XIII Simposio Internacional de Comunicacion Social Santiago de Cuba, 21-25 Jan 2013

A linguistic and gender approach to 1841 Tuscany population Census
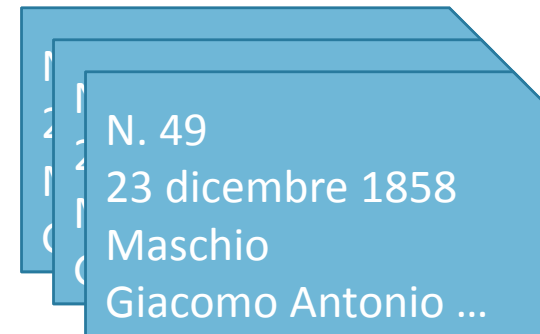
M. Sassi*, S. Biagioni** and G. Pardelli*

CNR, Pisa, Italy

* Istituto di Linguistica Computazionale "A. Zampolli",
** Istituto di Scienza e Tecnologie dell'Informazioni "A. Faedo"

The contest: the Census of 1841

The Census of 1841 in Tuscany was first official data registry which tried to describe Tuscan population as a whole on the basis of the Grand Duchy of Tuscany.

With the use of special ad hoc created forms all demographic and socioeconomic characteristics of families and single persons in the Grand Duchy of Tuscany were described.

COMPARTIMENTO PISANO

These data of Census supply a precious source for studies of all different aspects of the population.

Registers were generated and updated by priests, the only certainly educated widespread on territory, and included following information:

Name, Surname, Age, Gender, Marital status, Employment, Religion, Schoolarity, "Social status".

For each community and parish a full account is given *of homes, resident families, and composition of families including family servants.*

The contest:
from paper to machine-readable form

N. 49
23 dicembre 1858
Maschio
Giacomo Antonio …

During early 80:s the Department of Modern History of Pisa University charged the Computational Linguistic Institute of CNR with digitalization and the electronic processing of:

1. the data from the Census of 1841,
2. the data from "Catasto Leopoldino" [public registry of buildings and land ownership]

to enable statistical, demographical, historical, sociological and economic analysis.

The target and the method

- Terminological and lexical issues
- Gender related analysis of work and craftmanships

We have used a subset of data of Census, concerning four communities with own peculiar profile, in the province of Pisa (Italy): Bièntina, Càscina, Pontedera and San Giuliano Terme to develop a new project.

The total of the units (Male and Female) is 41.887, grouped
in 7.190 families, spread in 4 communities as follows:

Bientina	Cascina	Pontedera	SanGiuliano Terme	Totale
2.420	14.999	9.446	15.022	41.887

Women ▯ and men ▯ / communities →
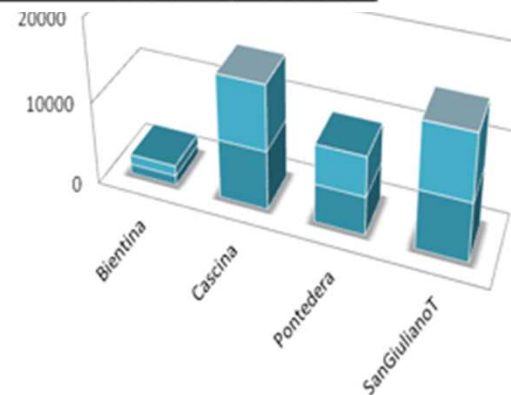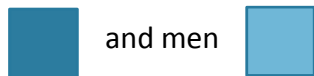
... The steps and the glossary and the taxonomies

The work is developed in 5 points:

a) retrieval of linguistic information from the Census of Tuscany of the
    19th century focused by the arts and craftsmanship more in use in
    families of that time,
b) gender division of works and craftsmanship,
c) observation of lexical disparity in the four communities and terminological
    curiosities of that historical period,
d) actually no longer existing craftsmanship,
e) diachronic analysis of communities.

In addition it has been given the English translation of the terms extracted from
the Census Corpus (Appendix Glossary).

Creation of ad-hoc taxonomies was also needed for the conceptual classification
of different jobs and trade types (agricultural, textile, trade, craftsmanship,
professions).

... The software used

The two software used for textual analysis and terms extraction are:
Text Power and DBT software for textual analysis and full-text querying.

The texts analysed with the DBT is available for use by other application
Procedures

DBT&Facets integrates the core textual search engine DBT with the
TextPower (TP) technology.

DBT-TextPower is based on NLP techniques and linguistic resources used
to create tools for the evaluation, analysis, classification and browsing of
Information.

http://dbtvm1.ilc.cnr.it/SitoDBT/
http://dbtvm1.ilc.cnr.it/Faccette/DBT_e_Faccette.dll

... *The methodology*

The extraction of implicit knowledge from the texts through which TP can enrich the documents, is a specialization of the "Facets" technology.
DBT-Corpus Query system of textual corpora, i.e. text collections of homogeneous texts structured, tagged and preanalysed and query-able with database methods.

The adopted methodology uses all of those capabilities, applying them to subsets of the data of the Census considered as a Corpus, structured on family records. Every analysis is based on ad-hoc created tables.

These tables concern works and craftsmanships which are sub-classed as exclusively or typically male, exclusively or typically female and shared or gender neutral works.

The results: Jobs arts and craftmanships in Pisa Province in XIX

Today's "use & throw" lifestyle was unconceivable then, and so ordinary life goods were repeatedly fixed and arranged in order to be again fit to everyday's use.

Whence over 70% of all craftsmanships was concerned and specialized primarily with such "Ars reparatoria" which enabled entire households to live on that income.

Which occupations where recorded in 1841 by the charge of the many parishes of the territory of *Granducato di Toscana* in 1841?

- In which sectors was female work employed?
- Did gender shared or neutral jobs exist ?

... Jobs arts and craftmanships in Pisa Province in XIX

The sample of the four communities sees [expectably] agricultural work ahead of any other activity.

The voice *colono/a* (equivalent: farmer) lists 583 female and 1695 male workers (about 4% of total population). To this single term we must add both multiword term like:

colono al monte (mountain farmer)
lavoratore/trice al campo (countryside worker)
lavora al campo (works in the countryside)
mezzajolo, or *mezzano*
contadino d'altri (farmer employee)

By adding up all these labels we reach total amount of 854 female and 2663 male operators.

The results: Verbs use in craftmanships naming

Multiword terms often include verbs and the most used are "andare" [to go] fare [to craft] and "operare"[to operate or to act as] and the graphic form that all these took might be very different.
For verb "andare" we have a basic significance "to move for working outside own home or land" but two quite distinct usages:

1. Forms like:
   a. *Va a far legna in macchia* (female), picks up [burning] wood in forest
   b. Va a scuola, for a (male) student
   c. *Va a raccogliere letame* (male), (goes to gather manure) in 1841 in Pontedera this was still a salable skill

2. Forms indicating works for third parties like:
   a. *Va a bottega di calzolaro* (male), indicates a male who goes to learn activity or help activity of a craftman who produces and repairs shoes;
   b. *Va a opera a fare erba* (female), weed gatherer for other people
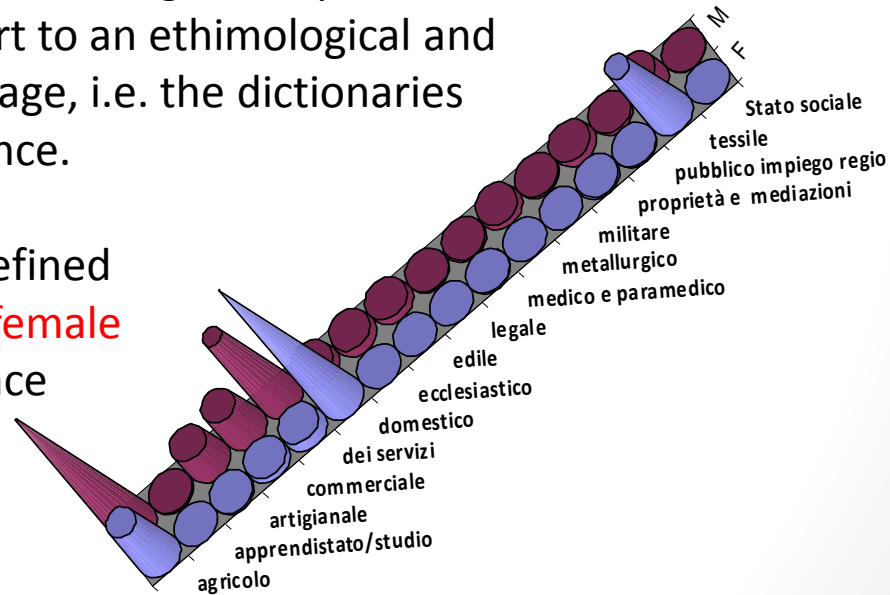
The results: Gender split of works in XIX in Pisa Province

In which job sectors were women involved?

An answer to this question required the categorization of all jobs arts and craftsmanships, which we made in 15 categories and related sub categories, as follows:

To generate taxonomies (or Facets) each job label has been analysed and inquired, and not all of them were at first sight comprehensible and exhaustive, thence we found necessary to resort to an ethimological and historical dictionary of Italian language, i.e. the dictionaries of Accademia della Crusca in Florence.

We associate to the sector labels defined above letters M for male and F for female in order to show the gender presence in different sectors in the present Figure and in the Table of the next slide.

...Gender split of works in XIX in Pisa Province

Women are mostly involved in the sectors n. 1, 5, 6 and 15 as shown in the table:

	Settore	Sector	F	M
1	agricolo	agriculture	854	2663
2	apprendistato/studio	apprenticeship and learning	3	30
3	artigianale	craftmanship	25	462
4	commerciale	trade & sale	85	605
5	dei servizi	services	160	1552
6	domestico	domestic work	2766	131
7	ecclesiastico	church	0	113
8	edile	building	1	109
9	legale	legal	0	18
10	medico e paramedico	medical and paramedical	10	32
11	metallurgico	metallurgy	0	107
12	militare	military	0	39
13	proprietà e mediazioni	ownership and intermediation	36	115
14	pubblico impiego regio	public administration	7	30
15	tessile	textile	1206	269

...Gender split of works in XIX in Pisa Province

Semantic aggregation procedure has necessarily been of heuristic type, which explicitly means that several terms were moved across sectors in order to reach a correct conceptual distribution.
Originally we had 24 job sectors, but we merged some of them to avoid dispersion of information so that finally we finish with 15 sectors.

Notably we were forced to add a special category for brokering activities which had a great relevance there and then. The entire phenomenology of work market entry has been classed under apprenticeship e.g.:

impara l'arte del carratore,
impara l'arte del calzolaro,
inizia a tessere canestre.

The results: A graphical view of jobs and kinship

Men and women reporting a job label are classed by their kinship relationship to family head and appear as bars - reference numbers are on left side - while job indication [yellow] are measured on the total of individuals [blue] and the reference number is on right side:
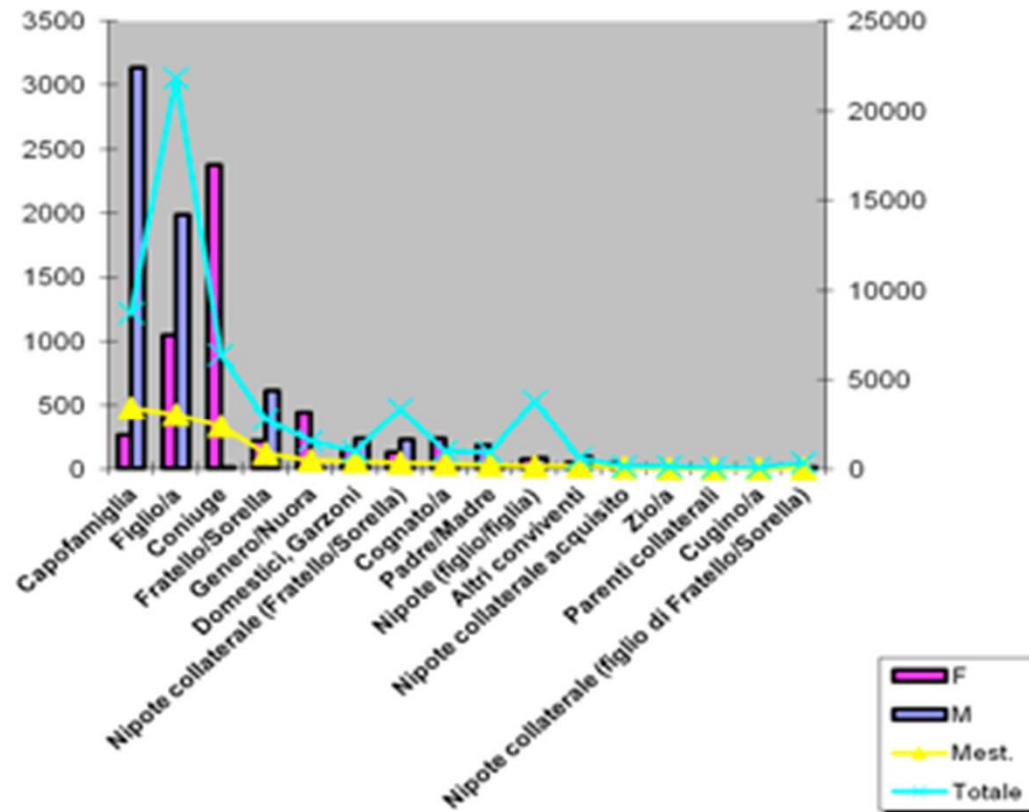
DBT&Facets and TP Technology applied to the Corpus

The contents are extracted by the procedures of
TP directly from free text, using their
linguistic-statistical approach
and elevated to the rank of "Facets".

The facets are so extended,
replacing them with the set of
all terminology annotations that
TP has marked automatically .

The text so enriched can be used by tools
for analysis, classification, query and
navigation that we have developed,
starting from our base engine DBT and named "DBT & Facets".

The result of the query "tessiera+canapajo" shows both
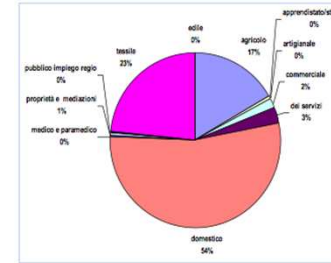the relations among Craftmanships and family names.
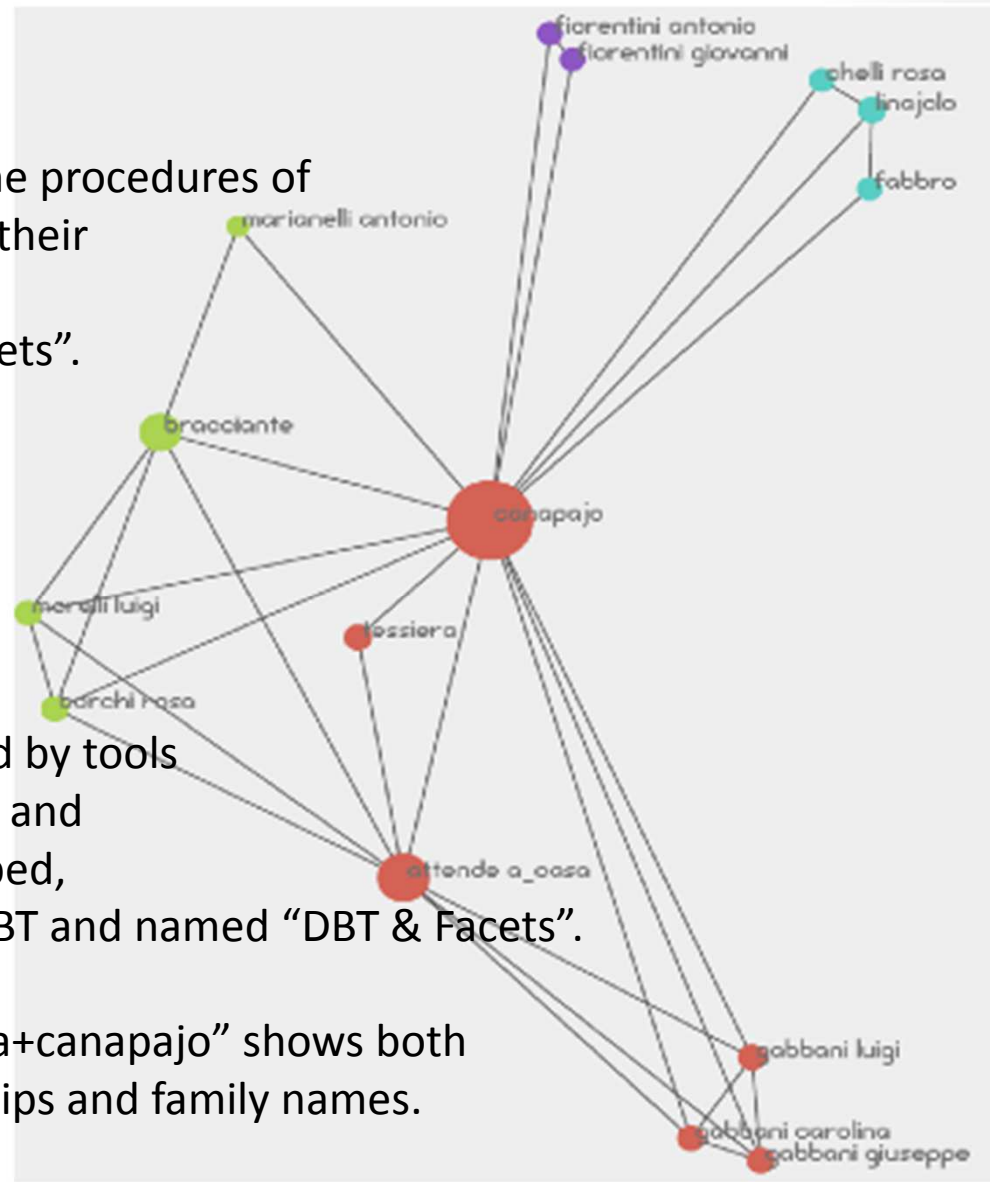
Conclusions

The study emphasizes that women have worked inside their home (the domestic work sector).

Our final graph shows clearly that domestic work is 54% of all female work,
textile work is 23% and agriculture 17% .
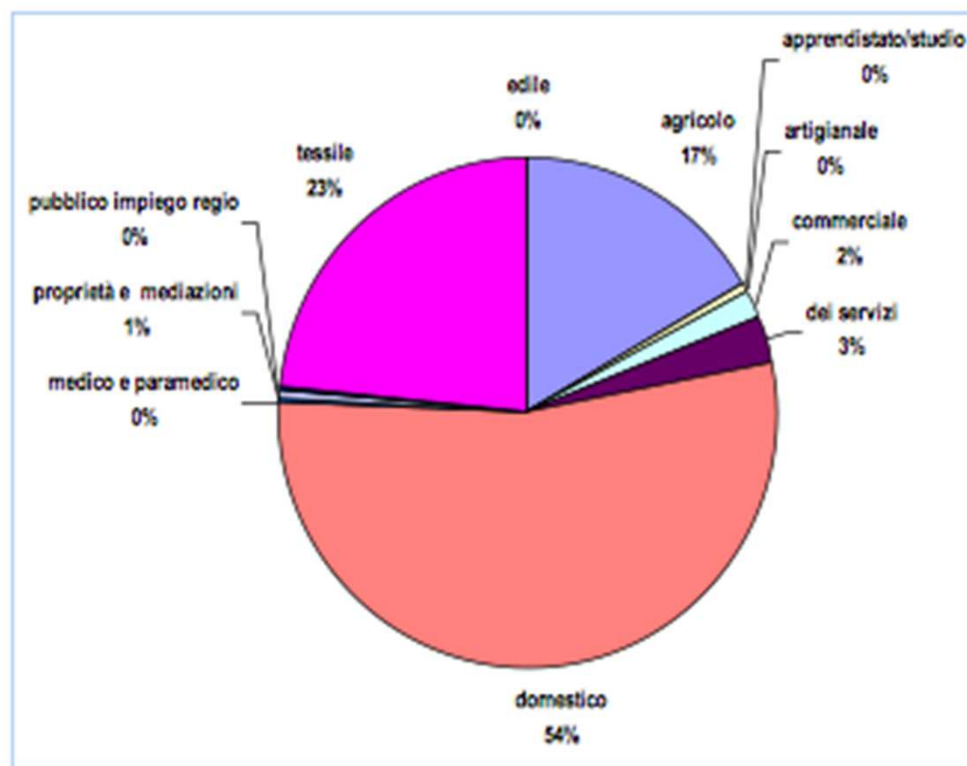
... *Conclusions*

Special relevance has the domestic work sector in which most of jobs are typically or exclusively **female**, like e.g.: *balia, domestica, fa le faccende a domicilio.*

The professional terms related to medical and veterinary work results totally unambiguous.
Service sector is a restricted version of actual tertiary and includes works like e.g.: *vetturale, vetturino, nocchiero*, (indicates the activity of conduction ship in and out from harbors), *stiratora, tintore/tintora*.

In the trade&sale sector we have extracted all related activities, even if the term in dictionaries presented semantic ambiguity between the producing craftsman and the seller like e.g.: *fusajo, chincagliere*.

... Conclusions
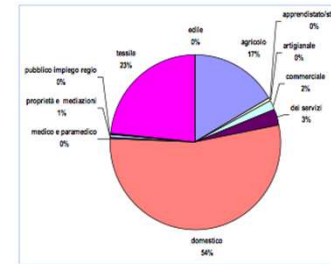
In the craftsmanship sector three subsectors have been created and considered as man's categories: metallurgy, textile works, and **masonry**. From each taxonomy all jobs in which female were admitted or dominating have been extracted. In that period most out of home works were **exclusively male.**

We can immediately see that agriculture is by far the dominating sector and also the most gender shared or neutral. Women and men shared countryside work (***bracciante* is gender free term**) and organized it (*fattore/essa, contadino/a, colono/a*).

The second gender shared sector is the textile craftsmanship which we see in terms: *tessitore/a, sarto/a*, while the threading worg named *filandaja/filandara/tessiera/tessitrice* is excusively female. Also commerce sector is shared and we have terms like *fornajo/a, fruttajolo/a, bottegajo/a*, commerce of tobacco (and salt) is instead **exclusively male.**

... Conclusions
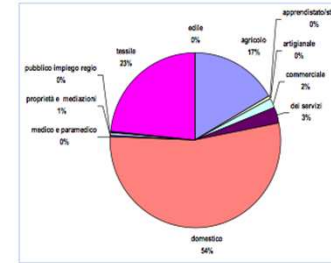
Medical sector has professions which are exclusive to each gender profesion of *medico* and *chirurgo* are **exclusively male** while *ostetrica* and *levatrice* are **exclusively female**.

Sectors military, clergy, legal and royal officials are **exclusively male**, we may observe that the terms *suora* and *monaca* are not listed in this census.

Obsolete terms

vetturale, vetturino,

fusajo,

nocchiero,

chincagliere,

calzettaja, canapajo, fiascaja, manifattrice di pèneri, seggiolaio, zuccherinajo