公益財団法人 日本対がん協会

がんホットライン・相談の分析

厚生労働省委託事業
「がんと診断された時からの相談支援事業」

～地域統括相談支援センター活性化策検討のための基礎調査～

日本対がん協会「がん相談ホットライン」データベース分析
公益財団法人日本対がん協会　Mynd, Inc.

2015 年 2 月

概 要

厚生労働省から 2014 年 6 月、「がんと診断された時からの相談支援事業」の委託を受けた公益財団法人日本対がん協会は、その事業の主目的とされた「地域統括相談支援センター」の活性化を図る方策を検討する手がかりをえるため、対がん協会内の「がん相談ホットライン」の過去三年間のデータから相談者のニーズを整理・分析した。

厚労省が「地域統括相談支援センター」事業を開始した 2011 年度当初の計画では、がん診療連携拠点病院の「がん相談支援センター」と連携しつつ、介護・就労・教育などがん患者・家族の方々が抱える様々な生活上の課題にワンストップで対応する組織――とされている。

これを受け、地域統括相談支援センターの役目を考えるに際してがん患者・家族の方々の「医療面以外の相談」にどのようなことがあるのかを定量的に把握することが必要と考えて、病院以外のところに置いた相談窓口の一つである「がん相談ホットライン」を一つの対象とした。

方法は主に①相談毎に付与された「相談区分」（複数）を統計処理②相談者と相談員の「主訴」と「対応」のテキストを対象に自然言語処理による解析――の 2 点。

その結果、医療面の相談が多い一方で、「医療面の相談」の中に日常生活を送る上での障害となる後遺症など日常生活と医療が複雑に関係する領域や心の不安の訴えが含まれ、これらにも複合的に対応する必要があることが改めて浮かび上がった。就労・経済的負担の問題、（対医療者を含む）人間関係の悩みの相談も一定数あった。

限られた予算の中で多岐にわたるニーズに対応する必要に迫られた「地域統括相談支援センター」を効率的に機能させるには、福祉、労働、法律、税、等を含めた地域ごとの各種の相談窓口との連携を密に図り、相談者の声をまず丁寧に「受け止め（寄り添い）」、自ら対応するだけでなく、相談内容を整理し、適切な相談孫口に「つなぐ」働きを持たせることが大切であると考えられた。

必要な時に必要な情報を必要な人に適切な形で届ける――こうした相談支援を柔軟に機能させるには、それなりの規模をもった中核的な組織が必要性なことも示唆されたといえよう。

内容

1　目的

この分析の目的は、地域統括相談支援センター活性化の方策を立案する手がかりとして、「がん相談ホットライン」の相談者のニーズを把握することである。

ここで言う「ニーズの把握」とは、相談者の相談したい内容が、医学的な質問なのか、経済的な問題なのか、公的支援を求めているのか、心理的な不安を訴えているのか、などの粗いレベルの量的／質的な理解であり、そのことによって、相談支援センターの活性化や、未来の施策の設計に役立てようとするものである。

分析については補遺 A「分析の手続き」のステップに従った。

2　相談データ

まず、日本対がん協会の電話相談「がん相談ホットライン」のデータベースに納められたデータに対し、協会スタッフおよび相談員も交えてデータ訊問(“Cross examination of Data”)を行ない、全データ項目の一次的な分析を行なった（補遺 A「分析の手続き」を参照)。

具体的には、データの取得方法、各データ項目の定義と意味、データの外れ値や誤差の検討、顕著な特徴の把握などである。これによって、目的に対して分析するべき対象データを策定した。

その結果、2011 年 7 月 1 日から 2014 年 11 月 1 日までの 32,629 件を分析対象とすることにした。それ以前のデータも存在するが、データベースのフォーマットや項目立ての一貫性のためと、直近の約三年間が現状の分析にふさわしいだろうと考えたからである。

各相談データは以下のような 41 個のフィールドを持つ:

相談番号, 相談担当, 最終相談担当, 相談日時, 相談時間, 相談者名, 相談者年 齢, 相談者年代, 相談者性別, 相談者との続柄, 患者名, 患者年齢, 患者年代, 患者 性別, 患者住所（都道府県）, 患者住所, 患者電話番号, 患者疾患部位, 患者疾患 部位（その他）, 転移, 状況, 症状, 治療場所, 相談区分１, 相談区分２, 相談区分３, 主訴, 対応, 備考, がん相談を知った場所, 資料, 備考２, メモ, 登録日, 編集 日, 再発, 相談者続柄（その他）, がん相談を知った場所（その他）, がん相談を 知った場所（リーフレットの入手場所）, 治療場所（その他）, リーフレットの入手場所(その他)

以上のフィールドについて、事前のデータ訊問および相談担当者へのインタヴューの結果、このうち「相談者のニーズを理解する」という目的のため分析できる主な項目として、「相談区分１」「相談区分２」「相談区分３」「主訴」「対応」の五つを分析対象とする。

（なお、相談者の性別、続柄、居住地域、相談時間、相談のがん部位など、背景的な基本統計は参考文献［1］などの年報に報告されており、開設以来、大まかな傾向は変化していない）

前の三つ「相談区分」の１から３は、その相談がどのような種類のものなのかを相談担当者が判断して 21 種類の分類ラベルの一つを記入したものである。

また後の二つ「主訴」と「対応」は相談者の相談内容とその対応を、電話相談を終えてからメモ、記憶、印象を基に相談担当者がテキスト起こししたもので、相談の実質的な内容が残された唯一のフィールドである。

これらのデータは相談担当者の主観によって記録されたものであるから、その信頼性については注意が必要である。詳しくは補遺 B に譲るが大まかに言って、事前のデータ訊問とインタヴューの結果から、相談担当者の訓練、分類のマニュアル化、知識共有などが徹底されており、この分析目的に対しては十分であると考えられる。

3　分析の方針

3.1　「相談区分」タグによる分析

各相談データに付加されている「相談区分」の１から３は、その相談の相談担当者が対応したあと、以下の 21 種類の分類のラベルをつけたものである。

「外来」,「入院・退院」,「転院」,「診断」,「治療」,「検診」,「検査」,「緩和ケ ア」,「告知・ＩＣ [1]」,「ＳＯ [2]」,「医療連携」,「在宅療養」,「施設設備・アクセ ス」,「医療者との関係」,「症状・副作用・後遺症」,「不安など心の問題」,「生き方・生きがい・価値観」,「就労・経済的な負担」,「家族・周囲の人との関係」,「補完代替療法」,「その他」

[1]IC: Informed Consent
[2]SO: Second Opinion

この相談区分の割り当ては、事前のデータ訊問と相談員へのインタビューからして、相談担当者の訓練、マニュアル化や知識共有が徹底されて おり、分析目的には十分な精度を持つと考えられる（補遺 B 参照)。

したがって、この分類を適切に集計することで、「相談のニーズ」の概要を知ることができると考えられる。

しかし、「相談区分」の１から３の記入のされ方に注意しなければならない。その相談の主題が一つであると判断された場合には「相談区分１」のみが記入され、もし二次的な区分にも属する要素があると判断された場合には「相談区分２」にも記入し、さらに他の要素も含まれると判断した場合には「相談区分３」にも記入する。

いろいろな立場からの集計方法が考えられるが、分析の目的 を鑑みて、以下の二つの視点から集計を行う。

- 「相談区分１」のみを持つ相談の相談区分を集計する
- 「相談区分１」「相談区分２」「相談区分３」を区別せず、全ての相談区分を集計する

前者は、一つの主要な相談テーマを持つ相談のみの集計であり、明確な相談内容がどのようなものであるかを分析できると考えられる。また後者は、その相談テーマが主要なものか副次的なものかは問わず、少なくとも相談したいことではあるという意味で、各テーマにどれくらいのニーズがあるかを表していると考えられる。

また、相談区分のラベルは上に示したように「その他」を除いて 20 種 類もあるので、相談ニーズを粗く把握したいという目的には必ずしも向 いていない。そこで、上の分類で明確に同じ種類のテーマであると考えられるもの（例えば、「診断」、「検診」、「検査」、「告知・ＩＣ」、「ＳＯ」らは医療診断の相談である）をまとめて粗視化することで、より明確に相 談ニーズを量的、質的に浮かび上がらせる。

3.2　「主訴」「対応」テキストによる分析

「主訴」と「対応」は、相談者の相談内容と相談担当者の対応のテキ ストデータである。具体的には、相談担当者が相談の対応を終えたあと、相談中にとったメモ、記憶、印象をもとにテキスト起こしをして記入する(補遺 B も参照)。文章の長さは相談一つあたり「主訴」と「対応」あわせて 300 字程度である。

この項目は相談ニーズがそのままに記されているという意味で貴重なデータであるが、自由記入された文章であるため、通常の統計処理では分析が難しい。ここでは自然言語処理技術を用いて、上の「相談区分」データの分析をサポートする。

具体的には、各相談の文章を形態素解析して語の単位に分解、トークン化し、ある相談区分に属する相談と相談全体での頻度情報の差に注目することで、統計的に有意に特徴的なものをキーワードとして取り出す(アルゴリズムの詳細は補遺 E を参照)。

これによって、その相談区分に特徴的なキーワードを見ることができ、相談区分の中で問題になっている相談者のニーズは何かを知る手がかりにできる。

4　相談区分タグによる相談者ニーズ分析

各相談区分への割り当てを行うためのマニュアルを精査し、相談員を含む日本対がん協会メンバーとのディスカッションの上で、20 項目から なる相談区分を以下の表 1 のように分類した（この検討過程については、 補遺 A と補遺 D を参照)。各項目の左にゴチック体で書かれた語がその粗 視化区分のラベルである。

ここで、ラベル＜診断＞と＜治療＞に分類された項目は医療に関する相談であり、対応としては適切な医療知識の提供が必要とされる。

また、ラベル＜病院＞に分類された項目は、入退院、通院、アクセスなど、病院施設に関する具体的な相談であり、このような知識が提供できるよう準備されていることが望ましい。

ラベル＜心の不安＞は、がんにまつわる精神的な不安の相談であり、真摯に主訴に耳を傾けてあげること、心のケア、コンサルティングなどの対応が予想される。相談件数の多い項目でもある。

ラベル＜症状・副作用・後遺症＞は相談件数の多い項目であり、また、相談区分の割り当てのためのマニュアル（参考文献 [2]）によれば、手術による日常生活への影響、排尿・排便障害や人工膀胱・肛門による日常生 活への影響、胃切除による食事への影響、治療後の健康管理など、医療

表 1: 各相談区分へのラベル付け

ラベル	相談区分
＜診断＞	「診断」「検診」「検査」「告知･IC」「SO」
＜治療＞	「治療」
＜病院＞	「外来」「入院・退院」「転院」 「施設設備・アクセス」
＜心の不安＞	「不安などの心の問題」
＜症状・副作用・後遺症＞	「症状・副作用・後遺症」
＜注目区分＞	「緩和ケア」「医療連携」「在宅療養」「医療者との関係」「補完代替療法」「就労・経済的な負担」「家族・周囲の人との関係」 「生き方・生きがい・価値観」

と日常生活の関係する領域の相談であり、医療情報に加え日常生活への影響を考慮した対応が必要である。

最後のラベル＜注目区分＞は必ずしも医療面だけとは限らない要素を含む様々な相談の項目群であり、どのような相談ニーズがあり、どのような対応が必要とされるか、注目して別に分析したいものである。

以上のラベルについて、「相談区分１」のみを持つ相談の件数ヒストグ ラムをとったものが以下である（数表については補遺 B を参照)。

「図 1」のヒストグラムから、相談の主題が比較的はっきりした相談 における相談ニーズの粗い分布が読み取れる。まず＜治療＞と＜診断＞をあわせた相談が全体の 40%を占めていて、直接的な医療相談が主要 なものであることが分かる。

残りの 60%は何らかの意味で医療面以外の要素を含んでいる。具体的には、＜病院＞に関する具体的情報が 5%強、＜心の不 安＞を訴える相談が約 20%、日常生活についての相談を含む＜症状・副作用・後遺症＞の相談が 20%、そして様々な相談項目を含む＜注目区分＞も十数%存在している。

以上は、はっきりとした一つの主題を持った相談のみのヒストグラムだが、以下(図 2)は「相談区分１」「相談区分２」「相談区分３」のいずれかに含まれた相談区分でヒストグラムをとったものである（数表については補遺 B を参照)。これは相談の主題ではなかったかも知れないが、とにかく相談の項目としてあがったテーマ全体の集計であり(ゆえに合計は

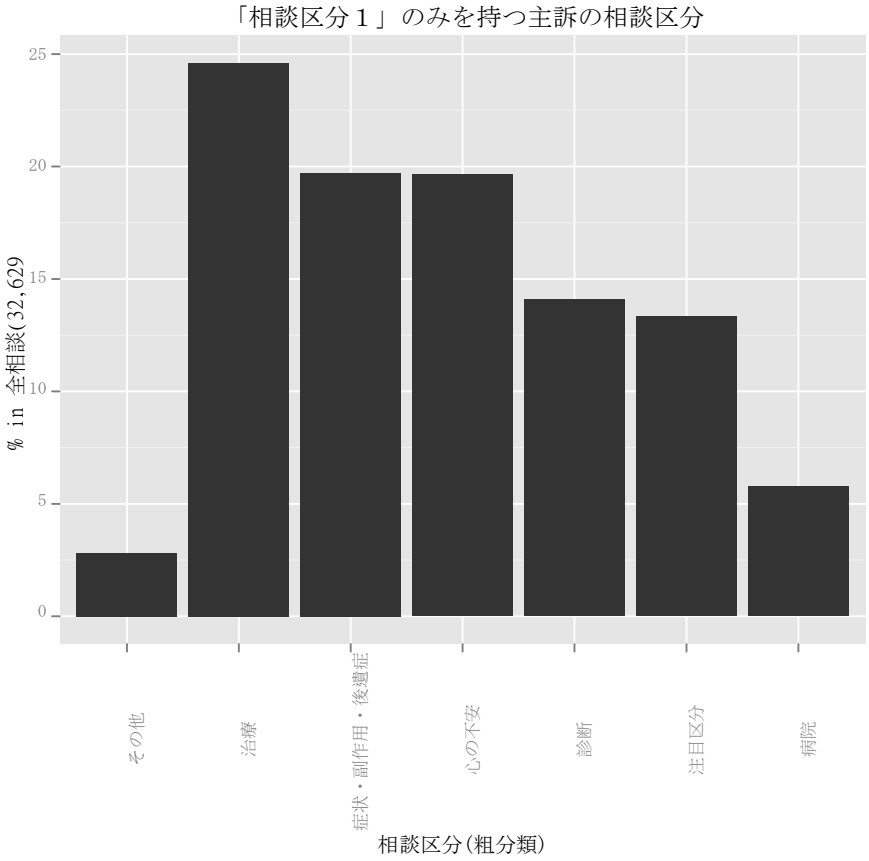

図 1: 「相談区分１」のみを持つ主訴での相談区分ヒストグラム

100%を越える)、潜在的な各相談ニーズの割合を表現していると考えられる。

「図 2」のヒストグラムによれば、８％程度を占める＜病院＞ラベルを除き、各ラベルの相談分類に 20～30%の大きなニーズがあることがわかる。

より具体的には、医療についての相談、診断についての相談、日常生活と医療が関係した相談、心の不安の訴え、そして以下で見る＜注目区分＞の雑多な相談、以上の五つの領域が同程度に需要を持っていると考えられる。

「図 3」のヒストグラムは、＜注目区分＞に含まれる相談区分だけに絞って、相談区分毎の件数を示したものである（Y軸は件数、数表については補遺 B 参照)。「相談区分１」だけを持つ相談に限った集計なので、1 つの主題がはっきりした相談での分布であると考えられる。

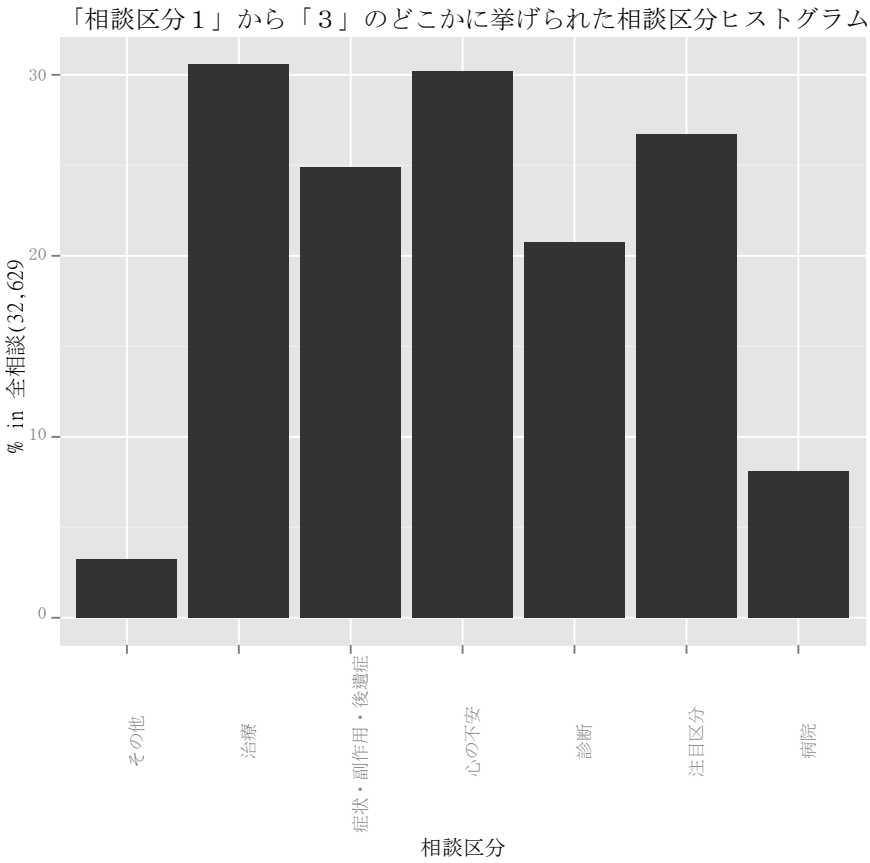

図 2: 「相談区分」１～３に挙げられた相談区分ヒストグラム

「図 3」のヒストグラムによれば、＜注目区分＞で最も多い相談項目は「就労・経済的な負担に関するものである。また「医療との関係」と「家族・周囲の人との関係がそれに次ぎ、この二つをあわせた人間関係についての相談が最も多いことが分かる。

「図 4」のヒストグラムは同じく＜注目区分＞の相談項目について「相談区分」の１から３のどこかに含まれている項目を集計したものである(数表については補遺 B 参照)。つまり、主題ではなかったかも知れないが、とにかく相談のテーマとして挙がったもの全ての集計である。

このヒストグラム(図 4)によれば、最も多いのは「医療者との関係」であり、全相談の８％というかなり大きなニーズを持つことがわかる。

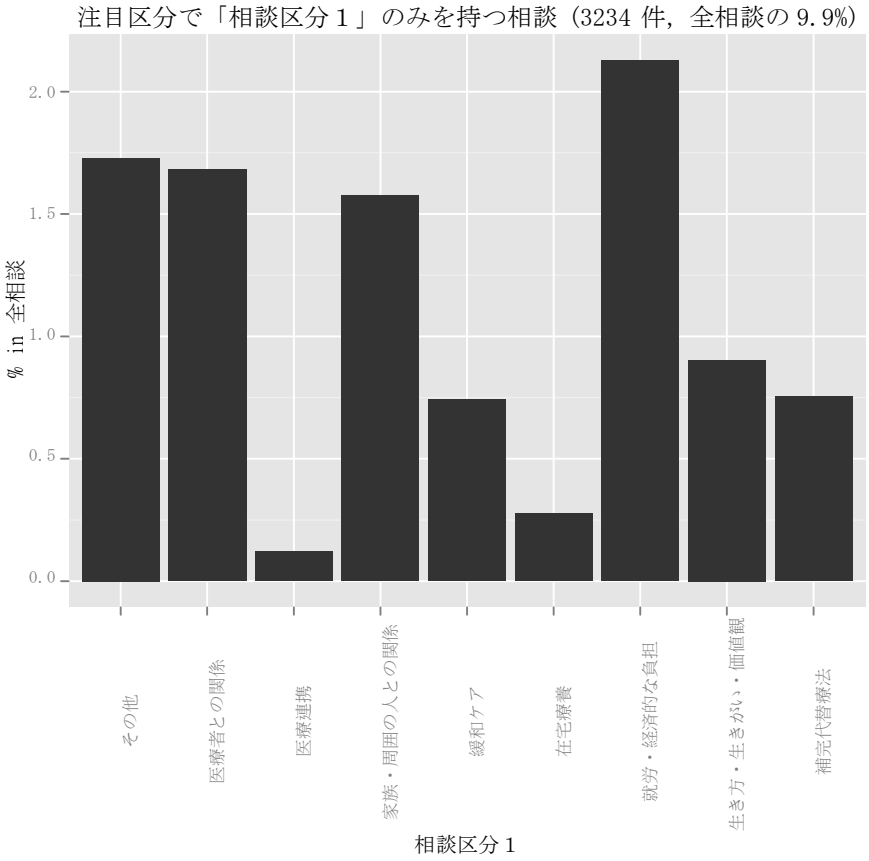

図 3: 「相談区分１」のみを持つ主訴の注目区分ヒストグラム

「家族・周囲の人との関係」も加えれば、人間関係についての相談は全相談の十数%という大きなニーズを持っている。「就労・経済的な負担」も全相談の数%を占めており、無視できない大きさである。

よって、「就労・経済的な負担」の相談区分、および人間関係についての二つの相談区分が、以下でのさらなる分析の対象になる。

5 相談テキストの自然言語処理によるニーズ分析

データベースに納められた各相談データの「主訴」と「対応」の項目は、相談者の相談内容と相談担当者の対応のテキストデータである。具

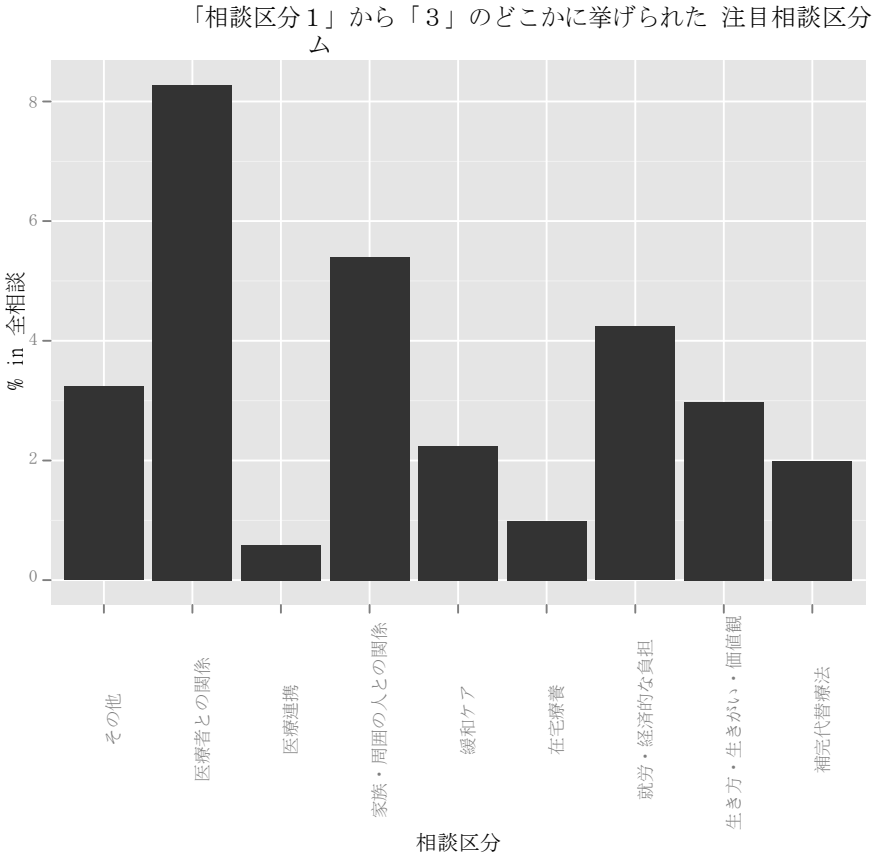

図 4: 「相談区分」１から３に挙げられた注目区分ヒストグラム

体的には、相談担当者が相談の対応を終えたあと、相談中にとったメモ、記憶、印象をもとに、テキスト起こしをして記入する。

この項目は相談ニーズがそのままに記されている貴重なデータであるが、自由記入された文書であるため、通常の統計処理では分析が難しい。

ここでは自然言語処理技術を用いて（詳細は補遺 E を参照)、上の「相談区分」データの分析をサポートする。

「主訴」「対応」の文章の長さの分布は以下のようになっていて（図 5)、多くの相談で一つあたり「主訴」「対応」あわせて 300 字程度しかない。

この文書長さは、相談一つ一つに対し高度な自然言語処理するには十 分ではないが、相談数自体は 3 万 6 千超とかなり多いため、相談の集合に おけるキーワードの抽出はある程度可能である。

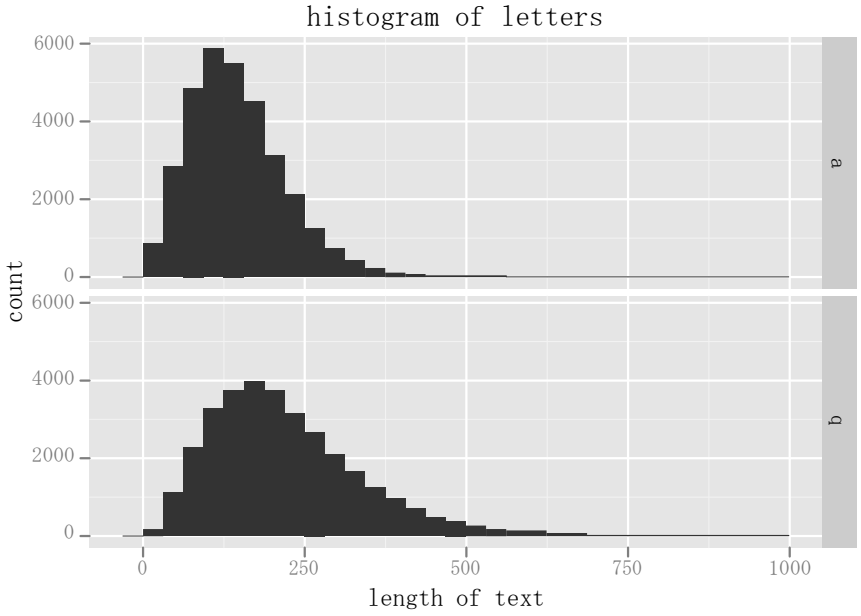

図 5: 「主訴」(q) と「対応」(a) の文章長さの分布

以下では、より詳しく性質を探りたい相談区分、あるいは、ラベルで整理された相談区分の集まりに対して、キーワード抽出を行ない、相談区分分析をサポートする。具体的には、注目区分として整理した相談区分の中で、比較的大きな割合を持ち、対応の方法が明確でない「就労・経済的な負担」、「医療者との関係」、「家族・周囲の人との関係」についてキーワード抽出を行う。なお、その区分の特徴をよりはっきりと抽出するために、「相談区分１」のみを持つ相談を対象とする。

その他の、「医療連携」、「緩和ケア」、「在宅療養」、「補完代替療法」については、相談目的の明確な相談としては全体の 1%以下 (図 3)、相談ニーズとしては全体の 2%程度以下であり (図 4)、またその相談に対して対応の仕方が比較的明確である。

「相談区分１」のみを持つ相談の中で、「就労・経済的な負担」に属するものは 694 件ある。以下がその「主訴」と「対応」から自動的に抽出 したキーワードの上位 100 個である。順序はその「強さ」の順になって いる (詳しくは補遺 E を参照)。

1: 高額, 2: 制度, 3: 保険, 4: 限度, 5: 医療費, 6: 療養, 7: 認定, 8: 適用, 9: 仕

事，10： 加入，11： 手当，12： 支払い，13： 経済的，14： 傷病，15： 生活保護，16： 費用，17： 傷病手当金，18： 収入，19： 診断書，20： 申請，21： お金，22： 健康保険，23:職場，24： 会社，25： 手続き，26： パート，27： 控除，28： 年金，29： 就職，30： 契約,31： 組合，32： 高額医療費，33： 会計，34： 就労，35： 国保，36： 生命，37： 適応，38:役所，39： ハローワーク，40： 概算，41： 窓口，42： 復職，43： 提出，44： 支援，45:申告，46： 給付，47： 退職，48： 保険証，49： 文化，50： 食費，51： 国民健康保険，52:復帰，53： 労働，54： 障害年金，55： 助成，56： 書類，57： 貸付，58： 返済，59： 健保，60： 社会保険，61： 減額，62： ローン，63： 経済的負担，64： 金額，65： SW，66： 休職，67： 差額，68： 社員，69： 請求，70： 就業，71： 支給，72： 休暇，73： 扶養，74： 雇用，75： コーナー，76： 国民，77： 受給，78： 支払，79： 滞納，80： いくら，81： 上司,82： 相談，83： ソリューション，84： タクシー運転手，85： 助成金，86： 医事，87:困窮，88： 社会福祉協議会，89： ハード，90： 対象，91： 分割，92： ソーシャルワーカー，93： 登録，94： 所得，95： 有給，96： 融資，97： 補助，98： 勤務，99： 生保，100:負担,...

これによれば、この区分に属する相談では、保険、年金、税金、公的支援、就職支援、借金返済などについて具体的な相談がされていることが分かる。この区分に属する相談が無視できない一定量がある以上は、こういった種類の問題に対し、少くとも正しい相談先に誘導できることが期待されるだろう。

「相談区分１」のみを持つ相談の中で、「医療者との関係」に属するものは549 件ある。以下がその「主訴」と「対応」から自動的に抽出したキーワードの上位 100 個である。

1： 患者，2： 転院，3： 不信感，4： 先生，5： 医師，6： 対応，7： 看護師長，8： 不満，9： 病院，10： 看護師，11： 怒り，12： 態度，13： 転勤，14： コミュニケーション,15： 思い，16： 質問，17： 医療者，18： 女医，19： 師長，20： 担当，21： 希望，22： 言い方，23： 主治医，24： 院長，25： やりとり，26： 外来看護師，27： メモ，28： 病棟,29： 外来，30： 妨害，31： テープ，32： 言葉，33： 相性，34： 診察，35： オンブズマン，36： トラブル，37： 親身，38： 医者，39： 信頼，40： 本意，41： スタッフ，42： 都庁，43： 一切，44： 変更，45： 話し合い，46： 同年，47： 職種，48： 担当医，49： 感情,50： 一緒，51： 手紙，52： キーパーソン，53： 納得，54： 後手，55： 挨拶，56： 要望,57： テラス，58： 業務，59： 経緯，60： 解決，61： 事務，62： 院内，63： 入院，64： 診察時，65： 率直，66： そこら，67： ドクハラ，68： 屈辱，69： 懇切，70： 損害，71： 止血剤，72： 無下，73： 精神的ダメージ，74： 豊洲，75： 錯乱状態，76： 傾聴，77： 修正，78： 故意，79： 線香，80： 大学病院，81： 担当看護師，82： 不在，83： 受付，84:最後，85： 事情，86： 批判，87： 死亡診断書，88： 苦情，89： 軟膏，90： パチンコ，91:医療事故，92： 安らか，93： 責任，94： 同席，95： 一方的，96： 医療過誤，97： 親切,98： +、，99： ESD，100： テトラミド，...

上のキーワードから推測されることは、これらの相談が主として、様々な種類の医療関係者（医師、看護師長、看護師、女医、主治医、院長、外来看護師、etc.） とのコミュニケーションにおける患者側のストレスに関係しているということである。

これらの相談に対応するには、相談者の気持を汲んで真摯に対応すること、コミュニケーションに関する心理的問題について的確なアドバイスができること、医療関係者側の立場も理解して説明できること、などが望ましい要素であると 考えられる。

「相談区分１」のみを持つ相談の中で、「家族・周囲の人との関係」」に属するものは 514 件ある。以下がその「主訴」と「対応」から自動的に抽出したキー ワードの上位 100 個である。

1: 自分, 2: 関係, 3: 家族, 4: 離婚, 5: イライラ, 6: 病気, 7: 兄弟, 8: 状況, 9:態度, 10: 面会, 11: 世話, 12: 傾聴, 13: 入院中, 14: 子ども, 15: 疲弊, 16: 気持ち, 17: 面倒, 18: 結婚, 19: 協力, 20: 相談者, 21: 母親, 22: 言葉, 23: 同居, 24:お互い, 25: 性格, 26: 義妹, 27: 父親, 28: 余命, 29: 状態, 30: 援助, 31: 感情, 32: 両親, 33: 関係性, 34: サポート, 35: 姉妹, 36: 周囲, 37: 距離, 38: 実家, 39:叔父, 40: 本人, 41: 長男, 42: 見舞い, 43: 喧嘩, 44: 精神的, 45: 学校, 46: 行動, 47: 苦労, 48: 機嫌, 49: 次男, 50: 高校生, 51: 現在, 52: 義母, 53: 接し方, 54:別居, 55: 女性, 56: 息子, 57: 生活, 58: 生徒, 59: 付き添い, 60: 精一杯, 61: 独身, 62: お金, 63: 警察, 64: 看病, 65: 愚痴, 66: 支え, 67: 一緒, 68: 友人, 69: 悩み, 70: おば, 71: 養護教諭, 72: 暮らし, 73: 手伝い, 74: 大切, 75: 子供, 76: 長女, 77: 心理, 78: 解決, 79: 八つ, 80: ご主人, 81: 彼女, 82: あなた, 83: 我慢, 84: 不満, 85: かかわり, 86: 家庭, 87: お母さん, 88: 相手, 89: 怒り, 90: 弱み, 91: 役員, 92: 非協力的, 93: 暴力, 94: 暴言, 95: 実母, 96: 当たり, 97: ストレス, 98: ツリー, 99: ホープ, 100: いざこざ, ...

上のキーワードから推測されることは、これらの相談が主として、家族を中心とする様々な人間関係の中でのストレスを原因としているものであるということである。これらの相談への対応では、相談者の気持に沿って真摯に訴えに耳を傾けることはもちろん、がん治療や診断をきっかけにした身近な人々との関係のストレスについて、心理的あるいは社会的な立場からアドバイスできることが望ましいだろう。

6　要約

地域統括相談支援センター活性化の方策を立案する手がかりとして、「がん相談　ホットライン」の過去三年間のデータを用いて、相談者のニーズを整理・分析した。分析の方法としては主に、相談員によって相談毎に付与された「相談区分」(複数) を統計処理する他に、相談者と相談員の「主訴」と「対応」のテキストデータを対象に自然言語処理による解析を行なった。

「相談区分」はその相談に主要なテーマから順に「相談区分１」、「相談区分２」、「相談区分３」まで必要に応じて付与されている。今回は、「相談区分１」だけを持つ相談に付与された相談区分と、「相談区分１」から「相談区分３」のどこかに含まれた全ての相談区分の二通りについて集計を行なった。前者は、一つのはっきりした相談テーマを持った相談に対する集計であり、相談者の間に強いニーズの分析であると考えられる。一方後者は、主要であるかどうかに限らず相談に挙げられたテーマに関する集計であり、相談者の中での潜在的なニーズの分析である。

前者による強い相談主題の集計によれば、主要なものは治療と診断に関する医療の相談であり、全体の約 40%を占めている。一方で、残りの 60%は何らかの意味で医療面以外の要素を含んでいると考えられる。具体的には、医療と日常生活の両方に関係する相談を多く含む症状・副作用・後遺症の相談が約 20%、心の不安を訴える相談が約 20%、病院施設に関する具体的な相談が約 5%、残りの約 15%がその他の相談である。この「その他」の相談で主要なものは、就労・経済的な負担の相談、および、医療者、家族、周囲の人との関係の相談である。

後者による潜在的な相談ニーズの集計によれば、医療についての相談、診断についての相談、日常生活と医療が関係した相談、心の不安の訴え、そしてその他の雑多な相談、以上の五つの領域が同程度に需要を持っていると考えられる。その他の雑多な相談では、上と同様にやはり、就労・経済的な負担の相談と医療者、家族、周囲の人との関係の相談が主要なものである。

この就労・経済的な負担の相談、および人との関係についての相談については、「相談区分１」だけを持つ相談に限定して、自然言語処理のテクニックを利用し、具体的内容である「主訴」「対応」のテキスト分析を行なった。

これによれば、就労・経済的な負担の相談においては、保険、年金、税金、公的支援、就職支援、借金返済などについて具体的な相談がされている。また、人との関係についての相談では、様々な種類の医療関係者（医師、看護師長、看護師、女医、主治医、院長、外来看護師、etc.）および家族や周囲の人とのコミュニケーションにおける、ストレスの相談が主要なものである。

以上の分析を総合すると、医療・診断に関するものや病院施設に関する質問など医療面の相談が大半であるものの、日常生活と医療が関係する領域の相談と、心の不安の訴えも同程度の大きなニーズを持つ。これらに対しては、医療面の相談と同程度に対応のためのリソースを用意する必要があると考えられる。また全相談の数パーセント程度だが無視できないものとして、就労・経済的負担の問題と（対医療者を含む）人間関係の悩みの相談がある。これらについてもそれぞれに対応できる態勢を用意する必要があると考えられる。

A 分析の手続き

分析の手続は以下のステップに従った。 今回の分析の目的は、相談データに対し何らかの集計を行なうことによって、相談者たちの「相談ニーズ」を定量的／定性的に把握することであるため、分析の主な課題はデータの記述的な分析になる。 今回の目的は、サンプリングされた部分的データから母集団の特徴的な値を推測する、という問題ではないため、通常、統計的分析で行なわれる「推測的データ解析（IDA; Inferential Data Analysis)」(確率モデルを仮定し、未知の母 数を推測する）のステップは含まれない。

なお、関係者を交えた検討のため使用したが、報告書本文部分には直接反映されなかった図表の中で、比較的に重要なものを参考として補遺 D に挙げておいた。

1. 分析目的、課題の定式化
   - 関係者と分析者の検討会議
2. データ訊問（CED; Cross Examination of Data）
   - 第一次の発見的データ解析（データの取得方法、定義の理解、外れ値、特徴の分析、分析対象と分析方法の策定、etc.）
   - 相談対応者へのインタヴュー
   - （データの「信頼性」についての議論は補遺 B を参照）
3. 記述的データ解析（DDA; Descriptive Data Analysis）
   - 得られたデータの要約と記述による分析
   - 分析目的にあわせた「相談区分」の集計
   - 分析目的にあわせた「主訴」「対応」のキーワード抽出（補遺 E 参照）
     - キーワード抽出のための確率モデル設計
     - モデルと自然言語処理によるキーワード抽出
4. 分析目的、課題へのフィードバック
   - 関係者と分析者の検討会議
   - 必要に応じて上記ステップ 1, 2, 3 に戻り、全手続を繰り返す

B データの信頼性についての議論

分析対象としたのは、

・ 「相談区分１」,「相談区分２」,「相談区分３」

・「主訴」,「対応」

の 5 項目だった。前の三つ「相談区分」の１から３は、その相談がどのような種 類のものか相談担当者が判断して 21 種類の分類ラベルの一つを記入したもので あり、また後ろの二つ「主訴」と「対応」は相談者の相談内容に対して、電話相 談が済んでから、メモ、記憶、印象を元に、相談担当者が相談内容と対応をテキ スト起こししたものである。

したがって、これらのデータは相談担当者の主観によって記録されたものであり、その信頼性について若干の注意が必要である。

問題になるのは以下の二点である。

1. データは（真の）相談内容を正しく記述しているか
2. 相談担当者の間に分析を歪めるような個人差がないか

厳密に言えば、1 番目の問題は元の相談が残っていないため確認できない。2 番目の問題も一つの相談を一人の相談員が対応する以上は比較できない。

この状況は、実験計画なしに既に存在してしまっているデータの分析に対して常に生じる問題であり、分析目的に十分な程度の確認をすること、その分析結果の限界を承知しておくことが重要である。

まず、1 番目の問題については、事前のデータ訊問によってデータ欠損や外れ値など非正常値がほとんどない（実際 32,629 件中 7 件のみ）こと、および、相談 担当者へのインタヴューによって、相談担当者の訓練、分類のマニュアル化、 知識の共有が徹底されていることの二点を確認した。これは、「相談者の ニーズを粗く把握する」という分析の目的には十分であると考えられる。

2 番目の問題についても同じく、相談担当者の訓練、分類のマニュアル化、知識の共有などの一様性への努力がなされていることを確認した。さらに、相談担当者別（18 名）の層別ランダムサンプリングによって記録データが妥当で あることを目視で確認した。

定量的な検証方法として、同じ分布からのサンプリングであることを帰無仮説として相談担当者間に有意な差がないことを検定することも考えられるが、相談担当者の間に一定量の判断の差があるのは当然であり、かなり粗い精度でも数学的な同分布性を期待することは現実的でない。むしろ、18 名の担当者のそれぞれの判断の差異が、集合知の意味で平均化されると考えるべきであろう。

我々の目的はデータを集計し、要求された視点から記述することであり、統計的推定／検定は行わないため、厳密な誤差評価はあまり問題にならない。しかし、一方では我々のデータ集計は、あくまで相談担当者の主観の集計であり、何らかのバイアスが含まれている可能性があることも、限界として承知しておかなければならない。例えば、相談担当者自身は健康であってがん患者ではないから、相談者の意図を判断する時に何らかのバイアスを持っているかも知れない。

C　各ヒストグラムの数表

表 2:　「相談区分１」のみを持つ相談の相談区分集計（件数の%表示は全 相談中の割合。よって合計は 100%未満）

ラベル	相談区分	件数	件数（%）
診断	診断	466	2821(8.6%)
	検診	389	
	検査	1309	
	告知・ＩＣ	87	
	SO	570	
治療	治療	4924	4924(15%)
病院	外来	290	1155(3.5%)
	入院・退院	62	
	転院	706	
	施設設備・アクセス	97	
心の不安	不安などの心の問題	3933	3933(12%)
症状・副作用・後遺症	症状・副作用・後遺症	3946	3946(12%)
注目区分	医療者との関係	549	2670(8.2)%
	医療連携	40	
	家族・周囲の人との関係	514	
	緩和ケア	242	
	在宅療養	90	
	就労・経済的な負担	694	
	生き方・生きがい・価値観	295	
	補完代替療法	246	
その他	その他	564	564(1.7%)

表 3: 「相談区分」１，２，３のどこかに現れた相談区分の集計（件数の%表示は全相談中の割合。よって合計は 100%を越える）

ラベル	相談区分	件数	件数（%）
診断	診断	1152	6774(21%)
	検診	556	
	検査	2798	
	告知・ＩＣ	271	
	SO	1997	
治療	治療	9975	9975(30%)
病院	外来	519	2639(8.0%)
	入院・退院	161	
	転院	1733	
	施設設備・アクセス	226	
心の不安	不安などの心の問題	9855	9855(30%)
症状・副作用・後遺症	症状・副作用・後遺症	8130	8130(25%)
注目区分	医療者との関係	2699	8714(27%)
	医療連携	191	
	家族・周囲の人との関係	1761	
	緩和ケア	731	
	在宅療養	324	
	就労・経済的な負担	1384	
	生き方・生きがい・価値観	973	
	補完代替療法	651	
その他	その他	1056	1056(3.2%)

D　その他の参考図表

図 6 は、相談区分１のみを持つ相談における、現状の相談区分全てを単純に 集計したヒストグラムである。

明らかに見てとれるのは、主旨が一つではっきりした相談におけるテーマは、現状の相談区分での「治療」、「症状・副作用・後遺症」、「不安などの心の問題」 の三つが主だということである。

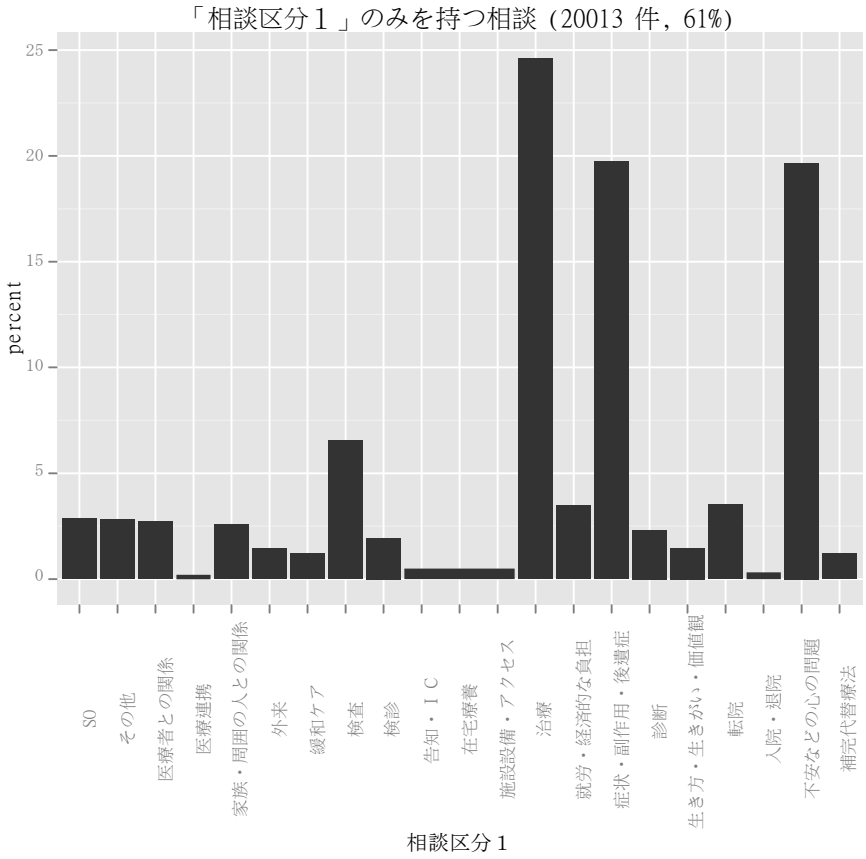

図 6: 「相談区分１」のみを持つ相談の相談区分ヒストグラム

図 7 と図 8 はそれぞれ、「相談区分１」と「相談区分２」のみを持つ相談にお ける、「相談区分１」および「相談区分２」の集計ヒストグラムである。

主テーマである「相談区分１」の分布は、前図 6 の「相談区分１」のみを持つ 相談と大差ない。しかし、副テーマである「相談区分２」の分布は、大きく「不 安などの心の問題」に偏っており、主な相談テーマが他にあったとしても、心の 不安を訴える相談者が多いことを示唆している。

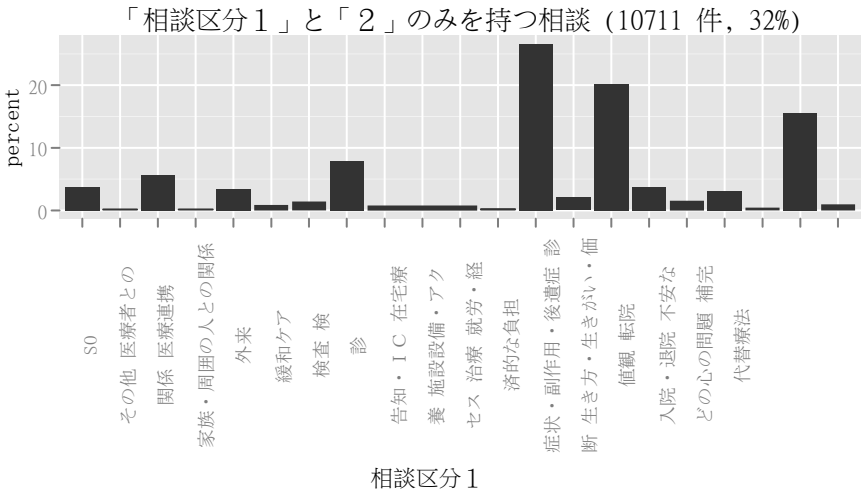

図 7:「相談区分」１と２のみを持つ相談での相談区分１のヒストグラム

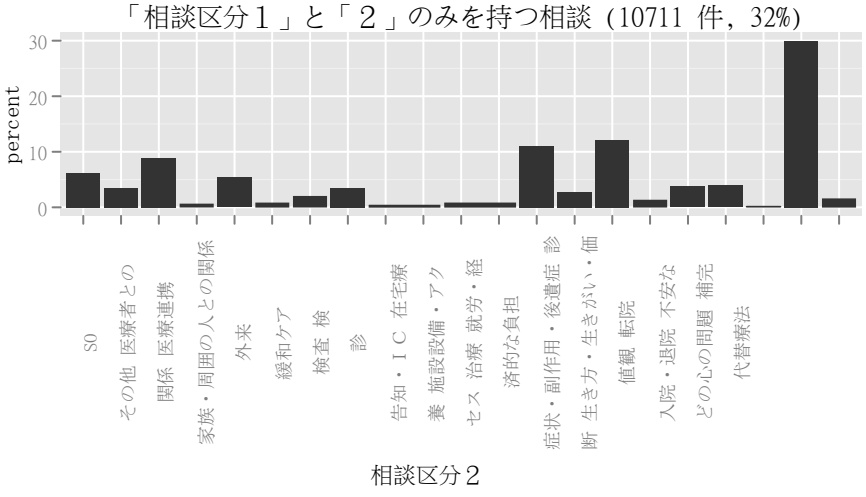

図 8:「相談区分」１と２のみを持つ相談での相談区分２のヒストグラム

図 9 は、「相談区分」１と２のみを持つ相談での相談区分１と２のヒートマッ プである。明るく示されているところほど、そこが示す「相談区分」１と２の組 合せで訴えた相談者が多い。

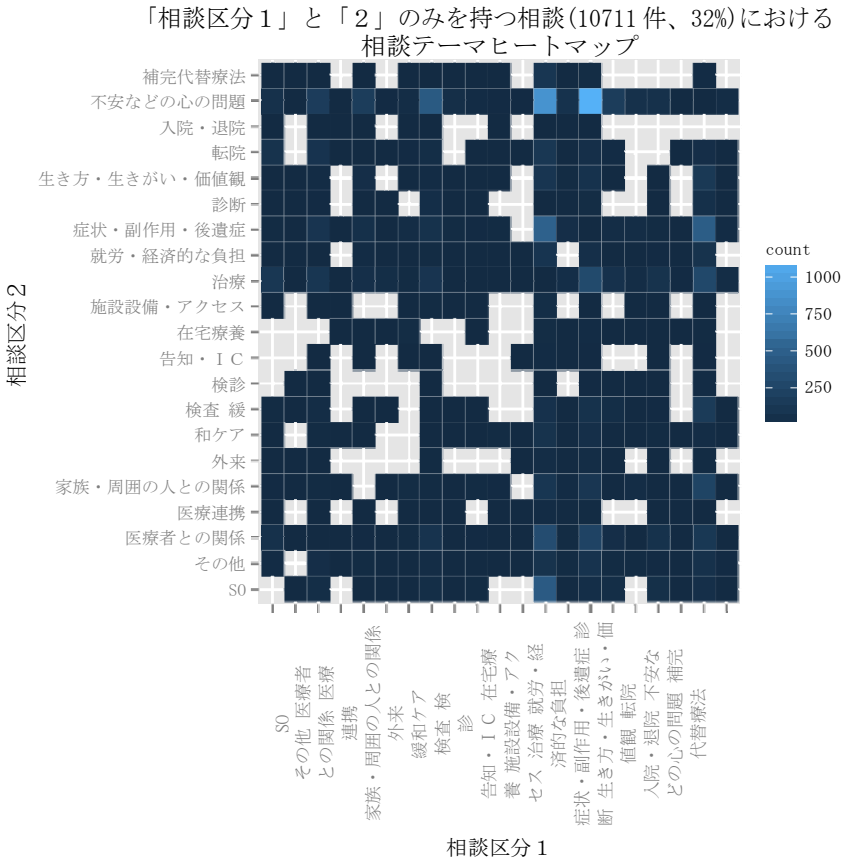

図 9: 「相談区分」１と２のみを持つ相談での相談区分１と２のヒート マップ

図 10 は、「相談区分１」のみを持つ相談における、「相談区分」とその粗視化 との関係をヒートマップにしたものである。明るく示されているところほど、そ こが示す粗視化区分が多いことを意味している。この図は相談区分の粗視化を 策定する段階で参考にした。

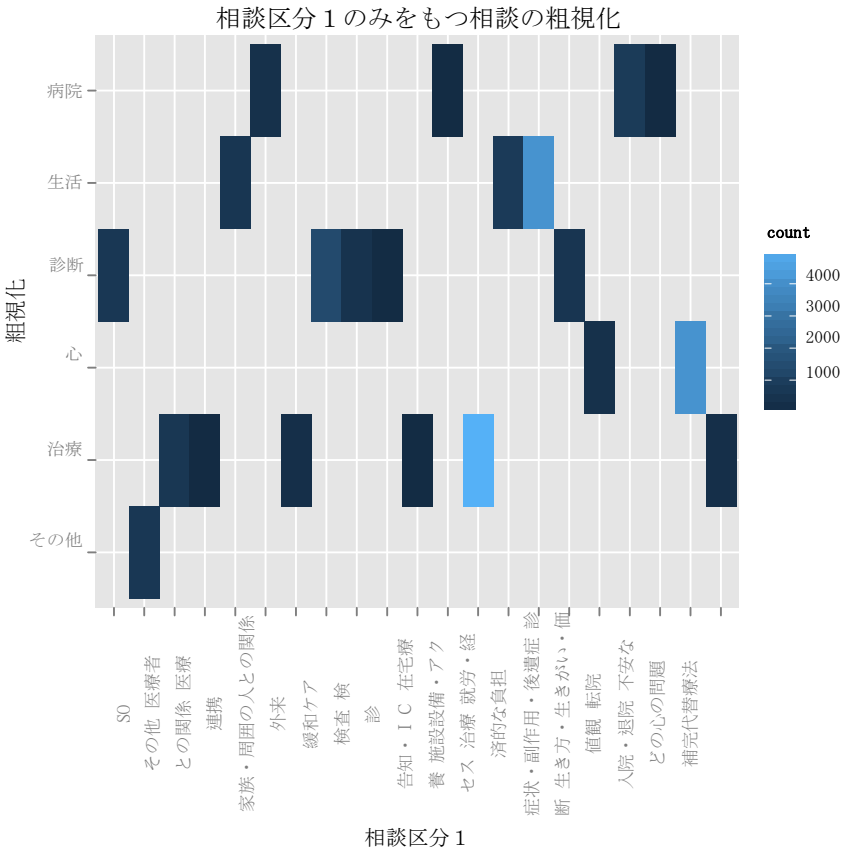

図 10: 「相談区分」１のみを持つ相談での相談区分とその粗視化区分のヒートマップ

E　文書頻度比較による分析の解説

「主訴」と「対応」のテキストデータの「分かち書き」、すなわち文章の語への分解は、以下のツールを主に用いて行なった。

・ MeCab (Ver.0.996)： 工藤拓氏 (Google/元奈良先端科学技術大学院大学)
による形態素解析器。

・ComeJisyo (Ver.5)：相良かおる氏（西南女学院大学）による MeCab 用の 医療用語辞書。医療情報提供： 千葉大学医学部付属病院、聖路加国際病院、 佐賀大学医学部付属病院。

さらに切り出した語について、分析のノイズになるものを除去する、複合語に再結合する、などの後処理（クレンジング）を独自のアルゴリズムで行なった。このような文書を構成する特徴的な語の取り出しをトークン化、その語をトー クンと言うが、以下、混乱がない限りトークンを単に「語」とも書く。

各トークンに対して、全相談の中でそのトークンを含む相談が占める割合を文書頻度 (document frequency) と呼ぶ。文書頻度は、その語が母集団となる文 書集合の中でどれくらい珍しいかを表していて、文書頻度が低いほどその語は 特徴的であると考えられる。

我々は、ある相談区分、あるいはいくつかの相談区分の集まりに属する相談に特徴的な語を選び出すために、その語 w の相談全体に対する文書頻度 $D(w)$ と、その特定の文書集合に対する文書頻度 $d(w)$ を比較する。もし、$D(w)$ に比 べて $d(w)$ が小さければ、その語 w はその文書集合に特徴的な語だろうと考え られる。

その「珍しさ」の強度を表すために、我々は自然言語処理で “bag of words” と呼ばれる確率モデルを用いる。すなわち、語順に関係なく各文書（各相談）を トークンの集まりであるとして、各トークンはその出現頻度を確率としてラン ダムに選ばれているものとする。今の場合、ある相談を一つランダムに選んだ ときに、その相談が語 w を含む確率は $q = 1/D(w)$ であり、相談を N 個とり だしたときに w を含む文書の個数は二項分布に従う。

今、問題の文書の部分集合に w を含む語が n 個含まれているとき $n = N/d(w)$ であり、我々はこの「珍しさ」を、二項分布の p-値で評価する。 つまり、二項分布 $B(q, n)$ の確率分布を $b(j; q, n)$ と書くとき、

$$P(w) = \sum_{j=n}^{N} b(j; q, N) = \sum_{j=n}^{N} \left(\begin{array}{c} N \\ j \end{array} \right) \left(\frac{1}{D(w)} \right)^{j} \left(1 - \frac{1}{D(w)} \right)^{N-j}$$

をもって、その語 w の強度とする。我々は、この $P(w)$ が小さいほど、その語 は一般的相談に現れる場合に比較してその文書集合に特徴的である、として特 徴語（キーワード）を選んだ。

参考文献

[1] 「がん相談ホットライン 2013 年度 年報」, 公益財団法人 日本対がん協会 相談支援室, 2014

[2] 「「相談内容区分」(新システム用) 2011/6/15 ミーティング資料」, 公益財 団法人 日本対がん協会, 2011