

『現代日本語書き言葉均衡コーパス』モニター公開データ（2009年度版）
を利用した研究成果における再配布の制限について

2009. 10. 1

大学共同利用機関法人人間文化研究機構国立国語研究所言語資源研究系
特定領域研究「日本語コーパス」データ班
山崎 誠(yamazaki@ninjal.ac.jp)

0. はじめに

本文書では、「現代日本語書き言葉均衡コーパス」モニター公開データ（2009年度版）を利用した研究成果における「再配布」の考え方について説明します。

以下、本文書で「データ」というのは、「BCCWJ モニター公開データ（2009年度版）」に収録された各サンプルのデータ（プレーンテキスト、XML ファイルの両方）の全体または一部を指します。

1. 再配布の定義

「データの再配布」とは、以下のような行為を指します。これらは例ですので、再配布に当たるかどうか判断に迷う場合は、お問い合わせください。

- (1)DVD の貸与や譲渡、売却
- (2)データ（の一部）を複製して他人に渡すこと
- (3)データ（の一部）を公衆送信可能な状態にすること（インターネット上での公開）
- (4)データ（の一部）を含むプログラムを公衆送信可能な状態にすること
（インターネット上での公開）
- (5)データ（の一部）を含む研究成果を発表すること（→具体的には、以下の4を参照）

2. データの格納場所

データは、ディスク上の以下の2個所のフォルダに格納されています。

- ・ BCCWJ2009_DATA
- ・ Himawari_1_3b03_BCCWJ

『ひまわり』のフォルダにもデータが格納されています。上記『ひまわり』のフォルダをコピーして他人に渡す等の行為もデータの再配布に当たりますので特にご注意下さい。

3. 再配布の禁止

大前提として、データそのものの再配布はできません。「再配布不可」という制限は、現時点での一時的なものではなく、「現代日本語書き言葉均衡コーパス」が完成したあとで予定している利用形態においても継続します。

4. データを利用した研究成果

データを利用した研究成果を発表する場合においても、データの再配布と認められる形態での発表はできません。例えば、データに独自のタグを付けて再配布するとか、データを組み込んだプログラムを再配布することなどが該当します。また、発表の主体を問わず、複数の発表を組み合わせた結果、再配布になるような形態での発表はできません。

発表形態・成果物の種類と発表の可否との関係を以下に例示します。以下に記載の無いものについては、必ずお問い合わせくださるようお願いいたします。

特定の語・文字の用法記述	○
語彙表・漢字表（注1）	○
Nグラム（注2）	○
ある語に対する共起語のリスト	○
KWIC（注3）	×
タグを追加したファイル	×
データを組み込んだプログラム	×

注1 語種，品詞，語構成，文字種の統計を含む。文脈をもたないもの。

注2 文字のNグラムは7文字以下。単語Nグラムは短単位で5単位以下。

いずれも高頻度順に並べた場合の順位で上位50%以内にとどめること。

注3 並べ替えることにより，サンプルがまるまる復元できるものが対象。

5. 共同研究における再配布

科研費などの共同研究でデータを利用する場合は，お手数ですが，メンバー全員が利用申請を行ってくださるようお願いいたします。

利用申請が認められた構成メンバー間であってもデータの再配布はできません。ただし，共同研究の計画遂行上の必要から，構成メンバーの間で「データの加工物」を共有して利用する場合は，以下の(1)(2)(3)の条件すべてを満たす場合に限り，他の構成メンバーに対して「データの加工物」を再配布することができます。

(1)期間が限定されている共同研究であること。

(2)「データの加工物」を共有するメンバーが共同研究の参加者であることが確認できること（科研費の分担者等）。

(3)対象者が多人数でないこと。

なお，上記の再配布を行った場合，以下の(4)～(6)についても順守してください。

(4)研究に差し支えない範囲で再配布の事実（研究課題名，「データの加工物」の作成者，再配布の対象者，「データの加工物」の概要）を下記の問い合わせ先に連絡すること。

(5)共同研究の期間が終了したら，すみやかに再配布を受けたデータの加工物を消去すること。

(6)複数の共同研究グループに所属している場合，ひとつの共同研究グループにおける「データの加工物」を他の共同研究グループの構成メンバーに再配布しないこと。

「データの加工物」の例は，前項4「データを利用した研究成果」で×の付いているものです。

6. 問い合わせ先

〒190-8561

東京都立川市緑町 10-2

国立国語研究所

「書き言葉コーパス」担当

電話：042-540-4300（代表）

FAX：042-540-4333（代表）

e-mail:kotonoha@ninjal.ac.jp

以上