ライフログ-アドバイスコーパスの仕様

岡照晃，粟村誉，仲村哲明，荒牧英治，河原大輔，黒橋禎夫

2016/4/10

1. はじめに

この資料は京都大学黒橋・河原研究室で構築した**ライフログ-アドバイスコーパス**のガイドラインである．このコーパスは，人々の健康な生活をサポートするための**健康アドバイス自動生成システム**構築を目的に作成されており，人々が実際に書いた日々のライフログ記事と，それに対し管理栄養士・運動トレーナーが執筆したアドバイス記事を含んでいる．ここでいう**ライフログ記事**とは「ある人がある1日の間に何を食べ，どんな運動を行なったか」という内容を記述した日記を指している．以下にライフログ記事の例を示す．

ライフログ記事の例：

朝ごはんは、ご飯と目玉焼きとみそ汁を食べました。 午前中は掃除や洗濯などの家事を行い、お昼はパスタを食べました（タラコ） 夕方自転車で買い物に行き、帰ってから夕食の支度。 寒かったので、今夜はお鍋を食べました。 食後にシュークリーム1個。 寝る前に軽くストレッチをするのが習慣です

本コーパスはXML形式で記述されており，文字コードはUTF-8（BOMなし），改行コードにはLF（¥n）を使用している．以降の章からは，このコーパスについて説明していく．

2. コーパスの概要

コーパスの概要は次の通りである．ルート要素である lifelog_advice_corpus 要素の直下には 複数の person 要素がある．1つの person 要素には，あるライフログ執筆者1人が書いたすべてのライフログ記事と，各ライフログ記事に対するアドバイス記事が含まれている．person 要素の直下には複数の daily_lifelog_advice 要素がある．1つの daily_lifelog_advice 要素には，1つの lifelog 要素と1つ以上の advice 要素が含まれている．lifelog 要素はライフログ1記事を格納しており，advice 要素はそのライフログ記事に対する（管理栄養士A，管理栄養士B，運動トレーナーのいずれかの）アドバイス 1 記事を格納している．1 つの lifelog 要素に対して複数人からのアドバイス記事が存在する場合は，daily_lifelog_advice 要素中に，その人数分の advice 要素が含まれている．以上をまとめると，本コーパスは全体として次のような構造となっている．

XML 概略：

```
<lifelog_advice_corpus>

<person>
<daily_lifelog_advice>
<lifelog>ここにライフログ記事本文があります．</lifelog>
<advice>
	ここに上記のライフログ記事本文に対するアドバイス記事があります．
	このタグ1つにつき専門家1人分のアドバイス記事が格納されています．
</advice>
<advice>
	上記のライフログ記事本文に対して複数人からのアドバイスがある場合は，
	同名のタグ（advice）が以降に連なります．
</advice>
</daily_lifelog_advice>

<daily_lifelog_advice>
<lifelog> ・・・ </lifelog>
<advice> ・・・ </advice>
<advice> ・・・ </advice>
</daily_lifelog_advice>
・・・
</person>

<person> ・・・ </person>
<person> ・・・ </person>
・・・
</lifelog_advice_corpus>
```

3．XML タグセット

本コーパスでは，ライフログ記事とそれに対するアドバイス記事へ XML によるマークアップを行なっている．そのための XML タグの一覧は，次の**表1**のとおりである．各タグで表される要素については続く各節で詳説する．

表１　XML タグセット

タグ名	説明	詳説する節番号
lifelog_advice_corpus	コーパス全体を表す．（ルート要素）	3．1
person	ライフログの書き手1人を表す．	3．2
daily_lifelog_advice	１人の書き手が執筆した1日分のライフログ記事と，それに対して管理栄養士や運動トレーナーが執筆したアドバイス記事のまとまりを表す．	3．3
lifelog	１人の書き手の執筆した1日分のライフログ記事を表す．	3．4
advice	１人の書き手の執筆した1日分のライフログ記事に対して，管理栄養士もしくは運動トレーナー1名が執筆した1つのアドバイス記事を表す．	3．5
l_s	ライフログ記事中の文を表す．	3．6
a_s	アドバイス記事中の文を表す．	3．7
health_concept	「ヘルスコンセプト」を表す．	3．8
a_subS	「文のタイプ」をアノテーションするために1文をさらに小さく分割した単位を表す．	3．9
fi	ライフログ記事中に記述された1食品名を表す．	3．10
fe	ライフログ記事中に記述された1食事イベントを表す．	3．11
ei	ライフログ記事中に記述された1運動名を表す．	3．12
ee	ライフログ記事中に記述された1運動イベントを表す．	3．13
q	ライフログ記事中に記述された食事や運動の量を表す．	3．14
t	ライフログ記事中に記述された食事や運動の時刻を表す．	3．15

３．１　lifelog_advice_corpus 要素

説明

コーパス全体を表す．ルート要素．

属性

なし．

直下に存在する子要素

person

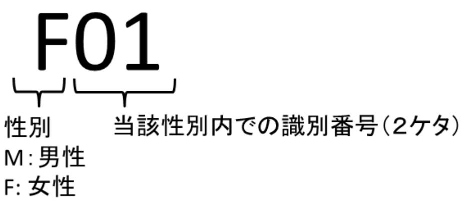
３．２　person 要素

説明

ライフログの書き手1人を表す．ライフログの書き手は，男女各10名．それぞれに後述するユニークな ID を設定している．

属性

id（必須）：　本コーパス中で，ライフログの書き手を識別するためのユニークな ID．形式は以下の通り．



sex（必須）：ライフログの書き手の性別．

problem（必須）：ライフログの書き手が抱える健康問題（自由記述）．

直下に存在する子要素

daily_lifelog_advice

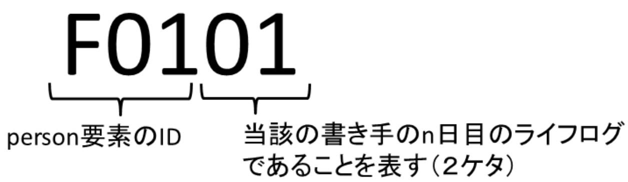
３．３ daily_lifelog_advice 要素

説明

１人の書き手の執筆した1日分のライフログ記事と，それに対して管理栄養士や運動トレーナーが執筆したアドバイス記事のまとまりを表す．

属性

id（必須）: 本コーパス中で，当該日のライフログ記事と，それに対するアドバイス記事のまとまりを，他の書き手もしくは他の日付と識別するためのユニークな ID．形式は以下の通り．



直下に存在する子要素

lifelog

advice

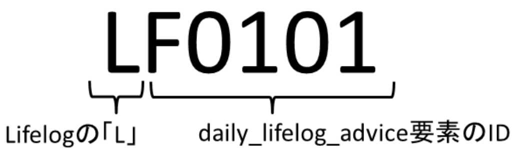
３．４ lifelog 要素

説明

１人の書き手の執筆した1日分のライフログ記事を表す．

属性

id（必須）:本コーパス中で，当該ライフログ記事を識別するためのユニークな ID．形式は以下の通り．



date（必須）: 当該ライフログ記事の執筆された日付（月＋日）

theme（必須）: 当該ライフログ記事のテーマ（自由記述）

title（必須）: 当該ライフログ記事のタイトル（自由記述）

直下に存在する子要素

l_s

health_concept

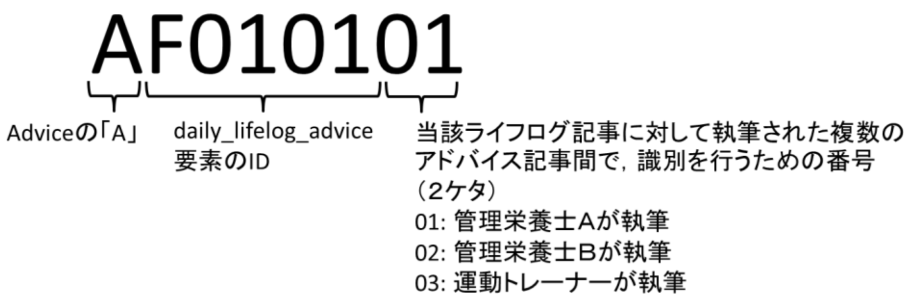
３．５ advice 要素

説明

当該ライフログ記事に対して1人の専門家（管理栄養士A，管理栄養士B，運動トレーナーのいずれか）が執筆したアドバイス記事を表す．

属性

id（必須）：本コーパス中で，当該アドバイス記事を識別するためのユニークな ID．形式は以下の通り．



adviser（必須）： 当該アドバイス記事を執筆した専門家を表す．値は管理栄養士A，管理栄養士B，運動トレーナーのいずれかである．

直下に存在する子要素

a_s

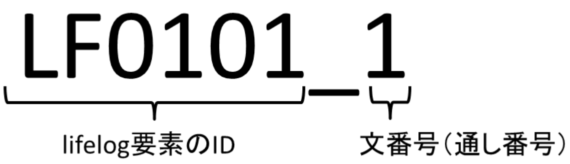
３．６ l_s 要素

説明

ライフログ記事中の文（ライフログ文）を表す．

属性

id（必須）: 本コーパス中で，当該ライフログ文を識別するためのユニークな ID．形式は以下の通り．



直下に存在する子要素

fi

fe

ei

ee

t

q

文字列

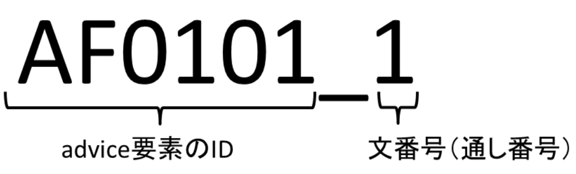
3．7　a_s 要素

説明

アドバイス記事中の文（アドバイス文）を表す．

属性

id（必須）：　本コーパス中で当該アドバイス文を識別するためのユニークな ID．形式は以下の通り．



link（任意）：　当該アドバイス文が書かれる基となったライフログ記事中の文番号を表す．文番号が複数ある場合は，半角カンマ区切りで並べて書かれる．ライフログ中の文番号は，l_s 要素の id の文番号（半角ハイフンより右の数字）と一致する．またライフログ記事中に対応する文はないが，内容から類推して生成されたアドバイス文には-1 が付けられている．また N 日前のライフログ記事に関連づいたアドバイス文である場合は，preN という値が付けられている．

直下に存在する子要素

a_subS

helth_concept

３．８ health_concept 要素

説明

当該ライフログ記事もしくはアドバイス記事中の文に付与された**ヘルスコンセプト**を表す．ヘルスコンセプトとは，本研究で定義した造語であり，ライフログ記事，アドバイス記事中の文を抽象化して捉えるために使用される．例えば，以下のライフログ文はいずれも<health_concept name=″**朝食をとる**″ value=″YES″/>というタグ（name と value を合わせて，「朝食を食べた」の意）として抽象化され，それぞれ文面や内容に差異はあっても，このタグの持つ意味の下で，すべて同一視される．

例１　今朝は朝食を食べた．

例２　朝ごはんはみそ汁と玉子焼き．

例３　今日は寝坊したので朝はトーストだけ．

例４　朝は昨夜の残り物ですませ，昼は抜き．

またアドバイス文に付与されたヘルスコンセプトは，ライフログ記事での抽象化と若干異なる．ライフログ記事に付与されたヘルスコンセプトがライフログ記事自体を抽象化だったのに対し，アドバイス文に付与されたヘルスコンセプトは当該アドバイス文の直接の抽象化ではない．アドバイス文に付与されたヘルスコンセプトは，あくまでそのアドバイス文が言及しているライフログ記事中のある部分（もしくは全体，または言及されていないがライフログ記事の内容から類推できる内容）を抽象化して付与されたものである．つまり，ライフログ記事をヘルスコンセプトへと一旦抽象化し，抽象化されたヘルスコンセプトからアドバイス文が生成された，という流れを想定したアノテーションとなっている．

ヘルスコンセプトに関する詳細は以下の[栗村ら, 2016]を参照してほしい．

栗村誉，岡照晃，荒牧英治，河原大輔，黒橋禎夫，ユーザのライフログに対する健康アドバイスの自動生成． 2016． 言語処理学会第 22 回年次大会発表論文集(NLP2016).

またヘルスコンセプトを利用したアドバイス生成の流れは以下の[岡ら, 2016]を参照してほしい．

岡照晃，栗村誉，荒牧英治，河原大輔，黒橋禎夫，おしゃべりけんこうノート：管理栄養士・インストラクターのアドバイスに基づく健康アドバイスシステム，言語処理学会第 22 回年次大会発表論文集(NLP2016).

属性

name（必須）: 当該ヘルスコンセプトの名前.

value（必須）: 当該ヘルスコンセプトの名前（name 属性）が実行されたか否か, もしくは不明かをYES, NO, UNK の3値で表す.

直下に存在する子要素

子要素を持たない.

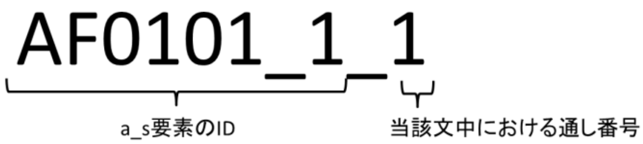
３．９ a_subS 要素

説明

本コーパスでは**文のタイプ**を1文をさらに小さく分割した単位でアノテーションしており, そこで用いている分割単位を表す.

属性

id（必須）: 本コーパス中で, この分割単位を識別するためのユニークな ID. 形式は以下の通り.



polarity（任意）: 当該分割単位がそのアドバイス記事の中でどういった役割を持っているか表したもので, 値には, 褒める（p）, 注意する（n）, 情報提示やあいづちなど(u), 将来への提案（f）の 4 つと, それらの複合（e.g., pf, cf, ...）をとる.

直下に存在する子要素

文字列

３．１０ fi 要素

説明

ライフログ記事中に記述された1食品名（風邪薬などの医薬品も含む）を表す. 構文解析システムKNP で解析した文節を単位としているため, 食品名を表す名詞句に連続して助詞や句読点も要素中には含まれる.

また特殊な例として, 食事イベントの生起と食事内容を同時に表す名詞句は, 後述の fe タグとともに, fi タグの付与が行われている. 例えば,「今夜は焼肉パーティー。」のような文に対しては,

「焼肉パーティー。」に fe タグが付与されるとともに，fi タグも付与される．

属性

fact（任意）：当該要素が食品の摂取を明示しているのか非摂取を明示しているのかを区別するための属性．デフォルトでは食品の摂取に言及しているとし，この属性は明示しないが，食品の非摂取に言及している場合は，neg という値で非摂取を明示化している．

直下に存在する子要素

fe

文字列

3．11　fe 要素

説明

ライフログ記事中に記述された1食事イベントを表す．構文解析システム KNP で解析した文節を単位としているため，食事イベントを表す単語列（「ペロリ」のようなオノマトペも含む）に連続して句読点も要素中には含まれる．また「する」や「なる」のように単独で意味がわからないものでも，食事イベントの発生に言及している場合はこの要素となる．タグ付与の方針として，「摂取した食品が記述されている文」には必ず<fe>タグを付与しているため，食事イベントの発生を明示する文節が存在しない場合，文末に空要素の<fe>タグ（<fe/>）を付与している．

また特殊な例として，食事イベントの生起と食事内容を同時に表す名詞句は， fe タグとともに，前述の fi タグの付与が行われている．例えば，「今夜は焼肉パーティー。」のような文に対しては，「焼肉パーティー。」に fe タグが付与されるとともに，fi タグも付与される．

属性

fact（任意）：当該要素が食品の摂取を明示しているのか非摂取を明示しているのかを区別するための属性．デフォルトでは食品の摂取に言及しているとし，この属性は明示しないが，食品の非摂取に言及している場合は，neg という値で非摂取を明示化している．

直下に存在する子要素

文字列

注）　子要素なしの場合もある．

３．１２　ei 要素

説明

ライフログ記事中に記述され1運動名を表す．構文解析システム KNP で解析した文節を単位としているため，運動名を表す名詞句に連続して助詞や句読点も要素中には含まれる．

また特殊な例として，特定の運動を実行した（または実行していない）ことが分かる名詞句などには，後述の ee タグの付与とともに，ei タグの付与も行われている．例えば，「歩く」の場合，

<ei><ee>歩く</ee></ei>

となる．また「参加したのは水泳大会。」のような文に対しても，「水泳大会。」に ee タグが付与されるとともに，ei タグが付与される．

属性

なし．

直下に存在する子要素

ee

文字列

３．１３　ee 要素

説明

ライフログ記事中に記述された1回の運動イベントを表す．構文解析システム KNP で解析した文節を単位としているため，食事イベントを表す単語列（「だらだら。」のようなオノマトペも含む）に連続して句読点も要素中には含まれる．また「する」や「なる」のように単独で意味がわからないものでも，運動イベントの発生に言及している場合はこの要素となる．タグ付与の方針として，「実行した運動が記述されている文」には必ず ee タグを付与しているため，運動イベントの発生を明示する文節が存在しない場合，文末に空要素の ee タグ（<ee/>）を付与している．

また特殊な例として，特定の運動を実行した（または実行していない）ことが分かる名詞句などには，前述の ei タグの付与とともに，ee タグの付与が行われている．例えば，「歩く」の場合，

<ei><ee>歩く</ee></ei>

となる．また「参加したのは水泳大会。」のような文に対しても，「水泳大会。」に ee タグが付与されるとともに，ei タグも付与される．

属性

fact（任意）：当該要素が運動の実行を明示しているのか非実行を明示しているのかを区別する

ための属性．デフォルトでは運動の実行に言及しているとし，この属性は明示しないが，運動の非実行に言及している場合は，neg という値で非実行を明示化している．

直下に存在する子要素

文字列

注） 子要素なしの場合もある．

３．１４　q 要素

説明

ライフログ記事中に記述された食事や運動の量を表す．構文解析システム KNP で解析した文節を単位として，記事内で書き手が摂取した（または摂取しなかった）食品の量や実行した（またはしなかった）運動の量を記述している文節に付与される．複合的な表現は1つにまとめてタグが付与されている場合もある（例えば「50回3セット」）．

属性

なし．

直下に存在する子要素

fi

fe

ei

ee

文字列

３．１５　t 要素

説明

ライフログ記事中に記述された食事や運動の時刻を表す．構文解析システム KNP で解析した文節を単位として，記事内で書き手が実行した（またはしなかった）食事イベントや運動イベントが発生した時刻を記述している文節に付与される．時刻を表す抽象的な食品名や運動名には，fi や ei とともに，t タグも付与されている（例えば「昼ご飯」や「朝練」が該当）．

属性

なし.

直下に存在する子要素

q

fi

fe

ei

ee

文字列

謝辞

このコーパスは，革新的イノベーション創出プログラム（COI STREAM）「活力ある生涯のための Last5X イノベーション」の支援を受けて作成されたものです．

運動アドバイスの付与には，フィットネスクラブ COSPA の協力を受けました．COSPA のインストラクターの方々の真摯な御対応に，心より感謝いたします．

文献

栗村 誉, 岡 照晃, 荒牧 英治, 河原 大輔, 黒橋 禎夫, 行間を読む健康アドバイス生成システムの実現に向けて. 2015. 情報処理学会 第 223 回 自然言語処理研究会 情報処理学会研究報告 Vol. 2015-NL223 No. 13.

栗村誉, 岡照晃, 荒牧英治, 河原大輔, 黒橋禎夫, ユーザのライフログに対する健康アドバイスの自動生成. 2016. 言語処理学会第 22 回年次大会発表論文集 (NLP2016).

岡照晃, 栗村誉, 荒牧英治, 河原大輔, 黒橋禎夫, おしゃべりけんこうノート：管理栄養士・インストラクターのアドバイスに基づく健康アドバイスシステム, 言語処理学会第 22 回年次大会発表論文集 (NLP2016).

Tetsuaki Nakamura, Takashi Awamura, Yiqi Zhang, Eiji Aramaki, Daisuke Kawahara and Sadao Kurohashi, Toward an Advice Agent for Diet and Exercise Based on Diary Texts, In proceedings of 2015 AAAI Spring Symposium Series-Ambient Intelligence for Health and Cognitive Enhancement.