

弊社各種コーパスについて、内容や定価のリストを、下記の通りご案内致します。

【(株)タイムヒル◆オリジナル制作・販売コーパス】(2016年秋版)

※それぞれ「非独占的著作物使用許諾(商用可)」契約での販売となります。

※転売はできません。再販目的のご購入については、案件ごとにご相談ください。

※定価はお客様のご購入目的や弊社とのお取引状況により、変更する場合がございます。

※下記コーパスの部分購入や、様々な加工を付加するサービスも行っております。

※従来どおり、請負制作でのコーパス構築も承ります。

コーパスは、大きく分けて、[自由発話多言語音声コーパス]と[多言語対訳テキストコーパス(翻訳コーパス)]に分かれます。前者を「大音泉(だいおんせん)」、後者を「大文嶺(だいぶんれい)」という製品名で販売しております。

これらについて、1) 定価 2) 共通仕様 3) 個別仕様、を下記に記します。

1) 価格(定価)について

① 「大音泉」(自由発話多言語音声 corpus)

【高音質音声コーパス】

●音声データのみ:

@ 5,000 円/1 話者 ※以下型番 01-04

@ 8,000 円/1 対話 ※以下型番 05-11

●音声データ+書き起こしテキスト:

@12,000 円/1 話者 ※以下型番 01-04

@20,000 円/1 対話 ※以下型番 05-11

【大規模音声コーパス】

●音声データのみ:

@ 30 円/1 発話 ※以下型番 12-17

●音声データ+書き起こしテキスト:

@100 円/1 発話 ※以下型番 12-17

② 「大文嶺」(多言語対訳テキスト corpus)

●日本語+英語+他1言語につき(3言語パラレル)

※日本語と英語のセットは無償提供です。

@ 30 円/1 論理行(3対訳)※以下型番 01

@ 50 円/1 論理行(3対訳)※以下型番 02

@ 250 円/1 論理行(3対訳)※以下型番 03

2) 共通 (全型番) 仕様について

① 「大音泉」(自由発話多言語音声 corpus)

【高音質音声コーパス】

〈標本化レート〉 48kHz/16bit

〈入力デバイス〉 SonyC357/ShureSM10

〈ファイル形式〉 .wav

〈発話スタイル〉 自由発話(事前原稿無し)

〈録音場所環境〉 ナレーションスタジオ

〈書起及びタグ〉 有り(フィラー等タグ含)

〈収録映像付き〉 一部有り

〈話者誓約書類〉 有り(生年月/生育地/両親母語等の話者情報及び商用利用&公開許諾)

【大規模音声コーパス】

〈標本化レート〉 16kHz/16bit

〈入力デバイス〉 i-Phone 4/4s/5/5s/6

〈ファイル形式〉 .wav

〈発話スタイル〉 自由発話(事前原稿無し)

〈録音場所環境〉 野外 及び 屋内

〈書起及びタグ〉 有り(フィラー等タグ含)

〈収録映像付き〉 無し

〈特徴〉 実際の旅行者などにその発話場面現場で収録を行ったもので、時間的制約などから、話者によって発話数が大きく異なる。

〈話者誓約書類〉 有り(生年月/生育地/両親母語等の話者情報及び商用利用&公開許諾)

② 「大文嶺」(多言語対訳テキスト corpus)

〈ファイル形式〉 タブ区切り CSV

〈フォーマット〉 分野(三階層)-日英~各語訳

〈対訳言語種類〉 簡体字中国語~繁体字中国語~韓国語~ベトナム語~インドネシア語
~タイ語~マレー語~フランス語~ドイツ語~スペイン語~イタリア語~オランダ語
~アラビア語~ロシア語~デンマーク語

3) 個別（型番ごとの）仕様について

① 「大音泉」（自由発話多言語音声 corpus）

01)SSMJJLS

〈発話種〉 独話

〈発話長〉 15-20 分/1 話者

〈言語種〉 日本語(各地)

〈話者種〉 日本人(男女-13 歳～80 歳)

〈話者数〉 100 人

〈テーマ〉 Life Story(私の人生)

02)SSMJJFV

〈発話種〉 独話

〈発話長〉 15-20 分/1 話者

〈言語種〉 日本語(各地)

〈話者種〉 日本人(男女-13 歳～80 歳)

〈話者数〉 100 人

〈テーマ〉 My Favorite(私の好きなもの)

03)SSMAAALS

〈発話種〉 独話

〈発話長〉 15-20 分/1 話者

〈言語種〉 米国語(各地)

〈話者種〉 米国人(男女-13 歳～80 歳)

〈話者数〉 100 人

〈テーマ〉 Life Story(私の人生)

04)SSMAAFV

〈発話種〉 独話

〈発話長〉 15-20 分/1 話者

〈言語種〉 米国語(各地)

〈話者種〉 米国人(男女-13 歳～80 歳)

〈話者数〉 100 人

〈テーマ〉 My Favorite(私の好きなもの)

05)SSDJJFR

- 〈発話種〉 対話
- 〈発話長〉 20-30 分/1 対話
- 〈言語種〉 日本語(各地)
- 〈話者種〉 日本人(男女-13 歳～80 歳)
- 〈話者数〉 100 人
- 〈テーマ〉 友人同士のおしゃべり

06)SSDJJSC

- 〈発話種〉 対話
- 〈発話長〉 20-30 分/1 対話
- 〈言語種〉 日本語(各地)
- 〈話者種〉 日本人(男女-13 歳～80 歳)
- 〈話者数〉 100 人
- 〈テーマ〉 店と客のやりとり

07)SSDJJTC

- 〈発話種〉 対話
- 〈発話長〉 20-30 分/1 対話
- 〈言語種〉 日本語(各地)
- 〈話者種〉 日本人(男女-13 歳～80 歳)
- 〈話者数〉 100 人
- 〈テーマ〉 教師と生徒の指導上のやりとり

08)SSDAAFR

- 〈発話種〉 対話
- 〈発話長〉 20-30 分/1 対話
- 〈言語種〉 米国語(各地)
- 〈話者種〉 米国人(男女-13 歳～80 歳)
- 〈話者数〉 100 人
- 〈テーマ〉 友人同士のおしゃべり

09)SSDAAASC

- 〈発話種〉 対話
- 〈発話長〉 20-30 分/1 対話
- 〈言語種〉 米国語(各地)

〈話者種〉 米国人(男女-13 歳～80 歳)
〈話者数〉 100 人
〈テーマ〉 店と客のやりとり

10)SSDAAJFR

〈発話種〉 対話
〈発話長〉 20-30 分/1 対話
〈言語種〉 米国語(各地)
〈話者種〉 米国人×日本人(男女-13 歳～80 歳)
〈話者数〉 50 人
〈テーマ〉 友人同士のおしゃべり

11)SSDJAJFR

〈発話種〉 対話
〈発話長〉 20-30 分/1 対話
〈言語種〉 日本語(各地)
〈話者種〉 米国人×日本人(男女-13 歳～80 歳)
〈話者数〉 50 人
〈テーマ〉 友人同士のおしゃべり

12)SSMJJSM

〈発話種〉 独話
〈発話長〉 3～30 秒(平均 8 秒)/1 発話
〈言語種〉 日本語(各地)
〈話者種〉 日本人(男女-13 歳～80 歳)
〈話者数〉 500 人以上
〈発話数〉 20,000 発話以上
〈テーマ〉 旅行及び日常会話文

13)SSMAASM

〈発話種〉 独話
〈発話長〉 3～30 秒(平均 8 秒)/1 発話
〈言語種〉 米国語(各地)
〈話者種〉 米国人(男女-13 歳～80 歳)
〈話者数〉 3,000 人以上
〈発話数〉 100,000 発話以上

〈テーマ〉 旅行及び日常会話文

14) SSMCTCSM

〈発話種〉 独話

〈発話長〉 3～30 秒(平均 8 秒)/1 発話

〈言語種〉 台湾中国語-北京標準語(各地)

〈話者種〉 台湾中国人人(男女-13 歳～80 歳)

〈話者数〉 3,000 人以上

〈発話数〉 100,000 発話以上

〈テーマ〉 旅行及び日常会話文

15) SSMKKKSM

〈発話種〉 独話

〈発話長〉 3～30 秒(平均 8 秒)/1 発話

〈言語種〉 韓国語(各地)

〈話者種〉 韓国人(男女-13 歳～80 歳)

〈話者数〉 1,000 人以上

〈発話数〉 30,000 発話以上

〈テーマ〉 旅行及び日常会話文

16) SSMFFFSM

〈発話種〉 独話

〈発話長〉 3～30 秒(平均 8 秒)/1 発話

〈言語種〉 フランス語(各地)

〈話者種〉 フランス人(男女-13 歳～80 歳)

〈話者数〉 800 人以上

〈発話数〉 25,000 発話以上

〈テーマ〉 旅行及び日常会話文

17) SSMSSSSM

〈発話種〉 独話

〈発話長〉 3～30 秒(平均 8 秒)/1 発話

〈言語種〉 スペイン語(スペイン本国各地)

〈話者種〉 スペイン人(男女-13 歳～80 歳)

〈話者数〉 800 人以上

〈発話数〉 25,000 発話以上

〈テーマ〉旅行及び日常会話文

② 「大文嶺」(多言語対訳テキスト corpus)

01) TBSMJACC~CT~KK~VT~IN~TH~ML~FF~GG~SS~II~DD~AR~RS~DM

〈論理行文単位〉フレーズ、短センテンス

〈論理行文数量〉20,000 行以上

〈単位文和字数〉平均 10 文字

〈分野文体特徴〉メール/メッセージ/LINE/SNS 等の短文

02) TBSSJACC~CT~KK~VT~IN~TH~ML~FF~GG~SS~II~DD~AR~RS~DM

〈論理行文単位〉センテンス

〈論理行文数量〉100,000 行以上

〈単位文和字数〉平均 20 文字

〈分野文体特徴〉ビジネスメール、旅行、医療などの現場での会話文

03) TBBLJACC~CT~KK~VT~IN~TH~ML~FF~GG~SS~II~DD~AR~RS~DM

〈論理行文単位〉ブロック単位

〈論理行文数量〉10,000 ブロック以上

〈単位文和字数〉平均 100 文字

〈分野文体特徴〉メール/メッセージ/LINE/SNS 等の文脈有る文章

(株)タイムヒル~IoT/M2M 事業・ASR/MT 事業・Global Research 事業・飲食/観光事業

日本本社：〒606-0912 京都市左京区松ヶ崎呼返町4-2

海外支店：シンガポール

海外拠点：アジア、アフリカ、中東、欧州、豪州、北米、中南米30カ国以上

本社電話：+81-75-702-8852 (日本国内からは075で始めて下さい)

時岡携帯：+81-90-6608-2956 (日本国内からは090で始めて下さい)

メール：tokioka@timehill.biz

ホームページ：<http://www.timehill.net>

※ホームページは2016年12月初旬公開予定です。