} (from clearance)_{pp} 頻度 2
 (acr mlg)_{sbj} (from altitude)_{pp} (deviate)_v
 → (excursion)_{np} (from clearance altitude)_{pp} 頻度 22
 (acr mlg)_{sbj} (return)_v → (land)_v 頻度 2
 (acr tcasii)_{sbj} (less than standard separation)_{obj} (have)_v → (system error)_{np} 頻度 2
 (acr tcasii ra)_{np} → (concept)_{obj} (see avoid)_v 頻度 3
 (acr tcasii ra)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 5
 (acr tcasii ra)_{np} → (pilot)_{sbj} (deviate)_v 頻度 3
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (have)_v
 → (system error)_{np} 頻度 2
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 4
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 3
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 3
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (with acr)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v → (system error)_{np} 頻度 2
 (acr tcasii ra)_{sbj} (less than standard separation)_{obj} (with acr)_{pp} (have)_v

→ (system error)_{np} 頻度 2
 (acr tcasii ra)_{np} (from aircraft)_{pp} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr tcasii ra)_{np} (from ghost target)_{pp} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr tcasii ra)_{np} (from vfr aircraft)_{pp} → (concept)_{obj} (see avoid)_v 頻度 2
 (acr tcasii ra)_{np} (with vfr)_{pp} → (concept)_{obj} (see avoid)_v 頻度 3
 (acr tcasii ra)_{np} (with vfr)_{pp} → (evasive action)_{sbj} (be taken)_v
 → (concept)_{obj} (see avoid)_v 頻度 2
 (acr tcasii ra)_{np} (with vfr aircraft)_{pp} → (concept)_{obj} (see avoid)_v 頻度 4
 (acr tcasii ra)_{np} (with vfr aircraft)_{pp} → (evasive action)_{sbj} (be taken)_v
 → (concept)_{obj} (see avoid)_v 頻度 2
 (acr tcasii ra)_{np} (with vfr sma)_{pp} → (concept)_{obj} (see avoid)_v 頻度 2
 (acr tcasii ra)_{np} (with vfr sma)_{pp} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr tcasii ra)_{np} (with vfr smt)_{pp} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (acr tcasii ta ra)_{np} (with vfr aircraft)_{pp} → (concept)_{obj} (see avoid)_v 頻度 2
 (acr wdb crew)_{sbj} (calculation)_{obj} (in etops)_{pp} (use bda)_v → (find)_v 頻度 2
 (acr wrong altimeter setting dsnded)_{np} (below assigned altitude)_{pp}
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (acr y)_{sbj} (less than standard separation)_{obj} (have)_v → (system error)_{np} 頻度 3
 (acr y)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 11
 (acr y)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr y)_{sbj} (standard separation)_{obj} (on occupied runway)_{pp} (land)_v
 → (system error)_{np} 頻度 2
 (acr y climb)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 24
 (acr y climb)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr y climb)_{sbj} (less than standard separation)_{obj} (from mlt)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (acr y descent)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 20
 (acr y descent)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (aircraft)_{np} → (continue)_v 頻度 2
 (aircraft)_{np} → (flight crew)_{sbj} (emergency)_{obj} (declare)_v 頻度 3
 (aircraft)_{np} → (flight crew)_{np} (in descent)_{pp} 頻度 2
 (aircraft)_{np} → (pilot)_{sbj} (deviate)_v 頻度 3
 (aircraft)_{np} → (reporter)_{sbj} (fly)_v 頻度 2
 (aircraft)_{np} → (reporter)_{sbj} (state)_v 頻度 2
 (aircraft)_{np} → (system error)_{np} 頻度 9
 (aircraft)_{sbj} (be)_v → (approach controller)_{sbj} (intervene)_v 頻度 2
 (aircraft)_{sbj} (be damaged)_v → (no injury)_{obj} (be)_v 頻度 2

(aircraft)_{sbj} (be stopped)_v → (runway)_{obj} (enter)_v 頻度 2
 (aircraft)_{sbj} (less than standard separation)_{obj} (through occupied altitude)_{pp} (climb)_v
 → (system error)_{np} 頻度 2
 (aircraft)_{sbj} (damage)_v → (error)_{sbj} (admit)_v 頻度 4
 (aircraft)_{sbj} (divert)_v → (land uneventfully)_v 頻度 2
 (aircraft)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (aircraft)_{sbj} (to gate)_{pp} (return)_v → (pax)_{sbj} (be removed)_v 頻度 2
 (aircraft descent)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (aircraft equipment problem)_{np} → (be)_v 頻度 2
 (aircraft equipment problem)_{np} → (emergency)_{obj} (declare)_v 頻度 3
 (aircraft equipment problem)_{np} → (from altitude)_{pp} (deviate)_v 頻度 3
 (aircraft equipment problem)_{np} → (discover)_v 頻度 2
 (aircraft equipment problem)_{np} → (to departure airport)_{pp} (return)_v 頻度 2
 (aircraft equipment problem)_{np} → (land)_{obj} (return)_v 頻度 3
 (aircraft equipment problem)_{np} → (down engine)_{obj} (shut)_v 頻度 2
 (aircraft equipment problem)_{np} → (engine)_{obj} (shut down)_v 頻度 5
 (aircraft equipment problem)_{np} → (aircraft)_{np} 頻度 2
 (aircraft equipment problem)_{np} → (aircraft)_{sbj} (to destination)_{pp} (continue)_v 頻度 2
 (aircraft equipment problem)_{np}
 → (aircraft)_{sbj} (to departure airport)_{pp} (return)_v 頻度 2
 (aircraft equipment problem)_{np} → (depressurization)_{np} 頻度 2
 (aircraft equipment problem)_{np} → (emergency)_{np} 頻度 2
 (aircraft equipment problem)_{np} → (emergency)_{sbj} (declare)_v 頻度 8
 (aircraft equipment problem)_{np} → (engine)_{sbj} (fail)_v 頻度 5
 (aircraft equipment problem)_{np} → (engine)_{sbj} (warning)_{obj} (fire)_v
 → (engine)_{obj} (shut down)_v 頻度 2
 (aircraft equipment problem)_{np} → (engine)_{sbj} (shut down)_v 頻度 2
 (aircraft equipment problem)_{np} → (fire)_{sbj} (go out)_v 頻度 2
 (aircraft equipment problem)_{np} → (flight)_{sbj} (divert)_v 頻度 2
 (aircraft equipment problem)_{np} → (flight crew)_{np} 頻度 2
 (aircraft equipment problem)_{np} → (flight crew)_{sbj} (emergency)_{obj} (declare)_v 頻度 3
 (aircraft equipment problem)_{np}
 → (flight crew)_{sbj} (to departure airport)_{pp} (return)_v 頻度 2
 (aircraft equipment problem)_{np} → (generator)_{sbj} (fail)_v 頻度 2
 (aircraft equipment problem)_{np} → (inflight engine)_{sbj} (shut down)_v 頻度 2
 (aircraft equipment problem)_{np} → (landing)_{sbj} (be uneventful)_v 頻度 2
 (aircraft equipment problem)_{np} → (reporter)_{sbj} (say)_v 頻度 2
 (aircraft equipment problem)_{np} → (takeoff)_{sbj} (abort)_v 頻度 3
 (aircraft equipment problem)_{np} → (uneventful landing)_{sbj} (follow)_v 頻度 2
 (aircraft equipment problem malfunction)_{np} → (takeoff)_{sbj} (abort)_v 頻度 2
 (aircraft y climb)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v

→ (system error)_{np} 頻度 2
 (aircraft y climb)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (altitude)_{sbj} (bust)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (altitude)_{sbj} (restriction)_{obj} (cross)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (altitude)_{sbj} (overshoot)_v → (traffic)_{obj} (climbing)_v 頻度 2
 (altitude)_{sbj} (overshoot)_v → (acr)_{sbj} (altitude)_{obj} (climbs)_v 頻度 2
 (altitude)_{sbj} (overshoot)_v → (wrong altimeter setting)_{np} 頻度 2
 (altitude)_{sbj} (in climb)_{pp} (overshoot)_v
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 4
 (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 2
 (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (clearance)_{sbj} (acknowledge)_v 頻度 2
 (altitude)_{sbj} (in descent)_{pp} (overshoot)_v
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 5
 (altitude)_{sbj} (in descent)_{pp} (overshoot)_v → (wrong altimeter setting)_{np} 頻度 2
 (altitude)_{sbj} (undershoot)_v → (altitude)_{sbj} (restriction)_{obj} (cross)_v 頻度 5
 (altitude)_{sbj} (undershoot)_v
 → (altitude crossing)_{sbj} (be restricted)_v 頻度 2
 (altitude)_{sbj} (undershoot)_v → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 4
 (altitude)_{sbj} (undershoot)_v → (potential conflict)_{np} 頻度 2
 (altitude)_{sbj} (in descent)_{pp} (undershoot)_v → (altitude crossing)_{sbj} (be met)_v 頻度 2
 (altitude)_{sbj} (in descent)_{pp} (undershoot)_v
 → (altitude crossing)_{sbj} (be restricted not met)_v 頻度 8
 (altitude)_{sbj} (in descent procedure)_{pp} (undershoot)_v
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 2
 (altitude)_{sbj} (in descent procedure)_{pp} (undershoot)_v
 → (wrong altimeter setting)_{np} → (non adherence)_{np} (to atc clearance)_{pp} 頻度 2
 (altitude crossing)_{sbj} (be restricted)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (altitude crossing)_{sbj} (be restricted not met)_v → (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (undershoot)_v 頻度 2
 (altitude crossing)_{sbj} (be restricted not met)_v → (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (in descent)_{pp} (undershoot)_v 頻度 3
 (altitude crossing)_{sbj} (be restricted not met)_v
 → (altitude)_{sbj} (undershoot)_v 頻度 2
 (altitude crossing)_{sbj} (be restricted not met)_v
 → (altitude)_{sbj} (in descent)_{pp} (undershoot)_v 頻度 3
 (altitude crossing)_{sbj} (be restricted not met)_v
 → (captain)_{sbj} (altimeters)_{obj} (fail to crosscheck)_v 頻度 2
 (altitude excursion)_{np} → (non adherence)_{np} (to atc clearance)_{pp} 頻度 3
 (altitude excursion)_{np} → (non adherence)_{np} (to atc clearance procedure)_{pp} 頻度 2
 (altitude excursion)_{np} → (potential conflict less than standard separation)_{np} 頻度 3

(annual inspection)_{sbj} (be been performed)_v
 → (not)_{sbj} (in aircraft logbook)_{pp} (enter)_v 頻度 2
 (approach controller)_{sbj} (intervene)_v → (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (overshoot)_v 頻度 2
 (approach controller)_{sbj} (intervene)_v → (altitude)_{sbj} (overshoot)_v 頻度 2
 (artcc artcc)_{np} → (system error)_{np} 頻度 2
 (artcc intrafac coordination)_{np} → (system error)_{np} 頻度 2
 (artcc radar controller)_{sbj} (intervene)_v → (advise)_v 頻度 2
 (artcc radar controller)_{sbj} (intervene)_v → (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (overshoot)_v 頻度 2
 (artcc radar controller)_{sbj} (intervene)_v → (altitude)_{sbj} (overshoot)_v 頻度 2
 (atc)_{sbj} (intervene)_v → (advise)_v 頻度 2
 (atc)_{sbj} (intervene)_v → (be assigned)_v 頻度 2
 (atc equipment problem icss position failure)_{np}
 → (communication)_{obj} (lose)_v 頻度 2
 (atc equipment problem radar communication power failure)_{np}
 → (radar and frequency failure)_{np}
 → (standby power)_{sbj} (on line)_{pp} (be come)_v 頻度 2
 (atc equipment problem radar communication power failure)_{np}
 → (standby power)_{sbj} (on line)_{pp} (be come)_v 頻度 2
 (atc flight crew miscom)_{np} (on altitude clearance)_{pp} → (reporter)_{sbj} (fly)_v 頻度 2
 (atc reporter)_{sbj} (allege)_v → (reporter)_{sbj} (be concerned)_v 頻度 3
 (atc reporter)_{sbj} (that tracon)_{obj} (claim)_v → (reporter)_{sbj} (accuracy and reliability)_{obj}
 (of system)_{pp} (question)_v 頻度 4
 (atct tracon)_{np} → (system error)_{np} 頻度 2
 (atx)_{np} → (pilot)_{sbj} (deviate)_v 頻度 4
 (atx)_{np} → (system error)_{np} 頻度 2
 (atx)_{sbj} (less than standard separation)_{obj} (from sma)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (autopilot)_{sbj} (altitude)_{obj} (fail to capture)_v → (altitude)_{sbj} (overshoot)_v 頻度 2
 (cabin fire)_{np} → (mlg)_{sbj} (divert to land)_v 頻度 2
 (captain)_{np} → (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (overshoot)_v 頻度 2
 (captain)_{np} → (altitude)_{sbj} (overshoot)_v 頻度 2
 (captain)_{np} → (flight crew behavior)_{sbj} (fatigue)_v 頻度 2
 (captain)_{sbj} (station)_{obj} (while leg discomfort)_{pp} (be assigned)_v
 → (captain)_{sbj} (in so)_{pp} (be sat)_v 頻度 2
 (captain)_{sbj} (report)_v → (however captain)_{sbj} (claim)_v 頻度 2
 (captain)_{sbj} (report)_v → (jumper)_{sbj} (in contact)_{pp} (be)_v
 → (however captain)_{sbj} (claim)_v 頻度 2
 (captain)_{sbj} (of mlg acr aircraft)_{pp} (respond)_v
 → (above assigned altitude)_{pp} (climb)_v 頻度 4
 (captain ss)_{sbj} (window)_{obj} (slide)_v → (acr lgt)_{sbj} (takeoff)_{obj} (abort)_v 頻度 2
 (center computer)_{sbj} (with art program)_{pp} (be not interact correctly)_v

→ (reporter)_{sbj} (this)_{obj} (occur)_v 頻度 2
 (clearance)_{sbj} (intercept)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (clearance)_{sbj} (problem)_{obj} (intercept)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (clearance landing)_{sbj} (receive)_v
 → (without authorization)_{pp} (land)_v → (frequency change)_{sbj} (miss)_v 頻度 2
 (clearance landing)_{sbj} (receive)_v → (frequency change)_{sbj} (miss)_v 頻度 2
 (commercial power failure)_{np}
 → (emergency generator)_{sbj} (not work)_{obj} (do)_v 頻度 3
 (commercial power failure)_{np} → (radar ots)_{np}
 → (emergency generator)_{sbj} (not work)_{obj} (do)_v 頻度 3
 (commuter aircraft)_{sbj} (tcasii ra)_{obj} (have)_v → (respond)_v 頻度 2
 (compressor stall)_{np} → (takeoff)_{sbj} (abort)_v 頻度 2
 (conflict)_{np} → (final 2)_{pp} 頻度 2
 (controller)_{sbj} (acr)_{obj} (for clearance runway)_{pp} (be reported)_v
 → (hold line)_{sbj} (runways)_{obj} (for 2)_{pp} (be common)_v 頻度 2
 (controller)_{sbj} (intervene)_v → (tell)_v 頻度 2
 (controller)_{sbj} (state)_v → (runway)_{obj} (enter)_v 頻度 2
 (controller)_{sbj} (state)_v → (aircraft)_{sbj} (be stopped)_v → (runway)_{obj} (enter)_v 頻度 2
 (controller training)_{np} (in progress)_{pp} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (controller training)_{np} (in progress)_{pp} → (system error)_{np} 頻度 3
 (cpr)_{np} → (pilot)_{sbj} (deviate)_v → (system error)_{np} 頻度 2
 (cpr)_{np} → (runway incursion)_{np} 頻度 2
 (cpr)_{np} → (system error)_{np} 頻度 7
 (cpr)_{sbj} (less than standard separation)_{obj} (through occupied altitude)_{pp} (climb)_v
 → (controller training)_{np} (in progress)_{pp} 頻度 2
 (cpr)_{sbj} (less than standard separation)_{obj} (through occupied altitude)_{pp} (climb)_v
 → (system error)_{np} 頻度 16
 (cpr)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (cpr)_{sbj} (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (be taken)_v 頻度 2
 (cpr)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 4
 (cpr descent)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 5
 (cpr descent)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (cpr descent)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (cpr y climb)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 3
 (cpr y climb)_{sbj} (less than standard separation)_{obj} (from cpr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (cpr y descent)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v

→ (system error)_{np} 頻度 2
 (emergency)_{sbj} (be declared)_v → (aircraft)_{sbj} (be evaced)_v 頻度 2
 (emergency)_{sbj} (be declared)_v → (flight)_{sbj} (divert)_v 頻度 2
 (emergency)_{sbj} (declare)_v → (land)_{obj} (divert)_v 頻度 2
 (emergency)_{sbj} (declare)_v → (land)_{obj} (return)_v 頻度 11
 (emergency)_{sbj} (declare)_v → (aircraft)_{sbj} (divert)_v 頻度 2
 (emergency)_{sbj} (declare)_v → (diversion alternate)_{np} 頻度 5
 (emergency)_{sbj} (declare)_v → (emergency crash)_{sbj} (be activated)_v 頻度 2
 (emergency)_{sbj} (declare)_v → (inflight engine)_{sbj} (shut down)_v 頻度 3
 (engine)_{sbj} (fail)_v → (off airport)_{obj} (land)_v 頻度 3
 (engine)_{sbj} (fail)_v → (to departure airport)_{pp} (return)_v 頻度 2
 (engine)_{sbj} (fail)_v → (land)_{obj} (return)_v 頻度 4
 (engine)_{sbj} (fail)_v → (aircraft)_{sbj} (to departure airport)_{pp} (return)_v 頻度 2
 (engine)_{sbj} (fail)_v → (emergency)_{sbj} (declare)_v → (land)_{obj} (return)_v 頻度 2
 (engine)_{sbj} (warning)_{obj} (fire)_v → (engine)_{obj} (shut down)_v 頻度 2
 (engine)_{sbj} (light)_{obj} (over)_v → (emergency)_{sbj} (declare)_v 頻度 2
 (engine)_{sbj} (quit)_v → (landing)_{obj} (force)_v 頻度 2
 (engine)_{sbj} (shut down)_v → (emergency)_{sbj} (declare)_v 頻度 2
 (erroneous penetration)_{np} (of airspace)_{pp} → (proximity)_{np} (of tca)_{pp} 頻度 4
 (error)_{sbj} (admit)_v → (mountainous terrain)_{obj} (designate)_v 頻度 2
 (error)_{sbj} (admit)_v → (from altitude)_{pp} (deviate)_v 頻度 6
 (error)_{sbj} (admit)_v → (from track)_{pp} (deviate)_v 頻度 2
 (error)_{sbj} (admit)_v → (off airport)_{obj} (land)_v 頻度 3
 (error)_{sbj} (admit)_v → (in progress)_{pp} (train)_v 頻度 2
 (error)_{sbj} (admit)_v → (aircraft)_{sbj} (damage)_v 頻度 3
 (error)_{sbj} (admit)_v → (night operation)_{np} 頻度 5
 (error)_{sbj} (admit)_v → (track)_{sbj} (from heading)_{pp} (deviate)_v 頻度 2
 (error)_{sbj} (be)_v → (flights)_{sbj} (course)_{obj} (turn)_v 頻度 2
 (evasive action)_{sbj} (be taken)_v → (see avoid)_v → (system error)_{np} 頻度 2
 (evasive action)_{sbj} (be taken)_v → (concept)_{obj} (see avoid)_v 頻度 50
 (evasive action)_{sbj} (be taken)_v → (non adherence)_{np} (to atc instruction)_{pp} 頻度 2
 (evasive action)_{sbj} (be taken)_v → (pilot)_{sbj} (deviate)_v → (system error)_{np} 頻度 2
 (evasive action)_{sbj} (be taken)_v → (system error)_{np} 頻度 45
 (evasive action)_{sbj} (climb)_{obj} (be taken)_v → (system error)_{np} 頻度 2
 (evasive action)_{sbj} (turn)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (evasive action)_{sbj} (turn)_v → (system error)_{np} 頻度 2
 (evasive action climb)_{np} → (from altitude)_{pp} (deviate)_v 頻度 2
 (evasive action climb)_{np} → (concept)_{obj} (see avoid)_v 頻度 5
 (evasive action climb)_{np} → (system error)_{np} 頻度 2
 (exhaustion)_{np} → (engine)_{sbj} (quit)_v 頻度 3
 (experimental aircraft engine)_{sbj} (quit)_v
 → (off airport emergency landing)_{sbj} (make)_v 頻度 2
 (flight attendant)_{sbj} (regard)_v → (to departure airport)_{pp} (return)_v 頻度 4

(flight attendant)_{sbj} (regard)_v → (aircraft)_{sbj} (to departure airport)_{pp} (return)_v 頻度 2
 (flight attendant)_{sbj} (regard)_v
 → (cockpit)_{sbj} (be notified)_v 頻度 2
 (flight attendant report)_{sbj} (regard)_v
 → (aircraft)_{sbj} (to departure airport)_{pp} (return)_v 頻度 2
 (flight attendant report)_{sbj} (pax)_{obj} (regard)_v → (aircraft)_{sbj} (to gate)_{pp} (return)_v
 → (pax)_{sbj} (be removed)_v 頻度 2
 (flight attendant report)_{sbj} (pax)_{obj} (regard)_v → (pax)_{sbj} (be removed)_v 頻度 2
 (flight crew)_{np} → (emergency)_{sbj} (declare)_v 頻度 2
 (flight crew)_{sbj} (deviate)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (flight crew)_{sbj} (deviate)_v → (non adherence)_{np} (to atc clearance)_{pp} 頻度 2
 (flight crew)_{sbj} (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (in descent procedure)_{pp} (overshoot)_v 頻度 2
 (flight crew)_{sbj} (from altitude)_{pp} (deviate)_v
 → (non adherence)_{np} (to atc clearance)_{pp} 頻度 2
 (flight crew)_{sbj} (tower)_{obj} (fail to contact)_v → (without clearance)_{pp} (land)_v 頻度 2
 (flight crew)_{sbj} (at assigned altitude)_{pp} (fail to level)_v
 → (from altitude)_{pp} (deviate)_v 頻度 2
 (flight crew)_{sbj} (without landing clearance)_{pp} (lands)_v
 → (without authorization)_{pp} (land)_v 頻度 2
 (flight crew)_{sbj} (receive)_v → (ra)_{np} 頻度 2
 (flight crew)_{sbj} (to tcas ra)_{pp} (respond)_v → (from altitude)_{pp} (deviate)_v 頻度 2
 (flight crew)_{sbj} (evasive action)_{obj} (in response)_{pp} (take)_v
 → (excursion)_{np} (from assigned altitude)_{pp} 頻度 2
 (flight crew)_{sbj} (onto runway)_{pp} (taxi)_v → (aircraft)_{sbj} (have to go around)_v 頻度 2
 (flight crew)_{sbj} (of acrl mlg aircraft)_{pp} (respond)_v
 → (from assigned altitude)_{pp} (deviate)_v 頻度 2
 (flight crew)_{np} (of mlg)_{pp} → (advise)_v 頻度 2
 (flight crew)_{sbj} (of mlg)_{pp} (return)_v → (land)_v 頻度 2
 (flight crew)_{sbj} (of mlg)_{pp} (return)_v → (takeoff)_{np} 頻度 2
 (flight crew)_{sbj} (of mlg)_{pp} (stop climbing)_v → (in response)_{pp} (be dsnded)_v 頻度 2
 (fo)_{np} → (from altitude)_{pp} (deviate)_v → (altitude)_{sbj} (overshoot)_v 頻度 2
 (fo)_{np} → (altitude)_{sbj} (overshoot)_v 頻度 2
 (fo)_{np} → (captain)_{np} → (from altitude)_{pp} (deviate)_v
 → (altitude)_{sbj} (overshoot)_v 頻度 2
 (fo)_{np} → (captain)_{np} → (altitude)_{sbj} (overshoot)_v 頻度 2
 (fo)_{sbj} (captain)_{obj} (ask)_v → (captain)_{np} 頻度 2
 (foreign acrl)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (foreign acrl)_{np} → (system error)_{np} 頻度 2
 (fuel)_{sbj} (dump)_v → (land)_{obj} (return)_v 頻度 2
 (fuel exhaustion)_{np} → (off airport)_{obj} (land)_v 頻度 3
 (fuel exhaustion)_{np} → (forced landing)_{np} 頻度 2

(hard landing)_{np} → (gear)_{sbj} (collapse)_v 頻度 2
 (hard landing)_{np} → (nose gear)_{sbj} (collapse)_v 頻度 2
 (heading)_{sbj} (from track)_{pp} (deviate)_v
 → (flight crew technique navigation)_{np} 頻度 2
 (heading)_{sbj} (from track)_{pp} (deviate)_v
 → (non adherence)_{np} (to atc clearance instruction)_{pp} 頻度 2
 (heading track position deviation)_{np} → (navigation error)_{np} 頻度 3
 (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (impostor flight attendant)_{sbj} (in acr uniform)_{pp} (be boarded)_v
 → (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (inflight engine)_{sbj} (shut down)_v
 → (land)_{obj} (return)_v 頻度 2
 (inflight engine)_{sbj} (shut down)_v → (and emergency)_{sbj} (declare)_v 頻度 2
 (inflight engine)_{sbj} (shut down)_v → (diversion alternate)_{np} 頻度 2
 (inflight engine)_{sbj} (shut down)_v → (emergency)_{sbj} (declare)_v 頻度 5
 (inflight engine)_{sbj} (shut down)_v → (flameout)_{np} 頻度 2

(instructor pilo)_{sbj} (report)_v → (error)_{sbj} (admit)_v 頻度 2
 (jumper)_{sbj} (in contact)_{pp} (be)_v → (however captain)_{sbj} (claim)_v 頻度 2
 (less than standard separation)_{np} → (pilot)_{sbj} (deviate)_v 頻度 3
 (less than standard separation)_{np} → (system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (less than standard separation)_{np} (with acr)_{pp} → (pilot)_{sbj} (deviate)_v 頻度 2
 (lgt)_{sbj} (tcasii ra)_{obj} (have)_v → (respond)_v 頻度 7
 (lft)_{np} → (pilot)_{sbj} (deviate)_v 頻度 3
 (main cabin door)_{sbj} (on beech 1900)_{pp} (open)_v
 → (emergency)_{sbj} (for aircraft)_{pp} (be declared)_v 頻度 2
 (maintenance)_{sbj} (remove)_v → (in logbook)_{pp} (fail to note same)_v 頻度 2
 (military)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (military)_{sbj} (near mid air collision)_{pp} (have)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (military)_{sbj} (less than standard separation)_{obj} (from military)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (military descent)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (mlg)_{sbj} (tcasii)_{obj} (have)_v → (ra)_{np} 頻度 4
 (mlg)_{sbj} (tcasii ra)_{obj} (have)_v → (climbs)_v 頻度 3
 (mlg)_{sbj} (tcasii ra)_{obj} (have)_v → (respond)_v 頻度 8
 (mlg)_{sbj} (tcasii ra)_{obj} (have)_v → (controller)_{sbj} (upset)_v 頻度 2
 (multiple runway operation)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 4
 (multiple runway operation)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (multiple runway operation)_{np} → (system error)_{np}
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (multiple runway operation)_{sbj} (runways)_{obj} (intersect)_v
 → (less than standard separation)_{obj} (from acr)_{pp} (have)_v 頻度 2
 (multiple runway operation)_{sbj} (runways)_{obj} (intersect)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (multiple runway operation)_{sbj} (runways)_{obj} (intersect)_v
 → (pilot)_{sbj} (deviate)_v 頻度 3
 (multiple runway operation)_{sbj} (runways)_{obj} (intersect)_v → (system error)_{np} 頻度 13
 (multiple runway operation)_{sbj} (runways acr)_{obj} (from acr y)_{pp} (intersect)_v
 → (system error)_{np} 頻度 2
 (multiple runway operation)_{sbj} (runways sma)_{obj} (from smt)_{pp} (intersect)_v
 → (system error)_{np} 頻度 2
 (multiple runway operation)_{sbj} (runways)_{obj} (parallel)_v
 → (evasive action)_{sbj} (be taken)_v 頻度 2
 (multiple runway operation)_{sbj} (runways)_{obj} (parallel)_v → (system error)_{np} 頻度 3
 (navigation error)_{np} → (heading track position deviation)_{np} 頻度 2
 (new airspace)_{np} (in effect)_{pp} → (some confusion)_{np} 頻度 3
 (new arrival)_{np} → (flight crew)_{sbj} (confuse)_v 頻度 2

(night operation)_{np} → (mountainous terrain)_{obj} (designate)_v 頻度 2
 (night operation)_{np} → (error)_{sbj} (admit)_v 頻度 3
 (no injury)_{np} → (only minor damage)_{np} (to aircraft)_{pp} 頻度 2
 (non adherence)_{np} (to atc clearance)_{pp} → (from altitude)_{pp} (deviate)_v 頻度 2
 (non adherence)_{np} (to atc clearance)_{pp} → (descent)_{sbj} (altitude)_{obj} (occupy)_v 頻度 2
 (non adherence)_{np} (to atc clearance)_{pp} → (pilot)_{sbj} (deviate)_v 頻度 8
 (non adherence)_{np} (to atc clearance instruction)_{pp}
 → (from altitude)_{pp} (deviate)_v 頻度 2
 (non compliance)_{np} (with faa handbook)_{pp}
 → (operational acceptable level)_{sbj} (of traffic)_{pp} (be exceeded)_v 頻度 2
 (not)_{np} → (separation)_{sbj} (be lost)_v 頻度 3
 (not)_{sbj} (be given)_v → (traffic)_{sbj} (sight)_v 頻度 2
 (offering)_{sbj} (with cabin service)_{pp} (to help)_v → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (offering)_{sbj} (with cabin service)_{pp} (to help)_v
 → (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (none)_{obj} (have)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (other flight attendant)_{sbj} (of impostor)_{pp} (become suspicious)_v
 → (impostor)_{sbj} (same thing)_{obj} (with other acrs)_{pp} (try to do)_v 頻度 2
 (pf)_{np} → (altitude crossing)_{sbj} (be restricted not met)_v
 → (captain)_{sbj} (altimeters)_{obj} (fail to crosscheck)_v 頻度 2
 (pf)_{np} → (captain)_{sbj} (altimeters)_{obj} (fail to crosscheck)_v 頻度 2
 (pilo land)_{np} → (gear)_{np} 頻度 2
 (pilot)_{sbj} (deviate)_v → (separation)_{obj} (lose)_v 頻度 2
 (pilot)_{sbj} (deviate)_v → (concept)_{obj} (see avoid)_v 頻度 7
 (pilot)_{sbj} (deviate)_v → (evasive action)_{sbj} (be taken)_v 頻度 7
 (pilot)_{sbj} (deviate)_v → (misunderstood clearance)_{np} 頻度 2
 (pilot)_{sbj} (deviate)_v → (situation)_{np} 頻度 3
 (pilot)_{sbj} (deviate)_v → (system error)_{np} 頻度 34
 (pilot)_{sbj} (deviate)_v
 → (tcasii)_{sbj} (less than standard separation)_{obj} (induce)_v 頻度 2
 (possible faa action)_{sbj} (on owner)_{pp} (be taken)_v → (into part 135)_{pp} (enter)_v 頻度 2
 (possible less than standard separation)_{np} → (possible system error)_{np} 頻度 2
 (possible system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 2

(possible system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (potential conflict)_{np} → (clearance)_{sbj} (intercept)_v 頻度 2
 (potential conflict)_{np} → (system error)_{np} 頻度 2
 (potential conflict)_{np} → (traffic)_{sbj} (sight)_v 頻度 2
 (potential conflict)_{sbj} (tcasii ra)_{obj} (resolve)_v
 → (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} 頻度 2
 (potential conflict)_{sbj} (tcasii ra)_{obj} (resolve)_v
 → (altitude excursion)_{np} 頻度 2
 (proximity)_{np} (of traffic)_{pp}
 → (potential conflict less than standard separation)_{np} 頻度 2
 (ra)_{np} → (climbs)_v 頻度 3
 (ra)_{np} → (respond)_v 頻度 4
 (radar and frequency failure)_{np} → (standby power)_{sbj} (on line)_{pp} (be come)_v 頻度 2
 (radar ots)_{np} → (emergency generator)_{sbj} (not work)_{obj} (do)_v 頻度 3
 (reporter)_{sbj} (acr)_{obj} (from acr y)_{pp} (allege)_v → (possible system error)_{np} 頻度 2
 (reporter)_{sbj} (contend)_v → (reporter)_{sbj} (state)_v 頻度 2
 (reporter)_{sbj} (fly)_v → (tell)_v 頻度 2
 (runway excursion)_{np} → (aircraft)_{sbj} (damage)_v 頻度 6
 (runway excursion)_{np} → (landing)_{sbj} (overshoot)_v 頻度 3
 (runway excursion)_{np} → (takeoff)_{sbj} (abort)_v 頻度 2
 (runway incursion)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (runway incursion)_{np} → (system error)_{np} 頻度 6
 (runway incursion)_{np} → (unauthorized runway entry)_{np} 頻度 6
 (same altitude)_{sbj} (be assigned)_v → (system error)_{np} 頻度 2
 (seneca atx pilo taxies off taxiway)_{np} (at hxd sc)_{pp}
 → (aircraft)_{sbj} (night)_{obj} (in ditch)_{pp} (end up spend)_v 頻度 2
 (sfo airport)_{np} (in compliance)_{pp}
 → (low visibility)_{sbj} (routes)_{obj} (during low visibility operations)_{pp} (taxi)_v 頻度 2
 (sma)_{np} → (concept)_{obj} (see avoid)_v 頻度 5
 (sma)_{np} → (evasive action)_{sbj} (be taken)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (sma)_{np} → (evasive action)_{sbj} (be taken)_v → (pilot)_{sbj} (deviate)_v 頻度 2
 (sma)_{np} → (language barrier)_{np} 頻度 2
 (sma)_{np} → (pilot)_{sbj} (deviate)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (sma)_{np} → (system error)_{np} 頻度 2
 (sma)_{sbj} (icing)_{obj} (encounter)_v → (emergency)_{obj} (declare)_v 頻度 2
 (sma)_{sbj} (near mid air collision)_{pp} (have)_v → (see avoid)_v 頻度 2
 (sma)_{sbj} (near mid air collision)_{pp} (have)_v → (concept)_{obj} (see avoid)_v 頻度 17
 (sma)_{sbj} (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (be taken)_v
 → (concept)_{obj} (see avoid)_v 頻度 5
 (sma)_{sbj} (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (turn)_v
 → (concept)_{obj} (see avoid)_v 頻度 2
 (sma)_{sbj} (than standard separation)_{pp} (have)_v → (system error)_{np} 頻度 2
 (sma)_{sbj} (airspace)_{obj} (near mid air collision)_{pp} (have)_v

→ (concept)_{obj} (see avoid)_v 頻度 2
 (sma)_{sbj} (less than standard separation)_{obj} (from aircraft)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (sma)_{sbj} (land)_v → (gear)_{np} 頻度 2
 (sma)_{sbj} (on occupied runway)_{pp} (land)_v → (pilot)_{sbj} (deviate)_v 頻度 2
 (sma)_{sbj} (on occupied runway)_{pp} (land)_v → (system error)_{np} 頻度 2
 (sma pilo)_{sbj} (out)_{obj} (of fuel)_{pp} (run)_v → (airport)_{obj} (land off)_v 頻度 2
 (sma y non adherence)_{sbj} (with sma)_{pp} (have)_v → (concept)_{obj} (see avoid)_v 頻度 2
 (smoke)_{np} (in cabin)_{pp} → (emergency)_{sbj} (declare)_v 頻度 2
 (smt)_{np} → (concept)_{obj} (see avoid)_v 頻度 2
 (smt)_{np} → (pilot)_{sbj} (deviate)_v 頻度 6
 (smt)_{sbj} (near mid air collision)_{pp} (have)_v
 → (concept)_{obj} (see avoid)_v → (evasive action)_{sbj} (be taken)_v 頻度 2
 (smt)_{sbj} (near mid air collision)_{pp} (have)_v → (evasive action)_{sbj} (be taken)_v 頻度 2
 (smt)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (smt descent)_{sbj} (less than standard separation)_{obj} (from acr)_{pp} (have)_v
 → (system error)_{np} 頻度 2
 (starvation)_{np} → (off airport)_{obj} (make)_v → (engine)_{sbj} (quit)_v 頻度 3
 (starvation)_{np} → (engine)_{sbj} (quit)_v 頻度 13
 (stuck mike)_{np} → (communication frequency)_{obj} (lose)_v 頻度 2
 (system error)_{np} → (acr)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (system error)_{np} → (acr)_{sbj} (tcasii ra evasive action)_{obj} (be taken)_v 頻度 2
 (system error)_{np} → (controller training)_{np} (in progress)_{pp} 頻度 2
 (system error)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 43
 (system error)_{np} → (evasive action descent)_{np} 頻度 2
 (system error)_{np} → (pilot)_{sbj} (deviate)_v 頻度 30
 (system error)_{np} → (potential conflict)_{np} 頻度 6
 (system error)_{np} → (potential conflict less than standard separation)_{np} 頻度 2
 (system error)_{np} → (runway incursion)_{np} 頻度 3
 (system error)_{np} → (tracon)_{np} 頻度 2
 (takeoff)_{sbj} (abort)_v → (no)_{sbj} (flaps)_v 頻度 2
 (takeoff)_{sbj} (abort)_v → (takeoff)_{sbj} (delay)_v 頻度 2
 (takeoff warning horn)_{sbj} (sound)_v → (takeoff)_{sbj} (abort)_v 頻度 2
 (tcasii)_{np} → (mlg)_{sbj} (descend)_v 頻度 2
 (tcasii)_{np} → (ra)_{np} 頻度 2
 (tcasii)_{np} → (reporter)_{sbj} (say)_v 頻度 2
 (tcasii)_{np} → (system error)_{np} 頻度 2
 (tcasii)_{sbj} (less than standard separation)_{obj} (induce)_v
 → (pilot)_{sbj} (deviate)_v 頻度 2
 (tcasii logic)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 2
 (tcasii ra)_{np} → (from altitude)_{pp} (deviate)_v → (altitude excursion)_{np} 頻度 6
 (tcasii ra)_{np} → (altitude excursion)_{np} 頻度 6

(tcasii ra)_{np} → (evasive action)_{np} 頻度 2
 (tcasii ra)_{np} → (evasive action)_{sbj} (be taken)_v 頻度 3
 (tcasii ra)_{np} → (flight crew)_{sbj} (respond)_v 頻度 2
 (tcasii ra)_{np} → (pilot)_{sbj} (deviate)_v 頻度 2
 (tcasii ra)_{sbj} (receive)_v → (from altitude)_{pp} (deviate)_v
 → (altitude excursion)_{np} 頻度 2
 (tcasii ra)_{sbj} (receive)_v → (altitude excursion)_{np} 頻度 2
 (tcasii ra climb)_{np} → (system error)_{np} 頻度 2
 (this)_{np} → (aircraft)_{sbj} (be)_v 頻度 4
 (this)_{np} → (opdev)_{np} 頻度 2
 (tracon)_{np} → (system error)_{np} 頻度 2
 (unable)_{sbj} (altitude)_{obj} (maintain)_v → (emergency)_{obj} (declare)_v 頻度 2
 (unauthorized penetration)_{np} (of airspace)_{pp}
 → (without authorization)_{pp} (land)_v 頻度 3
 (unauthorized penetration)_{np} (of airspace)_{pp} → (proximity)_{np} 頻度 2
 (unauthorized penetration)_{np} (of airspace)_{pp} → (proximity)_{np} (of arsa)_{pp} 頻度 9
 (unauthorized penetration)_{np} (of airspace)_{pp} → (proximity)_{np} (of ata)_{pp} 頻度 6
 (unauthorized penetration)_{np} (of airspace)_{pp}
 → (proximity)_{sbj} (of control)_{pp} (zone)_v 頻度 4
 (unauthorized penetration)_{np} (of airspace)_{pp} → (proximity)_{np} (of moa)_{pp} 頻度 2
 (unauthorized penetration)_{np} (of airspace)_{pp}
 → (proximity)_{np} (of restricted area)_{pp} 頻度 4
 (unauthorized penetration)_{np} (of airspace)_{pp} → (proximity)_{np} (of tca)_{pp} 頻度 11
 (vehicle)_{np} (on runway)_{pp} → (system error)_{np} 頻度 2
 (wdb)_{sbj} (tcasii ra)_{obj} (receive)_v → (respond)_v 頻度 2
 (weather)_{np} → (acr)_{sbj} (clearance)_{obj} (refuse)_v 頻度 2
 (wrong airport)_{sbj} (landing)_{obj} (approach)_v
 → (without authorization)_{pp} (land)_v 頻度 2
 (wrong altimeter setting)_{np} → (from altitude)_{pp} (deviate)_v 頻度 3
 (wrong altimeter setting)_{np} → (non adherence)_{np} (to atc clearance)_{pp} 頻度 2
 (wrong altimeter setting)_{np} → (tcasii ra)_{np} 頻度 2
 (wrong flap setting)_{np} → (takeoff)_{sbj} (abort)_v 頻度 2
 (wrong runway)_{sbj} (landing)_{obj} (approach)_v
 → (without authorization)_{pp} (land)_v 頻度 3