

Arbeitsbericht II des *Istitut Ladin Micurà de Rù:* lexikographische Projekte, Korrektursystem

Jürgen Runggaldier, Marco Forni,
Paolo Anvidalfarei

1. Das Korrektursystem im Internet

Im Jahr 2004 reifte die Idee einer neuen, internetbasierten Korrekturhilfe für das Ladinische. Nach Abschluss einer Konvention zwischen den beiden Kulturinstituten *Istitut Ladin Micurà de Rù* und *Istitut Cultural Ladin Majon di Fascegn* am 10. Februar 2005 war es für die Arbeitsgruppe von *Micurà de Rù* möglich, sich der wertvollen Vorarbeit des Fassaner Instituts zu bedienen. Auf Basis dieser Vorarbeit beauftragten beide Institute die Firma *Open Lab* aus Florenz mit der informatischen Umsetzung.

Majon di Fascegn hatte bereits 1999 mit der Realisierung einer Rechtschreibhilfe für das *Microsoft Office*-Paket begonnen, die 2002 für das Fassaladinische und 2003 für das *Ladin Standard* bereitgestellt werden konnte. Diese erste ladinische Rechtschreibhilfe wurde über die Firma *Expert System* in Modena als *Microsoft*-Rechtschreibhilfen lizenziert. Das *Windows*-basierte Korrektursystem entstand in enger Zusammenarbeit mit dem *Istituto Trentino di Cultura - Centro per la ricerca scientifica e tecnologica (ITC-Irst)* in Trient, das dafür ein eigenes Programm namens "Morpho" entwickelt hatte. Damit können die grammatikalischen Regeln der einzelnen ladinischen Idiome in Algorithmen gefasst und daraus automatisch Wortlisten (Deklination bzw. Konjugation der Adjektive, Substantive und Verben) generiert werden.

<	A	B	C	D	E	F
1	//4	radici	v-dé-1			
2	def	v-dé-1	v		1 irregolare	é
3	root	v-dé-1	begin	é		
4		d;	da;	daj;	daje	
5		nd;	nda;	ndaj;	ndaje	
6		sëurand;	sëuranda;	sëuradaj;	sëuradaje	
7		sëurast;	sëurasta;	sëurastaj;	sëurastaje	
8		st;	sta;	staj;	staje	
9	end					
10						
11						
12	decl	v-dé-1	begin			
13	modo	indicativo				
14	tempo	presente				
15	tipo	affermativo				
16						
17	ie	prima	é	d-é		
18	tu	prima	es	d-es		
19	ël	prima	à	d-à		
20	ëila	prima	à	d-à		
21	nëus	terza	on	daj-on		
22	vo	terza	ëis	daj-ëis		
23	ëi	prima	à	d-à		
24	ëiles	prima	à	d-à		
25	n	prima	à	d-à		
26						
27	end					
28						
29						
30	decl	v-dé-1	begin			
31	modo	indicativo				
32	tempo	presente				
33	tipo	interrogativo				
34						
35	ie	prima	é	d-é		

Abb. 1: Information, die dem Programm “Morpho” für die Generierung der Formen übergeben wird

Sabrina Rasom, die bereits Mitglied der Arbeitsgruppe der ersten fassaladinischen Rechtschreibhilfe war, stand bei der Einarbeitung in die Logik des “Morpho” beratend zur Seite. Für Gröden beschäftigten sich Maria Teresa Kostner und Jürgen RUNGALDIER (auch Redaktionsmitglied) und für das Gadertal Daria Valentin mit der Formalisierung der Grammatik dieser beiden Idiome. Die Redaktionsleitung für Gröden und das Gadertal lag bei Marco FORNI, Ansprechpartner im Institut für technische Fragen war und ist noch immer Paolo ANVIDALFAREI. Für die Programmierung dieses neuen *open-source* Systems zur Korrektur ladinischer Texte zeichnete die Firma *Open Lab* aus Florenz unter dem Projektleiter Carlo Zoli verantwortlich.

Das *Sistem de Corezion Ladin* wurde von den beiden ladinischen Kulturinstituten im April 2007 in St. Ulrich der Öffentlichkeit vorgestellt. Die zur Zeit (September 2008) vorliegende Version dieses internetbasierten Korrektursystems unterstützt die Korrektur der vereinheitlichten Schriftform des Grödnertal, des Gadertaler und



Abb. 2: Portal des Korrektursystems, Internet, 2007

des Fassaner Idioms sowie jene des *Ladin Standard* nach der vorläufig formulierten Orthographie, wobei die anderen dolomitenladinischen Idiome jederzeit in das System integriert werden können.

Das Korrektursystem ist ein erster Baustein der neu geschaffenen Plattform *ladinternet* <www.ladinternet.it>. Die Bezeichnung *ladinternet* stellt eine Verschmelzung von *Ladin* und *Internet* dar und symbolisiert somit die Verbindung von Tradition mit Neuem. Das Portal soll die gesammelte lexikographische Arbeit des Instituts aufnehmen, nach außen hin über das Internet sichtbar machen und interessierten Anwendern zur Verfügung stellen.

1.1 Die Formulierung von Regeln und Ausnahmen des Korrektursystems

Zunächst galt es, Adjektive, Substantive und Verben maschinenverständlich aufgeschlüsselt zu formulieren, also die Regeln der Orthographie in eine Computergrammatik zu übertragen. Hierzu wurde für jede der genannten Wortgruppen, insbesondere aber für die ersten beiden, ein Regelsystem zur automatischen Generierung ihrer Formen entwickelt. Die Logik der Deklinationen und Konjugationen orientierte sich dabei hauptsächlich an der Wortendung. Die daraus automatisch erstellten Listen wurden auf ihre Richtigkeit überprüft, wobei eine entsprechende Regel gegebenenfalls korrigiert, erweitert oder eingefügt werden musste. In Ausnahmefällen konnte übrigens auch ein einziges Wort eine eigene Regel bilden.

Beispiele aus der Liste: Substantiv: *cěsa, cěses*

Adjektiv: *bon, boni, bona, bones*

Verb: gesamtes Paradigma (siehe Tab. 1)

Präsens		Imperfekt		Futur		
<i>ie</i>	<i>é</i>	<i>ie</i>	<i>ove</i>	<i>ie</i>	<i>aré</i>	...
<i>tu</i>	<i>es</i>	<i>tu</i>	<i>oves</i>	<i>tu</i>	<i>ares</i>	
<i>ël</i>	<i>à</i>	<i>ël</i>	<i>ova</i>	<i>ël</i>	<i>arà</i>	
<i>ëila</i>	<i>à</i>	<i>ëila</i>	<i>ova</i>	<i>ëila</i>	<i>arà</i>	
<i>nëus</i>	<i>on</i>	<i>nëus</i>	<i>an</i>	<i>nëus</i>	<i>aron</i>	
<i>vo</i>	<i>ëis</i>	<i>vo</i>	<i>ais</i>	<i>vo</i>	<i>arëis</i>	
<i>ëi</i>	<i>à</i>	<i>ëi</i>	<i>ova</i>	<i>ëi</i>	<i>arà</i>	
<i>ëiles</i>	<i>à</i>	<i>ëiles</i>	<i>ova</i>	<i>ëiles</i>	<i>arà</i>	
<i>n</i>	<i>à</i>	<i>n</i>	<i>ova</i>	<i>n</i>	<i>arà</i>	

Tab. 1: Auswahl von Formen des Verbs *avëi* (“haben”)

Voraussetzung für die korrekte Interpretation war die Abspeicherung der Regeln im *Unicode UTF-8* Format auf einem PC mit *Microsoft Windows* Betriebssystem. Dieses Format wird auch in der Datenbank des Korrektursystems verwendet. Ein weiterer Grund für den Einsatz dieses Betriebssystems ist, dass das gesamte Paket auf *Windows*-Basis programmiert ist und auf entsprechenden Servern verarbeitet wird, womit eine wesentliche Arbeitserleichterung einherging.

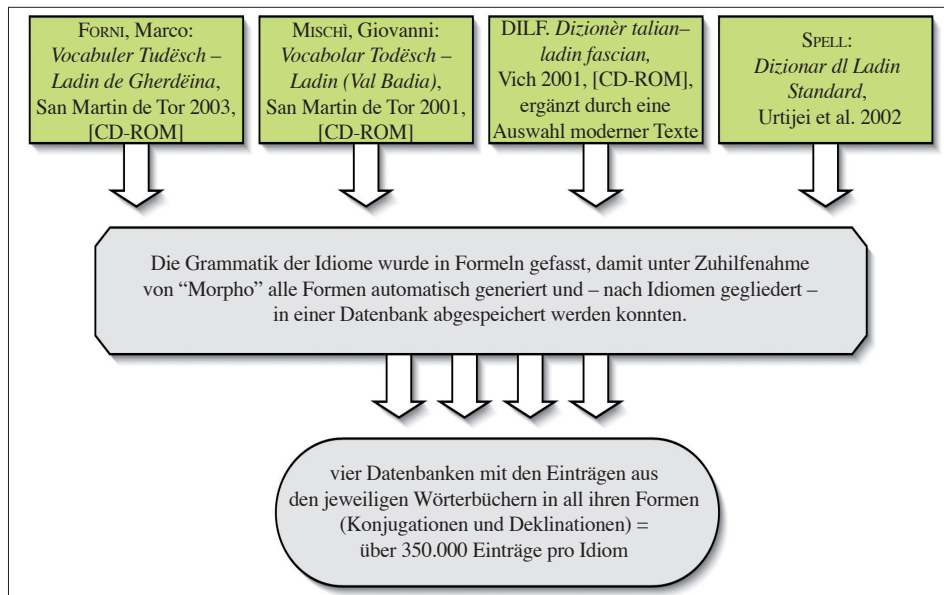


Abb. 3: Übertragung des lexikographischen Materials in die Datenbank des Korrektursystems

Das System wurde anhand bereits veröffentlichter Texte, aber auch anhand verschiedener Entwurfstexte und solcher aus der laufenden Übersetzungsarbeit des Instituts getestet. Einige vom System nicht erkannte Wörter sind auf “klassische” Schreibfehler zurückzuführen und gaben den Anstoß zur Entwicklung zusätzlicher Algorithmen, um eine Optimierung der Ermittlung der Korrekturvorschläge zu erzielen.

falsch	richtig	Bedeutung
<i>ël fesc</i>	<i>ël fej</i>	“er macht”
<i>artiset</i>	<i>artist</i>	“Künstler”
<i>cle</i>	<i>tle</i>	“Schlüssel”
<i>chl</i>	<i>chël</i>	“jener”
<i>jmauz</i>	<i>smauz</i>	“Butter”

Tab. 2: “Klassische” Schreibfehler und deren Korrektur

1.2 Die Datenbank auf einem Server

Das Korrektursystem ist internetbasiert und damit plattformunabhängig. Mit folgenden Internetprogrammen (*Browsern*) wurden Testläufe durchgeführt: *Safari*, *Mozilla Firefox* und *Microsoft Internet-Explorer*.

Der Vorteil dieser über das Internet zugänglichen Datenbank liegt nun darin, dass Verbesserungen am System (durch die Programmierer) sowie die Erweiterung und Korrektur der Wortschatzdatenbank (durch die Redaktion) im Hintergrund ablaufen, und alle Benutzer gleichzeitig davon profitieren, ohne Aktualisierungen auf ihrem Computer durchführen zu müssen.

1.3 Der Ablauf der Korrektur

Nach Auswahl des Idioms, nach dessen Regeln der Text zu korrigieren ist, erscheint die Benutzeroberfläche im gewählten Idiom. Der Text wird mit Hilfe der Kopieren/Einfügen-Funktion in das Textfeld des Korrektursystems übertragen und nach erfolgter Korrektur wieder zurück in das Textverarbeitungsprogramm verschoben. Der Text kann auch direkt im Korrektursystem verfasst und nach Abschluss der Korrektur, die nicht parallel zum Erstellen des Textes erfolgen muss, zur Formatierung in ein Textverarbeitungsprogramm übertragen werden. Beim Start der Korrektur werden zunächst die benötigten Wörter aus der Datenbank geladen. Anschließend erfolgt der für den Anwender sichtbare Teil der Abfrage. Während des Korrektur-

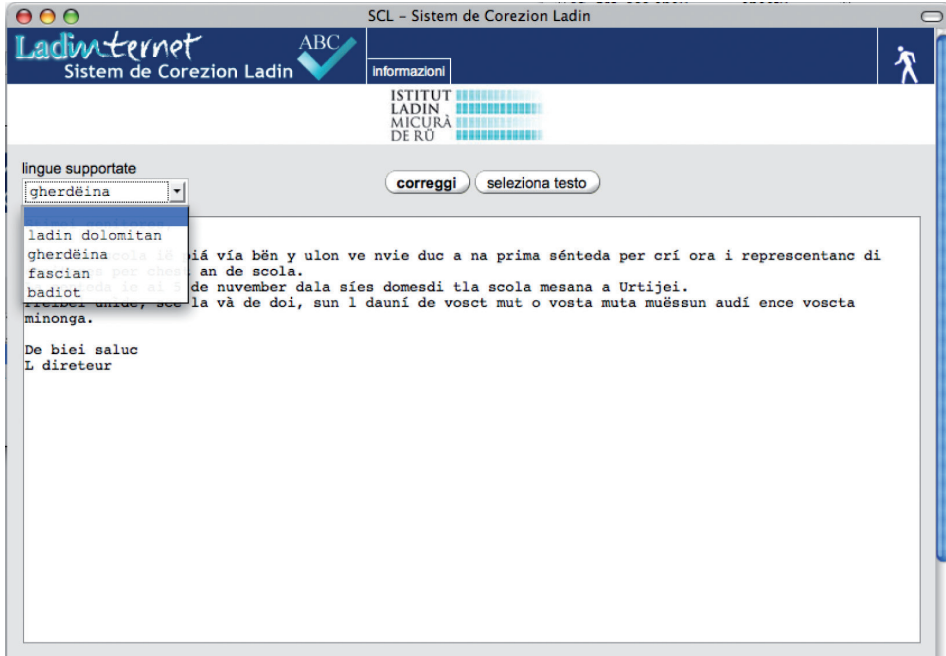


Abb. 4: Auswahl des Idioms für die Korrektur

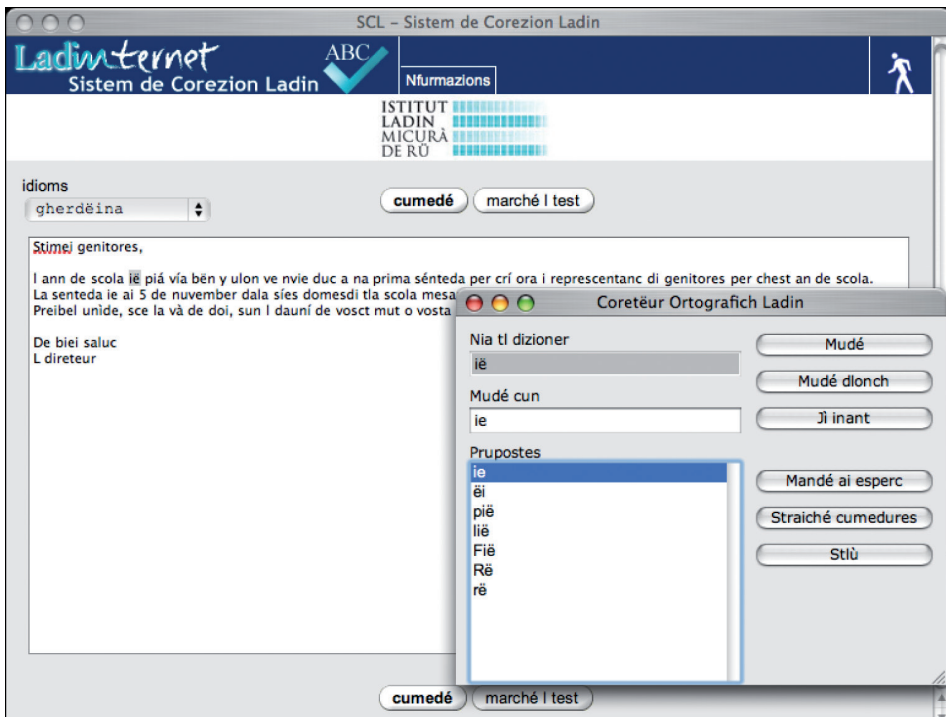


Abb. 5: Ablauf der Korrektur

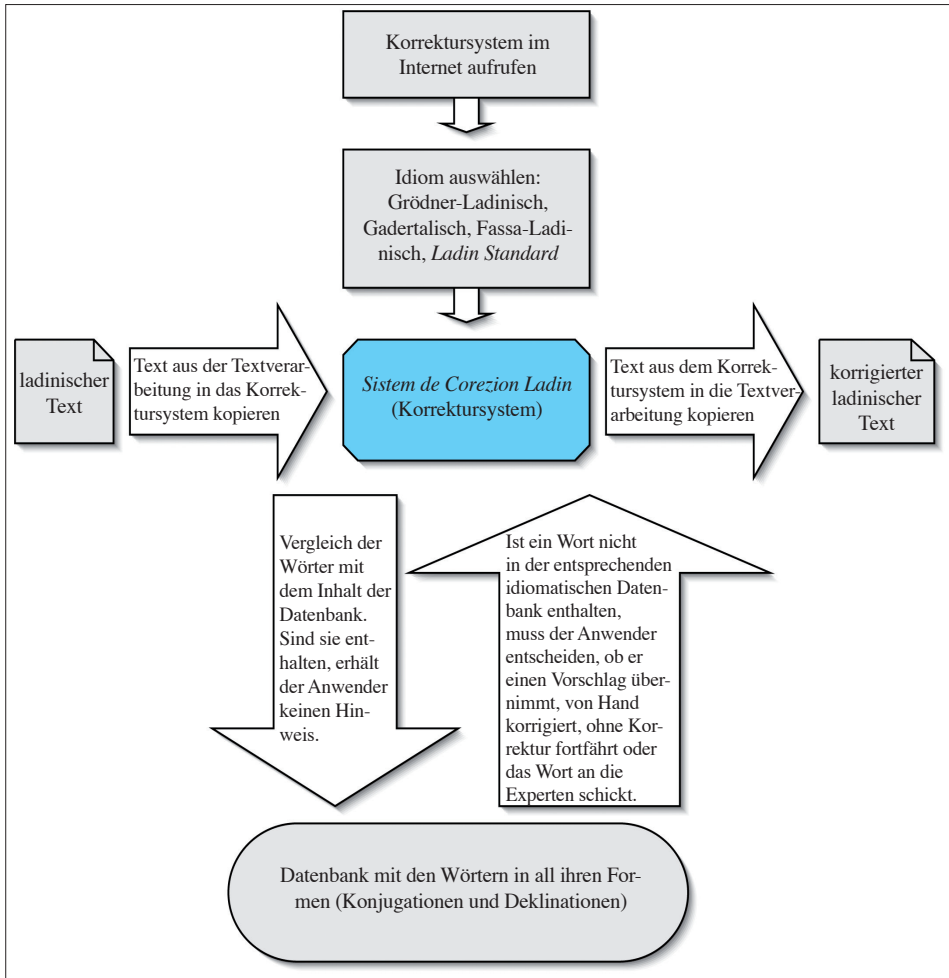


Abb. 6: Ablauf der Korrektur in der Online-Korrekturhilfe

vorgangs kann für jedes nicht erkannte Wort ein Vorschlag des Systems (sofern vorhanden) übernommen, das Wort übersprungen und für den weiteren Verlauf ignoriert, von Hand korrigiert oder an die Expertengruppe geschickt werden. Zudem können alle durchgeführten Korrekturen zurückgesetzt oder der Vorgang vorzeitig abgebrochen werden. Die an die Expertengruppe gesandten Wörter werden in einer Datenbank gesammelt und zu einem späteren Zeitpunkt von der Redaktion besprochen. Dabei wird festgelegt, ob es sich um tatsächlich fehlende Wörter handelt, die so oder in ähnlicher Weise geschrieben werden und deshalb aufzunehmen sind, oder ob es sich lediglich um fehlerhafte Schreibungen handelt, die außer Acht gelassen werden können. Der Administrator integriert die aufzunehmenden Wörter dann in die Wortschatzdatenbank des entsprechenden Idioms.

1.4 Die Editor-Version des Korrektursystems

Es ist geplant, auch eine erweiterte Version des Korrektursystems bereitzuhalten. Dem Anwender wird in dieser Version zusätzlich ein Editor zur Verfügung gestellt, in dem Texte verfasst und formatiert werden können. Folgende Formatierungsoptionen stehen dabei zur Auswahl: **fett**, *kursiv*, unterstrichen, verschiedene Schrifttypen und Schriftgrößen; auch das Einfügen von Tabellen wird möglich sein. Bei der Korrektur werden alle diese Elemente beibehalten; anschließend kann der erstellte Text direkt gedruckt werden.

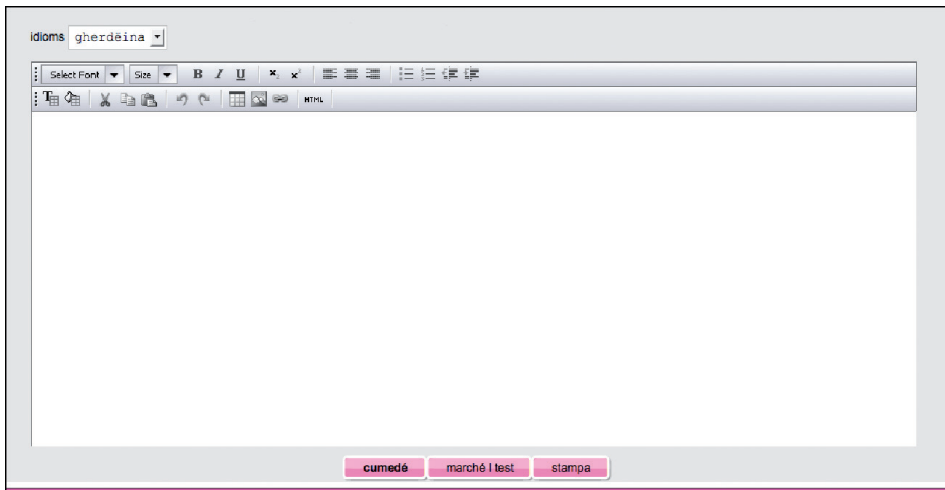


Abb. 7: Der Editor in der nächsten Version des Korrektursystems

1.5 Der Administrator

Der Administrator hat die Möglichkeit, das System selbst zu verwalten. Er kann Wörter zur idiomatischen Datenbank hinzufügen, korrigieren und löschen. Zudem bietet das in das System integrierte Modul zur automatischen Generierung der Konjugationen und Deklinationen die Möglichkeit, diese ohne Hilfe des Herstellers *Open Lab* in die Datenbank einzufügen. Das Resultat steht in Echtzeit allen Benutzern zur Verfügung, ohne dass zusätzliche Aktualisierungen auf den eigenen Computer geladen werden müssen.

2. Das in *Microsoft Word* integrierte Korrektursystem

Dem *Online*-Korrektursystem folgte bald eine erste, in ein Textverarbeitungsprogramm integrierte Version. Auch in diesem Anwendungsfall wird auf dieselben vier Datenbanken der *Online*-Version zugegriffen.

Im Herbst 2008 stand sowohl eine Testversion für *Windows* als auch für *Macintosh* zur Verfügung, die in dieser Form der endgültigen Version bereits sehr nahe kam. Beide Korrektursysteme sind auch mit dem neuesten Betriebssystem kompatibel und auf der letzten *Word*-Version einsetzbar.

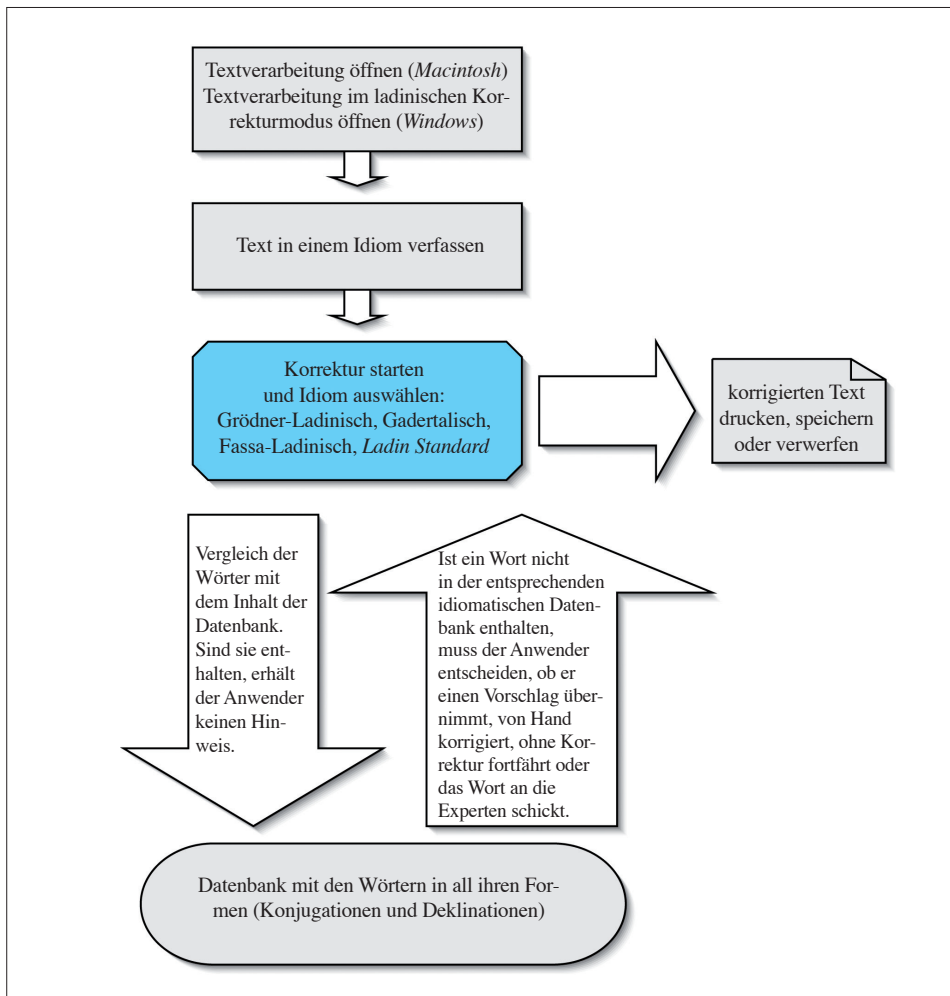


Abb. 8: Ablauf der Korrektur im Textverarbeitungsprogramm

Damit direkt im Textverarbeitungsprogramm korrigiert werden kann, muss ein Zusatzprogramm installiert werden, das die Schnittstelle zwischen *Word* und dem Server im Internet darstellt. Nach der selbsterklärenden Installation befindet sich auf dem *Display* ein eigenes *Icon* mit der das Korrektursystem gestartet werden muss (*Windows*). Die Verbindung zwischen der Textverarbeitung und dem Server wird über Makros hergestellt, die in dieser Datei aufgezeichnet sind. Makros sind kleine Programme, die eine Befehlsfolge für automatisierte Abläufe enthalten. Auf dem *Macintosh*-System ist die Korrekturhilfe hingegen durch ein *Script* in *Microsoft Word* eingebunden. Dies hat den Vorteil, dass es jederzeit geöffnet werden kann, und dann im aktiven Fenster die Korrektur durchführt.

Beim integrierten Korrektursystem muss der Text nicht mehr durch die Kopieren/Einfügen-Funktion in ein Korrekturfenster übertragen werden. Die Textformatierung sowie Tabellen bleiben erhalten und werden in die Korrektur miteinbezogen. Die Korrektur erfolgt für den Anwender ähnlich wie bei der programmeeigenen Rechtschreibhilfe. Aus technischer Sicht wird für jeden Absatz überprüft, ob die Wörter in der Datenbank enthalten sind; bei Bedarf werden Korrekturvorschläge unterbreitet. Auch im Falle der im *Microsoft Word* integrierten Version kann der Anwender einen der Vorschläge annehmen, mit der Korrektur fortfahren, das nicht erkannte Wort an die Experten schicken, den Korrekturvorgang abbrechen oder aber Korrekturen direkt im Text vornehmen.

3. Bibliographie

- FORNI, Marco: *Wörterbuch Deutsch – Grödner-Ladinisch. Vocabuler Tudësch – Ladin de Gherdëina*, San Martin de Tor 2002; CD-ROM: *ibid.* 2003.
- ISTITUT CULTURAL LADIN “MAJON DI FASCEGN”/SPELL (eds.): *DILF. Dizionario Italiano – Ladino Fasano / Dizionèr talian – ladin fascian*, Vich-Vigo di Fassa 1999; CD-ROM: *ibid.* 2001.
- MISCHÌ, Giovanni: *Wörterbuch Deutsch–Gadertalisch. Vocabolar Todësch–Ladin (Val Badia)*, San Martin de Tor 2000; CD-ROM: *ibid.* 2001, [neu bearbeitete Ausgabe].
- SPELL: *Dizionar dl ladin standard*, Urtijej et al. 2002.

Ressumé

La realizazion de n aiut de corezion per l ladin é n auter vare inant tl svilup di strumenc lessicografics. Chesta seconda relazion de laur sun i projec linguistics dl Istitut mostra l svilup dl sistem de corezion per i idioms dl ladin dles Dolomites. A pié via dal aiut ortografich per l fascian éson rovés al sistem che se baseia sun internet, te chel che al vegn integré al moment i trei idioms de valeda gherdeina, ladin dla Val Badia, fascian y la koiné scritta interladina ladin dolomitan. Dlongia i elemenc nia da vedei per l anuzadù dla generazion de listes de paroles y la memorisazion di dac sun n server zentral, végnel splighé ence l servisc dl sistem de corezion online sun <www.ladinternet.it> y dla verscion integreda tla elaborazion de tesé *Microsoft Word* sciche ence l decurs dla corezion te trames les verscions.