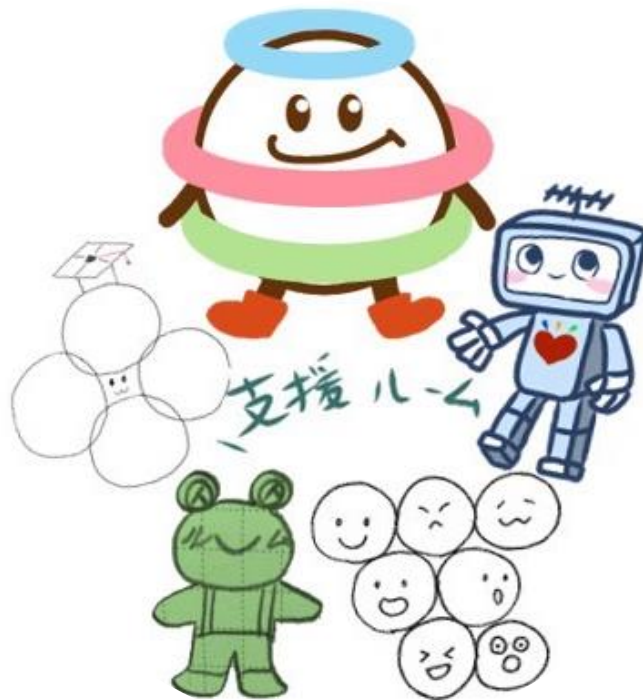


テキストデータ化 マニュアル

ver.2.0



2023年10月16日 小馬 作成

音声読み上げデータは、「見てわかりやすい」ものではなく、「読み上げてわかりやすい」ものを
目指すことが目標です。音声ソフトでどのように読まれるのかを考えながら作業しましょう。

このテキストデータをもとに、授業を受講したり、課題を提出したりします。表の表記や、挿入
されている画像などがどのように読まれるのか、漏れなくテキストデータになっているのか、
ズレや誤字がないか等に注意しながら進めてください。

校正にあたる際の大原則は「原本に忠実に」です。

途中で判断に迷うことや、分からないことがあれば遠慮なく支援ルームに相談してください。
また、このマニュアルの手順や、画像は作成当時のものです。それぞれの環境では別の表示に
なったり、もっと良いやり方が見つかったりするかもしれません。適宜修正や追記を行うので
ぜひ、報告・共有をしてください！よろしくお願いします！

作業工程

事前準備

- PDFから
- Word／Excelから

校正作業

- 文字の校正
- 表の修正
- 図／画像の注釈

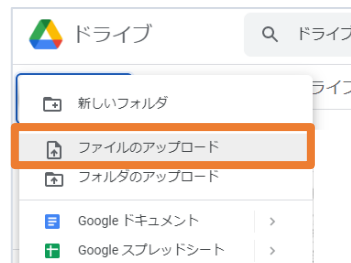
確認

- ページ数の表記
- 最終確認
- 報告・謝金処理

【事前準備:PDF から】

PDF 等の画像上の文字をデジタルの文字にする技術が OCR です。今回は Google ドライブを使った方法を紹介します。

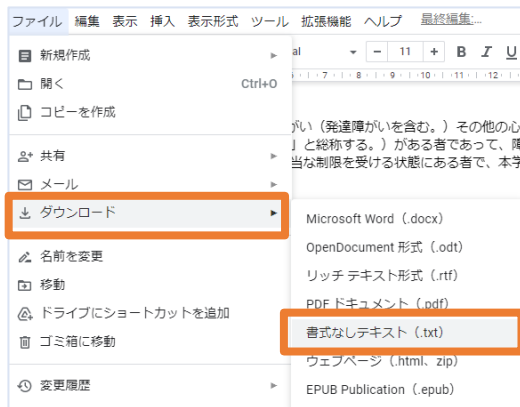
- ① PDF データをグーグルドライブにアップロードする。



- ② アップロードした PDF を開き、上部にある「Google ドキュメントで開く」から文字化する。



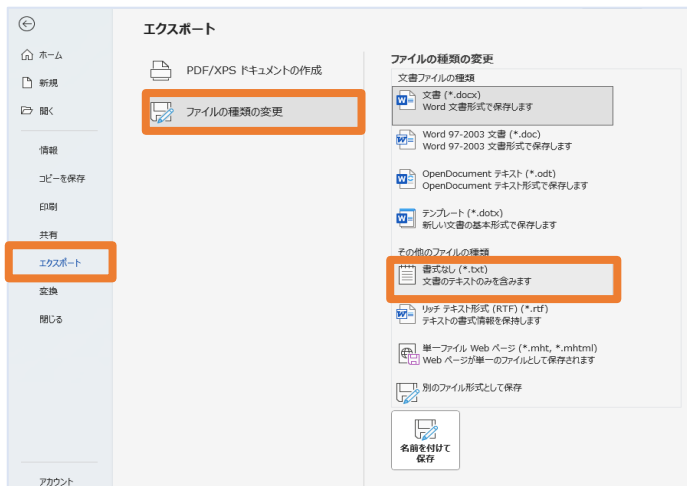
- ③ 開いた Google ドキュメントの
ファイル>ダウンロード
 >書式なしテキスト(.txt)
でテキストファイルで保存する。



【事前準備:Word や Excel から】

図表を PDF で貼っていたり、テキストボックスを多用したりしている Microsoft データは一度 PDF にしてから OCR でテキストデータ化の方が効率的だと思います。コピーしたファイルで作業し、元データは必ず残しておきます。

- ① ファイル>エクスポート
 >ファイルの種類の変更
 >書式なし(*.txt)
でテキストファイルで保存する。



事前準備 完了！

【校正作業:文字の校正】

テキストデータ化の一番大切なところです。

なお、元原稿にない文章の挿入は、注釈とわかるように(校注:)で記入するようにしてください。

- ① 人間は文脈で判断できる文字の差異も、OCR で処理した文字は似ている形の文字を混同してしまうことが多々あります。特に、手書き文字や荒いスキャンだと難しいです。

そのままでは、意味が通じませんし、読み方にも支障が出てきてしまうので、できる限り元の文章と見比べて校正を行います。

例:「タ方(ゆうがた)」→「タ方(たほう)」

間違いやすい文字(追加するのでぜひ報告してください！)

び	ぴ	人	入	待	侍
る	ろ	日	目	東	束
1(える)	1(いち)	O(おー)	0(ゼロ)	伸	仲

など。

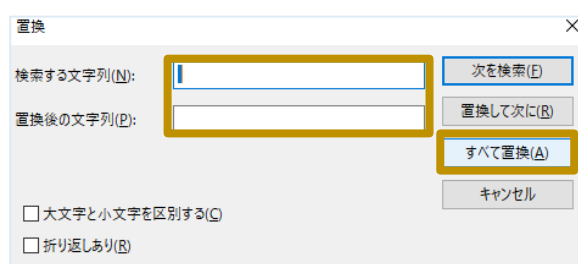
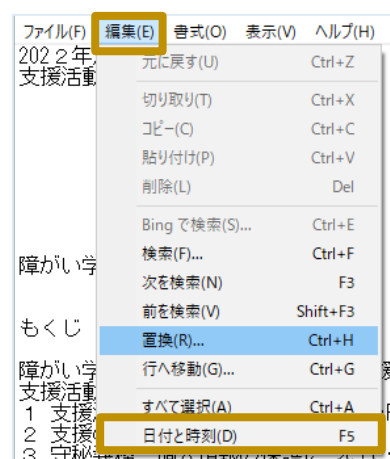
- ② 空白、意図しない改行の削除

OCR で処理した場合、意図しない場所にスペースが入りやすいです。また、データから取り込んでも、体裁を整える目的でスペースが入れられていることもあります。正しく読み上げられないため、 unnecessary スペースは消去するようにお願いします。

その際には、一つ一つ消すよりも、

「編集>置換」で「検索する文字列」にスペース、「置換後の文字列」には何も入力せずに「すべて置換」を行うと、空白がすべて消えるので便利です。

(ただし、空白が必要なところも消えてしまうので注意)



③ 半角・全角の統一

数字の表記は全角と半角でバラバラになってしまうことが多いです。読み上げの都合上、大きな意図がない場合、数字・英語はすべて半角で統一をお願いします。

その際には、②で説明している「置換」を使うと便利です。

④ 記号や省略記号の校正

記号は音声読み上げソフトによって、読み上げ方が異なります。また、省略して表記すると意図しない伝わり方になることもあるので修正を行います。

⑤ 読みの補記

人名などの固有名詞や英語読みなどは、思ったように読まれないことが多々あります。

読み上げソフトにもよるので、すべてを補完するのは難しいですが、誤読するであろうとわかるものについては(校注:____)で読みを補記するようにしてください。

省略・読みの補記について繰り返し登場するものについては、最初に登場するときに正式名(よみ)で表記、その後は読みのみの表記でもかまいません。

例)大教大(おおきょうだい)と読まれることがある。

1回目「大教大(校注:だいきょうだい)～」 2回目以降「だいきょうだい～～」

	本文	読み上げ(例)	修正後	修正後読み上げ
記号	※1	きごういち	注1	ちゅういち
	¥5,000 (全角表記)	ごぜろぜろぜろ	5,000 円 (半角表記)	ごせんえん
表記	2023/04/01	よんぶんのにせんにじゅうさんいち	2023 年4月1日	にせんにじゅうさんねんしがつついたち
	14:30	Fourteen thirty	14時30分	じゅうよじさんじゅつぷん
	DAISY 図書	でいーえーあいせすわいとしよ	デイジー図書 DAISY (でいじー)図書	でいじーとしよ
	日 時 (スペースあり)	ひ と き	日時	にちじ
省略	3日(月)	みっかつき	3日(月曜日)	みっかげつようび
	R5	あーるご	令和5年	れいわごねん

【校正作業:表の修正】

表は読みとり方が一定ではなく、ソフトではうまく読まれないことが多いもののひとつです。
表のはじめと終わりに、「(校注:表はじめ)(校注:表終わり)」の表示を入れ、それぞれの読み方に合わせて表記を工夫します。

なるべく、1 行に 1 項目になるようにしていきます。

(例)

買い物メモ			
品名	個数	値段(/1)	買った
りんご	2 個	¥150	✓
卵	6 個入り	¥100	
	8 個入り	¥130	✓
牛乳	1L	¥250	✓

このまま読み上げをすると…(漢字で表記してます)

買い物メモ 品名 個数 値段 いち 買った りんご にこ ひゃくごじゅう 卵 ろっ
こいり ひゃく はっこいり ひゃくさんじゅう 牛乳 one litre にひゃくごじゅう

のように、左上から右に読み上げになってしまうので、
テキストデータでは、このように表記すると、そのまま読み上げられるため、分かりやすい。

買い物メモ

(校注:表はじめ)

品名 りんご 個数 2 個 1 個当たりの値段 150 円 買った

品名 卵 個数 6 個入り 値段 100 円 買ってない

品名 卵 個数 8 個入り 値段 130 円 買った

品名 牛乳 個数 1リットル 値段 250 円 買った

(校注:表終わり)

【校正作業：図／画像の注釈】

パワーポイント資料やポスターなどは特に、表とは違い、図や画像が多く用いられていることが多い。

そのままでは情報が伝わらないので、表記の仕方を工夫したり、代替テキストを入れたりすることで対応する。

「(図)」、「(写真)」、「(イラスト)」を適宜入れ、どこに配置しているのかも表せると Good！

(例)修学支援ルームの学生作成のポスター

テキストデータの一例



(ポスタータイトル)

支援の輪を広げたいわ!! 学生スタッフの募集

(校注:左上の図)

つながりづくりを中心に、学びたいわ、目安箱カフェ、支援の拡大、目安箱がつながった図

(校注:図の右側に説明)

学びたいわ(勉強会) 月に1,2回障がいについての情報を共有しています。(気軽に参加していただいてOK!)

目安箱 気軽に相談できる場所を目指して、皆さんから寄せられた困りごとについて一緒に考えています!

目安箱かふえ〜(新聞) ちょっとした雑談を書いた親しみやすい新聞を発行中(不定期)

(校注:写真)交流スペースで学生が集う様子

(校注:写真の横で障がい学生修学支援ルームマスコットキャラクターのここわのイラストと、そこから吹き出し)Twitter、Instagramで活動日などを発信しています。ぜひ一度見学にお越しください!

校正作業 完了!

【確認: ページ数の表記】

元データを参考に、紙資料ならページ数、スライドならスライド番号を「(●ページ)、(スライド○)」などで入れる。キリが悪くても、ページが変わったときに入れる。ページが変わってすぐに入れる。

表などが複数ページにまたがる場合には、まとめて表記しても良い。

そのほか、大きくレイアウトがかわったりした際にも、注記を行えると Good !

表記の仕方は数字が最初。

紙資料	(1枚目)
本資料の場合はページ数	(12ページ)
パワーポイント資料	(1スライド)
複数ページにまたがる時	(3～4 ページ)

(例)

～～障害のある学生は、障がいの種類やその程度により個
(2 ページ)
人差が大きく、一人一人に様々なニーズがあります。このため～～

【確認: 最終確認】

- ☐ 誤植・脱字を確認した。
- ☐ 意図しない改行や空白がない。
- ☐ 数字や英語がすべて半角表記になっている。
- ☐ 省略・記号・英語・固有名詞が正しく表記されている。
- ☐ 図表、画像の注釈や修正を行い、その位置がずれていない。
- ☐ ページ数やスライド番号が表記されている
- ☐ テキストデータ形式(.txt)形式で保存されている

ここまでばっちりできたら、作業完了！

おつかれさまでした！

確認 完了！

資料の作成例① 表の資料

令和4年度 授業時間割

第1時限 (8:50~10:20)										第2時限 (10:25~12:05)									
授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属
月																			
火																			
水										C02014	病弱者の心理・生理・病理 2 (後)	平塚 大智	08-203						
木										C02006	視覚障害教育 2 (特)	正井	A-215						
金										C02016	発達障害指導論 2 (特)	湯浅	08-203						
										C02018	聴覚障害教育 2 (特)	井坂	08-203						
										C02001	特別支援教育総論 2 (後)	富永光	08-203						

第3時限 (12:35~14:25)										第4時限 (14:40~16:10)									
授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属
月										C02005	病弱教育 2 (特)	富永光	A-307						
火																			
水																			
木										C02007	発達障害教育 2 (特)	森田	A-216						
金										C02002	聴覚障害心理 2 (特)	西山	08-203						
										C02015	障害者の心理・生理・病理 2 (後)	湯浅	08-203						
										C02011	障害児研究法 B 2 (後)	金島	C7-101						

第5時限 (16:25~17:30)									
授業コード	授業科目名	等	教員名	所属	授業コード	授業科目名	等	教員名	所属
水	C02020	知的障害教育 2 (特)	今村	08-203					

(備考)

令和4年度 授業時間割

(前期)

月曜日

第1時限(8時50分から10時20分):授業なし

第2時限(10時35分から12時5分):授業なし

第3時限(12時55分から14時25分):授業なし

第4時限(14時40分から16時10分)

授業コード:C02005 講義名:病弱教育 単位数:2 教員名:富永光 講義室:A-307

火曜日

第1時限(8時50分から10時20分):授業なし

第2時限(10時35分から12時5分):授業なし

第3時限(12時55分から14時25分):授業なし

第4時限(14時40分から16時10分):授業なし

水曜日

第1時限(8時50分から10時20分):授業なし

第2時限(10時35分から12時5分)

変換者が表を見て、月曜1限はないから入力しなくていい！とすると、テキストデータだけを見た時にもともとデータがないのか、それとも抜け落ちてしまったのかが判断がつかないです。「ない」ことを伝えるのも重要です。

学生生活案内目次

	Page
はじめに	1
お知らせ	2
注意喚起	5
学内規則の遵守	
学生相談	7
01. 学生生活に関する窓口	9
02. 諸手続	11
03. 証明書の発行	
04. 学生証	13
05. 授業料	14
06. 授業料免除等	
07. 奨学金	15
08. 課外活動・その他施設	17
09. 附属図書館	19
10. 就職支援	20
11. 教育研究施設	21
12. 学生宿舎・下宿	22
13. 学生教育研究災害傷害保険(学研災)・ 学生教育研究賠償責任保険 (学研賠)	23
14. アルバイト	
15. 海外留学・国際交流	25
16. ボランティア活動	26
17. 大阪教育大学校友会	
18. 大阪教育大学教育振興会	27
19. 国民年金学生納付特例制度	28
20. 懲戒の標準	
国立大学法人大阪教育大学の 基本理念と基本目標	29
沿革	30
大阪教育大学学生規則	31
大阪教育大学学生懲戒規則	34
防災カード	巻末
インフォメーション	
緊急時の大学連絡先等	
学外の相談窓口	
学生生活に関すること	

はじめに

大学から学生への連絡等について

- 大学が学生に対して行う通知及び連絡は、Live Campus (LCU)・大教 Gmail (Google Workspace)・Moodle 及び ANPIC (災害情報)を使用します。
- 授業に関する連絡は、Live Campus (LCU)により行います。
- 構内放送による学生呼び出しは行いません。
- 郵便物や荷物は大学で受け取ることはできません。学生個人宛の郵便物等を、大学の住所に送らないでください。

掲示板の場所について

- 柏原キャンパス：
 - ・M 棟(大会館)：
 - 〈2F〉奨学金・授業料免除・就職・学研災保険・教育振興会等
 - 〈1F〉課外活動・ボランティア・同窓会等
 - ・N 棟(事務局棟)：〈3F教育実践支援ルーム前〉教育実習・介護等体験・学校インターンシップ・学校サポート活動
- 天王寺キャンパス：中央館1F

車両による構内への入構

- 柏原キャンパス
 - 車・バイク等で入構するには「許可」が必要です。許可された場合は「車両入構許可証」が交付されます。
 - 学生支援課学生企画係で手続きをしてください。
- 天王寺キャンパス
 - 車・バイクでの入構は原則禁止です。
 - 自転車で入構の場合は正門守衛室で「車両入構許可証」交付手続きをしてください。

正しい交通ルール・マナーを守り安全運転を心がけましょう。

大阪教育大学安否確認システム (ANPIC) について

大阪教育大学安否確認システム (ANPIC) は、大地震を始めとする大規模災害が発生した際に、本学の学生、教職員へ迅速に災害情報を発信し、受信者の安否状況の報告により安否確認を行うものです。大阪府及び隣接する府県で震度5以上の地震発生を気象庁が発表した際、安否確認のお知らせが LINE 等(事前に登録が必要)に自動で配信されますので、安否報告を必ず行ってください。QRコードを読み取れない場合は、以下の URL を入力してください。



<https://anpic-v6.jecc.jp/osaka-kyoiku/login/>

本人または父母等の、氏名・本籍・住所・電話番号を

変更した時の届出について

学生支援課奨学厚生係(天王寺キャンパス所属の学生は天王寺地区総務課)に届け出てください。手続には学生証が必要です。氏名の変更及び本籍の変更には、変更の事実を確認できる書類を持参ください。(※生協に届けただけでは大学の登録は変更されません。変更の際は「大学」と「生協」両方に届け出てください。)

(1 ページ)

(左側)

学生生活案内目次

はじめに 1ページ

お知らせ 2 ページ

注意喚起 5 ページ

学内規則の遵守 5 ページ

学生相談 7 ページ

01. 学生生活に関する窓口 9 ページ

02. 諸手続 11 ページ

03. 証明書の発行 11 ページ

04. 学生証 13 ページ

05. 授業料 14 ページ

06. 授業料免除等 14 ページ

07. 奨学金 15 ページ

08. 課外活動・その他施設 17 ページ

09. 附属図書館 19 ページ

10. 就職支援 20 ページ

11. 教育研究施設 21 ページ

12. 学生宿舍・下宿 22 ページ

13. 学生教育研究災害傷害保険(学研災)・ 学生教育研究賠償責任保険(学研賠) 23 ページ

14. アルバイト 23 ページ

15. 海外留学・国際交流 25 ページ

16. ボランティア活動 26 ページ

17. 大阪教育大学校友会 26 ページ

18. 大阪教育大学教育振興会 27 ページ

19. 国民年金学生納付特例制度 28 ページ

20. 懲戒の標準 28 ページ

冊子のページ数に対応するように、
テキストデータにも目次を入れます。
(冊子のページが途中から始まる場合は、
・表紙 ・表紙裏 ・次ページ
等を組み合わせながらページが変わったこ
とを伝えてください。

(右側)

はじめに

大学から学生への連絡等について

●大学が学生に対して行う通知及び連絡は、ライブキャンパス(LCU)・だいきょう G メール(グーグルワークスペース)・ムードル 及びアンピック(災害情報)を使用します。

●授業に関する連絡は、ライブキャンパス(LCU)により行います。

資料の作成例③ ポスターの資料



The poster is titled "サポート学生 募集!" (Support Student Recruitment!). It features a small robot character on the right side. The text includes the dates "4月25日、5月12日" and the time "12:20から", along with the location "柏原C 支援ルームにて 登録説明会開催!". Below this, a question is posed: "「すべての学生が、共に気持ちよく学び合う活動」に参加しませんか?". A row of five circles contains the following text: "PC テイク", "字幕 挿入", "音声 データ化", "移動 サポート", and "学生 活動". A paragraph of text follows, explaining the purpose of the support room and the recruitment process. At the bottom, there are QR codes, a Twitter/Instagram handle "@oku_syugakusien", and contact information for the support room, including an email address and phone numbers. A large, stylized character is also present at the bottom left.

わかる範囲でキャラクター名等は補足を入れます。ただし、主観が入りすぎないように注意しましょう。

また、QR コードが入っている場合には、URL に変換するようにしましょう。

(方法が分からないときは報告の時に伝えてください。)

(タイトル)

サポート学生募集！

(校注:タイトルの右側にロボットの支援ルームキャラクター:むーむのイラストがありその上に支援ルームという掛け声)

(校注:ロボットのイラストの左側に説明)

4月25日、5月12日12時20分から柏原C(キャンパス)支援ルームにて登録説明会開催！

「すべての学生が、共に気持ちよく学び合う活動」に参加しませんか？

PC テイク、字幕挿入、音声データ化、移動サポート、手話、学生活動

(校注:それぞれが丸で囲まれていてその中の字幕挿入のところに手を添えたカエルの支援ルームキャラクター:かえるーむのイラスト)

本学では、すべての学生が同じように教育を受けることができるよう取り組んでおり、障がいのある学生への修学支援活動を行う学生スタッフを募集しています。(時給1,100円)サポーター登録の後、それぞれの研修を受けることで支援活動に参加できます。学生活動や勉強会への参加も大歓迎！登録や活動については遠慮なく支援ルームにお問い合わせください。

登録は登録フォームまたは直接支援ルームへ

<https://forms.gle/xxbziHGwVgXog8ZW8>

(校注:この文章の右側に QR コードがありその右側に学生帽をかぶった支援ルームキャラクター:フォーレのイラスト)