Revision and Extension of the OIM Database – The Italianisms in German

Anne-Kathrin Gärtig
Paris Lodron Universität Salzburg
E-mail: anne-kathrin.gaertig@sbg.ac.at

Abstract

The paper presents the Osservatorio degli italianismi nel Mondo (OIM), an online database and a homonymous research project on Italianisms in various European target languages, and the revision of the existing data and their structure which has been underway since 2017.

The OIM is the digitized version of the Dizionario di italianismi in francese, inglese, tedesco (Stammerjohann et al. 2008). The paper focusses on the Italian loanwords in German registered in two opera, and outlines how further loans in this target language are being systematically integrated during the revision process, and how gaps and weaknesses in their lexicographical description are being filled.

Keywords: multilingual lexicography, specialized dictionaries, online lexicography, language contact, history of the Italian/German lexicon

1 Introduction: The OIM Database

With the turn of the millennium, we can see a turn in the lexicography of language contact. Besides classical dictionaries of foreign words, including foreign material in a recipient language, there have been various projects which, starting from the source language, have retraced the paths loan words have taken into one or more target languages, such as Görlach’s Dictionary of European Anglicisms (2001), van der Sijs’s Nederlandse woorden wereldwijd (2010) and the Lehnwortportal Deutsch (Institut für Deutsche Sprache), which offers access to lexicographical registration of Germanisms in dictionaries of Polish, Slovenian and Hebrew. Joining this group of dictionaries classified as “aktives polylaterales Sprachkontaktwörterbuch” (‘active, polylateral dictionary of language contact’) by Wiegand (2001: 125), Italian has been represented since 2008 by the Dizionario di italianismi in francese, inglese, tedesco (DIFIT) by Stammerjohann et al., registering Italianisms in the three European languages of French, English and German.

Concerning the number of speakers, Italian is clearly not comparable to the other source languages mentioned above. However, its status as a language of culture, which has been learned all over Europe beginning in the 14th and 15th centuries (cf. Christmann 1992), and affected domains of use such as trade and finance, art and architecture, music, gastronomy and so on like no other language, has made it an important source for borrowings in a multitude of languages all over the world. The DIFIT is the first work to register them systematically.

As the following figure, the entry for cappuccino shows, the DIFIT lemmatizes Italian source words, followed by information on their semantic fields or domains of use and one or more meaning descriptions. It then lists the form in which they were borrowed in the individual target languages within the microstructure. Following Gusmani (1986), a “borrowing (It. prestito, G. Entlehnung) is seen here as the result of the imitation of a foreign linguistic pattern” (Heinz & Gärtig 2014: 1100) and includes single words like G. Pizza, multiword expressions like G. Frutti di mare, formatives as the suffix
-issimo, but also loan translations as G. Großherzog (< It. granduca) and pseudo-loans (e.g. G. pico-bello), which are formatives that seem to be borrowed, but do not exist in the assumed source language, or exist with another meaning (cf. Winter-Froemel 2011: 44-45).

For each borrowing, the entry contains information on the grammatical category, a definition of its meaning(s) in the target language, the domain where it is used/the semantic field it belongs to, the first attestation and information on the type of borrowing (direct borrowing, calque, pseudo-loan, etc.).

The dictionary is based on lexicographical sources, i.e. mainly on the large reference dictionaries of the target languages, from which the borrowings from Italian were extracted: the OED for English, the Trésor de la langue française (TLF), the Grand Robert (GR) and the Dictionnaire étymologique de la langue française by Bloch and Wartburg (BW) for French. For the Italianisms in German, the authors had to face the problem that:

```
Il famoso Deutsches Wörterbuch dei fratelli Grimm contiene solo forestierismi molto antichi, da lungo integrati, mentre i Fremdwörterbücher [...] o sono selettivi, come il Deutsches Fremdwörterbuch (DFwb), o non danno dati, come Das Große Fremdwörterbuch (DuF), e sono selettivi anche i dizionari etimologici, come l’Etymologisches Wörterbuch der deutschen Sprache (Kluge 1995) e il Herkunftswörterbuch der deutschen Sprache (DuE), o storici, come il Deutsches Wörterbuch di Paul (Paul 2002) [...]. (Stammerjohann & Seymer 2007: 42) 1
```

This divergent lexicographical tradition leads to a lack of datings and attestations for a multitude of Italianisms in German as their target language, especially for newer borrowings arriving from the 20th century on.

Since 2014, the DIFIT data has been available online at www.italianismi.org (cf. Figure 2 for the entry s.v. cappuccino, cited in the printed version in Figure 1). The international project working on its transfer and on its extension is called Osservatorio degli italianismi nel mondo (OIM) and is hosted by the Accademia della Crusca. Its platform offers a web interface with various filters (cf. Figure 3), allowing for varied search options, e.g. for limiting the search to features such as Italianisms of certain grammatical categories, in only one or two recipient languages, of a certain domain of use or a limited period of borrowing, and so on. This creates a useful instrument that can be used for researching the phenomenon of borrowing and language contact (cf. the exemplary studies in Stammerjohann

1 ‘The Grimms’ famous Deutsches Wörterbuch contains only very old borrowings, which have long been integrated, while the Fremdwörterbücher [...] or are selective, as is the Deutsches Fremdwörterbuch (DFwb), or do not give datings, like Das große Fremdwörterbuch (DuF). The etymological dictionaries are also selective, e.g. the Etymologisches Wörterbuch der deutschen Sprache (Kluge 1995) and the Herkunftswörterbuch der deutschen Sprache (DuE), or they are historical, such as Paul’s Deutsches Wörterbuch (Paul 2002).’ On the process of the compilation of the DIFIT and the lexicographical sources, also see Heinz 2008.
& Seymer 2007 and in Heinz & Gärtig 2014). At present, the database contains 8,951 Italianisms derived from 4,660 Italian etyma, 3,145 of those in English, 2,525 in French and 3,281 in German.

Figure 2: OIM, s.v. cappuccino. Accessed at: http://www.italianismi.org/scheda.aspx?id=3963 [19/03/2018].

In 2014 a new project phase was given the green light, coordinated by Matthias Heinz, one of the authors of the DIFIT, and by Luca Serianni, who in the beginning of the new millennium had run a similar project on Italianisms in a much larger number of languages, which was, however, shelved in 2008 (cf. Serianni 2017; for more details on OIM and its new targets see Heinz 2017 & Pizzoli 2017). The new project, on the one hand, investigates Italianisms in further major European target languages. Since 2017, three working groups around Gloria Claveria Nadal, Yorick Gomez Gane, Gianluca Miraglia and Paolo Silvestri, Elżbieta Jamrozik and Zsuzsanna Fábián have been working on the integration of borrowings in Spanish, Portuguese, Catalan, Polish and Hungarian. On the other hand, in 2017 the project group also began to review the existing data and rethink some aspects of the database structure and its sources. This process is coordinated by Matthias Heinz and Lucilla Pizzoli and is being carried out mainly in Florence and Salzburg.

Figure 3: OIM, search mask. Accessed at: www.italianismi.org/ricerca-italianismi [19/03/2018].
This paper focusses on the data surrounding Italianisms in German, which the author is working on within the current project phase, shows problems and restrictions in the current form of the database, and illustrates how they are dealt with in the revision process.

2 The Representation of Italianisms in German

As mentioned above, OIM at present contains 3,281 Italianisms in the target language, German. Searching them with the database’s filters, one can gain a clear picture of how those borrowings are structured within German and can compare them to those in English and French. Figure 4 shows the division according to domains of use.

Borrowings from the field of music (e.g. *accelerando; Viola da gamba*) are by far the most frequent ones, followed by Italianisms in the area of gastronomy (e.g. *Brokkoli; Ravioli*), economy and finance (e.g. *Konto; netto*) and art and architecture (e.g. *Aquarell; Fresko*). The predominance of musical terminology is of course explainable by the important influence the Italian language has exerted in this field, but it is also due to the sources used by the DIFIT: In order to compensate for shortcomings in the general lexicography of German, lexica and studies on music were systematically taken into consideration, whilst works on other semantic fields were used far less extensively.3

The filter for the years of first attestation allows users to create a timeline of the periods with the highest influence of the Italian language on the German lexicon. As Figure 5 shows, peaks were reached in the 15th and 16th centuries, when Renaissance Italy became the major center of cultural expansion, and again and even stronger in the 19th century, in which most of the musical terms deriving from Italy were registered in German dictionaries for the first time. The broken line for the 20th century reveals a decreasing number of Italianisms, but also reveals the lack of precise, reliable data of first attestation for this period, which makes it impossible to specify a number: “va ricordato che le fonti non sempre forniscono datazioni, specie per il tedesco […]. Oltre a qualche datazione sommaria come ‘medio alto tedesco’, invece di un secolo preciso, la collocazione delle datazioni mancanti del tedesco è da
cercare soprattutto nel Novecento [...]." (Stammerjohann & Seymer 2007: 47) But as mentioned above, not only the newer Italianisms in German are registered without first attestation: The graph includes a total number of 2,311 borrowings, which means that almost one third of all the Italianisms in this recipient language in OIM do not have a dating, and the present system, in a chronological query, does not label them explicitly.

![Figure 5: Italianisms in German, chronology of first attestations in OIM (cf. Gärtig 2017: 374, based on Stammerjohann & Seymer 2007: 47-48).](image)

A very useful search option for a target language such as German, which presents a differentiated internal diatopic variation and is also a polycentric language, is the filter for Italianisms that have entered only one of its varieties. Aside from the unfortunate labelling as filter for special “uso dialectale”, ‘dialect use’, which doesn’t represent the linguistic reality, it shows that at present about 6% of the registered Italianisms belong to only one variety, most of them (119 or 64%) to Austrian German, 37 or 20% to Southern German varieties, 17 or 9% to Suisse German and 7 or 4% to the German of South Tyrol (cf. Gärtig 2017: 369-370). It is obvious that the South Tyrolian area is especially underrepresented, and the new project phase has to work on a balanced integration of its specific Italianisms.

An overall criticism of today’s version of the OIM, already remarked in a review on the DIFIT (Marazzini & Marello 2011: 164-165), is that it could represent a kind of museum-like lexicon, without information on the effective use of the individual Italianisms as found in Görlach (2001) which includes speaker’s surveys. Ten years after the publication of the DIFIT, another potential problem is related to the immediacy of its data, which has been the basis for the OIM until now. One just has to go to a modern café in a German-speaking area to see a Barista at work, ask for a Latte macchiato – or do you prefer Chai Latte? – and observe if he or she is able to produce the perfect Crema to find quite a number of Italianisms not registered in the actual database.

The review and extension process, thus, has to include the search for attestations and their datings, for information on the effective use of Italianisms in German. It should also create a balanced basis of sources that respects the current lexicon as well as the actual usage and frequency in certain geographical and semantic areas.

---

4 ‘It should be remembered that sources do not always provide dating, especially for German […]. In addition to some summary dating as Middle High German, instead of a precise century, missing German dating is to be found especially in the twentieth century […].’
3 Revision and Extension of the Database Structure and its Data

The fundamental question that has to be answered at the beginning of a revision regards the basic structure of the DIFIT, which uses mainly lexicographic works for sourcing Italianisms to be integrated. Should this decision be maintained, or should text corpora, readily available today and also relatively easily to incorporate into an online dictionary, be taken into consideration? One should bear in mind that the OIM structure should remain stable as it will be possible to add other source languages step by step, perhaps even languages of which the accessibility of corpora and lexicographic description greatly varies. It has been a strength of the DIFIT that its rather rigid structure and having been based on dictionaries guarantee a high level of comparability and reliability.

For this reason, the decision was reached to keep this basic structure, at least for now. New data will be mainly extracted from lexicographical sources, but corpora will be checked to gain attestations and information on the real use. The entries, however, shall be completed with fields on the “grado di stabilizzazione dei termini e la loro diffusione tra le varietà della lingua” (‘degree of stabilization of the terms and their diffusion among the varieties of the language’, Pizzoli 2017: 177) and with references to more detailed studies and/or lexicographic description in other (electronic) dictionaries.

Since the DIFIT went to press, there has been a notable increase in innovation within German lexicography. A substantial number of publications that can be used as sources for the OIM’s extension have emerged. To mention only a few, from 2011 to the present, there is the revised edition of the Deutsches Fremdwörterbuchs (DFWB), published until the letter i and accessible across the portal OWID, hosted by the IDS, where you can also find the electronic German dictionary elexiko and the online version of the Neologismenwörterbuch on the new and latest entries in the German lexicon. For a systematic investigation into Italianisms in common language, the new editions of Währig and of the various DUDEN dictionaries, in print and online, are very useful, as is the Digitales Wörterbuch der deutschen Sprache (DWDS), a digital information system on the German vocabulary with direct access to a vast part of its base corpora and to detailed word profiles and diachronic curves, elaborated since 2007 at the Berlin-Brandenburgischen Akademie der Wissenschaften. For the integration of Italianisms in singular varieties of German, the project can make use of the new edition of the Variantenwörterbuchs des Deutschen (2016). For those in South Tyrol, the linked dissertation of Abfalterer (2007) is of interest. For the verification of its entries, the Korpus Südtirol, developed at the EURAC in Bolzano, is used, whilst for the other entries the corpus platform of the DWDS is consulted.

3.1 Completion of the German Data

The work flow is divided into the revision, correction and completion of the data on Italianisms already present in OIM, and into the addition of new borrowings, extracted from the works mentioned above. For the revision process, an important aim is to guarantee correct and meaningful answers to research questions on the chronology of borrowing. For this purpose, all the entries at present listed without an attestation are extracted and looked up in newly available dictionaries. That has led for instance to attestations of Bruschetta, Calzone, Ciabatta or Papamobil (in DUDEN). If the dictionaries do not include a dating, it is searched for in corpora. The same process is carried out for Italianisms without datings in the OIM. Thanks to the DFWB, for example, it was possible to date Futurismus precisely to 1912, instead of the vague indication of 20th century in the OIM, Fiasco in the meaning of ‘failure’ to 1819 (instead of “sec. XIX” in the OIM) or Furore machen, partial calque of It. far furore, to 1830 (instead of 19th century, too) with the help of the DWDS corpus search.5

5 <https://www.dwds.de/r?q=Furore&corpus=public&date-start=1473&date-end=1900&genre=Belletristik&genre=Wissenschaft&genre=Gebrauchsliteratur&genre=Zeitung&format=kwic&sort=date_asc&limit=50> [22/03/2018].
Another desideratum for German, on which the current phase is working, is a uniform labelling of its varieties, indicated for the single Italianisms not found in the whole German speaking area. The labels, on a first level, should classify those borrowings as belonging only to one macro area (Germany/Austria/Switzerland/South Tyrol; for the areal conception cf. Ammon, Bickel & Lenz 2016: XXV-XXVII) and then, on a second level, add a finer classification for those units we have more precise information about. An example is *Bassena* (*< Fr. bassin < It. baccina*), classified as Austrian and as used particularly a Vienna by DUDEN (s.v.).

In addition to the improvement and completion of the data itself, as mentioned above, the integration of some new lexicographic information is planned. At present it is still being worked out how the real use and diffusion of single Italianisms in the target languages and in their varieties can be represented according to a scheme. In a similar way, the reliability of their attestation (whether they are registered in dictionaries, in reference corpora, in other corpora or only in single texts and contexts) should be represented in a way that can be queried by means of the web interface and that is comparable for the different target languages. This information can be useful when applied to quantitative analysis.

However, they do not provide in-depth information on the real use of single borrowings. As an online tool, the OIM should take advantage of the possibility to add qualitative information to the Italianisms for which they are available. For example, there could be bibliographical references to research on single arguments, similar to the corpus studies done by Rovere (2012) on *Cappuccino, Galleria* and *Bambini*. There could be links to portals like OWID or DWDS, that can provide diachronic curves, more precise etymological explanations, authentic examples and corpus hints or information on syntagmatic combinations, word formation processes and products. Finally, there could be analyses coming from the project itself, reassumed in info boxes. For example, at the present a master’s student from Salzburg, Katharina Kofler, is carrying out a sociolinguistic study on the real use of Italianisms in the German standard variety of Austria (supervisor of the thesis: M. Heinz). Using an online questionnaire, Kofler is asking if the single forms are known and if they are effectively being used by Austrian speakers, if their meanings are known and which geosynonyms are mostly used. Up to the present (end of March 2018) she has collected the answers and sociolinguistic data from more than 300 participants.

### 3.2 Addition of Italianisms in German

Parallel to the completion of the missing information on Italianisms already registered, the project phase intends to complete missing borrowings. This work is being done on three levels: borrowings in the common language, borrowings of certain domains of use and borrowings of certain diatopic varieties.

In the first group, Italianisms of the common language, mostly from the 20th and 21st centuries, have been added when they were not yet lexicographically described or not in use when the DIFIT was compiled. To collect them, by now all the entries marked as derived from Italian in the DUDEN Online-Wörterbuch and in the Neologismenwörterbuch have been extracted and compared to the stock in the OIM. The result was about 60 hits, almost all nouns, with the majority belonging to the semantic field of gastronomy: *Aceto balsamico* (with the variants *Balsamico, Balsamessig, Balsamicoessig*), *Caffè Latte, Chai Latte, Crema, Latte macchiato*, the Austrian *Melanzani, Pancetta, Pannacotta, Pecorino, Peperonata, Sugo, Vitello tonnato*; but there is also *Gabione ‘a cage of wire filled with rocks for use in civil engineering, road building and landscaping’* or the adjective *papabel*, belonging to the use of the Catholic Church and referring to a cardinal with the possibility of being elected Pope.
In order to prepare the new entry, for every loan a lexicographical sheet with the following information is prepared: Italian etymon – borrowed form in the target language with possible variants – pronunciation – type of borrowing – reliability of attestation – reference to bibliography and further literature – meaning – grammatical category and genus – domain of use / semantic field – source – dating – first attestation – diffusion – register.

The second step of systematic research aims at balancing the semantic fields of Italianisms in German. As mentioned previously, at the moment of the compilation of the DIFIT, only musical terminology was considered in major terms; this valuable work of excerption from specific lexicons and texts should be replenished with collections of terms belonging to other semantic areas and domains known to have been particularly influenced by the Italian culture and language, especially the areas of gastronomy, economy and finance, art and architecture. Currently another master’s student from Salzburg, Patricia Bagari, is writing her master thesis (supervisor: M. Heinz) on the Italianisms of the last mentioned area. She is collecting terms from Sandrart’s historical Teutsche Academie (1675), from dictionaries and lexica on arts, such as the Kunstlexikon (Hartmann 1996), the Lexikon der Kunst (Olbrich 2004), the Großes Bildwörterbuch der Architektur (Koepf & Binding 2005) and Reclam’s Wörterbuch der Architektur (2015), retracing their etymology and adding corpus studies. She has, at the moment, prepared about 80 entries to insert into the database.

In relation to Italianisms in single standard varieties of German, the focus at present is on the South Tyrolean area. An important source is Abfalterer (2007), who has also worked on the South Tyrolean part of the Variantenwörterbuch des Deutschen (2016), and whose work includes a glossary of 303 primäre Südtirolismen, 26 of which are to be classified as loanwords (e.g. ACI, Aranciata, Dopolavoro, Kondominium) and 98 as calques (Ausgeher, Autobüchlein, Berufsalbum, grüne Nummer) from Italian, using the OIM’s classification system (cf. Abfalterer 2007: 167-168). They have been integrated to the OIM database, using its fixed set of semantic labels, which sometimes differ from those used by Abfalterer, and have started to be checked for datings in the Südtirol Korpus. Many of the Italianisms, especially those created by loan translation, belong to the fields of administration, politics, finance, professions and the education system.


The possibilities to query the OIM database have to take into account this special situation of South Tyrol. In fact, mixing them with those of the other varieties without any remark could cause distortion in search requests. For example, it would create a much higher number of loan translations and administrative terminology. For this reason, it seems desirable to insert a filter which includes or excludes Italianisms used only in South Tyrol, or restricts the research only to them. In future studies, comparable solutions could have to be found for other polycentric languages among the target languages in OIM (cf. as an example Pierno 2017, with a study on Italianisms in the English and French used in Canada), and filters could work from top to bottom, making it possible to select entire target languages, entire standard varieties or single regional and dialectal varieties.

6 ‘One reason for the increased appearance of calques in administrative terminology is certainly to be found in the special situation of South Tyrol with its institutional superstructure by the Italian state. With the entry into force of the implementing regulations which established equality between German and Italian, it has become necessary to translate all terms from administration, finance, politics, construction and law. This is not an easy task and sometimes leads to peculiar word formations.’
4 Conclusion and Further Perspectives

For the target languages already included, such as German, which has been described here, the purpose of the revision can be summarized as follows:

- the completion of the supply of Italianisms, covering today’s common language, relevant domains of use and diatopic (standard) varieties;
- the completion of missing attestations and datings for all Italianisms based on the integration of recent documentation;
- the possibility to make search queries using filters for single diatopic (standard) varieties of the target languages;
- a representation of entries that includes details on the effective use of a single Italianism in the target language and on the reliability of the given data (whether taken from dictionaries/corpora, etc.);
- the creation of a technical mechanism that points in a stable way to more information on single Italianisms, such as bibliographic references or persistent URLs that lead to more detailed online dictionary hits or information from studies carried out in the context of the project.

In parallel to the revision process on Italianisms in German, French and English, the OIM project is working on an extension of the target languages for borrowings from Italian in concentric circles. Therefore an international research network, composed of scholars from the universities of Florence, Rome, Salzburg, Dresden, Warsaw, Budapest, Malta, Seville, Toronto and New York, has been built up. By the summer of 2018, the mentioned work on Spanish, Portuguese, Catalan, Hungarian and Polish should be completed, and there should be the new design of the website and database structure. In the following step, the circle of languages shall be expanded to those for which collections of Italianisms are already available, such as Maltese, a language with a very strong Italian impact.

References

DUDE


