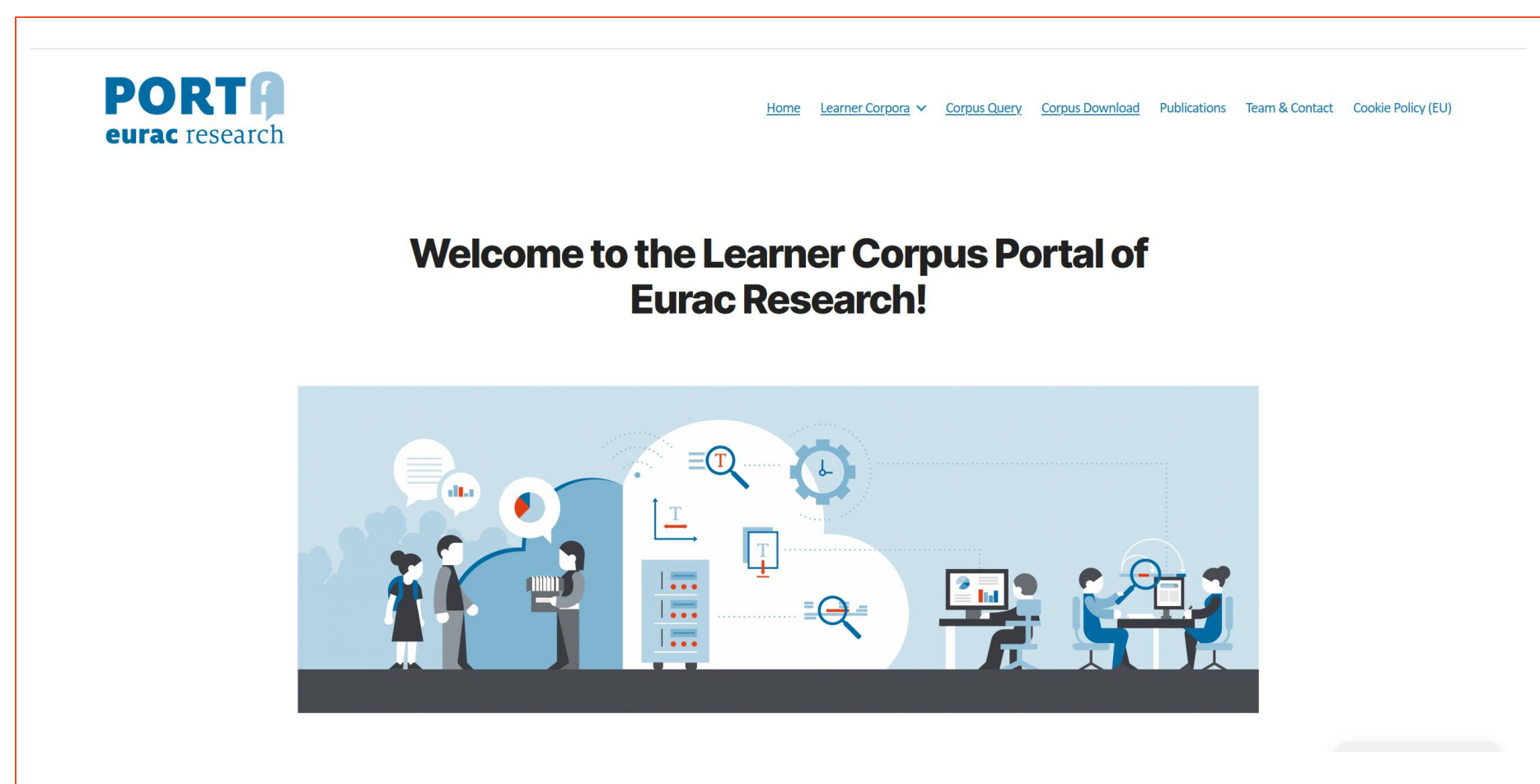


Was ist PORTA?

PORTA ist eine **Plattform**, auf der alle **Lernerkorpora** des Instituts für Angewandte Sprachforschung übersichtlich beschrieben werden. Zu jedem Korpus werden Angaben zur Größe, Anzahl der Texte und ihrer Autoren bzw. Autorinnen, Textsorte, Zielsprache und zum Jahr der Korpuserstellung bereitgestellt.

Über die Plattform kann auf alle Korpora zugegriffen werden. Hier befinden sich Links zur direkten Suche im jeweiligen Korpus über die **Suchoberfläche** von ANNIS sowie Links zum **Download** jedes Korpus über das Eurac Research Clarin Centre (ERCC), wo alle Daten in diversen Formaten verfügbar ist.

Eine vollständige **Dokumentation** zur Erstellung jedes Korpus ist ebenfalls über PORTA zugänglich. Dazu gehören die Aufgabenstellungen, die Transkriptionshandbücher, die Beschreibungen der verwendeten Annotationen sowie der jeweiligen Metadaten. Für Transparenz, welche Forschungsthemen pro Korpus bereits analysiert wurden, sorgen die Publikationshinweise am Ende jeder Korpusbeschreibung. Die Korpora werden regelmäßig aktualisiert. Informationen zu Korpus-Updates können ebenfalls über PORTA abgerufen werden.



Besonderheiten von PORTA

PORTA möchte möglichst viele **unterschiedliche Gruppen von Sprachlernenden in Südtirol** abdecken. Wir konzentrieren uns dabei auf junge Menschen und den öffentlichen Bildungsbereich. Die bisher gesammelten Texte stammen überwiegend aus Mittel- und Oberschulen aus Südtirol. Es wurden Texte von Schulen mit deutscher sowie italienischer Unterrichtssprache berücksichtigt.

Die gesammelten Korpora sind sich in mehreren Aspekten ähnlich, sodass die **Daten vergleichbar** sind, auch wenn sie in verschiedenen Projekten mit unterschiedlichen Forschungsfragen gesammelt wurden. So befinden sich in vielen Korpora argumentative Texte als Textsorte. Auch bei der Aufbereitung der Korpora wird darauf geachtet, vergleichbare Abläufe, Annotationen und Metadaten zu verwenden.

Mehrsprachigkeit ist ein zentrales Merkmal moderner Gesellschaften und ein Forschungsschwerpunkt unseres Instituts. Um der lebensweltlichen Realität der Sprachlernenden in Korpusanalysen gerecht zu werden, produzieren wir auch mehrsprachige Lernerkorpora, die das Sprachenrepertoire von Lernenden so weit wie möglich abdecken. Auch bei den Metadaten berücksichtigen wir diesen Aspekt.

Weiterentwicklung von PORTA

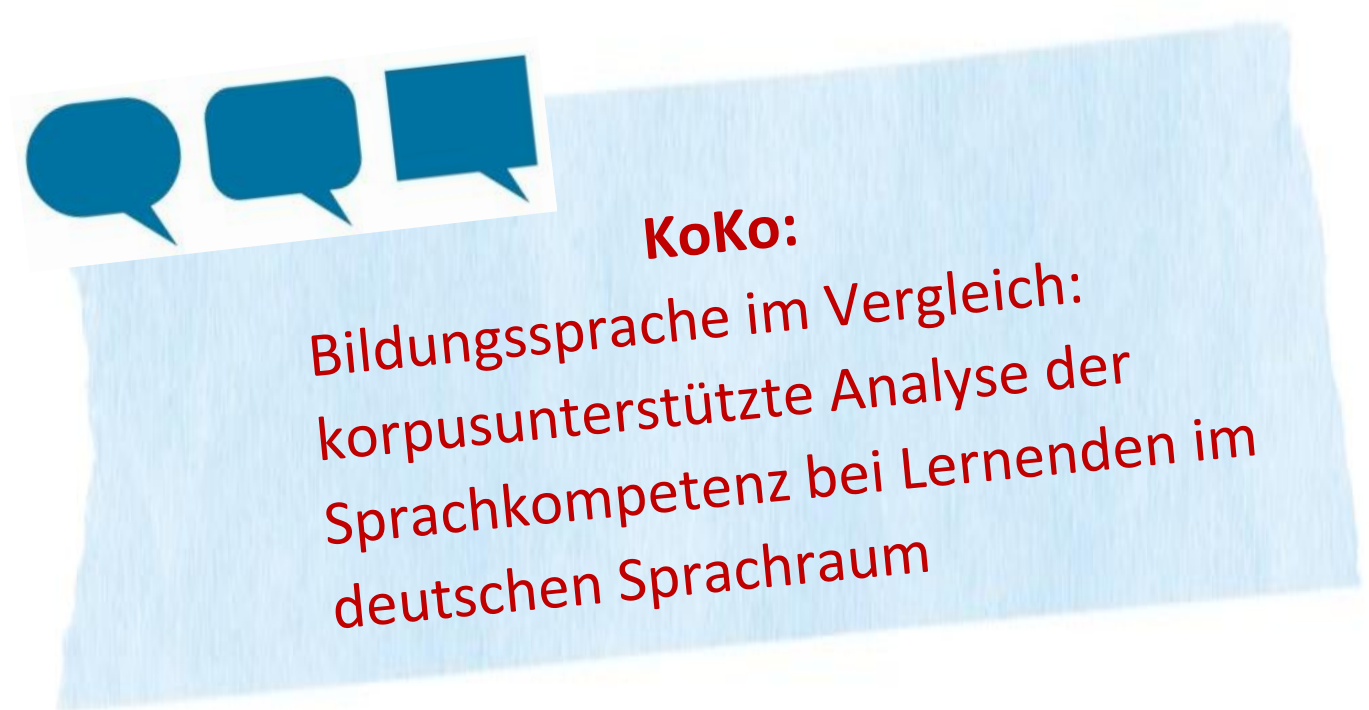
Das **Angebot an Informationen** rund um die zugänglichen Lernerkorpora wird zukünftig ausgebaut. Dazu gehören einerseits die Ergebnisse von Korpusanalysen, die für Folge-Analysen verwendet werden können. Dazu gehören andererseits auch technische Informationen darüber, wie die Lernerdaten

gesammelt, aufbereitet und verarbeitet wurden. Eine transparente Darstellung unserer Workflows ist das Ziel.

Das **Korpusangebot** wird laufend erweitert. Im Jahr 2025 werden weitere Lernertexte für die Kolipsi Korpus-Familie gesammelt, die das Angebot an L2-Lerner-

texten für Deutsch und Italienisch aus der Oberschule vergrößern werden. Außerdem arbeiten wir am neuen Korpus COMPLETT, das deutsche, italienische und ladinische Texte aus ladinischen Mittelschulen beinhalten wird, die zwischen 2024 und 2028 erhoben werden.

Welche Lernerkorpora gibt es auf PORTA?



Größe: 950.000 Tokens
Texte: 1.503 Argumentationen von 1.503 SuS deut. Oberschulen (DEU, AUT, STIR)
Sprache: Deutsch
Jahr: 2011



Größe: 340.000 Tokens
Texte: 2.246 Texte unterschiedlicher Textsorten von 2.246 Erwachsenen
Sprache: Deutsch, Italienisch, Tschechisch
Jahr: 2012



Größe: 1.171.000 Tokens
Texte: 5.871 Argumentationen und Narrationen von 2.993 SuS deut. und ital. Oberschulen
Sprache: Deutsch, Italienisch
Jahr: 2007/08, 2014/15



Größe: 64.000 Tokens
Texte: 332 Texte unterschiedlicher Textsorten von 141 SuS und 50 Erwachsenen
Sprache: Italienisch
Jahr: 2007, 2012

LEONIDE

Longitudinal **E**arner **c**Orpus **i**n **I**taliano, **D**eutsch and **E**nglish
A longitudinal trilingual corpus of young learners of Italian, German and English



Größe: 237.000 Tokens
Texte: 2.510 Argumentationen und Narrationen von 163 SuS deut. u. ital. Mittelschulen
Sprache: Deutsch, Italienisch, Englisch
Jahr: 2015-2018

Größe: 425.000 Tokens
Texte: 635 Argumentationen von 635 SuS ital. Oberschulen
Sprache: Italienisch
Jahr: 2022

