Ministry of Science and Technology
of The State of Israel

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Ministère de l'Europe
et des Affaires Etrangères
Ministère de l'Enseignement
Supérieur, la Recherche et
l'Innovation

France-Israel High Council for
Scientific and Technological Cooperation

Status Seminar on

"Digital Humanities"
&
"Computational modeling with functional and evolutionary genomics of infectious diseases"

PHC Maimonide 2018-2019

Program and abstracts

French Institute of Israel
Rothschild Boulevard 7, Tel Aviv

Tuesday, December 3rd 2019

08:30 Registration, Coffee and Cake

08:50-09:10 Greetings and Opening

Prof. Alexander Bligh
Chief Scientist, Ministry of Science and Technology, Israel
(AlexanderB@most.gov.il)

Dr. David Harari
Co-President, France-Israel High Council for Scientific and Technological Cooperation, Israel (doudyharari03@gmail.com)

Mr. Paul Furia
Scientific and Academic Cooperation Attaché & Deputy-director of the French Institute, French Embassy in Israel (p.furia@ambfr-il.org)

Session 1
Chairperson
Dr. Vincent Lemire
CRFJ Director
lemire.vincent@gmail.com

09:10 – 09:35 Exploring Hominin behavioral patterns through time and space: an analysis of 3d digital models of stone handaxes**.**

- **Dr. Naama Goren-Inbar,** The Hebrew University of Jerusalem, Israel (goren@cc.huji.ac.il)
- **Dr. Marianne Deschamps,** CNRS - UMR 5608, France. (mardesch1690@gmail.com)

09:35 – 9:50 Discussions

09:50 – 10:15 NLP Resources for Analysing Reactions to Major Events in Hebrew and French Social Media.

- **Prof. Yoav Goldberg**, Bar-Ilan University, Israel (yogo@cs.biu.ac.il)
- **Dr. Djamé Seddah,** Centre Inria Paris, France (djame.seddah@inria.fr)

10:15 – 10:30 Discussions

10:30 – 10:50 Coffee break

Session 2

Chairperson

Dr. Alex Altshuler,

Director for Social Sciences and Humanities, MOST

AlexA@most.gov.il

10:50 – 11:15 Tikkoun Sofrim: digitization of Hebrew manuscripts by HTR and adaptive crowdsourcing.

- **Prof. Tsvika Kuflik**, University of Haifa, Israel (tsvikak@univ.haifa.ac.il)
- **Dr. Moshe Lavee,** University of Haifa, Israel (mlavee@research.haifa.ac.il)
- **Prof. Daniel Stoekl Ben Ezra**, Université Paris 4 / CNRS / EPHE, France (Daniel.Stoekl@ephe.psl.eu)
- **M. Ben Kiessling**, Université Paris 4 / CNRS / EPHE, France

11:15 – 11:30 Discussions

11:30 – 11:55 The alphabet: history and development in the southern Levant during the first millennium B.C.

- **Prof. Esther Eshel**, Bar Ilan University, Israel (esther.eshel@biu.ac.il)
- **Dr. Michael Langlois**, Université de Strasbourg (Unistra), France (dida@mlanglois.com)

11:55 – 12:10 Discussions

Panel Discussions

Chaired by **Dr. Alex Altshuler**

Director for Social Sciences and Humanities, MOST

AlexA@most.gov.il

12:10 – 12:40 Digital Humanities in Israel: Challenges and Opportunities

Participants:

- **Dr. Liat Kozma**, The Hebrew University of Jerusalem and Israel Young Academy, Israel
- **Prof. Ophir Muenz-Manor**, The Open University of Israel
- **Dr. Sinai Rusinek**, University of Haifa and The Open University of Israel

12:40 –14:00 Lunch

Panel Conference
Chaired by **Dr. Iris Eisenberg**
Director for Life Science, MOST
IrisE@most.gov.il

14:00 – 14:30 Inherited Memories

- **Dr. Oded Rechavi.** Tel Aviv University (odedrech@mail.tau.ac.il)

Session 3

Co-Chairpersons
Dr. Robert Gardette
Program Officer for Biology and Health, France
robert.gardette@recherche.gouv.fr

Dr. Iris Eisenberg
Director for Life Science, MOST
IrisE@most.gov.il

14:30 – 14:55 Modeling and functional analysis of clonal evolution in Leishmania. Possible role of the non-coding RNA and the modification.

- **Dr. Shulamit Michaeli**, Bar Ilan University, Israel (Shulamit.michaeli@biu.ac.il)
- **Dr. Gerald Spaeth,** Institut Pasteur / INSERM, France (gerald.spaeth@pasteur.fr)

14:55 – 15:10 Discussions

15:10 –15:35 Transcriptome-wide RNA-RNA interaction networks in Staphylococcus aureus.

- **Dr. Hana Margalit**, the Hebrew University of Jerusalem, Israel (hanahm@ekmd.huji.ac.il)
- **Dr. Pascale Romby**, CNRS - UPR 9002, France (p.romby@ibmc-cnrs.unistra.fr)

15:35 – 15:50 Discussions

15:50-16:00 – Closing remarks

ABSTRACTS

Exploring hominin behavioral patterns through time and space: A morpho-techno-functional analysis of 3D digital models of stone handaxes

Abstract – Israeli and French side

Israeli PI – Dr. Naama Goren-Inbar,
The Hebrew University of Jerusalem, Israel
French PI – Dr. Marianne Deschamps,
CNRS, France

Our study explores the morphological, technological and functional variability of a prehistoric tool – the handaxe, a tool present in many stone assemblages from different regions and periods of Prehistoric times. As part of this study, multiple analyses reflecting different methodological approaches are performed on handaxes from Israeli and French archaeological sites and dated to different phases of the Lower and Middle Paleolithic periods, between 1.6 and 0.045 Ma. The main methodology of the study is high-resolution 3D geometric morphometric shape analysis. To accomplish this, all the artifacts included in the study (16 sites, 438 Israeli and 294 French handaxes) underwent digitization resulting in high-resolution 3D digital models. Following, several multivariate statistical analyses and procedures were performed on the digital models to provide a quantitative description of the shape variability both within and between assemblages. Additional analyses included a classical typo-technological attribute analysis describing the reduction sequence (*chaîne opératoire*) of the artifacts providing information on the function of the tools. The results of the study have provided significant insights with regard to some of the fundamental questions of human evolution and behavior during the Lower and Middle Paleolithic times. These include better understanding of the factors underlying variability in handaxe assemblages across time and space. Additionally, the results may help to elucidate the phylogenetic relations between the different cultural entities that produced the studied assemblages. Our results allow testing of several hypotheses regarding hominin dispersal routes, timing and the mechanisms of cultural transmission.

NLP Resources for Analyzing Reactions to Major Events in Hebrew and French Social Media

Abstract – Israeli and French side

Israeli PI – Prof. Yoav Goldberg,
Bar-Ilan University, Israel
French PI – Dr. Djamé Seddah,
Centre INRIA Paris, France

Social-media and the abundance of social-media text provide unprecedented opportunity to the study of human communication, human societies, human behavior and language change, in ways which were not possible before. Social media does not only reflect human activity, but also shape it. One such influence of social media on society is how they evolved to impact our perception of events. For instance, during the various Spring Revolutions, FACEBOOKusers were in the front line of the information war; more recently, during the November 2015 Paris Attacks, TWITTER was used to gather information about the victims and to offer shelter to those stranded by these attacks. Twitter was also used in 2014 to spread information about upcoming attacks on the Israeli population and their outcomes. These events generated a steady flow of global textual interactions, whose analysis can provide interesting insights about communication, communities, the formation of discourse around events, and reactions to crisis. The goal of this project is to study how language in social media can evolve over time, driven or not by global events. To do so, we decided to focus on Twitter data collected from 2014 and 2018, both in French and Hebrew and to explore the possibility to automatically detect language change (i) using either extra-linguistics properties, such as socio-demographics features and users' interest characterizion, as well as diachronic lexical information in a dynamic embedding framework or (ii) using a new vector analysis method that are surprisingly effective. Part of this project was also devoted to the understanding of recent contextual embeddings and language models that prove resistant to noise and brought a performance boost in various multilingual natural language processing tasks. In order to do so both teams developed (or extended) annotated data set for user-generated content and developed neurals models to handle code-switching data that are prevalent in user-generated content.

Tikkoun Sofrim: Automatic Transcription and Crowdsourcing Corrections of Medieval Hebrew Manuscripts of Midrash Tanhuma

Abstract – Israeli and French side

Israeli PI – Dr. Moshe Lavee,
University of Haifa, Israel
French PI – Prof. Daniel Stoekl Ben Ezra,
Université Paris 4/CNRS/EPHE, France

1. Overview

Tikkoun Sofrim is a multifaceted project, aimed at integrating machine learning based Handwritten Text Recognition (HTR) with citizen science correction for full textual availability of Hebrew manuscripts. In the next couple of years, with the expected completion of Ktiv, the immense project for digitization of Hebrew manuscripts, more than 10M of images will be available through the NLI. Our project lay the foundations for making this immensely important cultural heritage treasure available and computable not only as images, but also in enriched text format. We have chosen to experiment with a specific sub-set of medieval Hebrew manuscripts, which include Midrashic literature in the genre of Tanhuma-Yelamdenu. Based on homilies from Late Antiquity, these texts underwent a long process of creative transmission through the middle ages. Accordingly, we are facing today with an extremely rich and complicated corpus, which was never fully treated with a helpful and comprehensive print edition.

2. The process and its components

The text production is comprised of the following elements: ***Kraken*** - ML based HTR, **Tikkoun Sofrim** – Crowdsourcing platform for HTR correction. In addition, the following activities took place: computing aggregated text from crowd corrections and data modeling for future edition of the Tanhuma.

3. Results

a. Transcribed manuscripts - We have processed 4 manuscripts through different stages of our pipeline. I) MS Geneva 146 - crowdsourcing completed, BNF Paris 150 - Currently under crowdsourcing, Vatican 44 - HTR completed, crowdsourcing launch with new community engagement scheme towards the end of December, Parma 3123 - HTR completed. Oxford 3 – has not started yet.

b. Data modeling for future edition - We prepared a pilot critical edition for Tanhuma literature The edition is based on the following principles and technologies:

   i. CTS (Canonical Text Services) for citable texts. Each of the textual witnesses is published in a CTS digital library - based on CapiTaiS open-source software (Perseids, Perseus Digital Library, DH Universität Leipzig, École nationale des chartes, PSL).

    ii. Semantic web to model the links between related passages in the corpus. CapiTainS software was extended, to support the intertextual links.

    iii. TRAViz (eTRACES project, Universität Leipzig) - Text Re-use Alignment Visualization, to display related paragraphs in an interactive graph; Again extension to CapiTainS.

4. The way ahead

a. eScriptorium: A web-based platform for automatic analysis and manual annotation of inscribed objects with kraken at its core has been created and is continuously developed by the eScriptorium team in Paris.

b. Text reuse in Tanhuma: The future pipeline for creating full textual availability of heritage manuscripts will comprise of the following elements in Tikkouns Sofrim we mainly focused on text production, but also looked at elements of text enrichment which will enable future implementation of the text in user interfaces (see Fig. 1)

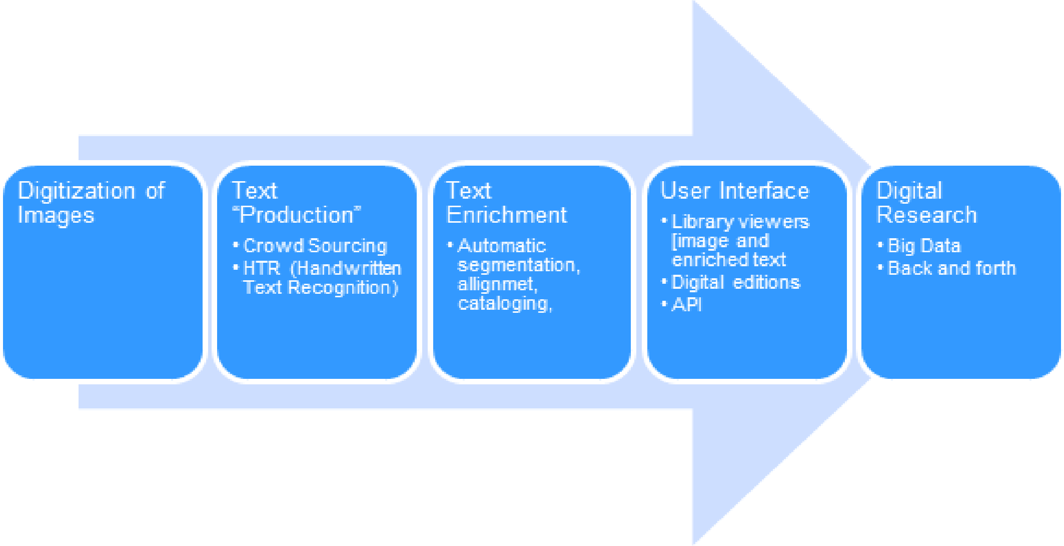

***Figure 1**. Tikkoun Sofrim and the larger vision. Tikkuon Sofrim dealt with the first two components while laying the foundation for the rest of the steps to be carried out in subsequent projects*

History and Development of the Alphabet in the First Millennium BCE Levant

Abstract – Israeli side

Israeli PI – Prof. Esther Eshel,
Bar Ilan University, Israel
French PI – Dr. Michael Langlois,
Université de Strasbourg (Unistra), France

In general, our research project seeks to study the history and evolution of the alphabet — especially Hebrew and Aramaic — in the First Millennium BCE Southern Levant in the light of new discoveries and with the help of new technologies. By shedding new light on ancient documents written in alphabetic scripts, the results of this research will benefit epigraphers, historians, archaeologists, historians of religions, philologists, linguists, and all who study the Orient through its texts.

Paleographical analysis requires a detailed examination of the inscription; the photographs that traditionally accompany an *editio princeps* do not suffice for such a task, as their low resolution usually prevents the close examination of a given letter, let alone its individual strokes.

All scientific fields face similar challenges, and the goal of this project is to revisit this nomenclature by establishing a set of factual criteria based on material evidence. For each letter of the alphabet, attention will be devoted to the shape, number and order of strokes. Factors such as angularity, writing utensil and technique, surface rigidity and speed will also be taken into account. The database will be designed in such a way that any unforeseen paleographical feature can later be modelled and recorded. Paleographical charts will be created, but not from handmade drawings, as is customary in epigraphy. Michael Langlois has therefore been developing a method to create paleographical charts based on actual photographs rather than drawings. Paleographical charts will be used to compare, side by side, occurrences of a letter within a single inscription (or section), so as to assess their consistency or identify various kinds and create ad hoc types.

The first corpus of inscriptions to be studied is the Aramaic Maresha collection of inscriptions, including 400 ostraca and inscriptions (dated to 3rd-2nd centuries BCE) entrusted to Esther Eshel for publication. In Maresha more than 1200 Greek and Semitic (mainly Aramaic) inscriptions were found. The number and diversity of the finds places Maresha among the most important and enriching ancient archeological sites in Israel.

Other projects done by us includes the following :

1. The Mesha inscription (9th century BCE): We used Reflectance Transformation Imaging (RTI) in the Louvre, and held a conference last November in Jerusalem.
2. Hebrew ostraca from the Jeselsohn collection (7th-6th c BCE). We were using Multispectral imaging (MSI), collaboration with André Lemaire.

History and Development of the Alphabet in the First Millennium BCE Levant

Abstract – French side

French PI – Dr. Michael Langlois,
Université de Strasbourg (Unistra), France
Israeli PI – Prof. Esther Eshel,
Bar Ilan University, Israel

The alphabet underwent considerable changes in the first millennium BCE, producing so-called “national” scripts such as Hebrew, Aramaic, or Phoenician. Our project seeks to track and account for some of these changes in the light of new archaeological findings and using new techniques such as Reflectance transformation imaging (RTI) and Multi-spectral imagine (MSI). We selected several periods, places and media throughout the first millennium BCE Levant: (1) 9th century Moab, with the Mesha stele; (2) 3rd and 2nd century Maresha, with newly found Aramaic ostraca; (3) Second Temple Judah, with the Dead Sea scrolls on parchment and papyrus. We are also investigating the social and technical environment that surrounds the production of these texts and may account for the accompanying evolution of the alphabet. One of the milestones of our project is the organisation of an international conference on the training of scribes in the biblical world; this conference took place in June 2019 and the proceedings will be published in 2020.

Modelling and functional analysis of clonal evolution in Leishmania - possible role of non-coding RNAs and RNA modification

Abstract – Israeli and French side

Israeli PI – Dr. Shulamit Michaeli,
Bar Ilan University, Israel
French PI – Dr. Gerald Spaeth,
Institut Pasteur/INSERM, France

Parasitic protozoa of the genus Leishmania are the etiologic agents of a spectrum of severe diseases known as leishmaniases that cause substantial human morbidity and are among the five most important parasitic diseases worldwide. The parasite shows a digenetic life cycle relying on two stages: motile extracellular promastigotes infect the mid-gut of the sand fly insect vector, while non-motile intracellular amastigotes proliferate inside phagolysosomes of mammalian macrophages. The success of Leishmania as a pathogenic microbe relies on its capacity to adapt to a variety of environmental fluctuations encountered in these hosts via an evolutionary process, with frequent copy number variations (CNVs) of individual genes or chromosomes linked to drug resistance or changes in tissue tropism. As a result, Leishmania genome instability represents THE major limiting factor in current anti-leishmanial treatment, with treatment failure or drug resistance observed for all current drugs [9]. Despite the relevance of Leishmania evolvability for disease outcome and chemotherapy, only little is known on mechanisms underlying adaptation and fitness gain of this important eukaryotic pathogen. Our proposal directly responds to this important knowledge gap with direct relevance for other disease-causing eukaryotic systems, such as fungi and cancer. Synergizing complementary expertise in experimental evolution, genomics and epitranscriptomics as part of this bi-national program, both partners have recently made a series of breakthrough discoveries that are currently prepared for publication and significantly advance the currently limited state of the art in Leishmania fitness gain. Partner 1 demonstrated that Leishmania fitness gain during culture adaptation is driven by the positive selection of chromosome and gene amplifications that are increasing the gene dosage of many non-coding RNAs. These data suggest that in the absence of transcriptional control, Leishmania has evolved an alternative ‘ecto-genomic’ mechanism of regulating gene expression levels, where the nuclear genome serves as a master copy for extra-chromosomal amplification. RNA-seq analyses of small RNAs and rRNA modification levels applied on evolving parasites conducted by Partner 2 supports this groundbreaking novel concept. These analyses established the first genome-wide repertoire of small RNAs in L. donovani and revealed major changes in the level of small nucleolar

RNA (snoRNA) expression during parasite fitness gain, which correlated with increased pseudouridine levels at over 10 positions located in the active site of the ribosome. These changes likely generate different types of ribosomes in response to environmental cues, which may affect translation much like it was observed in cancer cells or during differentiation of stem-cells. In conclusion, the "Maïmonide-Israel" program allowed us to synergize our complementary expertise and obtain important results that together support our main working hypothesis that Leishmania evolutionary adaptation is driven by highly dynamic, epistatic genome/RNome interactions. These interactions change the parasite phenotypic landscape in response to a given environment, which is the substrate for Leishmania fitness gain in culture, in experimental infection and in the field. Investigating this highly innovative working hypothesis through future funding (a joint ERC synergy grant has been submitted to the EC) will generate novel insight into regulatory mechanisms driving evolutionary adaptation in eukaryotic systems, including cancer, where the evolution of drug resistant phenotypes has been recently linked to pre-existing, intrinsic genetic fluctuations and adaptive transcriptional responses, thus echoing the genome/RNome interactions we revealed in our Leishmania system. Furthermore, frequent phenotypic shifts in virulence, tropism and drug susceptibility observed in Leishmania clinical isolates aggravate the often devastating clinical outcome of infection. Our proposal exploits the very same mechanisms causing these phenotypic shifts to uncover genome/RNome interaction networks as novel biomarkers that will inform epidemiological investigations conducted as part of the H2020 LeiSHield project coordinated by Partner 1 (www.leishield.org).

Unveiling the small RNA-target interaction network of a pathogenic bacterium

Abstract – Israeli side

Israeli PI – Dr. Hana Margalit,
Hebrew University of Jerusalem, Israel
French PI – Dr. Pascale Romby,
CNRS - UPR 9002, France

Pathogenic bacteria inhabit different ecological niches and experience rapidly varying life conditions. The adaptation to changing conditions involves substantial changes in the expression programs of many genes, and is tightly regulated. Deciphering the regulatory circuitry that underlies virulence is a pre-requisite for basic understanding of bacterial pathogenicity and for the development of antibacterial therapeutics. For many years, research was mainly focused on transcriptional factors, which were considered as the main regulators of virulence in bacteria. This has changed dramatically during the last two decades with the recognition that small non-coding RNAs (sRNAs), acting as post-transcriptional regulators of gene expression, play important roles in the control of central cellular processes, including virulence. This is particularly the case for the pyogenic and toxigenic extracellular pathogen *Staphylococcus aureus*, where several sRNAs have been integrated into complex regulatory circuits linking metabolism, stress response and virulence. An important step towards understanding the role of sRNAs in the cell and in the virulence of *S. aureus* is the determination of the sRNA-target interaction network. In my talk I will describe a new experimental-computational approach, RIL-seq, which we recently developed to decipher sRNA-target interactions at a global scale, and the novel knowledge that can be gained by its application.

Unveiling the small RNA-target interaction network of a pathogenic bacterium

Abstract – French side

French PI – Dr. Pascale Romby,
CNRS - UPR 9002, France
Israeli PI – Dr. Hana Margalit,
Hebrew University of Jerusalem, Israel

Our previous collaborative study has led to the discovery of an elaborate regulatory circuit, which plays an important role in the temporal expression of numerous virulence factors of *S. aureus*. Our current interdisciplinary study, involving extensive scientific interactions between our two teams, helps to set up innovative methodologies (MAPS, RIL-Seq, LIGR-Seq) to gain more insights into the functions of sRNAs. Using specific examples, I will illustrate the complexity of interactions between RNAs, RNA-binding proteins and transcriptional regulatory proteins to form complex and dynamic networks interconnecting stress responses, metabolism adaptation, and virulence. Ongoing experiments are expected to reveal the whole sets of RNA-RNA interactions at a global scale, providing important insights into the pathogenicity of *S. aureus*.