



## I CORPORA SEAH DI COMUNICAZIONE SPECIALIZZATA NEL SETTORE DELL'ARCHITETTURA E DELLE COSTRUZIONI. STRUTTURA, COMPILAZIONE E USI<sup>1</sup>

### 1 IL PROGETTO SEAH: CONTESTO E OBIETTIVI

L'Unione europea promuove da tempo scambi interuniversitari affinché studenti e studentesse possano arricchire il proprio percorso formativo in prospettiva internazionale, nell'intento di valorizzare il patrimonio multilingue e multiculturale degli Stati membri. Le difficoltà connesse all'apprendimento della lingua straniera possono tuttavia scoraggiare la partecipazione ai programmi, o limitare l'acquisizione di conoscenze teoriche, tecniche e professionali. Sebbene sia previsto almeno un livello B2 del QCER nella lingua del paese di accoglienza, spesso gli *incoming students* giungono nelle università ospitanti con un basso livello di competenza linguistica, convinti di poter utilizzare l'inglese in ambito accademico, soprattutto nei curricula scientifici. Tale problematica investe anche i corsi nel settore dell'Architettura e delle Costruzioni (AC) che includono una serie di ambiti tecnici definiti da tradizioni scientifico-culturali, pratiche professionali e quadri giuridici nazionali, per i quali non risulta sufficiente l'uso dell'inglese quale *lingua franca*.

Al fine di sviluppare risorse linguistiche innovative nella comunicazione accademico-disciplinare nel settore dell'AC è stato avviato nel 2020 un progetto multilingue finanziato dall'UE nel quadro del Programma Erasmus+ K203, dal titolo *Sharing European Architectural Heritage: Innovative language teaching tools for academic and professional mobility in Architecture and Construction* (SEAH), articolato in due fasi. La prima attiene alla compilazione di *corpora* specialistici comprendenti testi rappresentativi di alcuni sottodomini nel campo dell'AC in lingua francese, italiana, russa, spagnola e tedesca, mentre la seconda prevede lo sviluppo di moduli linguistici online per l'apprendimento del linguaggio accademico del suddetto settore nelle cinque

---

\* sara.piccioni@unich.it

\*\* mariapia.dangelo@unich.it

\*\*\* maria.ferro@unich.it

1 Il presente contributo è stato concepito ed elaborato in modo unitario dalle autrici, tuttavia sono stati redatti da Mariapia D'Angelo i §§ 1 e 4, da Sara Piccioni i §§ 2 e 3 e da Maria Chiara Ferro §§ 5 e 6. La bibliografia è comune.

lingue del progetto<sup>2</sup>. Le risorse elaborate sono liberamente accessibili dal sito di SEAH (<https://www.seahproject.eu/>) tramite una *open educational resources platform* nel duplice intento di sopperire alle necessità formative degli studenti in mobilità e delle categorie professionali (architetti, ingegneri, traduttori, interpreti, guide turistiche, ecc.) coinvolte con committenti stranieri, pubblici e privati nel campo dell'AC.

Introdotta il quadro teorico di riferimento, le metodologie (§ 1.1) e le finalità (§1.2) di SEAH, il presente contributo si soffermerà sui criteri e sulle procedure generali del *corpus design* (§2), con esemplificazioni della compilazione e impiego dei *corpora* per la lingua spagnola (§3), italiana (§4) e russa (§5).

### 1.1 Quadro teorico e premesse metodologiche

La comprensione e la produzione dei testi delle singole discipline è determinante per il raggiungimento degli scopi di studio degli apprendenti e implica l'acquisizione di quelle pratiche discorsive proprie dei generi testuali più tipici della *discourse community* di riferimento, anche in base all'intenzionalità dell'emittente e alla correlata accettabilità da parte dei destinatari. In tale contesto assumono particolare rilievo, sia per gli studi descrittivi (Swales 1990, 2004; Bhatia 1993) sia per le applicazioni didattiche, le indagini condotte nell'ambito della *genre analysis* che, disvelando il modo in cui i testi riflettono e costruiscono le comunità che ne fanno uso, rappresentano un potente strumento per l'individuazione dei bisogni comunicativi degli studenti non nativi (Hyland 2014).

Sul finire del secolo scorso, considerando da una parte le cifre caratteristiche del discorso accademico disciplinare evidenziate dalla *genre analysis*, e dall'altra l'esigenza di una formazione linguistica universitaria sempre più specializzata messa a fuoco nell'ambito della didattica dell'inglese accademico, si è pervenuti all'idea di distinguere l'*English for General Academic Purposes* (EGAP) attinente al lessico, alla morfosintassi, alle tipologie testuali, ai registri e all'organizzazione retorica comuni ai testi di tutte le discipline, dall'*English for Specific Academic Purposes* (ESAP), concernente gli aspetti linguistico-retorici delle singole pratiche discorsive disciplinari, spesso non esplicitate, ma determinate dal contesto epistemico di riferimento<sup>3</sup>. Nella stessa prospettiva dell'ESAP, i materiali prodotti per le cinque lingue del progetto SEAH sono elaborati sulla base di testi della comunicazione accademica disciplinare e specialistica

---

2 Nello specifico, il progetto SEAH è coordinato da Mariapia D'Angelo, mentre Sara Piccioni e Maria Chiara Ferro coordinano rispettivamente le sezioni relative allo sviluppo dei corpora specialistici (*Language corpora design, creation and distribution - French, German, Italian, Russian and Spanish for Architecture and Construction*) e dei materiali glottodidattici (*Developing and testing innovative language learning materials - French, German, Italian, Russian and Spanish for Architecture and Construction*). Al progetto partecipano le università di: Chieti-Pescara (Italia, ateneo capofila), Bordeaux Montaigne (Francia), Politecnico di Madrid (Spagna), Polotsk State University (Bielorussia) e Masaryk University (Repubblica Ceca).

3 Nel contesto dell'inglese per scopi accademici, tale suddivisione appare opportuna sia sul versante del lessico, che si è dimostrato variabile a seconda dell'ambito disciplinare (Hyland/Bondi 2006), sia per quanto attiene a citazioni, discorso indiretto, selezione delle strategie retorico-pragmatiche e dei generi testuali impiegati nei vari settori scientifici (Nesi/Gardner 2012).

nel settore dell'AC, per corrispondere alle reali esigenze linguistiche degli apprendenti e non ad una rigida sequenza di contenuti imposta da sillabi astratti (Hyland 2016: 20), che prevedono invece l'introduzione soltanto ad un livello B2 dei linguaggi specialistici e dei generi del discorso accademico. Questo è qui inteso come quel discorso che include varie tipologie testuali – orali e scritte – prodotte dai docenti, dagli studenti universitari, dai ricercatori per la diffusione dei propri lavori sia all'interno della comunità scientifica, sia all'esterno per la promozione e la divulgazione del sapere scientifico. Esso si declina in una complessa gamma di pratiche comunicative che variano non solo al variare del settore di studio, ma anche in ordine al grado di formalità e tecnicità in relazione alle coordinate pragmatiche del contesto situazionale.

Nei testi della comunicazione accademica è dunque possibile individuare una variazione “orizzontale”, in base ai contenuti, ed una variazione “verticale”, attinente al cambiamento del registro a seconda del grado di specializzazione dell'istanza emittente e ricevente, in un *continuum* ai cui estremi sono situati la costruzione del sapere scientifico-specialistico e la divulgazione di tale sapere (Berruto 1987). Sul piano verticale, i *corpora* SEAH comprendono un numero di generi testuali orali e scritti relativi ai domini della comunicazione accademica, didattica, professionale e divulgativa nel settore dell'AC. La variazione orizzontale attiene invece ai vari ambiti disciplinari e alle aree tematiche dei testi raccolti che possono variare nei cinque corpora, nell'ottica dell'adeguamento alle discipline maggiormente scelte dagli *incoming students* nei paesi di lingua francese, italiana, russa, spagnola e tedesca.

## 1.2 Impieghi e finalità dei corpora SEAH

I *corpora* compilati all'interno del progetto SEAH assolvono a molteplici funzioni connesse allo sviluppo dei materiali glottodidattici, quali l'identificazione e l'estrazione automatica di caratteristiche rilevanti sul piano lessicale, morfo-sintattico e retorico-pragmatico nel settore dell'AC, e l'individuazione di testi orali o scritti autentici da impiegare nelle attività stesse. Le strutture linguistiche e terminologiche esibite nelle unità didattiche sono infatti state estratte da grandi raccolte di testi autentici, prodotti da docenti e professionisti in contesti comunicativi reali; ciò ha facilitato in maniera significativa il reperimento dei materiali da parte dei ricercatori non esperti nelle discipline oggetto di studio e ha reso possibile lo sviluppo di attività didattiche costruite su fonti autentiche e motivanti, in quanto percepite dagli apprendenti come significative per il proprio percorso formativo.

Inoltre, secondo i principi del *Data-Driven Learning* (Johns 1991)<sup>4</sup> applicato all'insegnamento delle lingue per scopi accademici (Lee/Swales 2006), i moduli linguistici SEAH prevedono la consultazione diretta dei *corpora* da parte degli utenti stessi per il

---

4 Come è noto, si tratta di un approccio glottodidattico che considera l'apprendente come un 'ricercatore', al quale si chiede di verificare deduttivamente la validità di una regola tramite l'osservazione delle occorrenze di un determinato *token* in contesti diversi, oppure induttivamente formulando una regola a partire dai molteplici usi linguistici osservati nel *corpus*. A livello lessico-semantico, inoltre, tale approccio consente di far verificare direttamente agli apprendenti il valore polisemico di una parola, osservando il variare del significato a seconda dei contesti d'uso.

completamento di compiti guidati, allo scopo di potenziare la consapevolezza sui tratti peculiari del linguaggio disciplinare in questione (Flowerdew 2015).

## **2 PROGETTAZIONE, ARCHITETTURA GENERALE E COSTRUZIONE DEI CORPORA SEAH**

La costruzione dei cinque corpora SEAH si è svolta in tre fasi. Nella fase di progettazione i gruppi di lavoro di ciascuna lingua hanno realizzato una *needs analysis* per selezionare gli ambiti tematici, lessicali e disciplinari, nonché i generi testuali da includere in ciascun corpus. Questi sono stati stabiliti a partire da considerazioni circa gli utilizzi finali delle risorse create e, in ultima istanza, in base alle esigenze di apprendimento dei discenti cui sono rivolte le *open educational resources* elaborate. In consultazione con gli *stakeholder* del progetto (docenti nei Dipartimenti di AC, professionisti e docenti di lingua impegnati nell'insegnamento del linguaggio specializzato dell'AC), ogni gruppo ha stabilito:

- 1) Ambiti tematici da coprire (selezionandoli tra quelli cui prioritariamente sono esposti gli studenti di AC impegnati in programmi di mobilità)
- 2) Tipo di abilità linguistiche richieste (di comprensione/produzione scritta/orale)
- 3) Generi testuali che gli studenti e/o professionisti in mobilità sono chiamati a saper produrre o comprendere in lingua straniera.

Gli esiti della *needs analysis* hanno permesso di elaborare l'architettura generale del corpus, basata su una categorizzazione multilivello che tiene conto del canale di ricezione, del dominio e del genere testuale di ciascun testo. Per quanto riguarda il canale (*mode* nell'interfaccia di interrogazione, cfr. teoria del registro: Halliday/Hasan 1985), si è deciso di includere testi sia scritti, sia orali, rispettando per ciascuna lingua una proporzione di circa il 75% di testi scritti e il 25% di testi orali: lo sbilanciamento a favore dei testi scritti si deve al rilievo che la comunicazione scritta ha in contesti accademici.

Il dominio (*domain* nell'interfaccia) è stabilito a partire dalle principali componenti comunicative dei testi, in particolare dal ruolo e identità dei partecipanti alla comunicazione e dal proposito dei testi, mentre i generi (*genres*) vengono identificati – seguendo Swales (1990: 58) – come tipologie di eventi comunicativi caratterizzate da convenzioni strutturali, stilistiche e di contenuto ricorrenti e condivise all'interno di una comunità discorsiva. In base alla *needs analysis* condotta, si è deciso di rappresentare all'interno dei corpora SEAH i seguenti quattro domini comuni a tutte le lingue del progetto:

- 1) Dominio accademico (ACCAD): testi in cui ricercatori e/o studenti presentano risultati di ricerca alla comunità scientifica (ad es., articoli scientifici, monografie di ricerca, presentazioni a convegno scientifico, tesi di Master o dottorato, etc.).
- 2) Dominio didattico (DID): testi prodotti da ricercatori o docenti e rivolti a studenti (ad es., consegne di progetti/tesine/esercizi, manuali, lezioni, programmi di corsi universitari, dimostrazioni software, dispense, ecc.).
- 3) Dominio professionale (PROF): testi prodotti da professionisti del campo dell'AC rivolti ad altri professionisti e, in seconda battuta, a studenti e ricercatori/docenti (ad es., descrizioni di progetti, report di progetti, descrizioni di

materiali da costruzione, presentazione orale di progetti, materiale promozionale distribuito presso le fiere di settore, ecc.)

- 4) Dominio divulgativo (DIV): testi prodotti da ricercatori/docenti e/o professionisti e rivolti al pubblico generale (ad es., documentari, podcast, conferenze pubbliche, ecc.)

La Tab. 1 riporta la lista dei 22 generi selezionati nella fase di progettazione e diversamente distribuiti nei cinque corpora:

Genere	Canale	Dominio	Descrizione
monograph	scritto	ACCAD	monografia di ricerca
research article	scritto	ACCAD	articolo scientifico
thesis	scritto	ACCAD	tesi (di Masters o dottorato)
thesis abstract	scritto	ACCAD	abstract di tesi (di Master o dottorato)
course assignment	scritto	DID	istruzioni (scritte da un docente e rivolte agli studenti) per la realizzazione di progetti/esercizi/tesine richiesti per la valutazione di un corso universitario
handbook	scritto	DID	manuale didattico, libro di testo o dispensa
syllabus	scritto	DID	descrizione di corsi universitari
interview	scritto / orale	PROF / DIV	intervista scritta (in riviste, quotidiani, ecc.) o orale (radio, TV, ecc.) rivolta a professionisti (ad es., in una rivista specializzata) o al pubblico generale (ad es., intervista a un architetto famoso alla radio)
teaching material	scritto	DID	materiale didattico utilizzato in classe
material description	scritto	PROF	descrizione dei materiali da costruzione
project description	scritto	PROF / DID	descrizione di progetti scritta da professionisti o giornalisti specializzati (PROF) o da studenti (DID)
project report	scritto	PROF / DID	report di progetti elaborato da professionisti (PROF) o da studenti (DID)
trade shows	scritto	PROF	materiale promozionale distribuito presso fiere di settore e altri eventi commerciali
conference	orale	ACCAD	conferenza rivolta a un pubblico di accademici
roundtable	orale	ACCAD	dibattito/discussione tra accademici
thesis defense	orale	ACCAD	discussione di tesi (di Master o dottorato)
lecture	orale	DID	lezione frontale rivolta agli studenti
software demonstration	orale	DID	dimostrazione di software rivolta a studenti
documentary	orale	DIV	documentario rivolto al pubblico generale
podcast	orale	DIV	programma di radio o registrazione audio rivolta al pubblico generale
public lecture	orale	DIV	conferenza tenuta da un accademico o un professionista rivolta al pubblico generale
project pitch	Orale	PROF / DID	presentazione di un progetto in un contesto professionale (PROF) o didattico (DID)

Tab. 1. Generi inclusi nei corpora SEAH

Conclusa la progettazione, si è passati alla fase di raccolta, classificazione e immagazzinamento dei testi. Queste attività sono state portate a termine in maniera autonoma dai gruppi di lavoro di ciascuna lingua, che hanno provveduto anzitutto a recuperare i testi e a

salvarli in formato *.doc*, *.docx* o *.txt*. Quando il testo sorgente era un pdf, questo passaggio ha richiesto operazioni di “pulizia” dei testi per garantirne l’integrità formale, la qualità e la rilevanza ai fini del progetto (eliminando, ad esempio, tabelle, intestazioni di pagina, note e bibliografia finale). I testi orali sono stati dapprima trascritti automaticamente con l’ausilio della funzione Trascrizione della versione online di Word (Microsoft 365); successivamente, l’output della trascrizione automatica è stato sottoposto a *post-editing* manuale, correggendo gli errori del riconoscimento vocale e seguendo delle convenzioni stabilite per uniformare la trascrizione delle idiosincrasie proprie dell’oralità, quali ad esempio pause, interruzioni, frammenti non chiari, ripetizioni, errori, ecc.

I file così creati sono stati salvati su un drive condiviso, etichettando ciascun testo in base a un set di 10 classi di metadati (cf. Tab. 2), che permettono – in fase di interrogazione del corpus – il recupero automatico di testi e la creazione di subcorpora per ricerche mirate su specifici domini o generi.

Metadati	Descrizione
<i>Id</i>	codice alfanumerico di 8 cifre identificativo di ciascun testo (ad es., FR000001, DE000659, ecc.)
<i>Language</i>	lingua (DE, ES, FR, IT, RU)
<i>Mode</i>	scritto o orale
<i>Domain</i>	ACCAD / DID / PROF / DIV
<i>Genre</i>	generi selezionati dalla lista riportata in Tabella 1
<i>Title</i>	breve titolo identificativo dei contenuti del testo
<i>Author type</i>	categoria professionale dell’autore del testo (accademico, media, professionista, studente)
<i>Source</i>	informazioni utili per rintracciare il testo e le sue fonti (ad es., riferimento bibliografico)
<i>Size</i>	numero approssimativo di parole contenute nel testo
<i>Url</i>	solo per testi reperiti online, url della pagina web

Tab. 2. Metadati utilizzati per la classificazione dei testi all’interno dei corpora SEAH

La terza e ultima fase ha riguardato l’annotazione e indicizzazione del corpus, che sono state realizzate con il software di gestione di corpora Sketch Engine (Kilgarriff *et al.* 2014). Queste operazioni hanno permesso di corredare i corpora di lemmatizzazione e annotazione morfosintattica, attribuendo a ciascuna parola un’etichetta che ne identifica il lemma (o forma base di riferimento), la categoria morfosintattica cui appartiene e ulteriori informazioni morfologiche (genere e numero per sostantivi e aggettivi, tempi e persone verbali per i verbi, ecc.). Utilizzando le funzionalità di creazione dei corpora di Sketch Engine, lemmatizzazione e annotazione morfosintattica sono state realizzate con il software FreeLing (Padr /Stanilovsky 2012) per il francese, l’italiano e lo spagnolo, con RFTTagger (Schmid/Laws 2008) per il corpus in lingua tedesca e con MULTEXT-East (Erjavec 2012, 2017) per la lingua russa.

I corpora cos  annotati e indicizzati sono, infine, stati caricati e resi disponibili – dalla pagina <https://corpora.unich.it/sito/seah-corpora-it.html> – sulla piattaforma

NoSketch Engine (Rychlý 2007), strumento di interrogazione *open source* per la gestione dei corpora che consente in maniera semplice e intuitiva di generare elenchi di parole, recuperare collocazioni basate su diverse misure statistiche, o ancora studiare il comportamento di espressioni polirematiche più complesse, molto utile per osservazioni lessico-grammaticali.

### 3 CORPUS ES\_SEAH<sup>5</sup>

Il corpus di lingua spagnola si concentra su cinque ambiti tematici corrispondenti ad altrettante aree lessicali che saranno oggetto dei moduli di lingua sviluppati nella seconda fase del progetto, ovvero:

- Ristrutturazione
- Edifici ed opere di ingegneria civile
- Edifici non residenziali
- Edifici residenziali
- Materiali da costruzione

Le aree tematiche sono rappresentate da testi appartenenti ai generi e domini di seguito indicati:

Canale	Dominio	Genere	Token
scritto	ACCAD	monograph	65.910
	ACCAD	research article	5.118
	ACCAD	thesis	87.573
	DID	teaching material	10.310
	DID	course assignment	165.049
	PROF	project description	345.928
	PROF	project report	178.602
	PROF	material description	23.967
<b>Tot. scritto</b>			<b>882.457</b>
orale	ACCAD	thesis defense	13.532
	DIV	podcast	5.005
	DIV	public lecture	8.544
	PROF	project pitch	140.513
<b>Tot. orale</b>			<b>167.594</b>
<b>Tot. ES_SEAH</b>			<b>1.050.051</b>

Tab. 3. Composizione del corpus spagnolo ES\_SEAH

5 Della progettazione e raccolta testi del corpus di lingua spagnola si è occupata l'unità di ricerca della Universidad Politécnica de Madrid (Spagna), nelle persone di Paloma Úbeda Mansilla (Responsabile di unità), María Luisa Escribano Ortega e Elena Romero Cañabate, in sinergia con Claudia Colantonio dell'Università di Chieti-Pescara.

Come dimostrano i dati in tabella, rispetto all'architettura iniziale che prevedeva una proporzione del 75%-25% delle componenti scritte e orali, la composizione finale è sbilanciata a favore dei testi scritti, che rappresentano l'84% del corpus, mentre ai testi orali è riservato il 16%. La difficoltà a reperire testi adatti al livello di competenze linguistiche dei discenti, nonché le complessità legate al processo di trascrizione e *post-editing* hanno reso l'acquisizione dei testi orali particolarmente laboriosa, limitando le dimensioni di questa componente.

Per quanto riguarda la distribuzione dei domini e dei generi, la componente scritta è costituita prevalentemente da testi professionali (62%), in particolare 'descrizioni di progetto', mentre il dominio accademico (monografie e articoli scientifici) e didattico (consegne di progetto e materiale didattico) occupano rispettivamente il 18% e il 20%.

La componente orale consta quasi totalmente di testi professionali (84%, presentazioni di progetto), con proporzioni residuali di testi accademici (8%, discussioni di tesi) e testi divulgativi (8%, podcast e conferenze pubbliche).

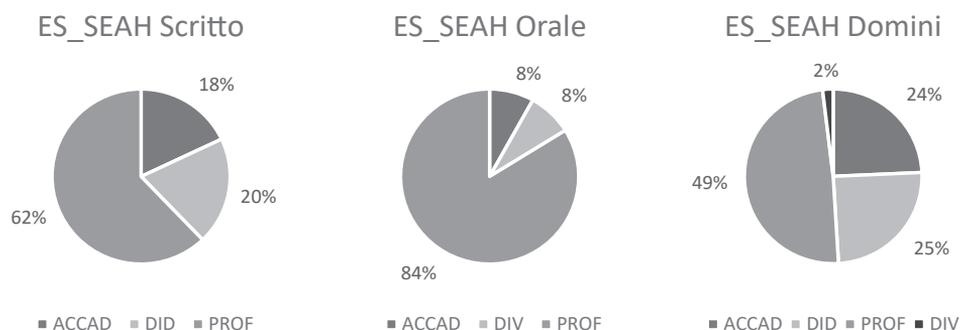


Fig. 1. Distribuzione dei domini nella componente scritta e orale e nel corpus ES\_SEAH nel suo insieme

Nel suo complesso, il corpus è costituito per metà (49%) da testi professionali, mentre la restante parte è divisa quasi perfettamente a metà tra dominio accademico (24%) e didattico (25%). La proporzione di testi divulgativi è minima (2%), rispondendo così alle esigenze di apprendimento del *target group* del progetto, costituito di futuri professionisti in formazione, cui si richiedono competenze linguistiche nella comprensione e produzione di testi professionali e accademici (per poter acquisire i contenuti specializzati della loro area di studio), nonché la capacità di gestire materiale didattico per inserirsi con successo nei corsi dell'università ospitante.

### 3.1 Uso del corpus ES\_SEAH: lessico dei materiali da costruzione

La realizzazione di materiali didattici in lingua spagnola si basa sia su usi indiretti del corpus, sia su usi diretti (McEnergy/Xiao 2011)<sup>6</sup>.

6 Gli usi indiretti riguardano lo sfruttamento del corpus per la selezione dei tratti lessico-grammaticali da includere nelle attività didattiche proposte, a partire da considerazioni legate alla frequenza di

Per questioni di spazio, tuttavia, nel presente articolo ci soffermeremo solo su alcuni usi diretti del corpus ES\_SEAH riconducibili alla pratica didattica del “teaching to exploit [corpora]” (Leech 1997), riferita alla elaborazione di materiali che incoraggino gli studenti a utilizzare i corpora in prima persona per guidarli nell’apprendimento di strutture linguistiche o elementi lessicali specifici e per dotarli delle competenze necessarie a utilizzare i corpora autonomamente. Questo approccio mira a promuovere un metodo didattico centrato sullo/a studente/ssa (McEnery et al. 2006), sui suoi bisogni di apprendimento, sulla sua iniziativa e curiosità, in linea con le proposte del *corpus-aided discovery learning* (ad es., Gavioli 2001, Huang 2011) e le esigenze del contesto di auto-apprendimento dei moduli didattici del progetto SEAH.

Nell’ambito dei materiali didattici per lo spagnolo, gli usi diretti del corpus vengono proposti all’interno della sezione *Mi glosario*, in cui allo studente si richiede la realizzazione di attività volte all’acquisizione di lessico specializzato (ad es., il lessico delle infrastrutture urbane) e di meccanismi morfosintattici particolarmente produttivi all’interno del linguaggio dell’AC (ad es., specifiche trasformazioni morfologiche per la formazione di parole).

Obiettivo degli esempi che si illustrano di seguito (tratti dalla sezione *Mi glosario* della prima unità) è quello di guidare gli studenti a costruirsi un glossario personale relativo ai materiali da costruzione impiegati nell’architettura sostenibile, utilizzando diverse funzionalità di ricerca del corpus per esplorare due strutture linguistiche produttive in questo senso: le sequenze ‘sostantivo + aggettivo’ (ad es., *piedra arenisca*, *vidrio traslúcido*, ecc.) e ‘sostantivo + *de* + sostantivo’ (ad es., *madera de cedro / pino / ecc.* o *panel de acero / aluminio / ecc.*).

Nella prima attività si chiede agli studenti di usare la sintassi di ricerca CQL per recuperare la sequenza ‘*piedra* + AGGETTIVO’ (Fig. 2), ordinando la risultante concordanza in ordine alfabetico in base all’aggettivo (Fig. 3).

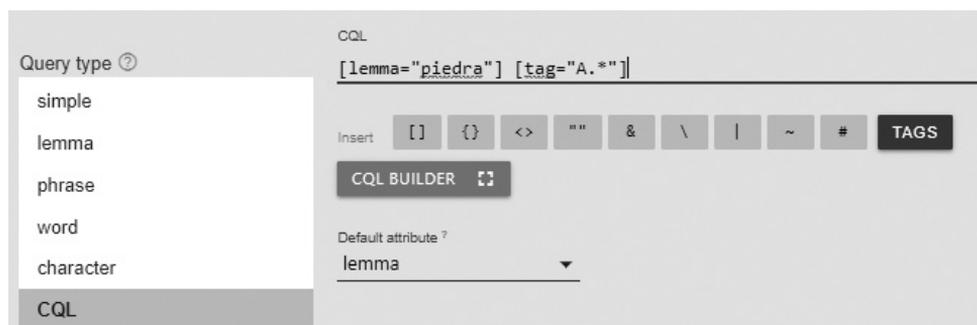


Fig. 2. Uso della sintassi CQL per il recupero della sequenza ‘*piedra* + AGGETTIVO’

---

elementi lessicali e strutture grammaticali nel linguaggio specializzato in analisi. Rientrano negli usi indiretti anche l’estrazione di frasi e testi dal corpus per la messa a punto di esercizi (ad es., di riempimento) sulla base di materiale autentico.

CONCORDANCE

cgj [emma+“piedra”][tag="A"] 65 (0.94 par.millon)

Details	Left context	KWIC	Right context
<input type="checkbox"/>	doc#31 streno, machihembrada	PIEDRA ARTIFICIAL	Se colocarán elementos de piedra artificial en alfeizares, coronamientos de muros...
<input type="checkbox"/>	doc#47 jecoes de superficie mayor de 2 m²	PIEDRA NATURAL	CON MORTERO DE CEMENTO COMO MATERIAL DE AGARRE...
<input type="checkbox"/>	doc#62 siperza visual y táctil confirma la sensación del exquisito gusto	PIEDRA TOSCA	PARQUE DE PIEDRA TOSCA...
<input type="checkbox"/>	doc#62 la sensación del exquisito gusto	PIEDRA TOSCA	El proyecto de las Bodegas i...
<input type="checkbox"/>	doc#62 eno, paisaje y construcción, orden y geometría irregular, desarrolladas en el Parque de	Piedra Tosca	en Les Preses (2003-2005), y en el proyecto simultáneo para el Parque de La Arboleda...
<input type="checkbox"/>	doc#67 ntada de negroCapa de compresiónDoble lámina asfálticaSablón compactado e=20cm	Piedra granítica	Ø20cm armada en seco con mallazo 15x15cm Ø10mmGravas graníticas de reparto ca...
<input type="checkbox"/>	doc#45 slamiento de ruidos según DB HR	Piedra natural	Solado 2 Hormigón visto...
<input type="checkbox"/>	doc#23 acabadas con encalado, y de 48°C en las zonas en las que se ha eliminado, dejando la	piedra arenisca	de color oscuro a la vista...
<input type="checkbox"/>	doc#67 isgo más excepcional de este descubrimiento fue el carácter megalítico de las losas de	piedra arenisca	, y especialmente su profusa decoración geométrica, esculpida en bajorrelieve...
<input type="checkbox"/>	doc#73 a	piedra arenisca	local...
<input type="checkbox"/>	doc#73 debidamente ligada a la hoja interior, se ejecuta con muros de 20 y 15 cm de grosor (la	piedra arenisca	local que define el paisaje construido de la isla)...
<input type="checkbox"/>	doc#74 y dos capas de mampostería: en el interior, bloques de hormigón claro; y en el exterior,	piedra arenisca	local de Marés...
<input type="checkbox"/>	doc#13 adrado construido --, y el man tenimiento es prácticamente nulo	piedra artificial	de nuestro tiempo, tiene además un valor simbólico...
<input type="checkbox"/>	doc#31	piedra artificial	en alfeizares, coronamientos de muros y petos, así como en peanas de puertas y venal...
<input type="checkbox"/>	doc#33 rtas y ventanas para doble acristalamiento 6+12+6	piedra artificial	con goterón y desague...
<input type="checkbox"/>	doc#34 rientación sur, y 5+6+6 tipo Climat con Planitherm	piedra artificial	con goterón y desague...
<input type="checkbox"/>	doc#35 ertas y ventanas para doble acristalamiento 6+6+6	piedra artificial	con goterón y desague...

Fig. 3. Concordanza della sequenza ‘piedra + AGGETTIVO’ ordinata alfabeticamente in base all’aggettivo

L’analisi della concordanza così ordinata facilita una visualizzazione degli aggettivi che modificano ‘piedra’, permettendo agli studenti di realizzare altre attività finalizzate all’ampiamiento del lessico e al suo consolidamento attraverso la sistematizzazione delle varie unità identificate in base a criteri semantici (Fig. 4):

En el cuadro de abajo escribe los adjetivos que se usan en combinación con “piedra”:

artificial, natural, tosca, ...

Clasifica los adjetivos según el tipo de propiedad que indican:

Material	Tono/color	Propiedad	Tipo
Piedra	blanca ...	porosa ...	granítica ...

Fig. 4. Attività di ampiamiento e consolidamento del lessico attraverso la sistematizzazione delle varie unità lessicali identificate nella concordanza

Per incoraggiare l’apprendimento autonomo, si chiede poi agli studenti di ripetere la procedura per esplorare il lessico relativo ad altri sostantivi (Fig. 5), in modo da facilitare l’acquisizione della routine di ricerca e di favorire ulteriormente l’ampiamiento lessicale:



#### 7.5. ¿Y QUÉ PROPIEDAD TIENEN LOS DEMÁS MATERIALES?

Close ▼

Ahora haz tú la búsqueda (siguiendo los pasos ilustrados en 3.1) con los siguientes materiales, para identificar adjetivos que determinan las propiedades de cada uno:

LADRILLO, query [lemma="ladrillo"] [tag="A.\*"]

claro, natural ...

VIDRIO, query [lemma="vidrio"] [tag="A.\*"]

doble, fijo...

ALUMINIO, query [lemma="aluminio"] [tag="A.\*"]

fijo, lacado...

Fig. 5. Attività per promuovere l'uso autonomo del corpus

## 4 CORPUS IT\_SEAH<sup>7</sup>

Allo stato attuale, le dimensioni del corpus di lingua italiana ammontano a 1.123.563 *token*, 968.408 parole e 292 testi (orali e scritti) concernenti le seguenti quattro aree tematiche, attorno alle quali sono stati sviluppati i materiali glottodidattici:

- Storia dell'architettura
- Urbanistica
- Restauro architettonico
- Architettura sostenibile

Nella Tab. 4 è indicata la suddivisione dei testi raccolti in base ai vari generi testuali e ai domini selezionati in corrispondenza dei criteri generali di *corpus design* sopra illustrati (§2):

<sup>7</sup> Della compilazione del Corpus IT\_SEAH si è occupata l'unità di ricerca dell'Università di Chieti-Pescara, nelle persone di Mariapia D'Angelo (Responsabile di unità), Paolo Petricca e Claudia Colantonio, in collaborazione con Claudio Varagnoli del Dipartimento di Architettura, che qui ringraziamo per la collaborazione nella selezione e nel reperimento dei testi raccolti.

Canale	Dominio	Genere	Token
scritto	ACCAD	monograph	52.306
	ACCAD	research article	350.032
	ACCAD	thesis	188.937
	ACCAD	thesis abstract	1.337
	DID	teaching material	97.305
	PROF	project description	143.494
	PROF	project report	18.102
	PROF	material description	92.693
<b>Tot. scritto</b>			<b>944.206</b>
Orale	DID	lecture	147.195
	DIV	public lecture	4.434
	DIV	podcast	27.588
	DIV	documentary	140
<b>Tot. orale</b>			<b>179.357</b>
<b>Tot. IT_SEAH</b>			<b>1.123.563</b>

Tab. 4. Composizione del corpus italiano IT\_SEAH

In analogia agli altri *corpora* del progetto, sul piano diamesico nel corpus IT\_SEAH dominano i testi scritti (79%) rispetto a quelli orali (21%), per ragioni connesse al reperimento di documenti autentici adeguati al livello di competenza linguistica degli apprendenti e alla laboriosità del processo di trascrizione. Al dominio accademico appartiene il 52% dei testi (monografie, articoli scientifici, tesi), al dominio didattico il 22% (lezioni, dispense, ecc.), il 23% è costituito da generi del dominio professionale (report di progetto, descrizione di progetto, descrizione di materiali), mentre solamente il restante 3% afferisce al dominio divulgativo (podcast, lezioni pubbliche, documentari), in quanto si è preferito dare priorità ai generi testuali con cui gli studenti in mobilità saranno maggiormente chiamati a confrontarsi (Fig. 6).



Fig. 6. Distribuzione dei canali e dei domini del corpus IT\_SEAH

Sempre in consonanza con le finalità del progetto, tra i testi orali raccolti predomina il genere della lezione accademica nel dominio didattico, al quale afferiscono l'82% dei testi, mentre il restante 18% è di carattere divulgativo e comprende documentari, podcast e lezioni pubbliche (Fig. 7). Nell'ambito dei testi scritti, infine, la disponibilità di materiali idonei attinenti alle aree tematiche sviluppate nelle unità didattiche di SEAH ha determinato il prevalere dei generi testuali appartenenti al dominio accademico (63%) riguardo a quelli del dominio professionale (27%) e didattico (10%). In generale, nel novero dei dodici generi testuali che figurano in IT\_SEAH, il più rappresentato è quello dell'articolo scientifico, una tipologia generalmente presente nei programmi di studio delle discipline scelte dagli *incoming students*, quali restauro architettonico e storia dell'architettura.



Fig. 7. Distribuzione dei domini nei testi scritti e nei testi orali del corpus IT\_SEAH

#### 4.1 Uso del corpus IT\_SEAH: unità lessicali polirematiche

Nell'ambito dell'italiano come lingua seconda non sono mai stati realizzati materiali glottodidattici *corpus-based* liberamente accessibili per l'insegnamento del linguaggio accademico per scopi speciali. Le risorse linguistiche sviluppate nel quadro del Progetto SEAH rappresentano al momento l'unica offerta di questo tipo in riferimento al macro-settore dell'AC. Come per le altre lingue del progetto (§ 3, 5), dal corpus specialistico IT\_SEAH è stata estratta la quasi totalità dei testi e delle frasi utilizzati nelle attività didattiche, inoltre sono stati elaborati esercizi che implicano la consultazione diretta del corpus da parte degli apprendenti in due sezioni di ciascuna Unit (*Vocabolario e grammatica* e *Il mio glossario*). A seguire riportiamo alcuni esempi di attività contenute nella Unit relativa al restauro architettonico che in un esercizio di comprensione orale presenta l'espressione polirematica "volta a foglio". In un esercizio della sezione *Vocabolario e grammatica* (Fig. 8) agli apprendenti viene chiesto di trovare con l'ausilio del corpus altri esempi di tipologie di volta (ad es., *volta a crociera, a botte, ecc.*).

Nel testo abbiamo incontrato il termine “volta a foglio”. In Italiano esistono molte espressioni formate da “sostantivo + preposizione + sostantivo”, ad esempio *sala da pranzo*, *volta a foglio*. Con il sostantivo “volta” si trovano molte combinazioni. Usa il corpus IT\_SEAH (<https://www.seahproject.eu/resources.php>), segui le istruzioni e trova almeno tre tipi di volta.

#### ISTRUZIONI:

- 1) scegli la funzione “CONCORDANCE” -> “Advanced”,
- 2) scegli query type: CQL
- 3) Copia e incolla: [lemma="volta"] [lemma="a"] [tag="N.\*"]
- 4) completa l’esercizio e scrivi almeno 3 tipi di “volta”:

volta a ....

The screenshot shows the Sketch Engine interface. The top navigation bar includes 'BASIC', 'ADVANCED', and 'ABOUT'. The 'ADVANCED' tab is active, showing the 'Query type' dropdown set to 'CQL'. The query input field contains the CQL query: `[lemma="volta"] [lemma="a"] [tag="N.*"]`. Below the query input, there are buttons for 'Insert', 'TAGS', and 'CQL BUILDER'. The 'Subcorpus' dropdown is set to 'none (the whole corpus)'. The 'Filter context' and 'Text types' dropdowns are also visible.

The bottom part of the screenshot shows the results of the query in a KWIC concordance view. The query is `cql [lemma="volta"] [lemma="a"] [tag="N.*"]` with 72 results (64.06 per million). The results are displayed in a table with columns for 'Details', 'Left context', 'KWIC', and 'Right context'. The table contains 13 rows of results, each showing a document ID, a snippet of text, and the KWIC concordance.

Details	Left context	KWIC	Right context
1	doc#2 ) leggera, e molto utilizzate sono appunto le	volte a crociera	, i pilastri a fascio che decorano i pilastri in m
2	doc#2 questi pilastri dove andavano a scaricare le	volte a crociera	</s><s> Tra un pilastro e l'altro di fatto la m
3	doc#11 upola centrale in ferro e vetro e anche delle	volte a botte	sulle vie principali d'accesso, e con lo stessc
4	doc#13 ealizzato non soltanto gli archi, ma anche le	volte a botte	(che sono veramente degli archi, diciamo co
5	doc#13 nente degli archi, diciamo così, allungati), le	volte a crociera	che praticamente sono le intersezioni di due
6	doc#13 e praticamente sono le intersezioni di due e	volte a botte	, e poi dopo anche la volta a cupola che di fa
7	doc#13 ri di due e volte a botte, e poi dopo anche la	volta a cupola	che di fatto diventava una vera e propria sen
8	doc#13 ria semisfera. </s><s> Importantissima è la	volta a cupola	realizzata per la copertura del Pantheon, che
9	doc#16 e venivano utilizzate. </s><s> Si parte dalla	volta a botte	, la più semplice, la volta a crociera come int
10	doc#16 i parte dalla volta a botte, la più semplice, la	volta a crociera	come intersezione di due volte a botte per ar
11	doc#16 la volta a crociera come intersezione di due	volte a botte	per arrivare poi alla volta a cupola. </s><s> I
12	doc#16 ione di due volte a botte per arrivare poi alla	volta a cupola	. </s><s> La volta a cupola in particolare, è r
13	doc#16 r arrivare poi alla volta a cupola. </s><s> La	volta a cupola	in particolare, è molto studiata nel periodo hi

Fig. 8. Uso della sintassi CQL per il recupero della sequenza ‘volta + a + SOSTANTIVO’

Nello svolgere l'esercizio sopra riportato, lo studente reperirà i seguenti esempi: *volta a crociera, volta a vela, volta a botte, volta a cupola, volta a sesto ribassato, volta a ventaglio, volta a lacunari, volta a calotta*. A seguire, l'apprendente dovrà ripetere la sintassi di ricerca con il sostantivo *casa* + *a* + sostantivo e con il sostantivo *pannello* + *di* + sostantivo, al fine di approfondire la costruzione polirematica più produttiva dell'italiano contemporaneo, ovvero quella formata da sostantivo + preposizione + sostantivo (D'Achille 2012: 163).

Infine, nella sezione *Il mio glossario*, un'altra attività di interrogazione di IT\_SEAH ha per oggetto un'altra struttura già incontrata nei testi della Unit, in cui *volta* corrisponde al participio passato del verbo *volgere*, seguito dalla preposizione *a* + verbo *all'infinito*. In tal caso verrà chiesto ai discenti di utilizzare una sintassi di ricerca simile alla precedente, sostituendo però il sostantivo con un verbo per rintracciare esempi di frase in cui il participio passato del verbo *volgere* ha valore finale (Fig. 9).

Attenzione: nell'esercizio precedente (2.6) hai letto "L'operazione volta a conservare..." e anche "qualsiasi intervento volto a rimettere in buono stato...". In questo caso "volto/a" non è un sostantivo, ma è il participio passato del verbo *volgere* ed esprime una finalità, ad esempio: "una scala esterna volta a collegare i due livelli degli edifici" = "una scala esterna che è finalizzata a collegare i due livelli degli edifici"

Usa il corpus IT\_SEAH (<https://www.seahproject.eu/resources.php>), segui le istruzioni e scrivi 5 esempi con il participio passato del verbo "volgere" ed esprime una finalità.

ISTRUZIONI:

- 1) scegli la funzione "CONCORDANCE" -> "Advanced",
- 2) scegli query type: CQL
- 3) Copia e incolla: [lemma="volta"] [lemma="a"] [tag="V.\*"]

Es. Una serie di operazioni volte a rilanciare il turismo....

una risposta progettuale volta a proteggere...

...

Fig. 9. Esempio di attività *corpus-based*

## 5 CORPUS RU\_SEAH<sup>8</sup>

Il corpus di lingua russa comprende materiali riconducibili a tre aree tematiche:

- Storia e teoria dell'architettura
- Progettazione architettonica e Design
- Architettura sostenibile

8 Dell'ideazione e realizzazione del Corpus di Lingua Russa si è occupata l'unità di ricerca dell'Università di Chieti-Pescara, nelle persone di Maria Chiara Ferro (Responsabile di unità) e di Natalia Guseva (Collaboratore), in sinergia con i partners della Polotsk State University (Bielorussia), guidati da Svetlana M. Ljasovič.

I testi selezionati a partire da questi argomenti appartengono ai domini e ai generi indicati qui di seguito:

Canale	Dominio	Genere	Token
scritto	ACCAD	monograph	248.039
	ACCAD	research article	105.743
	ACCAD	thesis abstract	111.699
	DID	handbook	363.482
	DID	teaching material	18.468
	DID	project report	9.548
	DID	project description	11.656
	PROF	trade shows	107.648
<b>Tot. scritto</b>			<b>976.283</b>
orale	ACCAD	thesis defense	38.091
	ACCAD	conference	25.531
	ACCAD	roundtable	5.485
	DIV	lecture	78.109
	DIV	podcast	41.514
	DIV	public lecture	68.921
	DID	project pitch	6.812
<b>Tot. orale</b>			<b>264.463</b>
<b>Tot. RU_SEAH</b>			<b>1.240.746</b>

Tab. 5. Composizione del corpus russo RU\_SEAH

La scelta di circoscrivere reperimento e collazione dei materiali testuali a tre sole aree tematiche è stata dettata dai limiti temporali imposti dal progetto. Il corpus è dunque passibile di ampliamenti, tramite l'introduzione di nuove sezioni tematiche e/o l'arricchimento della selezione di testi inventariati. Due i criteri fondamentali che hanno determinato la scelta dei temi: la rilevanza nel comparto professionale di riferimento e, in considerazione delle verosimili conoscenze del *target group*, la presenza nei piani di studio universitari. L'inclusione di materiali inerenti all'architettura sostenibile è altresì motivata dall'attualità di questo campo d'indagine e dagli sviluppi teorici e pratici che lo stesso sta conoscendo in Europa negli ultimi anni.

La dimensione del corpus ammonta a 1.240.746 tokens, ripartiti nelle due sezioni che lo compongono, testi scritti e testi orali, nella misura del 75% e del 25% rispettivamente. Il minore spazio riservato alla componente orale è dovuto agli oneri del processo di lavorazione del testo orale ai fini del suo inserimento nel corpus nelle fasi di trascrizione, controllo e pulizia della bozza ottenuta con software *speech to text*. Inoltre, anche in questo caso, la tipologia del *target group* è stata determinante: il corpus è inteso *in primis* come strumento al servizio di studenti e professionisti di architettura e ingegneria delle costruzioni che abbiano una conoscenza pregressa della lingua russa almeno di livello B1, che resta tuttavia un livello *low-intermediate*: il discorso

scientifico orale in lingua russa presenta maggiore complessità per l'apprendente straniero rispetto a quello scritto, potendo contemplare al suo interno marche colloquiali o espedienti linguistico-testuali di carattere espressivo-emozionale (Markova 2016: 111).

I documenti caricati sono 170, appartenenti a 15 generi testuali; tutte e 4 le tipologie d'autore contemplate (accademico, professionista, media, studente) sono rappresentate.

Nel corpus scritto tra i testi professionali troviamo monografie e testi impiegati in occasione di mostre e fiere di settore, i testi pedagogico-didattici sono rappresentati da manuali di insegnamento e sussidi didattici, nei *teaching materials* sono ricompresi dispense, programmi di studio, consegne d'esame, linee guida per la realizzazione di lavori di fine corso, guide dello studente; tra i testi scientifico-accademici sono stati selezionati articoli su riviste scientifiche, monografie e tesi.

Nel corpus orale troviamo anzitutto testi di carattere divulgativo, quali conferenze pubbliche, interventi tenuti in occasioni di divulgazione scientifica, podcast; in secondo luogo testi accademici (lezioni, interventi a convegni scientifici, discussione di tesi dottorali).

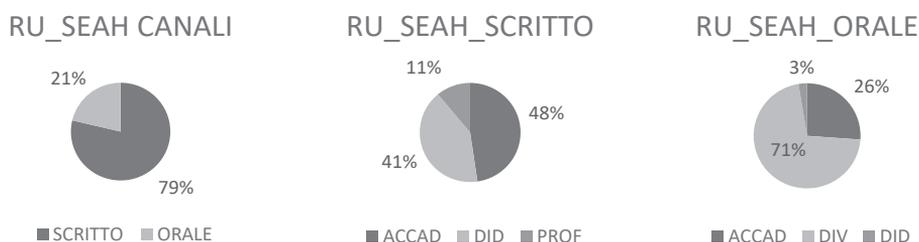


Fig. 10. Distribuzione dei canali (scritto e orale) e domini nel corpus RU\_SEAH

La versione attuale di RU\_SEAH presenta uno sbilanciamento a favore dei testi pedagogico-didattici e scientifici, che tuttavia appare in linea con gli scopi generali del progetto SEAH.

L'apporto del gruppo di lavoro della Polotsk State University (Bielorussia) è stato determinante sia nella fase di ideazione che in quella di realizzazione del corpus RU\_SEAH, in modo particolare per il reperimento di materiali pedagogico-didattici, quali consegne d'esame, piani di studio, elaborati di fine corso degli studenti, incluse alcune registrazioni, che sono stati attinti dall'archivio di quella università, grazie alla disponibilità dei colleghi dei dipartimenti di Architettura e Design e di Ingegneria.

### 5.1 Uso del corpus RU\_SEAH: sviluppo autonomo di glossari

La linguistica dei *corpora* ha prodotto i primi risultati per la lingua russa tra la fine del XX e l'inizio del XXI secolo (Biagini/Bonola/Noseda 2019: 35-36). Nonostante gli apprezzabili tentativi condotti fin dagli anni '60-'70 (cfr. Zacharov 2013), il *Corpus Nazionale della lingua russa* (*Nacional'nyj Korpus Russkogo Jazyka*, NKRJa, <http://ruscorpora.ru>) ha visto la luce a partire dal 2003 (Plungjan/Reznikova/Sičinava 2005; Sičinava 2005) e l'ultimo aggiornamento risale al 2021. Allo stato attuale degli studi

le potenzialità applicative del NKRJa vengono indagate sui versanti dell'analisi contrastiva tra il russo e una o più lingue seconde, e dell'approfondimento delle strutture morfosintattiche, sia per quanto attiene al loro uso, sia al fine di comprenderne con maggiore precisione la semantica. Anche l'utilità dell'impiego del *corpus* nell'ambito della didattica del russo come lingua straniera (*Russkij kak inostrannyj* – RKI) ha iniziato ad essere apprezzata recentemente sia in Russia (si pensi a Dobrušina 2005), che in Italia (si vedano, ad esempio, Biagini 2007 e Benigni 2010)<sup>9</sup>, e uno sviluppo delle applicazioni dei *corpora* in ordine all'insegnamento dei linguaggi di specializzazione non appare al momento essersi sviluppata.

Anche nella realizzazione dei moduli didattici SEAH per la lingua russa il corpus è stato utile anzitutto come archivio di materiali autentici specializzati cui attingere nella ricerca di testi *pivot* per l'elaborazione delle unità didattiche; in questo caso, allo scopo di meglio valutare il livello linguistico dei paragrafi individuati, l'esperienza delle autrici nella didassi RKI si è talora affiancata all'impiego di *Tekstometr*<sup>10</sup>. Inoltre, determinate attività proposte nei moduli didattici risultano *corpus-based* nel senso che presuppongono un utilizzo guidato del *corpus* da parte del discente.

In relazione al livello di partenza nella conoscenza della lingua russa da parte dei discenti e in considerazione degli aspetti linguistici che assumono maggiore rilievo nell'insegnamento delle lingue di specializzazione, l'uso del *corpus* si è rivelato proficuo per l'implementazione di attività di arricchimento lessicale, in vista della creazione di glossari personalizzati.

Il testo *pivot* scelto per l'elaborazione dei *teaching materials* relativi alla progettazione reca il sostantivo *fasad* (фасад, facciata). La consegna dell'esercizio guida lo studente a reperire nel *corpus* gli aggettivi con i quali il lemma si combina:

---

9 Poiché la disamina degli studi in queste direzioni eccede i limiti di interesse (e di spazio) del presente saggio, ci limitiamo qui a indicare, senza pretesa di esaustività, i nomi degli studiosi italiani che maggiormente si sono occupati dell'argomento, affinché il lettore interessato possa recuperare la loro bibliografia: V. Benigni, F. Biagini, A. Bonola, V. Nosedà.

10 Cfr.: <https://textometr.ru/>, software che fornisce una serie di informazioni, relative in modo particolare al lessico impiegato, utili a stabilire il livello linguistico di un testo in base alla classificazione QCER.

Utilizza il corpus di lingua russa (<https://www.seahproject.eu/resources.php>) per la ricerca del lessico collegato con il lemma фасад:

- vai a: Concordance > Advanced;
- scegli: Query Type: CQL;
- copia e incolla questa stringa: [tag="A.\*"] [lemma="фасад"];
- in tal modo potrai sapere con quali aggettivi viene impiegato il sostantivo "фасад";
- scrivi qui di seguito (nella forma del nominativo singolare) gli aggettivi che compaiono nei primi 40 risultati della ricerca. Conosci il loro significato? Se necessario, cercalo sul dizionario.

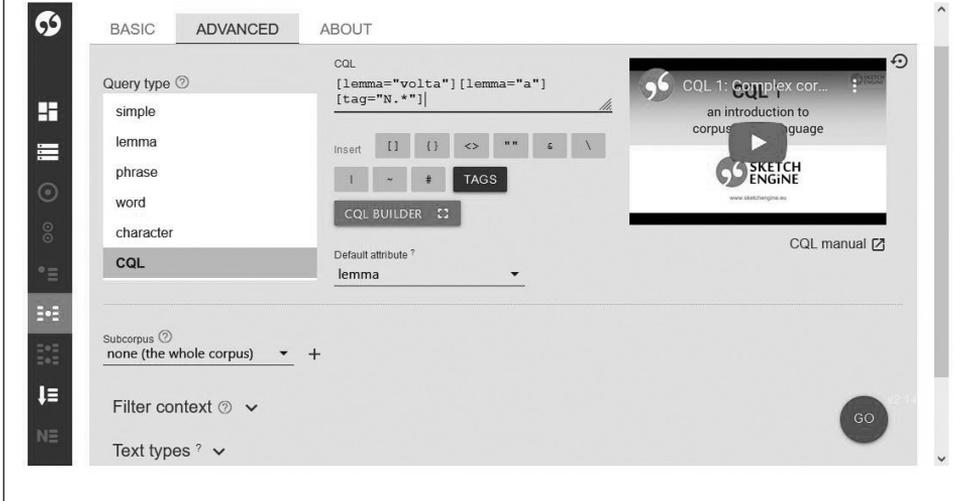


Fig. 11. Interrogazione del corpus tramite sintassi CQL per il recupero della sequenza “aggettivo+ фасад”.

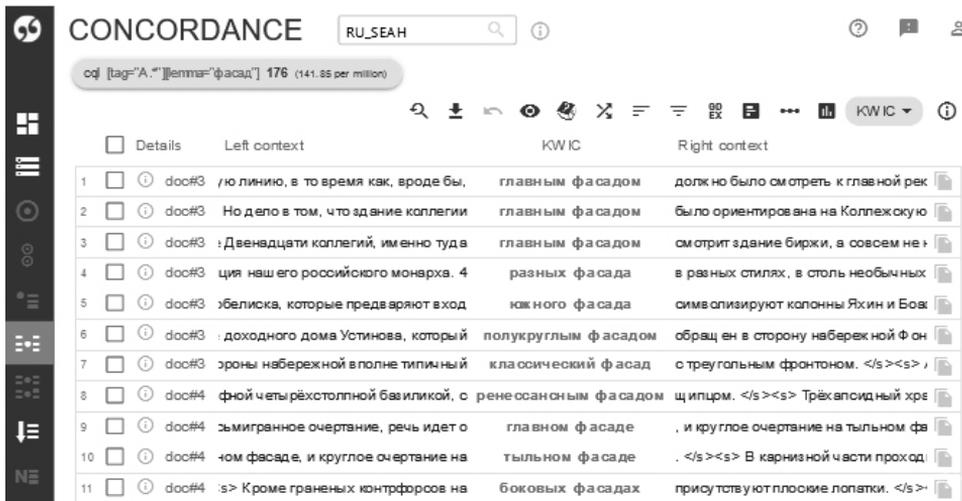


Fig. 12. Risultati 1-11 ottenuti tramite l'interrogazione del corpus sopra descritta

Il numero totale di occorrenze restituite dall'interrogazione è di 176; di conseguenza, l'attività può essere calibrata diversamente, nel senso di un maggiore ampliamento, in base alle finalità perseguite. I primi 40 risultati recano (escluse le ripetizioni) i 25 aggettivi qui riportati:

главный (glavnyj), разный (raznyj), южный (južnyj), <b>полукруглый</b> (polukruglyj), классический (klassičeskij), <b>ренессансный</b> (renessansnyj), <b>тыльный</b> (tyl'nyj), <b>боковой</b> (bokovoj), зелёный (zelėnyj), <b>вентируемый</b> (ventiruemyj), северный (severnij), <b>западный</b> (zapadnyj), задний (zadnij), <b>внутренний</b> (vnutrennyj), <b>утепленный</b> (uteplennyj), прекрасный (prekrasnyj), <b>остеклённый</b> (osteklennyj), <b>парадный</b> (paradnyj), уникальный (unikal'nyj), живой (živoj), <b>навесной</b> (navesnoj), новый (novyj), <b>стеклянный</b> (steklannyj), активный (aktivnyj), <b>внешний</b> (vnešnij)
---

Fig. 13. Risultato del compito assegnato

Considerato che nei contesti d'uso i 25 aggettivi così ottenuti compaiono quasi sempre in casi diversi dal nominativo e talvolta al plurale, la richiesta di elencarli risalendo al nominativo singolare, insieme all'utilità pratica di recuperare la forma presente nel dizionario nell'eventualità di dover rintracciare il significato del lemma, ha anche quella didattica di ripasso e consolidamento della flessione nominale, che costituisce uno degli obiettivi fondamentali della didassi RKI non solo nei livelli elementare (TEU/ A1) e di base (TBU/A2), ma anche in quello intermedio (I/B1). Nello specifico, per svolgere correttamente il compito assegnato, il discente dovrà ricordare e impiegare le regole che presiedono alla declinazione aggettivale, ivi incluse le variazioni di terminazione che si attuano in presenza di consonanti gutturali (qui la -k- di *klassičeskij*) e l'esistenza degli aggettivi in -n- debole (qui rappresentati da *zadnij* e *vnešnij*). Il compito comporta qui una maggiore complessità essendo applicato al lessico di specializzazione.

Il grassetto evidenzia i 13 aggettivi che non sono contenuti nel minimo lessicale del primo livello del TRKI (*Test po Russkomu jazyku kak inostrannomu* - Andrušina 2015), corrispondente al B1 del QCER, confermando l'utilità di una simile esercitazione in vista dell'ampliamento del bagaglio lessicale. L'insieme degli attributi rinvenuti nel *corpus* permette un ulteriore approfondimento nel senso di una organizzazione dei lemmi per gruppi semantici significativi nell'ambito AC, strategia di memorizzazione tra le più comuni. Si chiede, infatti, al discente di individuare quali tra i 25 lemmi recuperati si riferiscono ai seguenti ambiti:

Ambito semantico	Lemmi
Punti cardinali	южный, северный, западный
Stili architettonici	классический, ренессансный
Posizione	тыльный, задний, внутренний, внешний, боковой
Materiale	остеклённый, стеклянный

Tab. 6. Lemmi ordinati per ambito semantico

Come si vede, sia nella fase di reperimento dei lemmi tramite l'interrogazione del *corpus*, sia in quella di organizzazione dei risultati per ambiti semantici rilevanti nella sfera AC, attività di questo tipo concorrono in maniera apprezzabile ad accrescere e sistematizzare il bagaglio lessicale dell'apprendente. Nell'ottica poi di pervenire alla redazione di un glossario autonomo e opportunamente calibrato sui bisogni specifici, a partire dalla tabella così composta, il discente può in autonomia aggiungere altri aggettivi a lui noti nella lingua madre reperendone il traducevole russo: ad esempio, per l'ambito "Punti cardinali" manca l'aggettivo "orientale" (*vostočnyj*, восточный), in ordine alla posizione non compaiono gli aggettivi "superiore", "inferiore" e "anteriore", e così via.

## 6 CONSIDERAZIONI CONCLUSIVE E PROSPETTIVE DI SVILUPPO

Come si vede, i *corpora* presentati, annotati linguisticamente e resi disponibili gratuitamente online, risultano una risorsa linguistica inedita per le lingue del progetto nel campo dell'AC.

Innovativo è anzitutto il metodo con il quale sono stati prodotti: la consultazione di esponenti dei comparti professionali e dell'ambito accademico, propedeutica all'implementazione di SEAH, ha permesso una mappatura dei reali bisogni linguistici di architetti e ingegneri, pervenendo all'identificazione di quei domini, generi testuali e pratiche comunicative che costituiscono un ostacolo alla mobilità studentesca e professionale. Su queste premesse si auspica che i moduli multilingui *corpus based* sviluppati nella seconda fase del progetto contribuiscano a favorire l'apprendimento della lingua straniera di specializzazione nel settore dell'AC, incoraggiando studenti e professionisti a proiettare studi e attività in ambito internazionale.

Come esemplificato i *corpora* SEAH si rivelano un valore aggiunto per l'elaborazione dei materiali didattici per almeno due motivi. Da un lato, l'accesso a una grande raccolta di testi rappresentativi del linguaggio specializzato dell'AC permette il recupero di unità terminologiche difficilmente reperibili in materiali di consultazione della lingua generale (ad es., sp. *ladrillo silicocalcáreo*, *vidrio semirreflectante*, it. *volta a lacunari*, *volta a calotta*, ru. *bokovoj fasad*, *osteklënnij fasad*, ecc.); inoltre, la possibilità di osservare l'uso dei vari lemmi nel contesto della frase favorisce la formulazione di ipotesi e la deduzione di regole, che assisteranno gli studenti nell'acquisizione di altre unità lessicali e dei meccanismi morfosintattici in cui esse partecipano. Dall'altro lato, l'apprendimento di diverse funzionalità di ricerca del corpus (ad es., ricerche avanzate con sintassi CQL, ordinamento delle concordanze e elaborazione di liste di frequenza in base a diversi criteri, analisi di collocati, ecc.) apre la strada a un uso autonomo del corpus da parte dell'apprendente, che potrà consultare la risorsa per risolvere i dubbi linguistici cui di volta in volta si troverà a far fronte.

In una prospettiva più ampia, i *corpora* SEAH costituiscono un utile strumento per l'approfondimento della conoscenza del patrimonio architettonico francese, italiano, russo, spagnolo e tedesco, e delle pratiche professionali ad esso connesse, obiettivo al quale la politica culturale dell'Unione europea guarda con particolare interesse: la

valorizzazione delle singole culture nazionali, infatti, costituisce il presupposto per “operare a livello transnazionale e promuove[re] la circolazione transfrontaliera delle opere culturali e la mobilità degli operatori culturali” (<http://www.europacreativa-media.it/europa-creativa>). Le risorse linguistiche di SEAH contribuiscono ad una maggiore visibilità delle specificità delle espressioni e tradizioni architettoniche nazionali, facilitando la creazione di reti di collaborazione a livello europeo.

Oltre a ciò, in linea con le finalità dell'*open access*, suddetti *corpora* rappresentano una risorsa fruibile da parte di una varietà di utenti, ben oltre gli scopi del progetto SEAH, quali traduttori nel campo dell'AC, programmatori per la creazione di applicazioni di *natural language processing*, inclusi software di scrittura tecnica, nonché per la redazione di dizionari, glossari, ecc.

Infine, i *corpora* esistenti possono essere implementati progressivamente sia includendo altre aree tematiche dell'AC, sia contemplando settori disciplinari contigui (ad esempio “beni culturali”, “archeologia” sul fronte storico-artistico, i diversi comparti dell'ingegneria civile sul versante più tecnico, ecc.); la metodologia che presiede allo sviluppo dei materiali glottodidattici multilingui e *corpus-based* del progetto SEAH, dal canto suo, può essere in futuro trasferita in altri ambiti accademico-disciplinari e raggiungere gruppi diversi di apprendenti.

### Riferimenti bibliografici

- ANDRJUŠINA, Natal'ja Pavlovna *et al.* (2015) *Leksičeskij minimum po ruskomu jazyku kak inostrannomu. Pervyj sertifikacionnyj uroven'. Obščee vladenie*. Sankt-Peterburg: Zlatoust.
- ASTON, Guy (a cura di) (2001) *Learning with corpora*. Houston: Athelstan.
- BENIGNI, Valentina (2010) “Sposoby ispol'zovanija NKRJA v prepodavanii russkogo jazyka ital'janskim studentam.” In: *Informacionnye i kommunikativnye tehnologii v rusistike: sovremennnoe sostojanie i perspektiva. III meždunarodnaja virtual'naja naučno-praktičeskaja konferencija*. Erevan: Limush, 18-22.
- BERRUTO, Gaetano (1987) *Sociolinguistica dell'italiano contemporaneo*. Firenze: La Nuova Italia.
- BHATIA, Vijay K. (1993) *Analysing Genre: Language Use in Professional Settings*. London/New York: Longman.
- BIAGINI, Francesca (2007) “Sostavlenie upražnenij na upotreblenie ‘začem’ i ‘počemu’ dlja studentov-inostrancev pri pomošči NKRJa.” In: *Nacional'nyj korpus russkogo jazyka i problemy gumanitarnogo obrazovanija*, Moskva: Izdatel'stvo TEIS, 101-108.
- BIAGINI, Francesca/Anna BONOLA/Viviana NOSEDA (2019) “Il corpus parallelo italiano-russo del NKRJa. Progetto di ampliamento, applicazioni e sviluppi”. In: M. C. Bragone/M. Bidovec (a cura di), *Il mondo slavo e l'Europa. Contributi presentati al VI Congresso Italiano di Slavistica. Torino, 28-30 settembre 2016*. Firenze: FUP, 35-45.
- D'ACHILLE, Paolo (2012) *L'italiano contemporaneo*, Bologna: il Mulino.

- DOBRUŠINA, Nina Rolandovna (2005) “Kak ispol'zovat' Nacional'nyj korpus ruskogo jazyka v obrazovanii?” In: *Nacional'nyj korpus russkogo jazyka: 2003–2005*. Moskva: Indrik, 308–329.
- ERJAVEC, Tomaž (2012) “MULTEXT-East: Morphosyntactic Resources for Central and Eastern European Languages.” *Language Resources and Evaluation* 46/1, 131–142.
- ERJAVEC, Tomaž (2017) “MULTEXT-East”. In: N. Ide/J. Pustejovsky (a cura di), 441–462.
- FLOWERDEW, Lynne (2015) “Corpus-based research and pedagogy in EAP: From lexis to genre.” *Language Teaching* 48/1, 99–116.
- GAVIOLI, Laura (2001) “The learner as researcher: Introducing corpus concordancing in the Classroom” In: G. Aston (a cura di), 108–37.
- HALLIDAY, Michael A. K./Ruqaiya HASAN (1985) *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin University Press.
- HINKEL, Eli (a cura di) (2011) *Handbook of Research in Second Language Teaching and Learning*. London/New York: Routledge.
- HYLAND, Ken (2014) “English for Academic Purposes”. In: Constant Leung/Brian V. Street (a cura di), *The Routledge Companion to English Studies*. London: Routledge, 392–404.
- HYLAND, Ken (2016) “General and specific EAP”. In: Ken Hyland/Philip Shaw (a cura di), *The Routledge handbook of English for Academic Purposes*. London: Routledge, 17–29.
- HUANG, Li-Shih (2011) “Corpus-aided language learning.” *ELT Journal* 65/4, 481–484.
- HYLAND, Ken/Marina BONDI (a cura di) (2006) *Academic Discourse across Disciplines*. Bern: Peter Lang.
- IDE, Nancy/James PUSTEJOVSKY (a cura di) (2017) *Handbook of Linguistic Annotation*. New York: Springer.
- JOHNS, Tim (1991) “From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning.” *English Language Research Journal* 4, 27–45.
- KILGARRIFF, Adam/Vít BAISA/Jan BUŠTA/Miloš JAKUBÍČEK/Vojtěch KOVÁŘ/Jan MICHELFEIT/Pavel RYCHLÝ/Vít SUCHOMEL (2014) “The Sketch Engine: Ten Years On.” *Lexicography – Journal of ASIALEX* 1/1, 7–36.
- LEE, David/John SWALES (2006) “A Corpus-Based EAP Course for NNS Doctoral Students: Moving from Available Specialized Corpora to Self-Compiled Corpora”. *English for Specific Purposes* 25/1, 56–75.
- LEECH, Geoffrey (1997) “Teaching and language corpora: A convergence.” In: A. Wichmann/S. Fligelstone/T. McEnery/G. Knowles (a cura di), 1–23.
- MARKOVA, Valentina Alekseeva (2016), *Stilistika russkogo jazyka. Teoretiko-praktičeskij kurs*, Moskva: LENAND.
- McENERY, Tony/Richard XIAO (2011) “What Corpora Can Offer in Language Teaching and Learning”. In: E. Hinkel (a cura di), 364–380.

- McENERY, Tony/Richard XIAO/Yukio TONO (2006) *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- NESI, Hilary/Sheena GARDNER (2012) *Genres Across the Disciplines: Student Writing in Higher Education*. Cambridge: Cambridge University Press.
- PADRÓ, Lluís/Evgeny STANILOVSKY (2012) “FreeLing 3.0: Towards Wider Multilinguality.” In: Nicoletta Calzolari *et al.* (a cura di), *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. Istanbul: ELRA, 2473–2479.
- PLUNGJAN, Vladimir Aleksandrovič/Tat’jana Isidorovna REZNIKOVA/ Dmitrij Vladimirovič SIČINA (2005) “Nacional’nyj korpus ruskogo jazyka: obščaja charakteristika.” *Naučno-tehničeskaja informacija. Ser. 2* 2005/3, 9-13.
- RYCHLÝ, Pavel (2007) “Manatee/Bonito – A Modular Corpus Manager.” In: *1st Workshop on Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University, 65–70.
- SCHMID, Helmut/Florian LAWS (2008) “Estimation of Conditional Probabilities With Decision Trees and an Application to Fine-Grained POS Tagging”. In: *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. Manchester: Coling, 777–784.
- SIČINA Dmitrij Vladimirovič. (2005) “Nacional’nyj korpus ruskogo jazyka: očerk predystorii.” In: *Nacional’nyj korpus ruskogo jazyka: 2003–2005. Rezul’taty i perspektivy*. Moskva: Indrik, 21-30.
- SWALES, John M. (1990) *Genre Analysis: English in Academic and Research Setting*. Cambridge: Cambridge University Press.
- SWALES, John M. (2004) *Research genres. Exploration and applications*. Cambridge: Cambridge University Press.
- WICHMANN, Anne/Steven FLIGELSTONE/Tony McENERY/Gerry KNOWLES (a cura di) (1997) *Teaching and language corpora*. London: Longman.
- ZACHAROV, Viktor (2013) “Corpora of the Russian Language.” In: I. Habernal, V. Matoušek (a cura di), *Text, Speech and Dialogue: Proceedings of the 16th International Conference, TSD 2013, Plzen, Czech Republic, September 1–5, 2013. (Lecture Notes in Artificial Intelligence, 8082)*. Berlin/Heidelberg: Springer-Verlag, 1-13.

#### Riassunto

### I CORPORA SEAH DI COMUNICAZIONE SPECIALIZZATA NEL SETTORE DELL’ARCHITETTURA E DELLE COSTRUZIONI. STRUTTURA, COMPILAZIONE E USI

La mancanza di competenze nel linguaggio accademico-disciplinare costituisce spesso un ostacolo alla mobilità degli studenti. Questo è particolarmente vero nel campo dell’Architettura e delle Costruzioni (AC), in cui il percorso formativo comprende una serie di sotto-domini tecnici che sono spesso definiti da pratiche professionali, tradizioni culturali e quadri giuridici specifici di un dato paese. Con l’obiettivo di favorire la partecipazione ai programmi di scambio, il progetto Erasmus+ SEAH (*Sharing*

*European Architectural Heritage: Innovative language teaching tools for academic and professional mobility in Architecture and Construction*) mira a creare corpora specializzati nel campo dell'AC e moduli linguistici *open access* basati sui suddetti corpora in lingua francese, tedesca, italiana, russa e spagnola. Il contributo presenta il quadro teorico di riferimento, le metodologie e le finalità del progetto SEAH, soffermandosi sui criteri e sulle procedure generali del *corpus design*, con esemplificazioni della compilazione e impiego dei *corpora* per la lingua spagnola, italiana e russa.

**Parole-chiave:** corpora SEAH; glottodidattica *corpus-based*, linguaggio accademico-professionale dell'Architettura e delle Costruzioni, *data-driven learning*, mobilità studentesca

#### Summary

#### THE SEAH SPECIALIZED CORPORA FOR ARCHITECTURE AND CONSTRUCTION: DESIGN, COMPILATION AND USE

The lack of proficiency in technical language required for the study of specialist fields often represents an obstacle to student mobility. This is especially true in the field of Architecture and Construction (AC), where professional training covers a number of technical sub-fields which are often defined by professional practices, cultural traditions and legal frameworks specific to a given country. The Erasmus+ SEAH project (*Sharing European Architectural Heritage: Innovative language teaching tools for academic and professional mobility in Architecture and Construction*) aims at creating specialized corpora in the field of AC and open access corpus-based language learning materials in French, German, Italian, Russian and Spanish with a view to promoting wider student participation in mobility programmes. This paper presents the theoretical framework, methodologies and aims of the SEAH project, focusing on the general criteria and processes that have informed corpus design, giving examples of the compilation and use of the corpora for Spanish, Italian and Russian.

**Keywords:** SEAH corpora, corpus-based language teaching, academic and professional language of Architecture and Construction, data-driven learning; student mobility

#### Povzetek

#### SPECIALIZIRANI KORPUSI SEAH NA PODROČJU ARHITEKTURE IN GRADBENIŠTVA: STRUKTURA, SESTAVA IN UPORABA

Pomanjkanje ustreznih kompetenc v zvezi z akademskim in strokovnim jezikom pogosto predstavlja oviro za mobilnost študentov. To še posebej velja na področju arhitekture in gradbeništva, kjer proces izobraževanja vključuje vrsto specialističnih podpodročij, ki jih določajo razne strokovne prakse, kulturne tradicije in specifični pravni

okviru posameznih držav. Namen projekta Erasmus+ SEAH (*Sharing European Architectural Heritage: Innovative language teaching tools for academic and professional mobility in Architecture and Construction*), ki želi spodbuditi sodelovanje v programih izmenjav, je izgradnja specializiranih korpusov na področju arhitekture in gradbeništva ter učnih gradiv v odprtem dostopu, izdelanih na osnovi teh korpusov v francoščini, nemščini, italijanščini, ruščini in španščini. V članku so predstavljeni temeljni teoretski okvir, metodologija ter cilji projekta SEAH. Posebna pozornost je posvečena kriterijem in postopku zastavitve korpusa; zgledi, ki ponazarjajo sestavo in uporabo korpusov, so v španščini, italijanščini in ruščini.

**Ključne besede:** korpusi SEAH, korpusno podprto poučevanje jezika, akademski in strokovni jezik na področju arhitekture in gradbeništva, učenje na osnovi podatkov, študentska mobilnost