



Línguas indígenas no mundo digital

Inventário de recursos
e lacunas

OEI

Línguas indígenas no mundo digital. Inventário de recursos e lacunas

Autor: Daniel Prado

© Organização dos Estados Ibero-Americanos para a Educação, a Ciência e a Cultura (OEI)

Bravo Murillo, 38
28015 Madrid, Espanha

www.oei.es

Secretário-Geral: Mariano Jabonero

Direção-Geral do Multilinguismo e Promoção da Língua Portuguesa e Espanhola: Ana Paula Laborinho

Coordenação: Mónica García

Apoio técnico: Rut Sánchez

Edição e paginação: OEI

Relatório final: 15 de abril de 2023

Edição digital: 21 de fevereiro de 2024

O material desta publicação foi compilado a partir das actividades realizadas no âmbito do projeto "Celebrar as Línguas". As opiniões expressas neste documento são da exclusiva responsabilidade dos seus autores. Os editores e patrocinadores desta obra não estiveram envolvidos na sua elaboração e as suas opiniões não correspondem necessariamente às expressas nas várias secções desta obra.

Índice

Índice de conteúdos	2
Contextualização.....	3
Inventário dos recursos e das lacunas	4
Inventário dos recursos digitais	6
Considerações iniciais	7
Sobre a classificação dos dados estudados.....	7
Relativamente a algumas das rubricas estudadas	8
Relativamente ao material de estudo	11
Resultados do inventário dos recursos.....	13
Análise do inventário.....	14
Principais aplicações.....	22
Análise de lacunas digitais.....	26
Introdução	26
Resultados do inventário de lacunas	27
Principais aplicações.....	27
Recursos digitais na Web e em aplicações móveis.....	28
Panorama das lacunas identificadas	31
Inventário de entidades ou iniciativas	32
Introdução	32
Resultados do repertório.....	32
Relevância das entidades estudadas no seu conjunto.....	33
Natureza das entidades estudadas.....	34
Âmbito territorial das entidades estudadas	34
Atividade principal ou tema das entidades	34
Recomendações finais.....	36
Recomendação 1: Acesso à Internet e ao telemóvel.....	36
Recomendação 2: Literacia digital da população indígena.....	37
Recomendação 3: Educação e formação.....	37
Recomendação 4: Identificação, formação e apoio aos activistas digitais	37
Recomendação 5: Participação ativa das comunidades	38
Recomendação 6: Fornecimento de material para a digitalização.....	38
Recomendação 7: Sectores prioritários para a digitalização.....	39
Recomendação 8: Ação prévia para as línguas menos utilizadas.....	40
Conclusões.....	41



Contextualização

Durante séculos, as línguas indígenas do continente foram ofuscadas, ao ponto de serem invisíveis na vida quotidiana, com exceção da prática oral dos seus falantes ou, mais especificamente, dos vários tratados de investigação sobre as culturas indígenas. Para além desta indiferença em relação às línguas indígenas, a urbanização, as migrações, a destruição de habitats, a falta de inclusão no sistema educativo ou nos canais de informação, entre outros factores, conduziram ao desaparecimento de muitas das línguas faladas no continente (até ao dobro, segundo algumas estimativas) e à extinção progressiva da grande maioria das restantes línguas, vítimas de uma espiral descendente em que *quanto menos visível é uma língua, menos é falada; e quanto menos é falada, menos visível é*.

"As línguas indígenas podem recuperar a visibilidade através do universo digital".

No entanto, desde há várias décadas, diversas iniciativas comunitárias, associativas, universitárias e estatais¹, têm permitido que várias línguas sejam progressivamente integradas na vida pública e educativa de várias regiões da América luso-hispânica, aumentando a sua visibilidade, ainda que timidamente para uma minoria e escassamente para a grande maioria delas.

O chegada da Internet poderia ter acelerado essa visibilidade, mas a falta de protocolos, fontes, teclados e outros elementos tecnológicos e, sobretudo, a falta de acesso à Internet e de literacia digital das populações indígenas, aumentaram o fosso entre as línguas dominantes e as línguas indígenas, tendo estas últimas sido integradas muito tarde e lentamente na rede das redes. Sabemos que a falta de conteúdos, *corpus*, de material didático e de ferramentas diversas é um travão à inclusão digital das línguas indígenas, mas não sabemos qual é a profundidade desse fosso, nem temos uma noção global do que são esses conteúdos.

Embora existam numerosos recursos digitais² sobre as línguas indígenas, a sua pesquisa é dificultada pela enorme dispersão, pela falta de sistematização e pela ausência de repositórios que permitam um acesso fácil a todas as línguas. Embora existam alguns catálogos que reúnem informações valiosas para determinados grupos de línguas, nenhum oferece um panorama geral que facilite uma política de preservação e desenvolvimento das línguas do continente. A invisibilidade das línguas indígenas na Internet é agravada pela dificuldade de acesso às suas fontes primárias.

¹ Tanto o Estado nacional como as entidades regionais (províncias, estados ou departamentos) e as entidades municipais ou locais.

² Acessíveis *online* ou simplesmente para consulta.



Inventário de recursos e lacunas

Por esta razão, foi criado um diretório de recursos linguísticos *online*, bem como realizada uma análise das lacunas digitais para todas as línguas indígenas da América Luso-Hispânica, o que permitirá emitir uma série de recomendações sobre as políticas a implementar para as integrar no mundo digital.

A proposta inclui um inventário pormenorizado dos recursos institucionais envolvidos nos programas de revitalização das línguas indígenas, a que se pode recorrer para procurar maior cooperação e implementação.

Estes elementos foram recolhidos e analisados entre julho de 2022 e abril de 2023, pelo que a informação recolhida não pretende ser exaustiva ou permanente. No entanto, a sua utilidade pode ser reforçada por uma atualização constante que permita avaliar a evolução das línguas indígenas na Internet e, assim, avaliar políticas dinâmicas para a sua integração no universo digital.

Para a compilação dos dados optou-se por trabalhar com tabelas do tipo folha de cálculo (Microsoft Excel, OpenOffice Calc, Google Sheets), compatíveis com qualquer sistema de gestão de bases de dados, de forma a facilitar uma eventual publicação *online* dos conteúdos.

Figura 1: Conteúdo dos ficheiros principais



Fonte: elaboração própria



O sistema concebido é composto por 3 ficheiros principais e 1 ficheiro de apoio.

1. Ficheiro de *recursos linguísticos*. É composto por dois quadros principais:
 - Recursos da Internet (web, redes sociais, blogues, etc.)
 - Aplicações e ferramentas digitais (sistemas operativos, burótica, aplicações móveis, etc.)
2. Ficheiro de *recursos institucionais*. É composto por um quadro principal e um quadro derivado:
 - *Entidades* (OIG, ONG, organismos públicos, sindicatos, universidades, colectividades, etc.)
 - *Individuos* que podem ser consultados no âmbito deste projeto (este quadro não será público)
3. Ficheiro de *inventário de lacunas*. Tabela de dados que mede, língua a língua, as suas lacunas relativas. Estas são avaliadas em termos de comparação com as línguas indígenas que se encontram numa situação mais favorável.
4. Arquivo de *línguas* (arquivo de apoio). É constituído por tabelas de classificação e informações sobre as línguas abrangidas.

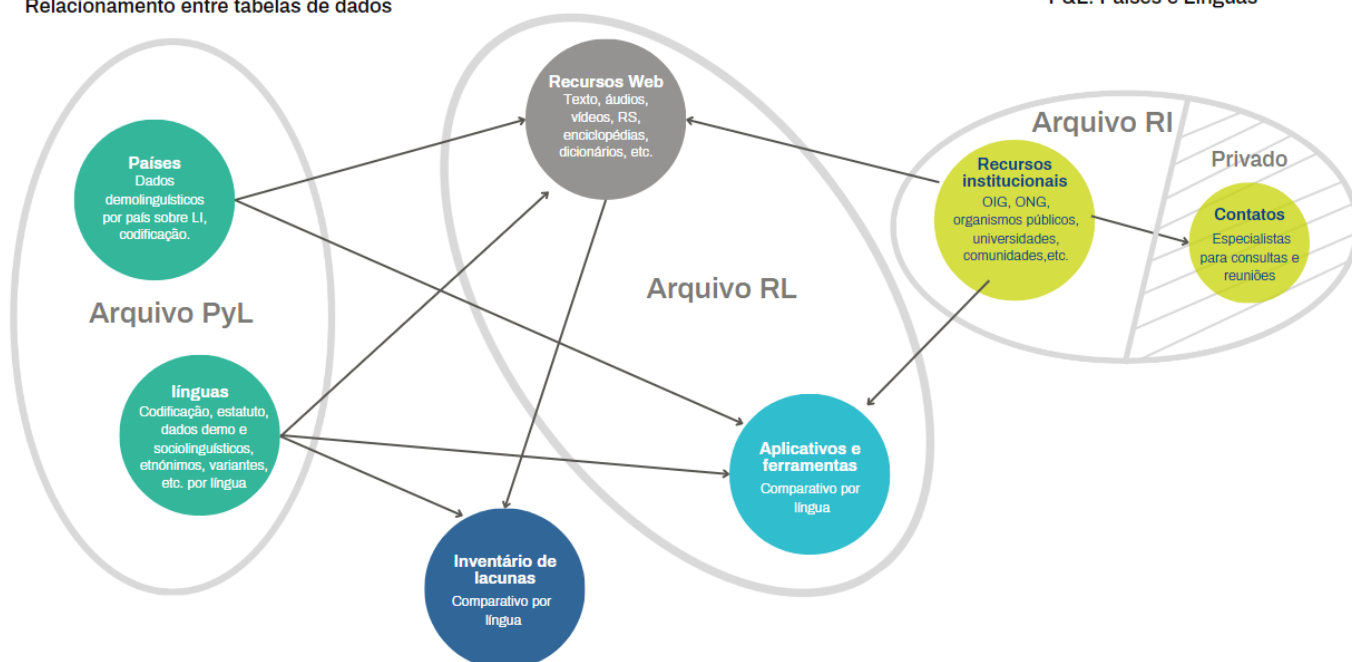
A *figura 1* especifica o conteúdo dos três ficheiros principais, separados por folhas. A *figura 2* especifica as relações entre as tabelas de dados.

Figura 2: Relações entre tabelas de dados

Gestão de dados

Relacionamento entre tabelas de dados

(*)RI: Recursos institucionais
RL: Recursos linguísticos
P&L: Países e Línguas




Fonte: elaboração própria



Inventário dos recursos digitais

A base do projeto consiste num inventário dos recursos linguísticos das línguas indígenas acessíveis na Internet ou nas suas referências³, bem como em informações sobre os mesmos para facilitar a sua utilização em qualquer política de revitalização e desenvolvimento destas línguas.

Este relatório tem como objetivo sintetizar graficamente os resultados da investigação e explicar tanto o desenvolvimento do trabalho como a metodologia aplicada. Os dados recolhidos são anexados sob a forma de uma folha de cálculo, que pode ser exportada para diferentes formatos para uma possível publicação *online* e uma fácil atualização.

-  Para maior clareza na leitura deste documento, optámos por nos referir a qualquer sítio eb, blogue, página (por exemplo, Facebook), canal (por exemplo, YouTube), conta (por exemplo, Twitter), aplicação móvel ou de ambiente de trabalho, ou qualquer outro meio digital que tenha sido inventariado como um "**recurso**". Do mesmo modo, designaremos por "**base de dados**" o registo completo dos recursos compilados.



³ Nos casos em que apenas está disponível a referência de obras editadas em papel.



Considerações iniciais

Sobre a classificação dos dados estudados

A fim de valorizar a informação recolhida sobre cada recurso e permitir uma síntese gráfica pormenorizada, os campos enumerados no *quadro 1* foram utilizados na folha de cálculo correspondente.

Quadro 1. Campos utilizados na descrição dos recursos encontrados

CAMPOS	TIPO DE CAMPO	DESCRIÇÃO DO DOMÍNIO
Nome do recurso	Texto	Nome habitual definido pelos produtores do recurso.
Código ISO	Texto	Código da língua utilizada, de acordo com a norma ISO 639-3. Este elemento é utilizado para a correcta gestão da denominação das línguas utilizadas.
Língua de chegada	Texto	Nome da língua a que o recurso se refere.
Descrição do recurso	Texto	Breve informação textual sobre o recurso descrito.
Línguas de interface	Texto	Informa sobre a língua principal de acesso ao recurso.
Atualizado (Sim ou Não)	Lógico	Se o recurso parece estar ou não atualizado ou se não existe qualquer informação disponível.
Tipo de promotor	Texto	Indicação do tipo de promotor.
Descrição / família	Lógico	Se o recurso fornecer informações descritivas sobre a língua.
Estrutura	Lógico	Se contiver informações descritivas sobre a sua estrutura (gramática, fonologia e fonética, ortografia, escrita, etc.).
Léxico	Lógico	Se contiver informações terminológicas ou lexicográficas.
Referencial	Lógico	Se contiver pormenores sobre livros, artigos, documentos, etc., em línguas indígenas ou sobre o estudo de línguas indígenas (se a informação estiver imediatamente disponível).
Textos	Lógico	Se contiver textos em línguas indígenas (documentos, literatura, técnicas, diálogos, etc.)
Áudio	Lógico	Se contiver ficheiros de som.
Visual	Lógico	Se contiver ficheiros visuais (incluindo fotografias).
Mapa / Localização	Lógico	Se contiver mapas ou informações sobre a região onde a língua é falada.
Sócio-demolinguística	Lógico	Se contiver informações sobre os seus falantes, o estatuto da língua, a utilização da língua, etc.
Ensino	Lógico	Se tratar ou contiver métodos de ensino de línguas
Ferramentas	Lógico	Se contiver ferramentas linguísticas digitais ou se é uma ferramenta linguística utilizável
Número de entradas	Numérico	Se for acessível e pertinente para o recurso, deve ser indicado o número de entradas lexicais ou terminológicas.
Interesse	Opções	Será avaliado o interesse que o recurso pode ter para os seus falantes ou para ações de promoção e desenvolvimento linguístico.
URL	Texto	Ligação à página direta do recurso
Comentários	Texto	Observações pertinentes, se necessário, sobre a utilidade ou a pertinência do recurso
Data	Data	Data de consulta do recurso
País	Texto	País indicado no recurso ou, na sua falta, o país de origem da língua em causa, tal como indicado no sítio <i>Ethnologue</i> (*).



Relativamente a algumas das rubricas estudadas

a. *Nome do recurso*

Sempre que possível, procurou-se incluir o nome exato indicado no sítio *web* ou, no caso de uma página específica para o recurso, o nome dessa página precedido do nome genérico do sítio *web*. Em muitos casos, esta informação era pouco clara, apresentava vários nomes ou era demasiado extensa. Noutros casos, faltava informação ou era difícil de encontrar. O autor teve de fazer escolhas que podem não coincidir com a intenção dos criadores dos recursos.

Se o nome do recurso for idêntico para diferentes línguas, é indicado o nome genérico do sítio *web* e acrescentada menção da língua correspondente. No caso de existirem diferentes canais de difusão do sítio *web* ou da aplicação, é indicada essa particularidade (por exemplo, página Facebook de ..., canal YouTube de ..., etc.).

b. *Codificação*

Como já foi referido, a norma ISO 639-3 foi utilizada para a estruturação final dos dados devido à sua praticidade, embora não tenha sido escolhida pela sua exatidão. É sabido que esta norma pode não satisfazer muitos interlocutores: alguns consideram que exclui certas línguas, uma vez que as classifica como variantes, enquanto outros acreditam que cria uma divisão artificial ao agrupar variantes linguísticas com códigos diferentes dentro da mesma língua. No entanto, esta norma internacional foi mantida porque é a única que permite um tratamento estatístico para uma análise global.

É importante notar que foi necessário tomar decisões para classificar algumas línguas:

- 🔍 Quando um recurso se refere a uma língua que não existe na norma ISO, optou-se por utilizar o código pertencente ao seu antepassado mais próximo (língua ou macro-língua).
- 🔍 Quando um recurso se refere a uma macro-língua (como o quechua, aimará, zapoteca, nahuatl, maia, entre outras) que não tem um código ISO na última norma publicada (ISO 639-3), tiveram de ser utilizados códigos especiais.

c. *Língua abordada*

Esta rubrica indica o nome da língua indígena abrangida pelo registo, tal como aparece no recurso, mesmo que não seja o nome mais comum. Vários recursos referem-se a variedades específicas, que foram registadas na base de dados tal como são, mesmo que isso não se reflecta nas estatísticas finais que têm em conta a língua de que são variantes.

Numa minoria de casos, nesta rubrica registaram-se "Várias" para os pedidos que envolvem várias línguas de um país, região ou família linguística, e "Todas" para todas ou a maioria das línguas da Ibero-América.



Seguindo a lógica da aplicação da norma 639-3, o nome espanhol de cada língua foi utilizado para as estatísticas finais, com base na língua que o Ethnologue⁴ indica estar associada à codificação ISO. O Ethnologue é um sítio de referência cujos promotores presidem ao comité internacional da norma ISO 639-3. Embora tanto a norma ISO como o Ethnologue possam não satisfazer muitos interlocutores pelas razões já invocadas, não foi encontrada melhor fonte que permita uma classificação universal para facilitar a produção de estatísticas coerentes.

d. Língua de interface

Nesta rubrica, foi registada a língua que serve de interface principal do recurso, e não a língua a que a aplicação se refere. Foram considerados sete casos diferentes:

- Espanhol / Português / Inglês / Francês, quando a língua de interface é um destes.
- Língua indígena, quando a língua de interface é a mesma que a língua de estudo ou o conteúdo do recurso.
- Múltipla, no caso de acesso a diferentes línguas, incluindo a língua indígena. A estrutura inicial do ficheiro não previa uma repartição destas últimas ocorrências, o que se aconselha para futuras actualizações, mas estima-se que menos de 5% das aplicações com interface multilingue incluem uma língua indígena.
- Outros, quando se trata de recursos com interface noutras línguas.

e. Atualização

Este campo regista todos os recursos que foram recentemente actualizados ou que permitem actualizações sistemáticas. Os recursos que não apresentam indícios de actualização ou aqueles para os quais não é possível determiná-la são indicados como “não actualizados”.

f. Promotor

No que diz respeito ao promotor, exceto nos casos em que não foi possível detectá-lo, é classificado da seguinte forma:

- Promotor individual: quando se trata de uma iniciativa de uma ou algumas pessoas fora da estrutura institucionalizada.
- Promotor coletivo ou comunitário: no caso de iniciativas comunitárias, cooperativas, sindicatos ou qualquer tipo de coletividade, institucionalizada ou não.
- ONG ou organizações não governamentais: esta designação foi escolhida para qualquer entidade que indique claramente que tem o estatuto de associação não

⁴ www.ethnologue.com



pública, fundação ou ONG. Caso contrário, se tiver características semelhantes às de uma ONG mas não o indicar, é considerada colectiva ou comunitária.

- Público: todas as entidades públicas ou para o público, seja local, regional, nacional ou internacional.
- Universidade: quando se trata de um estabelecimento de ensino superior, de um centro de investigação ou de uma academia ou similar.
- Empresa: quando se trata de um empreendimento comercial.

g. *Léxico*

É importante notar que várias aplicações são abusivamente designadas por tradutores quando, na realidade, são apenas ferramentas terminológicas ou lexicográficas. Embora o nome dos seus criadores tenha sido respeitado no campo *do nome*, esta distinção é devidamente clarificada no campo da *descrição*.

No que diz respeito ao campo *de entrada*, o número de palavras ou termos tratados pela aplicação deve ser introduzido, desde que esta informação esteja facilmente disponível. Infelizmente, na grande maioria dos casos, não é esse o caso.

h. *Interesse (avaliação de recursos)*

Esta rubrica contém uma avaliação da aplicação em estudo que pode ser interpretada como arbitrária por muitos interlocutores.

A verdade é que, para avaliar plenamente um recurso, seria necessário conhecer cada uma das línguas estudadas, o que é uma tarefa difícil.

É por isso que o interesse do recurso foi estimado, não pela sua qualidade, mas pela quantidade de conteúdo relevante ou pela sua utilidade relativa. Esta última foi estimada de acordo com a ausência ou presença de aplicações semelhantes para a mesma língua.

Esta avaliação pode variar em função das características sócio-demográficas de cada língua, mesmo quando dois recursos são idênticos em quantidade ou utilidade, mas têm perfis sócio-demográficos diferentes.

Assim, vídeos, áudios ou aplicações para línguas ameaçadas ou línguas com poucos falantes serão considerados de elevado interesse, dada a sua singularidade. Por outro lado, material semelhante será considerado de baixo interesse para línguas que já dispõem de recursos de maior qualidade ou que já foram suficientemente descritas ou estudadas.

Do mesmo modo, o material que corresponde a uma variante regional com poucos recursos será considerado de grande interesse, mesmo que a *língua materna* disponha de recursos superiores. A mesma situação se verificará no caso das línguas transfronteiriças, quando houver poucos falantes num dos países em que a língua é falada, mesmo que exista melhor material para a mesma língua num país vizinho.



- ▶ Pelas razões acima expostas, pede-se a compreensão do leitor e aconselha-se a considerar esta avaliação como meramente indicativa e a necessitar de revisão por equipas que dominem as línguas estudadas.

i. País

Nesta secção, é dada prioridade ao país mencionado no recurso estudado. No entanto, quando este não é claramente especificado, foi utilizada a informação sobre o país de origem da língua referenciada segundo o *Ethnologue*.

- ▶ Embora a comunidade indígena muitas vezes não considere que as suas línguas têm um país, esta informação pode ser de interesse para as autoridades nacionais, provinciais ou regionais dos países em causa, caso estas implementem políticas de apoio ou promoção das línguas em questão.

Relativamente ao material de estudo

a. Sobre a designação das línguas:

A maioria das línguas é designada por nomes diferentes, consoante se trate de *endónimos* ou *exónimos*, da região ou do país onde são faladas, da variante linguística em causa, do período em que foram registadas ou das diferenças ortográficas devido à falta de normas consensuais, entre outros factores.

Por conseguinte, decidiu-se utilizar o nome referenciado na página de origem, embora possa ser objeto de discussão em função do interlocutor. No entanto, por razões metodológicas, tanto para a compilação de estatísticas como para o tratamento dos recursos em que um nome específico não é claramente especificado, decidiu-se utilizar o nome espanhol mais comum.

b. Sobre os programadores não nativos das páginas ou aplicações:

A grande maioria dos dados obtidos provém de bases de dados internacionais que contêm informações sobre as línguas indígenas e não nas línguas indígenas. Estas fontes dispõem de um número consistente de recursos recolhidos junto das populações indígenas (áudios, vídeos e textos), bem como de descrições das línguas e de numerosos documentos de referência. Quase todas elas são ONG europeias ou norte-americanas (RisingVoices, SIL, GlobalRecords, ELAR, Glottolog, etc.).

- ▶ É importante notar que muitas destas fontes estrangeiras, e algumas com implementação local, estão ligadas a diferentes igrejas cristãs, pelo que geralmente fornecem, para além de diverso material de interesse informativo ou utilizável na criação de ferramentas, traduções ou vocabulário da Bíblia, bem como áudios ou vídeos com orações e cerimónias religiosas em línguas indígenas. Neste contexto, existe uma grande quantidade de material bíblico em diferentes línguas indígenas e suas variedades, mas optou-se por registar o menor número possível por língua, devido à repetição de recursos que se considera não acrescentar valor ao estudo.



c. *Sobre a sustentabilidade das iniciativas:*

É importante notar que um grande número de projectos identificados foi abandonado ou mantém o estatuto de projeto após vários anos. Com exceção de um número marginal de projectos susceptíveis de evoluir favoravelmente no futuro, a maioria não foi incluída neste inventário. Estimamos que 20% a 30% dos projectos anunciados são atualmente inexistentes ou simplesmente não avançaram. Um dos exemplos notáveis incluem o anúncio em 2013 da tradução da interface do Firefox para cerca de 20 línguas indígenas do México, Guatemala e Colômbia, mas a interface só está atualmente disponível em 5 dessas línguas. O mesmo acontece com as chamadas incubadoras da Wikipédia, que incluem cerca de 20 enciclopédias em línguas indígenas das Américas, que estão à espera de editores.

d. *Sobre a multiplicidade das fontes:*

Uma vez que o material deste estudo foi concebido para facilitar a pesquisa por língua, tipo de conteúdo e tipo de suporte, foram apresentadas as seguintes situações:

- Muitos recursos da mesma fonte foram tratados individualmente quando esta fonte abrange diferentes línguas ou diferentes meios de comunicação para a mesma língua. Em quase todos os casos, foi indicado o URL específico para cada língua, suporte ou conteúdo, e não o URL global. Nalguns casos marginais, esta discriminação não foi possível, quer por não existirem URLs diferenciados, quer por conterem pouco conteúdo para cada língua. Nestes casos, será indicado o mesmo URL para diferentes línguas, meios ou conteúdos.
- Em várias ocasiões, foram mencionadas redes sociais associadas ao sítio *web* principal do projeto ou da aplicação, mas optou-se por omitir esses suportes (mesmo que as suas características permitissem a sua integração) quando não acrescentam valor em relação ao sítio *web* original.

e. *Utilização de repositórios maiores:*

A maior parte dos recursos provém de 13 portais ou sítios *web* que contêm dados sistemáticos sobre todas as línguas do mundo ou, pelo menos, sobre grandes grupos de línguas. As informações extraídas destes sítios constituem cerca de dois terços dos recursos detectados, uma vez que foi indicado o URL específico de cada língua abrangida por cada repositório, para que seja possível efetuar uma pesquisa por língua. A informação disponível nestes repositórios é de grande valor para as línguas ameaçadas, pois inclui numerosos textos, áudios, vídeos, referências, bem como informação bastante completa sobre aspectos demográficos e sociolinguísticos, estrutura, família, localização, variantes linguísticas, etc. O Anexo 1 contém uma lista destes repositórios, com indicação dos seus promotores, do interesse para o inventário e do tipo de recurso que pode ser encontrado em cada um deles.



Resultados do inventário dos recursos

Foram inventariados e descritos mais de **5.500** recursos⁵ de cerca de **1.000** sítios ou aplicações para cerca de **800** línguas⁶. No entanto, para obter as estatísticas finais, preferiu-se por depurar e agrupar as línguas em cerca de **400**, de acordo com os seguintes critérios:

- ▶ As línguas com menos de 2 falantes foram excluídas porque, embora os seus recursos tenham sido catalogados na base de dados, o seu desaparecimento indica que os recursos encontrados não seriam úteis para a construção de novas ferramentas digitais.
- ▶ Embora as línguas crioulas tenham sido consideradas no inventário, o crioulo haitiano e o patois jamaicano foram excluídos porque, embora amplamente falados (o primeiro na República Dominicana e o segundo na Costa Rica e no Panamá), a grande maioria dos recursos encontra-se no Haiti, na Jamaica, na América do Norte e na Europa, mas não nos países ibero-americanos.
- ▶ Por vezes, os produtores dos recursos são muito específicos na indicação da língua que está a ser discutida (por exemplo, Quechua de Pomabamba / Zapotec de Las Delicias / Nahuatl da Huasteca Veracruzana), mas na maioria dos casos são mais genéricos (Quechua, Zapotec, Nahuatl). Por esta razão, preferiu-se agrupar as diferentes variantes ou línguas do grupo estatisticamente por macro-língua.

⁵ Para uma maior clareza, a maioria dos valores citados no presente documento foram arredondados, com exceção dos que constam dos quadros.

⁶ É importante recordar que o número de línguas estudadas se baseia nos critérios linguísticos específicos do Ethnologue, que nem sempre coincidem com os de outras fontes. Recomenda-se, portanto, que o leitor retenha apenas as ordens de grandeza.



Análise de inventário

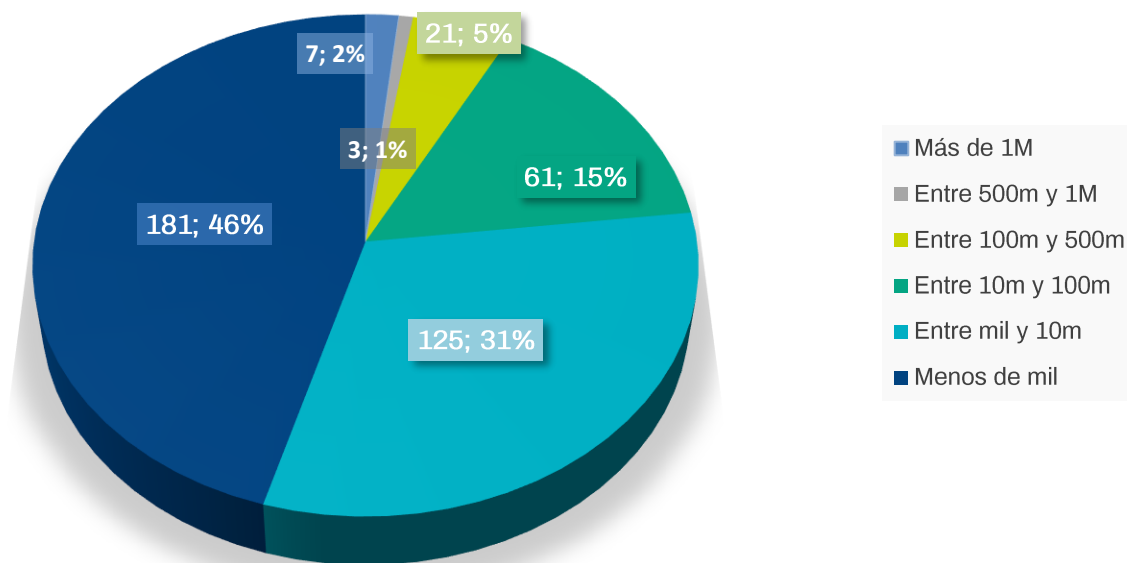
a. Introdução

Antes de mais, recordemos as grandes diferenças populacionais entre as línguas da América Latina. Situações como a do quéchua (7 milhões segundo o Ethnologue, 8 a 10 segundo outras fontes) ou do guarani (cerca de 6,5 milhões segundo o Ethnologue⁷) são excepcionais no continente, onde apenas cerca de 30 línguas são faladas por mais de 100.000 pessoas e mais de 600 estão ameaçadas ou claramente em vias de extinção, incluindo cerca de 350 pertencentes a macro-línguas. Para mais informações, ver *quadro 2* e *figura 3*.

Quadro 2 e Figura 3: Número de línguas por número de habitantes

NÚMERO DE LÍNGUAS	FALANTES DE MAIS DE	FALANTES DE MENOS DE
	1 000 000	
	500 000	1 000 000
21	100 000	500 000
	10 000	100 000
125	1 000	10 000
	0	1 000

Línguas por falante



Fonte: elaboração própria

⁷ Outras fontes indicam até 12 milhões, incluindo os que falam a língua como segunda língua.

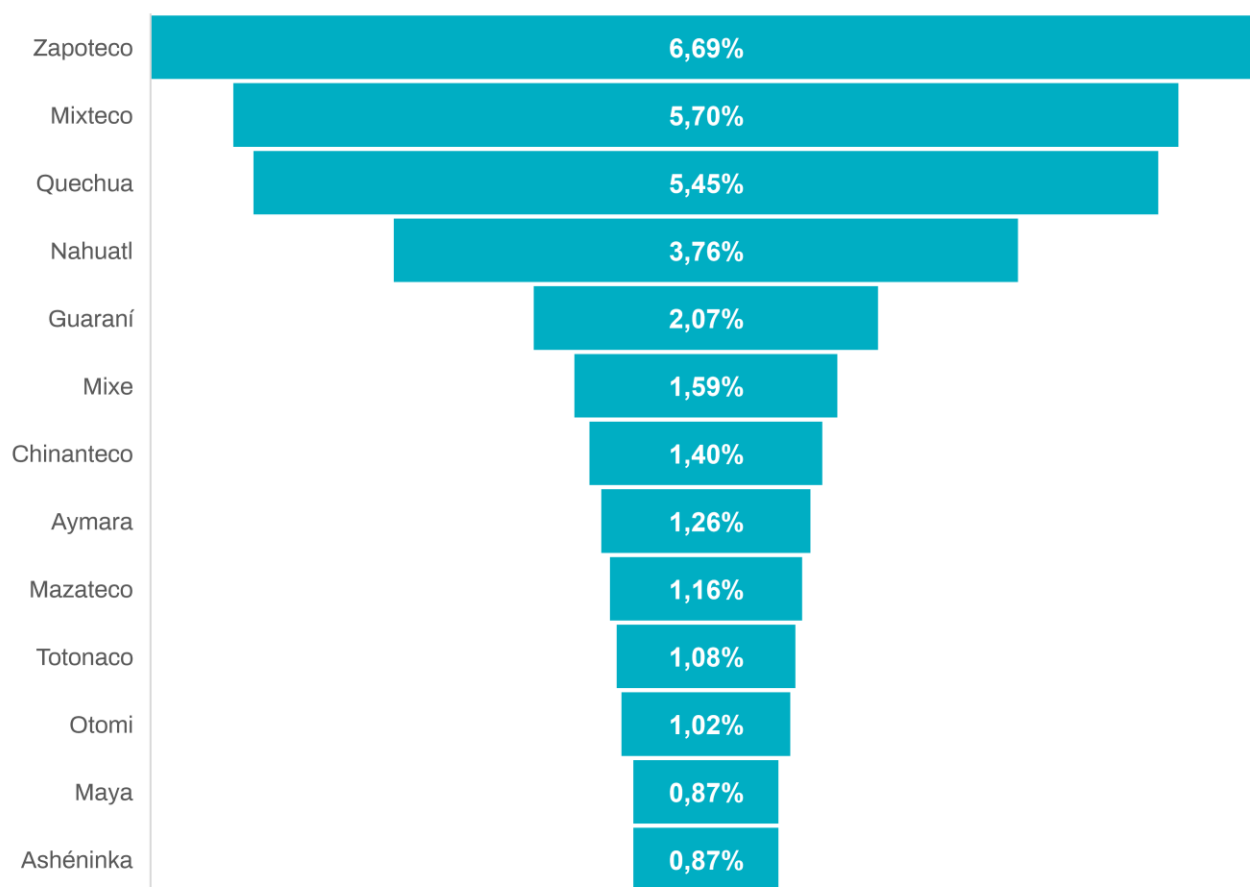


b. Distribuição por língua

Não é de surpreender que as línguas com mais de 100.000 falantes tenham, em média, 62 recursos disponíveis por língua, enquanto que para as línguas com menos falantes, o rácio diminui para uma média de 8 recursos por língua.

No entanto, é importante notar que foram encontrados mais recursos para o conjunto de variedades do zapoteco e do mixteco, línguas que têm muito menos falantes do que o quéchua ou o guaraní. A explicação pode estar no grande número de entidades públicas, comunitárias, associativas e universitárias no México - além das universidades americanas interessadas nessas línguas - que serão discutidas no capítulo *Inventário de entidades*. A *Figura 4* mostra a distribuição dos recursos em percentagem entre as principais línguas da região.

Figura 4: Línguas com o maior número de recursos



Fonte: elaboração própria

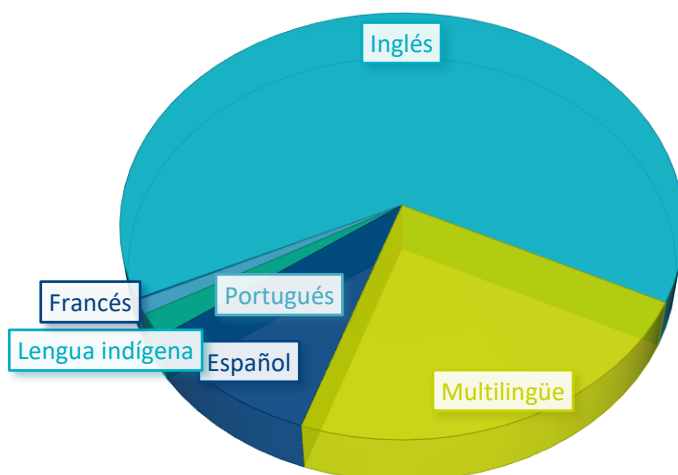


c. Língua de interface

Encontrou-se um número reduzido de recursos cuja única língua de interface são as línguas indígenas, embora existam várias páginas *multilíngues* que as incluem. Verifica-se também uma elevada presença de recursos em língua inglesa, dado que são a maioria dos sítios estrangeiros a descreverem as línguas (nomes, família, localização, situação sócio-demográfica, gramática, léxico, etc.) ou divulgarem registos sonoros e/ou visuais das línguas, sobretudo as que têm menos falantes. A *Figura 5* e o *Quadro 3* mostram a distribuição por interface.

Quadro 3 e Fig. 5 - Número de recursos por língua de interface

INTERFACE	NÚMERO RECURSOS
Inglês	3 500
Multilíngue	1 300
Inglês	500
Língua indígena	100
Português	
francês	8
Outra língua	1



Fonte: elaboração própria



d. *Atualização*

Dos recursos encontrados, **5.100** parecem ter sido actualizados recentemente. Para os restantes (menos de **10%** do conjunto) não foi detectado qualquer movimento recente. Note-se que as aplicações móveis (**150**), mesmo que pela sua natureza não possam ser actualizadas, são consideradas como actualizadas por serem recentes ou relativamente recentes.

e. *Promotores*

Para cada recurso encontrado, foi indicado o promotor, responsável, proprietário ou divulgador. Assim, foram encontrados **2.200** recursos produzidos ou promovidos por academias, instituições de ensino superior ou de investigação; **1.500** por ONG (ou Associações, Fundações, etc.); **700** por comunidades ou colectividades locais ou regionais; **600** por indivíduos ou grupos informais; **250** por organismos públicos (organizações intergovernamentais, estados nacionais, províncias ou estados federais, regiões, municípios, etc.) e **150** por empresas privadas. Para **100** recursos não foi possível detetar o tipo de promotor. *Ver figura 6.*

Fig. 6 - Promotores ou gestores de recursos por categoria



Fonte: elaboração própria



f. **Conteúdo**⁸

Em resumo, foram identificados diferentes tipos de conteúdos nos recursos encontrados. A maior parte deles (cerca de 4 400) contém uma breve descrição da língua, enquanto cerca de 2 100 tratam da estrutura da língua (gramática, fonologia, morfologia) e cerca de 1 000 contêm recursos lexicais ou terminológicos. Foram também encontrados recursos com referências bibliográficas ou hiperligações (mais de 4 000), recursos textuais (cerca de 2 800), recursos áudio (mais de 2 000) e recursos com vídeos e fotografias (cerca de 2 100).

Além disso, foram encontrados recursos que incluem mapas ou indicações precisas das áreas em que a língua é falada (cerca de 3.300) e que cobrem todas as características sociolinguísticas (3.300) e demo-linguísticas (3.400).

Em termos de recursos para a aprendizagem de línguas, foram identificados cerca de 250, a maioria dos quais destinados a um público alófono e poucos às populações-alvo.

Foram ainda encontrados mais de 1700 recursos que incluem ou lidam com ferramentas linguísticas, tais como dicionários automáticos, tradutores automáticos, aplicações móveis, análise e síntese linguística, software de escritório, enciclopédias digitais, entre outros. Ver *Figura 7*.

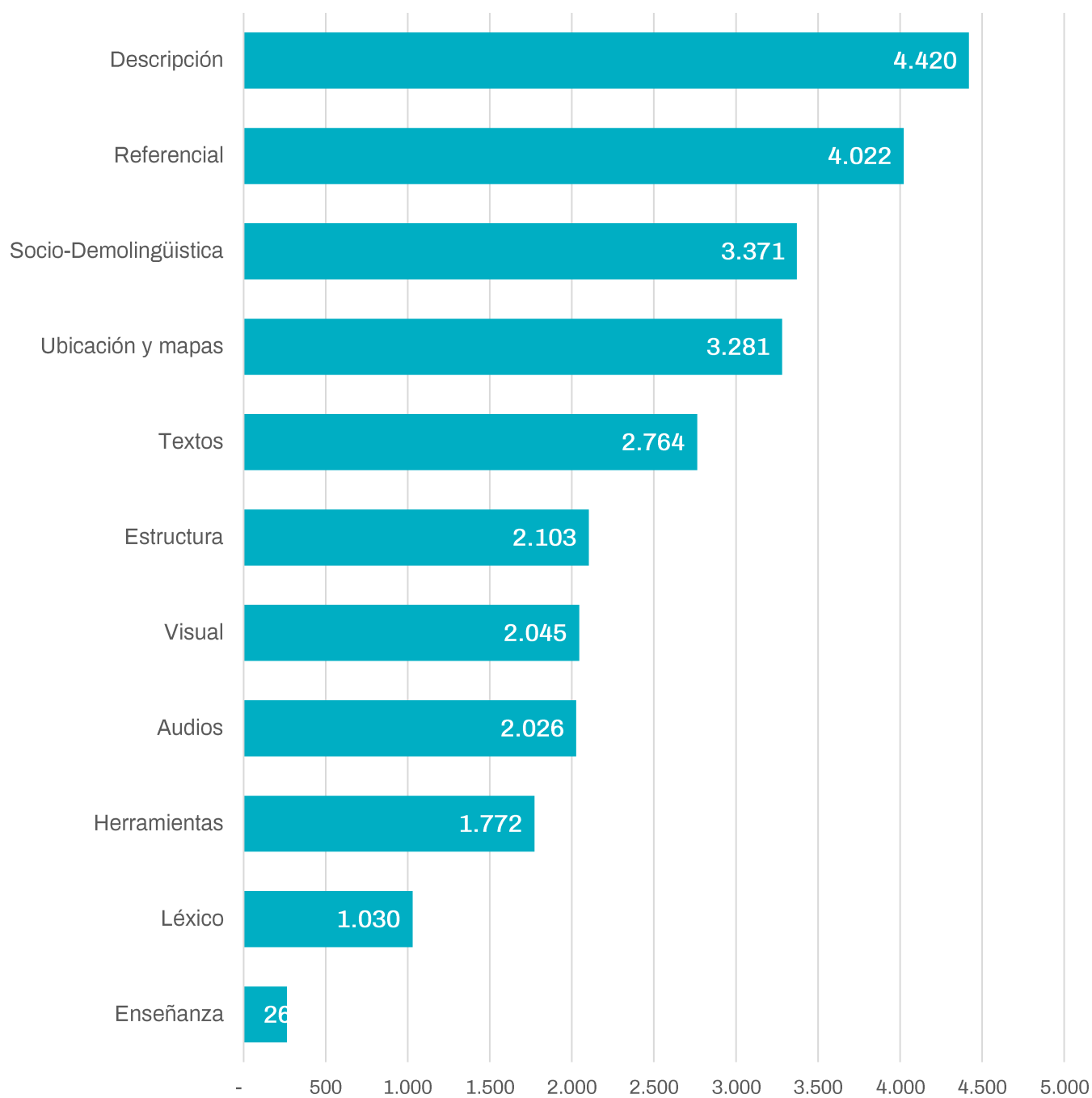
Procurou-se criar um registo sobre a quantidade de entradas (palavras ou termos) dos *recursos lexicais* encontrados (dicionários, vocabulários, glossários, etc.), mas na maioria dos casos esta informação não estava claramente indicada. Apenas 12 páginas o mencionavam de forma clara, pelo que só foi possível contabilizar cerca de 700.000 entradas, o que está longe de refletir as entradas reais, dado o número significativo de páginas ou aplicações encontradas a este respeito.

⁸ Os valores indicados neste parágrafo não devem ser somados, uma vez que podem existir vários conteúdos no mesmo recurso.



Figura 7 - Conteúdo principal dos recursos detectados

Fonte: elaboração própria





Fonte: elaboração própria

g. Países

Como já foi referido, o objetivo desta rubrica é enumerar os países em que o recurso foi criado. No entanto, em muitos casos em que esta informação é omitida ou pouco precisa, optou-se por indicar o país de origem da língua.

O mais notável é a proporção de recursos encontrados para as línguas do **México em comparação** com as línguas **sul-americanas no seu conjunto**. É também no México que se encontram a maior parte dos recursos linguísticos avançados, como a tradução automática, interface das principais aplicações, as aplicações móveis, os dicionários electrónicos, etc.

O Brasil é o país sul-americano com mais recursos para as línguas indígenas na América do Sul devido à sua elevada diversidade linguística. No entanto, esses recursos são produzidos principalmente por alofones e quase não existem aplicações de engenharia linguística. Isto deve-se à baixa densidade populacional das línguas indígenas no Brasil, com um máximo de 20 000 falantes para as línguas mais faladas segundo o Ethnologue, ou 35 000 segundo fontes brasileiras. Ver a *Tabela 4* e as *Figuras 8 e 9* para mais informações sobre os recursos por área geográfica.

Figura 8 - Recursos por zona geográfica



Fonte: elaboração própria



Quadro 4 e Figura 9 - Percentagem de recursos por país

PAÍS	NÚMERO RECURSOS
México	1 873
Brasil	1 189
Peru	568
Colômbia	493
Bolívia	276
Guatemala	273
Paraguai	159
Venezuela	125
Equador	115
Resto	567



Fonte: elaboração própria

h. Avaliação dos recursos

Convém recordar que, como já foi referido, esta rubrica está sujeita a uma análise mais aprofundada por parte dos falantes das línguas mencionadas. Deve ser interpretada como uma indicação quantitativa na maioria dos casos (com maior conteúdo) ou qualitativa no caso de instrumentos linguísticos de particular interesse para o futuro das línguas em causa.

Considerou-se que quanto menos recursos linguísticos existirem para uma língua, mais interessantes são os recursos encontrados, inclusive os mais simples. Isto leva a considerar que uma grande maioria dos recursos é de interesse "muito elevado" para as línguas menos utilizadas ou com menos recursos disponíveis (cerca de 3.300 para as línguas com menos de 10.000 falantes), devido à sua raridade. Por outro lado, e no caso das línguas com mais falantes e recursos, apenas serão anotados como de "alto interesse" aqueles com conteúdo particularmente notório ou os considerados avançados em matéria de ferramentas linguísticas (apenas 130 para as línguas com mais de 100.000 falantes). Ver a *Figura 10*, que apresenta em pormenor a qualidade do recurso em termos de segmentos linguísticos, de acordo com o número de falantes.

Figura 10 - Avaliação do interesse dos recursos

JUROS	MAIS DE 1 M	ENTRE 500 M E 1 M	ENTRE 100 M E 500 M	ENTRE 10 M E 100 M	MENOS DE MIL
Abaixo de	52	24	33	13	23
Médio	81	30	77	54	158
Elevado	95	50	139	184	378
Muito elevado	52	20	56	713	3287



Principais aplicações

Paralelamente ao estudo dos recursos em sítios *web* e aplicações móveis, foi realizado um levantamento exaustivo da presença das línguas indígenas da América Luso-Hispânica nas aplicações consideradas mais utilizadas a nível mundial, como portais, redes sociais, sistemas de tradução, motores de busca, enciclopédias e sistemas operativos, entre outros. Este levantamento incluiu uma análise detalhada da língua de interface e das ferramentas linguísticas disponíveis para cada uma dessas línguas.

DUAS DESTAS APLICAÇÕES MERECEM SER DESTACADAS:

Quanto à aplicação de inteligência artificial de utilização gratuita ChatGBT1 , observou-se que, embora tenha a capacidade de traduzir um grande número de línguas do mundo, incluindo cerca de 50 línguas indígenas da América Luso-Hispânica, os resultados para a maioria das línguas, especialmente as menos faladas, foram decepcionantes após uma dupla verificação do sistema. O que era de esperar devido à escassez de corpus acessíveis online para as línguas mencionadas. No entanto, recomenda-se que se faça um acompanhamento da aplicação nas versões futuras, já que este inventário de recursos foi concluído pouco tempo depois da aplicação ter sido lançada.



No que toca ao sítio Web da **W3Techs**¹ , é importante notar que, embora não disponha de ferramentas ou interface noutra idioma senão o inglês, considerou-se útil incluí-lo neste estudo pois este mede a presença das línguas do mundo nos 10 milhões de sítios *web* mais utilizados. No entanto, de acordo com a W3Techs, apenas o **guarani** e o **quechua** se encontram entre as 200 línguas com mais conteúdos.



O quadro 5 apresenta o conjunto das aplicações estudadas.

Quadro 5 - Total das principais aplicações estudadas

APLICAÇÕES ESTUDADAS			
Aikumi	Pesquisa no Google	Inverter	WeChat
Alexa	Tradução do Google	Acadêmico	Whatsapp
Apertium	GoogleMaps	SDL Trados	Wikilivros
Bing	Tradutor de MI	Sinal	Wikipédia
Blogger	Instagram	Siri	Wikiquote
ChatGBT	teclado iOS (apple)	Skype	Wikisource
Cromado	Lexilogos	Snapchat	Wikiversidade
Coursera	LibreOffice	Systran	Wikcionário
DMOZ	Logótipos	Telegrama	Wordpress
DuckDuckGo	Mensageiro	TikTok	Yahoo
Duolingo	Microsoft Office	Twitter	Yandex
Facebook	OpenOffice	Udemy	YouTube
Firefox	Perspectivas	Viber	Zoom

É importante notar que, entre estas aplicações multilingues, as mais ambiciosas têm normalmente interfaces e ferramentas linguísticas disponíveis em 10 línguas da América do Norte e das Caraíbas, 100 línguas da Ásia e da Oceânia, 100 línguas europeias e pouco menos de 40 línguas africanas. No entanto, apenas 17 destas aplicações encontraram a presença de 12 línguas indígenas da América Luso-Hispânica.

Para além do ChatGBT, a aplicação com o maior número de línguas na América de língua espanhola e portuguesa é o Aikumi⁹, um sítio *web* de prática linguística que oferece 15 línguas indígenas, embora a maioria delas tenha muito poucos praticantes.

Por outro lado, a Wikipédia¹⁰ e o Wikcionário¹¹, sítios multilingues por excelência, apenas disponibilizam conteúdos em 5 línguas indígenas, para além das 20 línguas em processo de criação nas "incubadoras", como já foi referido. Ver *quadro 6*.



⁹ <https://amikumu.com/statistics/>

¹⁰ http://meta.wikimedia.org/wiki/List_of_Wikipedias/es

¹¹ <https://meta.wikimedia.org/wiki/Wiktionary>



Quadro 6 - Línguas presentes nas principais aplicações - excluindo ChatGBT

CANDIDATURA	NATUREZA DO PEDIDO	PRESENÇA DA LÍNGUA INDÍGENA	LÍNGUAS EM CAUSA
Aikumi	Sítio Web das línguas	Métodos linguísticos	Quechua, Guarani, Maia, Aymara, Nahuatl, Bribri, K'iche', Nhengatu e Totonac
Bing	Motor de busca e tradutor	Interface e ferramentas linguísticas	Maia e Otomi
Cromado	Navegador	Interface e ferramentas linguísticas	Quechua, Guarani e Aymara
Cortana	Sistema operativo	Interface e ferramentas linguísticas	quíchua
Duolingo	Sítio de métodos linguísticos	Métodos linguísticos	Guarani
Facebook	Redes sociais	Interface	Quechua e Guarani
Firefox	Navegador	Interface	Guarani, Kaqchikel e Triqui
Tradutor gratuito	Tradutor	Tradução automática	quíchua
GBTchat	Inteligência artificial	Tradução automática	47 línguas
Pesquisa no Google	Motor de busca	Interface e ferramentas linguísticas	Quechua, Guarani e Aymara
Google Traduction	Tradutor	Interface e ferramentas linguísticas	Quechua, Guarani e Aymara
Lexilogos	Dicionários do mundo	Dicionários	Quechua, Maia e Nahuatl
Messageiro	Mensagens	Interface	Quechua, Guarani e Aymara
Microsoft Office	Automatização do escritório	Interface e ferramentas linguísticas	quíchua
Sinal	Mensagens	Interface	quíchua
W3Techs	Medidor de tráfego da Internet	Presença da língua na Internet	Quechua, Guarani
Wikipédia	Enciclopédia	Interface e artigos	Quechua, Guarani, Aymara e Náhuatl
Wikcionário	Dicionário enciclopédico	Interface e artigos	Quechua, Guarani, Aymara e Náhuatl

As línguas para as quais o ChatGPT propõe a tradução estão listadas na *Tabela 7*, embora a maioria delas se encontre numa fase inicial.



Tabela 7 - Línguas presentes no ChatGBT

TRADUÇÃO -COM RESERVA DOS RESULTADOS- NO CHATGBT			
Achi	Ixil	Mixteca	Tojolabal
Ashéninka	K'iche'	Nahuatl	Totonaco
Awajún	Kaqchikel	Ngäbere	Triqui
Aymara	Kuna	Otomi	Tz'utujil
Chinantec	Mãe	Paez	Tzeltal
Guarani	Mapudungun	Poqomchi	Tzotzil
Chol	maya	Purepecha	Wayuu
Chontal	Mazahua	Q'eqchi'	Wichí
Emberá-Chamí	Mazateca	quíhua	Yanomamö
Huastec	Miskito	Shuar	Zapoteca
Huichol	Mixe	Toba	<i>Outros</i>

TRADUÇÃO - COM RESERVA DOS RESULTADOS - EM CHATGBT			
Achi	Ixil	Mixteca	Tojolabal
Ashéninka	K'iche'	Nahuatl	Totonaco
Awajún	Kaqchikel	Ngäbere	Triqui
Aymara	Kuna	Otomi	Tz'utujil
Chinantec	Mãe	Paez	Tzeltal
Guarani	Mapudungun	Poqomchi	Tzotzil
Chol	maya	Purepecha	Wayuu
Chontal	Mazahua	Q'eqchi'	Wichí
Emberá-Chamí	Mazateca	quíhua	Yanomamö
Huastec	Miskito	Shuar	Zapoteca



Análise das lacunas digitais

Introdução

Com base numa classificação do tipo de recursos necessários para uma melhor inclusão digital, foram revistas todas as ferramentas e recursos linguísticos em falta ou insuficientes para as línguas indígenas que poderiam complicar a sua difusão, promoção e desenvolvimento.

De acordo com o conjunto de informações recolhidas, a maioria das línguas indígenas ibero-americanas tem pouca ou nenhuma presença digital, especialmente em comparação com a situação das línguas da Ásia, África e, sobretudo, da Europa.

Uma análise global revela uma grande ausência de recursos linguísticos que pode ser resumida da seguinte forma:

- Presença limitada (um punhado de línguas) nas principais aplicações de utilização generalizada (redes sociais, motores de busca, sistemas operativos, burótica, etc.).
- Pouco conteúdo produzido por falantes nativos, exceto nas línguas mais densamente povoadas.
- Quase todos os recursos encontrados são clássicos e carecem de criatividade e inovação: vídeos e áudios com objectivos etnológicos ou culturais; dicionários electrónicos pouco ergonómicos; gramáticas estáticas; falta de *corpus* de texto *online*; falta de enciclopédias e, quando estas existem, com muito poucos artigos, etc.
- Embora tenham sido encontrados muitos métodos de ensino de línguas, estes são bastante clássicos e destinam-se principalmente a falantes não nativos.
- As aplicações móveis encontradas estão orientadas, principalmente, para as necessidades básicas (educação, prevenção, dicionários, sensibilização cultural, etc.).
- Existem poucas interfaces linguísticas autóctones, predominando as línguas europeias.
- Há uma ausência quase total de ferramentas linguísticas elaboradas: análise e síntese linguística, tradutores automáticos, reconhecimento de voz, correção de texto, legendagem, etc.

Em geral, há uma grande falta do chamado "ativismo digital", exceto nas línguas mais densamente povoadas, especialmente na Mesoamérica, apesar de algumas iniciativas nacionais e internacionais de formação de activistas (como a Global Voices) ou de criação de conteúdos (cooperação espanhola e suíça, Wikipédia, Firefox, OpenOffice, etc.).



Recursos digitais na Web e em aplicações móveis

Uma vez que as lacunas nas línguas indígenas não serão homogêneas devido a factores como o peso demográfico, a situação socioeconómica, a localização, o acesso à Internet e as políticas públicas, serão apresentadas estatísticas de grupo nesta análise, mas a base de dados em anexo apresenta as lacunas, língua a língua.

i Note-se que qualquer material que não esteja digitalizado e acessível, como descrições, gramáticas, fonologia, léxicos e textos em papel ou em formatos que não sejam da Internet, não será incluído neste inventário, uma vez que foi realizado exclusivamente num ambiente digital (web, redes sociais, aplicações móveis e ferramentas de escritório).

A determinação da carência ou insuficiência de recursos, que é difícil de definir para um grupo tão grande de línguas, baseou-se numa comparação com as línguas da região mais preparadas, aquelas a que chamamos *línguas digitalmente activas*. Assim, de acordo com o número de recursos encontrados por língua, foi estabelecido um coeficiente que nos permite classificar esses recursos em 4 fases representadas por 4 cores diferentes nos quadros deste capítulo:

- Aceitável (a verde nas caixas)
- Insuficiente (amarelo)
- Escassas (a vermelho)
- Ausente ou quase ausente (a preto)

Para resumir os resultados, propomos um reagrupamento por categoria das línguas em função da sua presença ou ausência digital:

a. *Línguas digitalmente activas*

Este é o nome dado a uma minoria de línguas da região que, além de possuir um número significativo de recursos linguísticos primários, estão presentes em aplicações móveis, interfaces ou ferramentas que permitem a sua utilização em meios digitais (sistemas operativos, sistemas de escritório, tradutores, dicionários automáticos, interfaces, etc.). Podemos considerar estas línguas como aquelas que apresentam uma inserção promissora no mundo digital, mas ainda insuficiente,: *quechua, guarani, zapoteca, aimará, mixteca e nahuatl*.

Embora as línguas indígenas digitalmente activas disponham de recursos muito mais numerosos - e mais consistentes - do que as restantes, a situação está longe de ser ideal. Uma comparação com línguas de densidade populacional igual ou inferior, como o basco, o estónio, o esloveno ou o islandês, e mesmo com as que têm contextos socioeconómicos semelhantes, como o bambara (África), o havaiano e o maori (Oceânia), o ilocano e o quirguiz (Ásia), revela que estão mais presentes digitalmente do que as maiores línguas indígenas ibero-americanas.

É por isso que vamos comparar as lacunas de recursos do conjunto de línguas indígenas entre si, lembrando que o teto do conjunto é muito mais baixo quando comparado com línguas de outros continentes.

Para as línguas *zapotecas*, encontrámos uma quantidade considerável de recursos em quase todos os segmentos estudados, mais uma vez em comparação com as outras línguas



indígenas, embora haja menos disponibilidade de métodos de ensino de línguas. Para as línguas *quechua* e *mixteca*, encontramos vários segmentos com falta de recursos, particularmente, mas não só, em termos de *corpus* textuais e audiovisuais.

No que diz respeito às línguas *Nahuatl*, *Guarani* e *Aymara*, vários segmentos apresentam défices muito importantes. Ver *quadro 9*.

Quadro 9 - Recursos linguísticos digitalmente activos e lacunas

Lengua / Aplicación	Zapoteco	Quechua	Mixteco	Nahuatl	Guaraní	Aymara
Descripción	●	●	●	●	●	●
Estructura	●	●	●	●	●	●
Léxico	●	●	●	●	●	●
Referencial	●	●	●	●	●	●
Textos	●	●	●	●	●	●
Audios	●	●	●	●	●	●
Videos	●	●	●	●	●	●
Mapas	●	●	●	●	●	●
SocioDem	●	●	●	●	●	●
Enseñanza	●	●	●	●	●	●
Herramient	●	●	●	●	●	●

*Número de recursos: **aceitável** / **insuficiente** / **escasso** / **inexistente ou quase inexistente**

b. Línguas digitalmente inactivas

Designaremos desta forma as línguas para as quais encontramos recursos produzidos pelos seus próprios falantes, ainda que de forma muito dispersa. Para estas línguas, foi possível inventariar vários recursos linguísticos *online*, como dicionários e léxicos, gramáticas, fraseologia, ortografias, *corpus* textuais e audiovisuais, bem como algumas aplicações móveis e vários programas de ensino de línguas.

Embora algumas destas línguas, como o *maia*, o *totonac*, o *k'iche'* e o *kachikel*, tenham interfaces ou ferramentas em algumas aplicações muito utilizadas, como as redes sociais, os motores de busca e os *browsers*, estas são meramente *ad hoc*. Nesta categoria encontramos um pouco menos de 100 línguas, maioritariamente mesoamericanas, para além do *mapudungun*.

Das línguas menos presentes neste grupo, mas para as quais existem múltiplas iniciativas de digitalização de conteúdos, apenas foram encontrados recursos significativos para o *Mixe*, o *Chinantec*, o *Maia* e o *Totonac*, todas línguas da Mesoamérica. No entanto, este material não parece cumprir os requisitos para a inclusão digital das línguas.

Embora tenham sido encontradas várias iniciativas de ensino de línguas para as principais línguas deste grupo - como o *Mixe*, o *Chinantec*, o *Maia*, o *Otomi*, o *Mapudungun*, o *Ashéninka*,



o *Chatino* e o *Cakchiquel* - é difícil encontrar recursos suficientes para poder abordar uma política estruturada de digitalização do conjunto das línguas.

O *quadro 10* apresenta apenas as 60 línguas deste grupo com mais recursos, mas recordamos que as restantes podem ser consultadas na base de dados em anexo.

Quadro 10 - Deficiências das línguas digitalmente fracas

Lengua / Aplicación	Mixe	Chinanteco	Maya	Mazateco	Totonaco	Otomí	Maguolungun	Asíáninka	Chontal	Chatino	K'iche'	Tepehuan	Popoloca	Kacchikel	Tarahumara	Mejiteca	Triqui	Amuzgo	Tzotzil	O'egachi'	Tunebo	Huave	Witoto	Purepecha	Aweajun	Matlazincas	Popoluca	Tzeltal	N
Descripción	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Estructura	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Léxico	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Referencial	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Textos	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Audios	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Videos	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Mapas	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
SocioDem	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Enseñanza	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Herramienta	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

Lengua / Aplicación	Nahuatl	Achuar-Shiwiar	Mazahua	Mam	Cora	Chol	Wichi	Tz'utujil	O'anjóbal	Kumiai	Jakel'eko	Chuj	Surui	Katukina	Shipibo-Conibo	Chorote	Asurini	Bora / Miranha	Kuna	Cuicateco	Wampís	Ticuna	Chocholec	Itza'	Akatéko / Kanjibál	Pame	Huichol	Inga	A
Descripción	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Estructura	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Léxico	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Referencial	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Textos	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Audios	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Videos	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Mapas	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
SocioDem	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Enseñanza	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Herramienta	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

* N.º de recursos: **Aceitável** / **Inadequado** / **Escasso** / **Ausente ou quase ausente**

c. Línguas digitalmente passivas

Quanto às restantes línguas autóctones da região (cerca de 300), é raro encontrar conteúdos produzidos pelos seus próprios falantes e poucas ferramentas linguísticas. No entanto, existem numerosas fontes exógenas que nos podem orientar na obtenção de *corpus* utilizáveis tanto em políticas de salvaguarda e preservação, como na criação de material pedagógico para programas de Educação Bilingue Intercultural (EIB). Esta categoria inclui também as línguas em perigo iminente, para as quais tais recursos poderiam servir apenas como material patrimonial.

É evidente que as línguas que designamos por "digitalmente passivas" não dispõem dos recursos necessários para se integrarem no mundo digital. Se é certo que existem descrições



destas línguas, bem como gramáticas, fonologia, léxicos, textos, etc., não encontramos esse material no ciberespaço, pelo menos em formatos utilizáveis.

Panorama das lacunas identificadas

Em geral, constatamos uma ausência ou quase ausência de recursos para a maioria das línguas indígenas da região. Se analisarmos por segmento estudado, verificamos que apenas entre 1 e 3 línguas (consoante o segmento estudado) dispõem de recursos em quantidade aceitável; entre 6 e 17 línguas (idem) dispõem de recursos classificados como insuficientes; entre 40 e 66 línguas (idem) dispõem de recursos considerados escassos. A grande maioria das línguas (entre 306 e 339, consoante os segmentos) não dispõe de recursos (ou dispõe de recursos muito limitados). O quadro 11 mostra o número de línguas para as quais os recursos são aceitáveis, insuficientes, escassos ou quase inexistentes, por segmento.

Quadro 11 - Lacunas por segmento estudado

SEGMENTO / RECURSOS	ACEITÁVEL	INSUFICIENTE	ESCASSEZ	QUASE AUSENTE
<i>Descrição</i>	1	7	43	339
<i>Estrutura</i>	2	8	55	325
<i>Léxico</i>	3	10	40	337
<i>Referencial</i>	1	7	63	319
<i>Textos</i>	3	6	60	321
<i>Áudios</i>	1	9	63	317
<i>Vídeos</i>	2	7	66	315
<i>Mapas</i>	1	7	76	306
<i>Parceiro - Demonstração</i>	1	7	57	325
<i>Ensino</i>	3	17	35	335
<i>Ferramentas</i>	3	8	40	339





Inventário de entidades ou iniciativas

Introdução

O objetivo deste repertório é enumerar as entidades, projectos ou iniciativas dedicadas - total ou parcialmente - ao cuidado, preservação, dinamização, divulgação, promoção, desenvolvimento, etc., das línguas indígenas.

Para a elaboração deste diretório, começou-se por recolher informação em repositórios nacionais, provinciais, regionais, universitários, etc., bem como ligações a múltiplos sítios ou fóruns dedicados à **defesa, promoção ou difusão dos direitos sociais e culturais** das populações indígenas. O resultado foi um diretório de entidades do espaço ibero-americano - bem como de outros países - de mais de **2.250** onde existem iniciativas interessadas nas populações indígenas do continente.

i Depois de estudar os recursos obtidos um a um, passámos a analisar os recursos dedicados ou interessados na defesa, promoção e difusão das línguas indígenas, obtendo um número de **350** entidades que incluem acções ou vontades neste sector.

Resultados do repertório

A lista contém todas as informações necessárias para conhecer o âmbito da sua atividade, os seus objectivos, os seus promotores, a sua natureza, bem como o seu sítio *web* (caso exista) e todos os dados de contacto necessários.

É importante notar que só insistimos no preenchimento dos campos relevantes nos casos em que detectámos que a entidade **tem vocação para colaborar** em trabalhos específicos sobre **línguas indígenas**. Só foi recolhida informação de outras entidades com baixo perfil linguístico quando essa informação estava facilmente disponível, mas tal não foi a norma.

No que diz respeito às entidades relevantes para o projeto, foram feitos esforços para registar o máximo de informação possível, mas esta nem sempre está disponível. Em todo o caso, será sempre indicado o **nome**, o **tipo de entidade**, a sua **relevância** para o projeto e o seu **tema** principal. Em *quase todos* os casos, o **endereço URL** foi indicado, exceto nos raros casos em que não foi encontrado. Os campos do diretório são apresentados no *quadro 12*

Tabela 12 - Aspectos estudados para cada entidade detectada

CAMPOS ENTRADOS



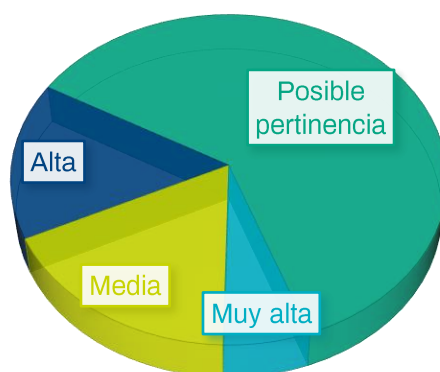
Nome da entidade	Populações indígenas em causa ⁽¹⁾
Acrónimo	País em causa ⁽¹⁾
Tipo de entidade	Região em causa ⁽¹⁾
Se for dependência de outro ⁽¹⁾	Endereço URL ⁽²⁾
Entidade superior ⁽¹⁾	Dados de contacto (cargo, endereço, endereço postal, etc.) ⁽²⁾
É governamental ou não? ⁽²⁾	Tema principal da entidade
Trabalha em coordenação com outra entidade ⁽¹⁾	Utilidade para o projeto
Está ligado a outro grupo, federação, etc. ⁽¹⁾	Comentários ⁽¹⁾
Âmbito (internacional, nacional, regional, etc.) ⁽¹⁾	Fonte ⁽¹⁾
Língua em causa ⁽¹⁾	Dispõe de recursos linguísticos? ⁽²⁾

*Os campos a negrito e sem nota foram sempre preenchidos. Os outros campos foram preenchidos conforme foi pertinente⁽¹⁾ ou se a informação estava disponível. ⁽²⁾

Relevância das entidades estudadas no seu conjunto

Como indicado acima, o trabalho consistiu numa recolha inicial de todos os tipos de organizações envolvidas em acções a favor da população indígena. Das 2.250 entidades, quase 1.900 foram descartadas por não indicarem actividades a favor das línguas, mas sim dos direitos, do desenvolvimento ou da cultura. As 300 restantes indicam acções em favor das línguas indígenas ou têm vocação para se envolver nelas. Estas últimas foram classificadas de acordo com o seu interesse para possíveis projectos de promoção e desenvolvimento a favor da inclusão digital das línguas indígenas. Destas, 60 entidades ou iniciativas destacam-se como tendo um potencial alto (50) ou muito alto (10) para serem contactadas ou incluídas em projectos de desenvolvimento digital das línguas indígenas. Ver figura 11.

Figura 11 - Relevância para as acções de digitalização de línguas indígenas

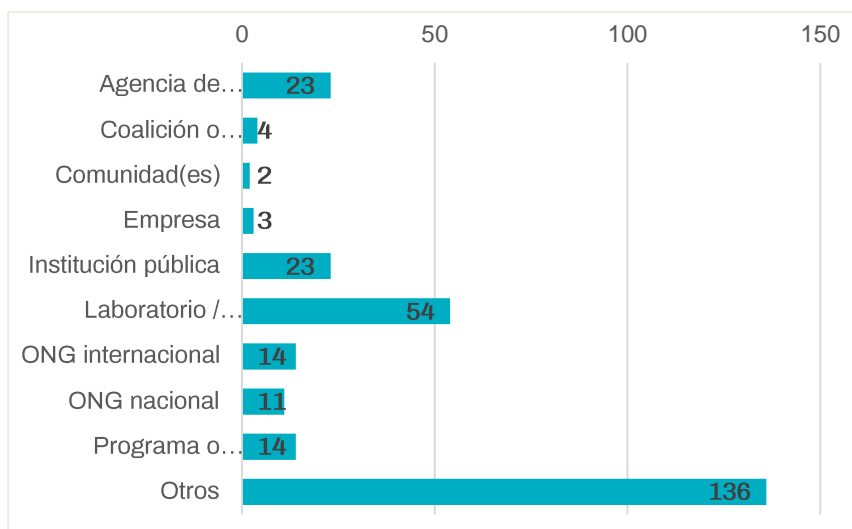




Natureza das entidades estudadas

Figura 12 - Tipo de entidade (relevante)

A maioria das entidades inventariadas são associações, sindicatos, comunidades, fundações, coligações, centros culturais ou sociais, constituídos por indígenas. Entre as demais, encontramos instituições públicas internacionais, nacionais, regionais ou municipais, além de ONGs e centros de investigação -



universidades, laboratórios ou institutos. Todas elas interessadas - total ou parcialmente - nos direitos, culturas e línguas indígenas.

Fonte: elaboração própria

Encontramos também uma variedade de projectos, programas ou coligações internacionais com interesse nas culturas indígenas, bem como iniciativas individuais ou empresariais, muitas vezes em formato exclusivamente digital (redes sociais, fóruns, portais, etc.).

A Figura 12 mostra a distribuição do tipo de entidade quando são relevantes para as acções linguísticas. Verifica-se uma percentagem significativa de organismos públicos (nacionais, provinciais ou internacionais), com 83 instituições entre as relevantes.

Âmbito territorial das entidades estudadas

Na medida em que a informação estava acessível, foi registado o âmbito territorial do trabalho da entidade. A Figura 13 mostra uma distribuição homogénea entre entidades de âmbito internacional, apenas nacional ou apenas regional (o que inclui o alcance local ou municipal).





Atividade principal ou tema das entidades

Por último, todas as entidades foram classificadas de acordo com os seus principais centros de ação. Sem surpresa, as entidades mais relevantes para os objetivos deste estudo são as que se dedicam às TIC e às línguas e culturas, como se pode ver no *Quadro 13*, onde as entidades mais relevantes estão listadas na última coluna.

Quadro 13 - Atividade ou tema principal

TEMA PRINCIPAL	TOTAL	PERTINENTE
Audiovisual, Comunicação e TIC	55	21
Cultura	13	4
Cultura e educação	18	5
Direitos	46	12
Direitos e educação	25	0
Direitos e proteção social	1803	0
Direitos, cultura e educação	14	8
Desenvolvimento	7	3
Desenvolvimento social. Identidade. Cultura	3	1
Desenvolvimento e educação	4	2
Desenvolvimento, cultura e línguas	1	1
Desenvolvimento, cultura, educação e saúde	3	3
Educação	14	7
Educação e línguas	3	2
Educação, cultura, direitos	3	0
Infância	1	0
Língua e cultura	37	34
Línguas	29	16
Línguas e culturas	4	2
Ambiente	11	2
Ambiente e terras	3	0
Social e cultural	4	2
Outros ou não identificados	150	150



Recomendações finais

A análise dos recursos encontrados e das insuficiências detectadas exige uma série de recomendações de ação para preencher progressivamente as lacunas e apoiar sistematicamente a integração de todas as línguas indígenas da América Luso-Hispânica no universo digital. Estas recomendações tentarão abranger todo o leque de acções que podem ser planeadas, sem cair na tentação de invocar os múltiplos instrumentos internacionais de preservação, promoção e desenvolvimento que têm sido recordados em diversos fóruns e documentos sobre o assunto e que devem servir de inspiração para todos os países. No entanto, é importante notar que esses instrumentos levaram à concepção da *Década das Línguas Indígenas (2022-2032)*¹², promovida pela UNESCO e aprovada pelos seus Estados-Membros.

Da mesma forma que a investigação realizada teve em conta a heterogeneidade de situações no conjunto das línguas da região, estas recomendações também terão em conta essa disparidade, catalogando o tipo de recomendação de acordo com as diferentes etapas que as línguas ibero-americanas poderiam seguir com vista a uma integração digital que ponha fim à agonia de muitas delas, e permita aos utilizadores da grande maioria o seu uso de acordo com as suas necessidades.

As recomendações que se seguem destinam-se especificamente ao que designámos por *línguas digitalmente activas e digitalmente inactivas*. Para todas as outras línguas, algumas destas recomendações seriam igualmente adequadas, mas é necessário começar por acções mais estruturais, explicadas no final deste capítulo.

1. Acesso à Internet e ao telemóvel

Muitas populações indígenas não têm acesso à Internet ou não podem usufruir partido dela por falta de meios económicos para a ligação e o equipamento (telemóveis, tablets, etc.).



As políticas públicas devem abordar estas deficiências, tal como recomendado por numerosos instrumentos internacionais e à luz das estatísticas que indicam que as populações indígenas da região estão muito menos conectadas do que o resto da população¹³.

¹² <https://es.unesco.org/news/proximo-decenio-lenguas-indigenas-2022-2032-se-centrara-derechos-humanos-sus-hablantes-0>

¹³ Ver o relatório de 2021 do Banco Mundial: <https://www.bancomundial.org/es/topic/poverty/lac-equity-lab1/ethnicity/ethnicity-education> (acedido em abril de 2023).



2. Literacia digital da população indígena

Como já foi repetido em numerosos fóruns internacionais, uma política de acesso à Internet não serve de nada se não for acompanhada de campanhas literacia digital em que a população seja um criador ativo de conteúdos e não consumidor passivo.



de
um

3. Educação e formação



O cerne da questão em matéria de revitalização das línguas através da digitalização é a falta de formação em técnicas digitais. Esta formação deve ser abordada a partir de diferentes ângulos e para diferentes sectores da população indígena. Eis os cursos de formação mais urgentes:

- Formação para a criação de ferramentas fundamentais: análise e síntese linguística, tradução automática, síntese do discurso, correctores ortográficos e gramaticais, inteligência artificial, etc.
- Formação em TIC de fonte aberta de uma forma que não dependa - ou favoreça sem troca – de interesses comerciais.
- Formação em matéria de criação de jogos de vídeo, de preferência em contextos localizados.
- Formação para a criação de sítios *web* ou blogues, motivando os próprios utilizadores da língua.
- Formação para edição no universo Wikimedia (nomeadamente Wikipédia e Wikcionário). Esta formação é dada gratuitamente por voluntários da Wikimedia, como tem acontecido com as línguas já integradas e as que estão em "incubadoras".
- Formação para a criação de livros electrónicos, a fim de digitalizar novos conteúdos, bem como obras existentes apenas em papel.

4. Identificação, formação e apoio aos activistas digitais

Este inventário de recursos permite concluir que a maioria das iniciativas indígenas se concretizaram graças ao impulso inicial dos chamados *activistas digitais*, pessoas (na sua maioria indígenas ou ligadas às suas culturas) que estão motivadas para realizar acções de digitalização das línguas indígenas. A *Rede de Activistas Digitais de Línguas Indígenas*^[1], cujos membros, criadores de numerosos recursos digitais indígenas, puderam ser contactados, tanto para iniciar actividades como para formar outros activistas, foi incluída no inventário de entidades.



^[1] Programa iniciado pela ONG Global Voices / Rising Voices



5. Participação ativa das comunidades



O inventário mostra que a grande maioria dos recursos digitais provém de entidades externas ao meio indígena, o que, apesar das boas intenções, não tem evitado a extinção ou a redução do uso das línguas indígenas. No entanto, constatamos que as línguas mais dinâmicas são aquelas em que os próprios falantes propõem ações para a sua revitalização e promoção.

É importante salientar que os recursos mais inovadores foram obtidos graças ao apoio de organismos públicos, ONGs e/ou universidades, mas, ao mesmo tempo, nenhuma política de digitalização de línguas indígenas pode ser levada a cabo sem consulta aos falantes, uma vez que esta nem sempre é vista de forma satisfatória pela população. Há povos que consideram a digitalização como uma *apropriação* da sua cultura, outros que devem pedir a autorização de toda a comunidade, outros para quem o carácter sagrado da sua língua os leva a tomar precauções, etc.

Por outro lado, é importante notar que a maioria dos esforços feitos para criar conteúdos linguísticos parece ter como objetivo uma população externa. É claro que a promoção e a divulgação da língua no mundo exterior só podem contribuir para o seu crescimento, mas é mais importante que os primeiros beneficiários sejam as pessoas que falam a língua, para que esta não perca o seu valor aos seus olhos. É sabido que uma língua é abandonada pelos seus falantes quando estes se apercebem de que não lhes serve para nada em termos de educação, informação, vida prática, entretenimento, etc.

6. Fornecimento de material para a digitalização

Os *activistas digitais* devem estar equipados com material informático e audiovisual (câmaras, ferramentas de edição e podcasts). Hoje em dia, vemos que a imagem e o som são cruciais em termos de motivação e promoção, e seriam um excelente monstrosário para as línguas indígenas, nomeadamente através das principais redes sociais (Youtube, TikTok, Instagram, Facebook, etc.), bem como de blogues ou sítios *web*.





7. Sectores prioritários para a digitalização



O objetivo final é que cada língua seja capaz de satisfazer todas as necessidades dos seus falantes. No entanto, é preciso começar pelo que tem maior probabilidade de ter um impacto imediato. Estes domínios são aqui indicados. É importante recordar que os conteúdos devem priorizar o acesso ao conhecimento dos próprios falantes da língua, como já foi referido.

- Produção de livros electrónicos (*e-books*), digitalizando assim material recente ou publicado apenas em papel.
- Criação de artigos da Wikimedia e do Wikcionário.
- Integração do Firefox e do OpenOffice-LibreOffice. Esta integração pode ser efectuada com o apoio de voluntários das comunidades que promovem estas ferramentas de código aberto.
- Criação de tutoriais audiovisuais:
 - Para a criação de conteúdos linguísticos
 - Para a criação de conteúdos didácticos
 - Para a criação de conteúdos especializados
 - etc.
- Integração nos universos Google, Facebook e Microsoft. Esta atividade só pode ser realizada em conjunto com os seus produtores, o que implicará a apresentação de projectos a longo prazo apoiados por universidades.
- Integração em diferentes redes sociais (Whatsapp, Telegram, Twitter, Instagram, TikTok, Tumblr, Instagram, Soundcloud, Spotify) tanto para propor interfaces (em contacto com os seus produtores) como para motivar a criação de conteúdos em línguas indígenas.
- Criação de redes sociais autóctones.
- Integração em plataformas de jogos e criação de jogos virtuais originais.
- Criação de aplicações móveis:
 - Campanhas de prevenção
 - Rádio e televisão *online*
 - Diagnóstico médico *online*, eventualmente com sistema de tradução
 - Ensino à distância (métodos *online* ou *offline*)
 - Tradutores
 - Jogos
 - etc.
- Técnicas de criação de livros electrónicos (*e-books*) através da digitalização de material recente apenas publicado em papel.
- Materiais didácticos para EBI (educação bilingue intercultural).



8. Acções prévias para as línguas menos utilizadas

As recomendações acima referidas aplicam-se igualmente às línguas pouco activas do ponto de vista digital, mas é necessário começar por completar a estruturação das línguas para as quais faltam frequentemente elementos de base e integrá-las adequadamente no mundo digital.



i Esta recomendação inspira-se sinteticamente no guia elaborado por Marcel Diki-Kidiri no livro "Como assegurar a presença de uma língua no ciberespaço"¹⁴ publicado pela Unesco e pela União Latina em espanhol e português, que o leitor é convidado a consultar para mais pormenores.

Devem ser previstas iniciativas para o desenvolvimento de recursos linguísticos indispensáveis, tais como:

- Análise e transcrição fonológica
- Análise e transcrição prosódica, nomeadamente para as línguas tonais
- Alfabeto e ortografia
- Gramática
- Dicionários
- Seleção de textos

A língua deve então ser equipada com ferramentas de tratamento informático, por forma a adaptá-las a qualquer tipo de plataforma:

- Mapa de caracteres
- Fonte(s) tipográfica(s)
- Teclado virtual
- Programas de processamento de *corpus*

Por último, deve proceder-se à elaboração de recursos culturais e ao seu condicionamento (registo e digitalização de textos, sons e imagens).

¹⁴ https://unesdoc.unesco.org/ark:/48223/pf0000149786_spa



Conclusões


Este estudo permite **preencher uma lacuna** na informação sobre a evolução digital das línguas indígenas da América Luso-Hispânica. Identificámos os **recursos linguísticos** destas línguas no ciberespaço, bem como as **limitações que impedem o seu pleno desenvolvimento**. Embora não tenhamos tentado compilar informação exaustiva, pensamos ter reunido dados suficientes para fazer um diagnóstico preciso da situação das línguas indígenas no mundo digital e para chegar a uma avaliação detalhada das suas limitações.

Embora já soubéssemos que as línguas estudadas não tinham a presença digital necessária para a sua plena utilização em todos os domínios requeridos pelos seus falantes, constatámos a **brecha profunda** que separa as línguas indígenas das línguas dominantes na utilização do ciberespaço. Esta brecha diminuiu para o grupo de línguas mais faladas, mas não ao ritmo desejado, e ainda menos para a grande maioria das línguas menos faladas.

Este repertório é um **retrato** da situação das línguas indígenas num determinado momento. Não podemos esquecer que o ciberespaço é um universo em constante evolução, no qual surgem novos recursos e outros se extinguem rapidamente. As tecnologias também avançam a um ritmo vertiginoso, e o que hoje é comum amanhã será obsoleto. É por isso que é extremamente **necessário atualizar este repertório periodicamente**, a fim de analisar com precisão a evolução digital das línguas em questão.

Este repertório de recursos e lacunas linguísticas, ao preencher uma importante lacuna na perceção da presença digital das línguas indígenas, torna-se um **importante instrumento** para a elaboração de *políticas de promoção, desenvolvimento e visibilidade dessas línguas*.

Por outro lado, o **inventário das entidades** permite identificar os melhores **intervenientes** com os quais se podem desenvolver mais eficazmente essas políticas ou acções. As populações indígenas, o sector universitário e as entidades públicas (locais, regionais, nacionais ou internacionais) identificadas poderão trabalhar em conjunto para obter os melhores resultados, aproveitando a experiência de várias entidades associativas nacionais e internacionais já activas no terreno.

 Gostaríamos que este estudo e os dados que o geraram fossem periodicamente actualizados e disponibilizados *online*, não só para serem utilizados na conceção de acções específicas, mas também para aumentar a visibilidade de todas as línguas da região. A sua integração num sítio *web* de acesso público serviria como um portal para o universo de dados linguísticos digitalizados do conjunto das línguas indígenas da Ibero-América.

Como resultado deste trabalho, a OEI dispõe de informações sobre as línguas no ciberespaço que incluem, entre outros dados interessantes:

- Inventário dos recursos existentes por língua
- Identificação de questões específicas
- Estado nos sítios *web*
- Nível de interesse em relação ao tema das línguas e do ambiente digital
- Alcance de aplicação territorial
- Diretório de entidades que trabalham neste domínio

OEI






Organización de Estados
Iberoamericanos

Organização de Estados
Ibero-americanos



C/ Bravo Murillo 38
28015 Madrid, Espanha
Tel.: +34 91 594 43 82
Fax: +34 91 594 32 86

oei.int

-  [Organização dos Estados Ibero-Americanos](#)
-  [Paginaoei](#)
-  [@EspacioOEI](#)
-  [@OIS_Espaço](#)
-  [Organização dos Estados Ibero-Americanos](#)